Genetic Variant of SARS-CoV-2 Isolates in Indonesia: Spike Glycoprotein Gene

Arif Nur Muhammad Ansori\textsuperscript{1,2,3}, Viol Dhea Kharisma\textsuperscript{2,3,4}, Sahal Sabilil Muttaqin\textsuperscript{5}, Yulanda Antonius\textsuperscript{2,6} and Arli Aditya Parikesit\textsuperscript{7*}

\textsuperscript{1}Doctoral Program in Veterinary Science, Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia. \textsuperscript{2}Division of Molecular Biology and Genetics, Generasi Biologi Indonesia (GENBINESIA) Foundation, Gresik, Indonesia. \textsuperscript{3}Department of Biotechnology, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, Malang, Indonesia. \textsuperscript{4}Master Program in Biology, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Malang, Indonesia. \textsuperscript{5}Faculty of Biology, Universitas Gadjah Mada, Yogyakarta, Indonesia. \textsuperscript{6}Department of Biology, Faculty of Biotechnology, University of Surabaya, Surabaya, Indonesia. \textsuperscript{7}Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences, Jakarta, Indonesia.

Abstract

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), a novel coronavirus and the primary causative agent of coronavirus disease 2019 (COVID-19), first occurred in China and rapidly spread worldwide. The government of the Republic of Indonesia confirmed its first two cases of COVID-19 in March 2020. COVID-19 is a serious illness with no efficacious antiviral medication or approved vaccine currently available. Therefore, there is a need to investigate the genome of SARS-CoV-2. In this study, we characterized SARS-CoV-2 spike glycoprotein genes from Indonesia to investigate their genetic composition and variability. Overall, ten SARS-CoV-2 spike glycoprotein gene sequences retrieved from GenBank (National Center for Biotechnology Information, USA) and the GISAID EpiCoV database (Germany) were compared. We analyzed nucleotide variants and amino acid changes using Molecular Evolutionary Genetics Analysis (MEGA) X and analyzed gene similarity using the LALIGN web server. Interestingly, we revealed several specific mutation sites, however, there were no significant changes in the genetic composition of SARS-CoV-2 spike glycoprotein genes, when compared to the Wuhan-Hu-1 isolate from China. However, this is a preliminary study and we recommend that molecular epidemiology and surveillance programs against COVID-19 in Indonesia be improved.

Keywords: Coronavirus, COVID-19, Genetic composition, Mutation, SARS-CoV-2

*Correspondence: arli.parikesit@i3l.ac.id

(Received: May 10, 2020; accepted: May 26, 2020)


© The Author(s) 2020. Open Access. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License which permits unrestricted use, sharing, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
INTRODUCTION

The Chinese government first reported a novel pneumonia-causing disease in Wuhan in December 2019\(^1\). The causative agent was identified and named severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses (ICTV)\(^2\). This new virus has rapidly spread across China and to many other countries across the world, including Indonesia\(^3,4\). The World Health Organization has named the illness caused by SARS-CoV-2 as coronavirus disease 2019 (COVID-19)\(^5\).

According to an online interactive dashboard hosted by the Center for Systems Science and Engineering at Johns Hopkins University (Baltimore, USA), which tracks reported cases of COVID-19 in real-time\(^6\), more than 4 million people have been infected by SARS-CoV-2 worldwide, with more than 14 000 cases in Indonesia alone. Currently, there are three coronaviruses that cause illness in humans: severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2\(^7\).

Taxonomically, coronaviruses belong to the Coronaviridae family in the order Nidovirales, with examples in four distinct genera: Alphacoronavirus, Betacoronavirus, Deltacoronavirus, and Gammacoronavirus\(^8\). The structural proteins are encoded by four genes, specifically the envelope (E), nucleocapsid (N), membrane (M), and spike glycoprotein (S)\(^7,9-10\). Previous studies have shown that the spike glycoprotein plays a crucial role in binding to receptors on the host cell. Therefore, this protein is a key target for a number of antiviral therapies and a promising antigen for generating vaccines formulated against SARS-CoV, MERS-CoV, and SARS-CoV-2.

Molecular epidemiology research is a crucial tool in the surveillance of newly emerging and reemerging viruses\(^11-12\). Indonesia was the eighth country in Southeast Asia after Brunei, Cambodia, Malaysia, Myanmar, Singapore, Thailand, and Vietnam to report the whole-genome sequences of SARS-CoV-2 in the region. Both Callaway (2020) and Shang et al. (2020) have shown that vaccines are being developed against SARS-CoV-2 by various research groups worldwide\(^13-14\). Similarly, Al-Tawfiq (2020) has discussed other potential therapeutic options for COVID-19\(^15\) and both remdesivir and chloroquine are capable of effectively inhibiting SARS-CoV-2 in \textit{in vitro} assays\(^16\). Despite these promising treatment options, COVID-19 remains a serious disease with no proven effective antiviral medication or approved vaccine available. Therefore, there is an urgent need to investigate the genome of SARS-CoV-2. In this study, we characterized SARS-CoV-2 spike glycoprotein genes from Indonesia in order to investigate their genetic composition and the similarity between different gene isolates.

MATERIALS AND METHODS

\textbf{SARS-CoV-2 Isolates}

SARS-CoV-2 spike glycoprotein gene (3822 bp) sequences were obtained from GenBank (National Center for Biotechnology Information, USA) and the Global Initiative on Sharing All Influenza Data (GISAID) EpiCoV database (Germany) (Table 1).

\textbf{Genetic Composition Analysis}

We analyzed the genetic composition of SARS-CoV-2 spike glycoproteins (both nucleotide variants and amino acid changes) using Molecular Evolutionary Genetics Analysis (MEGA) X\(^12,17\). We used the Wuhan-Hu-1 isolate as a reference gene, according to Sekizuka et al. (2020)\(^18\).

\textbf{Similarity Analysis}

We analyzed the similarity of SARS-CoV-2 spike glycoprotein genes using the LALIGN web server (The SIB Swiss Institute of Bioinformatics, Switzerland) with an E-value threshold of 10.0. The LALIGN program is based on an algorithm first described by Huang and Miller\(^19\).

RESULTS AND DISCUSSION

Coronaviruses infect both animals and humans and lead to various illnesses, including neurological, enteric, and respiratory diseases. There are four distinct genera of coronaviruses: Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus\(^8\). SARS-CoV, MERS-CoV, and SARS-CoV-2 are three highly pathogenic coronaviruses, capable of infecting humans, that emerged in 2002, 2012, and 2019, respectively\(^3,4\). In this study, the genetic compositions and sequence similarities of nine Indonesian SARS-CoV-2 spike glycoprotein genes

---


### Table 1. SARS-CoV-2 isolates obtained from the GenBank and GISAID EpiCoV databases

<table>
<thead>
<tr>
<th>No</th>
<th>Accession ID</th>
<th>Virus Name</th>
<th>Origin</th>
<th>Submitting Institution</th>
<th>Host</th>
<th>Specimen Source</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MN908947.3 (Reference)</td>
<td>Wuhan-Hu-1 (Wuhan)</td>
<td>China</td>
<td>Shanghai Public Health Clinical Center and School of Public Health, Fudan University, Shanghai</td>
<td>Homo sapiens</td>
<td>Unknown</td>
<td>Reference genome</td>
</tr>
<tr>
<td>2</td>
<td>EPI_ISL_435281</td>
<td>JKT-EIJK0141 (Jakarta)</td>
<td>Indonesia</td>
<td>Eijkman Institute for Molecular Biology, Ministry of Research and Technology/ National Agency for Research and Innovation of the Republic of Indonesia</td>
<td>Homo sapiens</td>
<td>Nasopharyngeal and Oropharyngeal swab</td>
<td>22x</td>
</tr>
<tr>
<td>3</td>
<td>EPI_ISL_435282</td>
<td>JKT-EIJK0317 (Jakarta)</td>
<td>Indonesia</td>
<td>Eijkman Institute for Molecular Biology, Ministry of Research and Technology/ National Agency for Research and Innovation of the Republic of Indonesia</td>
<td>Homo sapiens</td>
<td>Nasopharyngeal and Oropharyngeal swab</td>
<td>1,480x</td>
</tr>
<tr>
<td>4</td>
<td>EPI_ISL_435283</td>
<td>JKT-EIJK2444 (Jakarta)</td>
<td>Indonesia</td>
<td>Eijkman Institute for Molecular Biology, Ministry of Research and Technology/ National Agency for Research and Innovation of the Republic of Indonesia</td>
<td>Homo sapiens</td>
<td>Nasopharyngeal and Oropharyngeal swab</td>
<td>8,082x</td>
</tr>
<tr>
<td>5</td>
<td>EPI_ISL_437187</td>
<td>EJ-ITD853Sp (Surabaya)</td>
<td>Indonesia</td>
<td>Institute of Tropical Disease, Universitas Airlangga</td>
<td>Homo sapiens</td>
<td>Sputum</td>
<td>764x</td>
</tr>
<tr>
<td>6</td>
<td>EPI_ISL_437188</td>
<td>EJ-ITD3590NT (Surabaya)</td>
<td>Indonesia</td>
<td>Institute of Tropical Disease, Universitas Airlangga</td>
<td>Homo sapiens</td>
<td>Nasopharyngeal and Oropharyngeal swab</td>
<td>96x</td>
</tr>
<tr>
<td>7</td>
<td>EPI_ISL_437189</td>
<td>JKT-EIJK01 (Jakarta)</td>
<td>Indonesia</td>
<td>Eijkman Institute for Molecular Biology, Ministry of Research and Technology/ National Agency for Research and Innovation of the Republic of Indonesia</td>
<td>Homo sapiens</td>
<td>Nasopharyngeal and Oropharyngeal swab</td>
<td>2,256x</td>
</tr>
<tr>
<td>8</td>
<td>EPI_ISL_437190</td>
<td>JKT-EIJK02 (Jakarta)</td>
<td>Indonesia</td>
<td>Eijkman Institute for Molecular Biology, Ministry of Research and Technology/ National Agency for Research and Innovation of the Republic of Indonesia</td>
<td>Homo sapiens</td>
<td>Nasopharyngeal swab</td>
<td>5,297x</td>
</tr>
<tr>
<td>9</td>
<td>EPI_ISL_437191</td>
<td>JKT-EIJK03 (Jakarta)</td>
<td>Indonesia</td>
<td>Eijkman Institute for Molecular Biology, Ministry of Research and Technology/ National Agency for Research and Innovation of the Republic of Indonesia</td>
<td>Homo sapiens</td>
<td>Nasopharyngeal swab</td>
<td>2,112x</td>
</tr>
<tr>
<td>10</td>
<td>EPI_ISL_437192</td>
<td>JKT-EIJK04 (Jakarta)</td>
<td>Indonesia</td>
<td>Eijkman Institute for Molecular Biology, Ministry of Research and Technology/ National Agency for Research and Innovation of the Republic of Indonesia</td>
<td>Homo sapiens</td>
<td>Nasopharyngeal and Oropharyngeal swab</td>
<td>5,759x</td>
</tr>
</tbody>
</table>
were determined using sequences obtained from GenBank and the GISAID EpiCoV database (Table 2-4). Many researchers worldwide have previously reported mutations in the SARS-CoV-2 genome\textsuperscript{10,20-21}, however, there has not previously been a study of the SARS-CoV-2 genome from Indonesian isolates.

The coronavirus spike glycoprotein mediates membrane fusion and viral entry into host cells and is therefore the primary target for many neutralizing antibodies (Fig. 1). The spike glycoprotein has two domains, S1 and S2, where S1 is responsible for binding the virion to ACE2 on the host cell membrane\textsuperscript{21}. Several antiviral drugs and vaccines have been developed which target the spike glycoprotein. Du et al. (2009)\textsuperscript{12} demonstrated the effectiveness of several of these antiviral therapies, including small interfering RNAs, protease inhibitors, ACE2 blockers, fusion blockers, spike glycoprotein inhibitors, neutralizing antibodies, and spike glycoprotein cleavage inhibitors in in vitro studies. In addition, a number of techniques have been used to generate vaccines using all or part of the spike glycoprotein as an antigen. These include the use of recombinant receptor binding domain protein, viral vectors, full-length S protein, recombinant spike glycoprotein, and spike protein DNA-expressing vectors. Thus, it is very important to investigate the genetic composition and sequence similarity of this protein.

Interestingly, despite reporting several mutation sites in this study, we demonstrate that there is no significant change in the genetic composition of SARS-CoV-2 spike glycoprotein genes. Nucleotide variants in SARS-CoV-2 spike

### Table 2. Nucleotide mutation sites in the SARS-CoV-2 spike glycoprotein

<table>
<thead>
<tr>
<th>No.</th>
<th>Virus Name</th>
<th>Nucleotide Position</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>224</td>
</tr>
<tr>
<td>1</td>
<td>Wuhan-Hu-1</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>JKT-EIJK0141</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>JKT-EIJK0317</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>JKT-EIJK2444</td>
<td>T</td>
</tr>
<tr>
<td>5</td>
<td>EJ-ITD8535p</td>
<td>C</td>
</tr>
<tr>
<td>6</td>
<td>EJ-ITD3590NT</td>
<td>C</td>
</tr>
<tr>
<td>7</td>
<td>JKT-EIJK01</td>
<td>C</td>
</tr>
<tr>
<td>8</td>
<td>JKT-EIJK02</td>
<td>C</td>
</tr>
<tr>
<td>9</td>
<td>JKT-EIJK03</td>
<td>C</td>
</tr>
<tr>
<td>10</td>
<td>JKT-EIJK04</td>
<td>C</td>
</tr>
</tbody>
</table>

### Table 3. Amino acid mutation sites in the SARS-CoV-2 spike glycoprotein

<table>
<thead>
<tr>
<th>No.</th>
<th>Virus Name</th>
<th>Amino Acid Position</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>76</td>
</tr>
<tr>
<td>1</td>
<td>Wuhan-Hu-1</td>
<td>T</td>
</tr>
<tr>
<td>2</td>
<td>JKT-EIJK0141</td>
<td>T</td>
</tr>
<tr>
<td>3</td>
<td>JKT-EIJK0317</td>
<td>T</td>
</tr>
<tr>
<td>4</td>
<td>JKT-EIJK2444</td>
<td>I</td>
</tr>
<tr>
<td>5</td>
<td>EJ-ITD8535p</td>
<td>T</td>
</tr>
<tr>
<td>6</td>
<td>EJ-ITD3590NT</td>
<td>T</td>
</tr>
<tr>
<td>7</td>
<td>JKT-EIJK01</td>
<td>T</td>
</tr>
<tr>
<td>8</td>
<td>JKT-EIJK02</td>
<td>T</td>
</tr>
<tr>
<td>9</td>
<td>JKT-EIJK03</td>
<td>T</td>
</tr>
<tr>
<td>10</td>
<td>JKT-EIJK04</td>
<td>T</td>
</tr>
</tbody>
</table>
glycoprotein genes described in this study include JKT-EIJK2444 (224C>T), EJ-ITD3590NT (347C>G; 1841A>G; 2031G>T), JKT-EUK01 (141T>C; 1864G>T), and JKT-EUK04 (1715C>T; 2464C>T) (Table 2). In addition, we also analyzed amino acid changes in the SARS-CoV-2 spike glycoprotein including JKT-EIJK2444 (T76I), EJ-ITD3590NT (S116C; D614G; Q677H), JKT-EUK01 (V622F), and JKT-EUK04 (T572I; L822F) (Table 3). Finally, the interval score of similarity between each of the SARS-CoV-2 spike glycoproteins was between 99.9% and 100% (Table 4).

Previous studies investigating gene variability in SARS-CoV-2 samples include the work of Sekizuka et al. (2020), who performed whole-genome sequencing of SARS-CoV-2, directly from PCR-positive clinical specimens18. This was conducted in order to generate a haplotype network analysis of the Diamond Princess cruise ship outbreak, using the Wuhan-Hu-1 sequence as a reference. Additionally, Castillo et al. (2020) reported a phylogenetic analysis of the first four SARS-CoV-2 cases in Chile, analyzing nucleotide variants, amino acid changes, and sequence similarity23. While our study complements these previous works, several shortcomings remain, including the relatively small number of isolates studied, the methodology used for whole-genome sequencing, and the quality coverage of SARS-CoV-2 genomes isolated from Indonesia.

As new information on SARS-CoV-2 is published daily, new concepts and frameworks must constantly be adopted. Currently, the GISAID EpiCoV database and Tang et al. (2020) have established three subtypes of SARS-CoV-2 based on nucleotide variants that produce amino acid changes: S, G, and V24. The mutation rate of viruses is considerably higher than most other biological entities, including prokaryotes and eukaryotes. This is especially true of RNA-based viruses such as SARS-CoV-2, Ebola and dengue, due to hydroxyl groups in RNA that act as catalytic sites for mutation. This advanced mutation rate, leads to enhanced virulence and a higher capacity for adaptive evolution25-26. While Tang et al. (2020) have suggested that SARS-CoV-2 exhibits the characteristic high mutation rate of an RNA virus24, in fact, the mutation rate of SARS-CoV-2 and other coronaviruses might be slightly lower than other RNA viruses because of its genome-encoded
exonuclease. Regardless, its high mutation rate increases the potential of this zoonotic viral pathogen to adapt to efficient transmission from human to human and potentially allows it to become more virulent.

The genomic characteristics of SARS-CoV-2 are significantly different than either SARS-CoV or MERS-CoV. A previous study reported that the homology of SARS-CoV-2 with the bat coronavirus isolate, RaTG13, was 96%. Interestingly, another study reported that the homology of SARS-CoV-2 with a pangolin coronavirus was 99%. From these results, it could be suggested that pangolins act as an intermediate host between bats and humans.

This study of genomic variants of SARS-CoV-2 isolated from Indonesia is crucial for future investigations into the pathogenesis, prevention, and treatment of SARS-CoV-2. Development of this genomic data is vital work that will facilitate vaccine design, epidemiological investigations, viral detection, functional analysis, and evaluation of treatment options.

Outbreaks of SARS-CoV-2 have led to a state of medical and economic emergency worldwide. Therefore, understanding the characteristics of the SARS-CoV-2 genome and developing systems to monitor SARS-CoV-2 during the pandemic are critical steps for controlling this disease. The identification of genotypes connected to specific geographic and temporal infectious clusters suggests that genomic data can be used to track and monitor the transmission of SARS-CoV-2. Therefore, the rapid discovery of genetic variants of SARS-CoV-2 is necessary for a streamlined response to the COVID-19 outbreak. Similarly, identifying specific SARS-CoV-2 variants and connecting them using a molecular epidemiology approach would allow researchers to determine the origin of a specific variant and monitor its transmission. This could be an important tool in controlling the outbreak.

CONCLUSION

In summary, there was no significant difference between the SARS-CoV-2 spike glycoprotein gene sequences found in Indonesia and the Wuhan-Hu-1 isolate from China. However, this was only a preliminary study and we recommend expanding molecular epidemiology and surveillance programs to monitor COVID-19 in Indonesia.

ACKNOWLEDGMENTS

This study was supported by the Directorate General of Higher Education, Ministry of Education and Culture of the Republic of Indonesia; the Institute of Research and Community Empowerment (LPPM) of the Indonesia International Institute for Life Sciences.
(I3L); and Generasi Biologi Indonesia (GENBINESIA) Foundation, Indonesia. We thank Editage for editing the manuscript.

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

AUTHORS’ CONTRIBUTION
All listed authors made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

ETHICS STATEMENT
This article does not contain any experiments using human participants or animals performed by any of the authors.

AVAILABILITY OF DATA
Not applicable.

REFERENCES
21. Yin C. Genotyping coronavirus SARS-CoV-2: methods and implications. Genomics. 2020;S0888-


