Full-length genome characterization and phylogenetic analysis of SARS-CoV-2 virus strains from Yogyakarta and Central Java, Indonesia

.Gunadi Corresp., 1 Hendra Wibawa Corresp., 2, Marcellus 1 Mohamad Saifudin Hakim 3, Edwin Widyanto Daniwijaya 4, Ludhang Prapitita Rizki 3, Endah Supriyati 5, Dwi Aris Agung Nugraningsih 6, Afiahayati 7, Siswanto 8, Kristy Iskandar 9, Nungki Anggorowati 10, Alvin Santoso Kalim 1, Dyah Ayu Puspitarani 1, Kemala Athollah 1, Eggi Arguni 11, Titik Nuryastuti 3, Tri Wibawa 3

1 Pediatric Surgery Division, Department of Surgery, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
2 Disease Investigation Center, Wates, Ministry of Agriculture, Yogyakarta, Indonesia
3 Department of Microbiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
4 Department of Microbiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada /UGM Academic Hospital, Yogyakarta, Indonesia
5 Centre of Tropical Medicine, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
6 Department of Pharmacology and Therapy, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
7 Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia
8 Department of Physiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/UGM Academic Hospital, Yogyakarta, Indonesia
9 Department of Child Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/ UGM Academic Hospital, Yogyakarta, Indonesia
10 Department of Anatomical Pathology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
11 Department of Anatomical Pathology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

Corresponding Authors: Gunadi, Hendra Wibawa
Email address: drgunadi@ugm.ac.id, hendra.wibawa@pertanian.go.id

Background

Recently, SARS-CoV-2 virus with the D614G mutation has become a public concern due to rapid dissemination of this variant across many countries. Our study aims were 1) to report full-length genome sequences of SARS-CoV-2 collected from four COVID-19 patients in the Special Region of Yogyakarta and Central Java provinces, Indonesia; 2) to compare the clade distribution of full-length genome sequences from Indonesia (n=60) from March to September 2020; and 3) to perform phylogenetic analysis of SARS-CoV-2 complete genomes from different countries, including Indonesia.

Methods

Whole genome sequencing (WGS) was performed using next-generation sequencing (NGS) applied in the Illumina MiSeq instrument. Full-length virus genomes were annotated using the reference genome of hCoV-19/Wuhan/Hu-1/2019 (NC_045512.2) and then visualized in UGENE v. 1.30. For phylogenetic analysis, a dataset of 88 available SARS-CoV-2 complete genomes from different countries, including Indonesia, was retrieved from GISAID.

Results

All patients were hospitalized with various severities of COVID-19. Phylogenetic analysis revealed that one and three virus samples belong to clade L and GH. These three clade GH virus samples (EPI_ISL_525492, 516800 and 516829) were not only located in a cluster with SARS-CoV-2 genomes from...
Asia but also those from Europe, whereas the clade L virus sample (EPI_ISL_516806) was located amongst SARS-CoV-2 genomes from Asia. Using full-length sequences available in the GISAID EpiCoV Database, 39 of 60 SARS-CoV-2 (65%) from Indonesia harbor the D614G mutation.

**Conclusion**

These findings indicate that SARS-CoV-2 with the D614G mutation appears to become the major circulating virus in Indonesia, concurrent with the COVID-19 situation worldwide.
Full-length genome characterization and phylogenetic analysis of SARS-CoV-2 virus strains from Yogyakarta and Central Java, Indonesia

Gunadi1*, Hendra Wibawa2*, Marcellus3, Mohamad S. Hakim4, Edwin W. Daniwijaya5, Ludhang P. Rizki6, Endah Supriyati7, Dwi AA. Nugrahaningsih8, Afiahayati9, Siswanto10, Kristy Iskandar11, Nungki Anggorowati12, Alvin S. Kalim13, Dyah Ayu Puspitarani14, Kemala Athollah15, Eggi Arguni16, Titik Nuryastuti17, Tri Wibawa18, on behalf of the Yogyakarta-Central Java COVID-19 study group

1 Pediatric Surgery Division, Department of Surgery, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia; drgunadi@ugm.ac.id
2 Disease Investigation Center, Wates, Yogyakarta, Ministry of Agriculture Indonesia; hendra.wibawa@pertanian.go.id
3 Pediatric Surgery Division, Department of Surgery, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia; marcelluskorompis@gmail.com
4 Department of Microbiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia; m.s.hakim@ugm.ac.id
5 Department of Microbiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada /UGM Academic Hospital, Yogyakarta, Indonesia; m.edwin.d@ugm.ac.id
6 Department of Microbiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia; ludhang@ugm.ac.id
7 Centre of Tropical Medicine, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia; endah.supriyati@worldmosquito.org
8 Department of Pharmacology and Therapy, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia; dwi.aris.a@ugm.ac.id
9 Department of Computer Science and Electronics Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia; afia@ugm.ac.id
10 Department of Physiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/UGM Academic Hospital, Yogyakarta, Indonesia; sis12@ugm.ac.id
11 Department of Child Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/ UGM Academic Hospital, Yogyakarta, Indonesia; kristy.iskandar@mail.ugm.ac.id
12 Department of Anatomical Pathology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia; nungki@ugm.ac.id
13 Pediatric Surgery Division, Department of Surgery, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia; alvin_2495@yahoo.com
14 Pediatric Surgery Division, Department of Surgery, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia; raden.ajeng.d@mail.ugm.ac.id
15 Pediatric Surgery Division, Department of Surgery, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia; kemala.athollah@mail.ugm.ac.id
Abstract

Background: Recently, SARS-CoV-2 virus with the D614G mutation has become a public concern due to rapid dissemination of this variant across many countries. Our study aims were: 1) to report full-length genome sequences of SARS-CoV-2 collected from four COVID-19 patients in the Special Region of Yogyakarta and Central Java provinces, Indonesia; 2) to compare the clade distribution of full-length genome sequences from Indonesia (n=60) from March to September 2020; and 3) to perform phylogenetic analysis of SARS-CoV-2 complete genomes from different countries, including Indonesia.

Methods: Whole genome sequencing was performed using next-generation sequencing applied in the Illumina MiSeq instrument. Full-length virus genomes were annotated using the reference genome of hCoV-19/Wuhan/Hu-1/2019 (NC_045512.2) and then visualized in UGENE v. 1.30. For phylogenetic analysis, a dataset of 88 available SARS-CoV-2 complete genomes from different countries, including Indonesia, was retrieved from GISAID.

Results: All patients were hospitalized with various severities of COVID-19. Phylogenetic analysis revealed that one and three virus samples belong to clade L and GH, respectively. These three clade GH virus samples (EPI_ISL_525492, 516800 and 516829) were not only located in a cluster with SARS-CoV-2 genomes from Asia but also those from Europe, whereas the clade L virus sample (EPI_ISL_516806) was located amongst SARS-CoV-2 genomes from Asia. Using full-length sequences available in the GISAID EpiCoV Database, 39 of 60 SARS-CoV-2 (65%) from Indonesia harbor the D614G mutation.
Conclusion: These findings indicate that SARS-CoV-2 with the D614G mutation appears to become the major circulating virus in Indonesia, concurrent with the COVID-19 situation worldwide.

Keywords: COVID-19; D614G mutation; Indonesia; phylogenetic analysis; SARS-CoV-2; whole genome sequencing
Introduction

In December 2019, an outbreak of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) causing Coronavirus Disease 2019 (COVID-19) was detected in Wuhan, China and has become a global pandemic, including Indonesia [1,2].

In Indonesia, the first two COVID-19 cases were reported on March 2nd, 2020. Since then, the confirmed cases have been continuously increasing although several public health measures, involving isolation of confirmed patients and community wide containment in addition to strictly enforced personal health protocols, were conducted to halt transmission events [3]. Tragically, on November 19, 2020, Indonesia recorded 478,720 COVID-19 infections and 15,503 deaths [3]. This situation means that Indonesia has reported the second most confirmed COVID-19 cases in the South East Asia countries after the Philippines, yet has the highest number of deaths caused by COVID-19 among other South East Asia countries [3].

Just as many other countries, the detection of SARS-CoV-2 in suspected people is mainly based on reverse transcriptase real-time polymerase chain reaction (RT-PCR). The supply of PCR reagents, trained lab personnel and the availability of laboratories with sufficient biocontainment levels are major challenges of SARS-CoV-2 detection in developing countries, such as Indonesia [4]. Therefore, it is not surprising that the tested people per week is still lower than the World Health Organization (WHO) standard [3].

Recently, SARS-CoV-2 with the D614G mutation became the most frequently detected globally, including South East Asia region [5,6]. Interestingly, SARS-CoV-2 with the G614 variant had significantly higher infectious titers than the original D614 virus, and COVID-19 patients with the G614 variant had a higher viral load than patients without the mutation [5]. A recent study showed that the SARS-CoV-2 with the G614 variant revealed increased infectivity, competitive fitness, and transmission than the wild-type D614 virus in human airway epithelial cells and hamster [7]. However, this mutation was not associated with the severity of COVID-19 [5,6].

Here, we aimed: 1) to report full-length genome sequences of SARS-CoV-2 collected from four COVID-19 patients in the Special Region of Yogyakarta and Central Java provinces, Indonesia; 2) to compare the clade distribution of full-length genome sequences from Indonesia (n=60) from March to September 2020; and 3) to perform phylogenetic analysis of SARS-CoV-2 complete genomes from different countries, including Indonesia.
Materials & Methods

Classification of COVID-19 severity

We determined the COVID-19 severity according to the WHO classifications: 1) mild, symptomatic COVID-19 patients without evidence of hypoxia or pneumonia; 2) moderate, clinical signs of pneumonia (i.e., fever, cough, dyspnea, fast breathing) but not severe pneumonia, including blood oxygen saturation levels (SpO₂) ≥90% in room air; 3) severe, clinical signs of pneumonia plus one of the conditions as follows: respiratory rate >30 breaths/minute, severe respiratory distress, or SpO₂ <90% in room air; and 4) critical, Acute Respiratory Distress Syndrome (ARDS), sepsis, or septic shock, other complications such as acute pulmonary embolism, acute coronary syndrome, acute stroke, and delirium [8].

Virus samples

All virus samples were collected from hospitalized patients with COVID-19 from June-August 2020 in Yogyakarta and Central Java provinces. Samples were collected from nasopharyngeal swabs and then directly put into viral transport media (DNA/RNA Shield™ Collection Tube with Swab, Zymo Research, CA, United States). Samples were sent to the Department of Microbiology and Laboratorium Diagnostik Yayasan Tahija World Mosquito Program, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada and the Disease Investigation Center, Wates, Yogyakarta for SARS-CoV-2 virus detection using Real-Q 2019-nCoV Detection Kit (BioSewoom, Seoul, South Korea) with LightCycler® 480 Instrument II (Roche Diagnostics, Mannheim, Germany).

Whole genome sequencing

Total viral RNA was extracted from 15 original samples (nasopharyngeal swabs) using a QiAMP Viral RNA mini kit (Qiagen, Hilden, Germany), followed by double stranded cDNA synthesis using Maxima H Minus Double-Stranded cDNA Synthesis (Thermo Fisher Scientific, MA, United States), and then purified by a GeneJET PCR Purification Kit (Thermo Fisher Scientific, MA, United States). The Nextera DNA Flex for Enrichment using Respiratory Virus Oligos Panel was used for library preparations, and whole genome sequencing was performed using next generation sequencing (NGS) applied in the Illumina MiSeq instrument (Illumina, San Diego, CA, United States) with Illumina MiSeq reagents v3 150 cycles (2 x 75 cycles). Among
149 15 samples that were analyzed by NGS, only four samples showed good data for further
150 bioinformatics analysis. The paired reads were trimmed for quality and length and assembled by
151 mapping to the reference genome from Wuhan, China (hCoV-19/Wuhan/Hu-1/2019, GenBank
152 accession number: NC_045512.2) using Burrow-Wheeler Aligner (BWA) algorithm embedded
153 in UGENE v. 1.30 [9]. Single nucleotide polymorphism (SNP) was identified based on the
154 number of high confidence base calls (consensus sequence variations of the assembly) that
155 disagree with the reference bases for the genome position of interest. These variations were then
156 exported to a vcf file and visualized in MS Excel. All four full-genome sequences of SARS-
157 CoV-2 had the following accession IDs: EPI_ISL_516800, EPI_ISL_516806, EPI_ISL_516829,
158 and EPI_ISL_525492 [10].

160 Genome annotation and phylogenetic analysis
161 Full-length virus genomes were annotated using the reference genome of hCoV-19/Wuhan/Hu-
162 1/2019 (NC_045512.2). For phylogenetic analysis, a dataset of 88 available SARS-CoV-2 virus
163 genomes from different countries, including those from Indonesia, was retrieved from GISAID
164 (Acknowledgment Table provided in Supplementary Data). Instead of using all available
165 sequences in GISAID, we used sequences of several viruses representing SARS-CoV-2 clades
166 from some countries that have complete genome, high-coverage, and no stretches of ‘NNNN’ for
167 the phylogenetic dataset. Sequence alignment was performed using the MAFFT program server
168 for multiple nucleotide sequence alignment (https://mafft.cbrc.jp/alignment/server/). A
169 phylogenetic tree was constructed using 29,400 nucleotide length starting from the ORF1ab open
170 reading frame of SARS-CoV-2 using a maximum likelihood statistical method with 1,000
171 bootstrap replications and selected the best-fitting substitution model (GTR+G+I) for the dataset.
172 All the analyses were performed in Molecular Evolutionary Genetics Analysis version 10
173 (MEGA X) software [11]. Because the purpose of this phylogenetic analysis was to determine
174 the evolutionary relationships between our virus samples and the other SARS-CoV-2 viruses, the
175 tree was rooted to the oldest virus, hCoV-19/Wuhan/Hu-1/2019.

177 Ethical Approval
178 The Medical and Health Research Ethics Committee of the Faculty of Medicine, Public Health
179 and Nursing, Universitas Gadjah Mada/Dr. Sardjito Hospital approved this study
(KE/FK/0563/EC/2020). Written informed consent was obtained from all participants before joining in this study.

Results

Whole genome sequences of SARS-CoV-2 from the Special Region of Yogyakarta and Central Java provinces, Indonesia

All patients were classified as moderate COVID-19, except patient-4 as a mild case (Table 1 and Suppl. Table 1). The details of case presentations are described in the Supplementary Table 1. WGS revealed that the virus sample collected from patient-1 (hCoV19/Indonesia/YO-UGM-781481/2020, ID: EPI_ISL_516829), patient-2 (hCoV19/Indonesia/YO-UGM-202449/2020, ID: EPI_ISL_516800), and patient-3 belonged to the GH clade, while those from patient-4 (hCoV19/Indonesia/YO-UGM-200927/2020, ID: EPI_ISL_516806) showed the L clade. Moreover, WGS of virus from patient-1, patient-2, and patient 3 showed 9 amino acid mutations in 6 proteins, including NSP3 (P679S), NSP12 (P323L, A656S), NSP13 (M576I), spike (D614G), NS3 (A54V, Q57H, A99S), and NP (Q160R); 4 amino acid mutations in 4 proteins: NSP3 (P822L), NSP12 (P323L), Spike (D614G), and NS3 (Q57H); and 5 amino acid mutations in 5 proteins: NSP3 (P822L), NSP12 (P323L), Spike (D614G), NS3 (Q57H), and NS7a (H73Y), respectively; whereas those from patient-4 consisted of only one mutation in the NSP5 protein (M49I) (Table 1).

The genome-wide SNPs and amino acid variations of our samples are shown in Table 2 and 3, respectively (positions referred to the reference sequence: NC_045512.2). Not all SNPs cause amino acid changes in our samples.

Clade distribution of full-length genome sequences from Indonesia

Whole genome sequencing revealed that one virus (hCoV19/Indonesia/YO-202449/2020, EPI_ISL_516800) had a complete SARS-CoV-2 genome (29,903 nt). Although the other three virus samples were shorter due to incomplete UTRs at either the 5’ or 3’, they possessed full-length and complete open reading frames (ORFs) with a size of 29,409 nt consisting of 11 genes (ORF1ab, S, ORF3a, E, M, ORF6, ORF7a, ORF7b, ORF8, N, ORF10).
Next, we compared the clade distribution of full-length genome sequences from Indonesia (n=60) from March to September 2020. Based on the collection data, most (39/60, 65%) virus genomes contained the D614G mutation representing clade G (2), GR (7), and GH (30) (Fig. 1). From March to April 2020, clade L was dominant. On the other hand, there has been an increase in the detection of clade GH since April 2020 until now.

Phylogenetic analysis
Phylogenetic analysis of whole genome sequencing showed that three virus samples (EPI_ISL_525492, EPI_ISL_516800 and EPI_ISL_516829) clustered amongst viruses from the clade GH from multiple countries across Asia, Middle East, and Europe (Fig. 2). In particular to EPI_ISL_516829, this virus showing an extended branch length to the ancestral human SARS-CoV-2 virus, hCoV-19/Wuhan/Hu-1/2019. On the other hand, one virus sample (EPI_ISL_516806) was situated between clade L viruses mainly from Asia (China, Malaysia, Indonesia, India, United Arab Emirates, and Japan) (Fig. 2)

Discussion
The present study reports four full genomes of SARS-CoV-2 from patients with COVID-19 in Yogyakarta and Central Java Provinces, Indonesia. Phylogenetic analysis showed that 3 of 4 samples collected in June and August 2020 were clustered within SARS-CoV-2 viruses belonging to GH clade. One of these (EPI_ISL_516829) displayed a longer branch length compared to the other viruses indicating a greater evolutionary distance to the ancestral virus, hCoV-19/Wuhan/Hu-1/2019. This was confirmed by the molecular characterization showing that more SNPs and amino acid mutations were found within this virus genome. The phylogenetic analysis also revealed that one virus sample collected in May 2020 (EPI_ISL_516806) was in evolution closely related to the other SARS-CoV-2 viruses belonged to the L clade, including those from Indonesia which were collected in the early months of disease transmission during the pandemic between March and April 2020.
Our finding corresponds with the situation in Indonesia, showing that during the early pandemic only two clades, O and L, were detected, and the latter clade was more dominantly found from COVID-19 cases. However, since the first detection of clade GH in April 2020, this virus was more frequently detected than the previous circulating clades. Based on the data available in GISAID, 60 virus samples representing five clades have been detected from COVID-19 cases in Indonesia up to September 2020 (based on full-length genome and submission date): L (20), O (1), G (2), GR (7), and GH (30) [10]. Whether this pattern correlates with the increase in the number of COVID-19 cases recently in Indonesia has to be investigated further. Interestingly, a similar situation was found in some countries in North America [12] and Africa [13], which also detected more SARS-CoV-2 virus strains belonging to clade GH than to the other clades. An increase in SARS-CoV-2 detection conveys the D614G mutation concurrent with the recent global situation of COVID-19 [10].

The D614G mutation dominates globally approximately 77,818/96,215 (~81%) full genomes submitted at GISAID until September 18, 2020 [10]. Three of four (75%) SARS-CoV-2 in our case series also consisted of D614G. According to phylogenetic tree and sequence distribution analysis, it has been suggested that the dominating D614G globally is caused by a positive selection [5], while the dominating D614G in Europe is due to a founder effect [14]. Whether which mechanism occurs in Indonesia is difficult to conclude since only limited full genomes were submitted to GISAID until the submission date of the end of September 2020 (n=60) [10].

The virus with the D614G mutation in Indonesia was first detected in April 2020 in Surabaya, East Java [10], followed by other provinces, including Yogyakarta, Central Java, West Java and Banten. Clade L was mostly detected in Jakarta (7/20) and Surabaya, East Java (7/20), followed by Papua (3/20) (Fig. 1) [10].

It has been reported that COVID-19 patients with the D614G mutation have a higher viral load than patients infected by SARS-CoV-2 without mutations [5]. The patients with D614G had a \( C_T \) value lower than one patient without the mutation (Table 1).

Interestingly, patients infected with SARS-CoV-2 bearing D614G mutations showed moderate COVID-19, while the patient without mutations suffered from mild symptoms. These differences might be associated with the small sample size of our study (n=4) compared with previous studies (n=999 [5], 175 [15], and 88 [16]). Moreover, the severity of COVID-19 is affected by many factors, including age, sex, presence of comorbidities, and patients’ immune responses.
Further study with a larger sample size and involving risk factors for COVID-19 severity is mandatory to determine the association between the D614G mutation and the severity of COVID-19, particularly in Indonesia. Among GH clades, they also consisted of different mutations in addition to the variants that determine the clade name (Table 1). It has already been reported that the D614G variant is almost always accompanied by three other variants: a C-to-T change in the 5'UTR, a silent c.3307C>T variant, and P323L [4]. All GH clade samples in the present study also contained P323L (Table 1).

Notably, whole genome sequencing is of practical importance to determine virus variants and clades and is associated with particular geographic disseminations to decide clinical and political approaches at the regional and local levels [12]. Moreover, whether the differences in the case fatality rate and viral spread or transmission among different countries/regions are affected by differences in the virus clade [18] needs to be further studied.

Our study has some limitations, including that the samples sequenced are only 0.02% (60/290,000) of all confirmed cases in Indonesia. These facts should be considered in the interpretations of our findings, especially about epidemiological patterns (e.g. increase vs. decrease of frequency) of a particular clade in Indonesia. Another limitation of our study is we do not have any data from epidemiological tracing, nor data concerning detection (how far, how fast) to determine what is the size of the cluster where each patient belongs, or what is the transmission pattern of the cluster.

Conclusions

We report the full-genome sequence characterization and phylogenetic analysis of SARS-CoV-2 from Indonesia. SARS-CoV-2 with the D614G mutation appears to become the major circulating virus in Indonesia, which is concurrent with the COVID-19 situation worldwide. Further study with a larger sample size is necessary to investigate whether the dominating SARS-CoV-2 bearing the D614G mutation is due to a positive selection or a founder effect or some other mechanism and to explore the role of the D614G mutation in the pathogenesis and virulence of SARS-CoV-2.

Acknowledgements
We thank the Collaborator Members of the Yogyakarta-Central Java COVID-19 study group: Kurnianto (RSUP Dr. Soeradji Tirtonegoro), Indah Juliana (RSUP Dr Soeradji Tirtonegoro), Beby Dewi Sartika (RSUD Nyi Ageng Serang), Ardorisye Sapta Fornia (RS PKU Gamping), Dwiki Afandy (Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada (FK-KMK UGM)), Susan Simanjaya (FK-KMK UGM), William Widitjiarso (FK-KMK UGM), Aditya Rifiq Fauzi (FK-KMK UGM), Safitriani (PT. Pandu Biosains), Muhammad Taufiq Soekarno (PT. Pandu Biosains) and Sri Fatmawati (FK-KMK UGM). Gunadi, Marcellus, Dwi Aris Agung Nugrahaneingsih, Kristy Iskandar, Nungki Anggorowati, Alvin Santoso Kalim, Kemala Athollah, Dyah Ayu Puspitarani are members of the Genetics Working Group (Pokja Genetik), Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada. We gratefully acknowledge the authors, the Originating and Submitting Laboratories for their sequence and metadata shared through GISAID. All submitters of data may be contacted directly via www.gisaid.org. The Acknowledgments Table for GISAID is reported as Supplementary material.

References
1. World Health Organization. https://www.who.int/news-room/detail/27-04-2020-who-timeline-covid-19 Accessed on June 14, 2020.
2. Phelan AL, Katz R, Gostin LO. The novel coronavirus originating in Wuhan, China: challenges for global health governance. JAMA. 2020 323, 709–710, doi:10.1001/jama.2020.1097
3. World Health Organization. https://covid19.who.int/table Accessed on October 10, 2020.
4. Younes N, Al-Sadeq DW, Al-Jighefee H, Younes S, Al-Jamal O, Daas HI, Yassine HM, Nasrallah GK. Challenges in laboratory diagnosis of the novel Coronavirus SARS-CoV-2. Viruses. 2020; 26;12:582.
5. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi E, Bhattacharya T, Foley B, Hastie K, Parker M, Partridge D, Evans C, Freeman T, de Silva T. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 2020, 182, 812–827, doi: 10.1016/j.cell.2020.06.043
6. Nguyen TT, Pham TN, Van TD, Nguyen TT, Nguyen DTN, Le HNM, Eden JS, Rockett RJ, Nguyen TTH, Vu BTN, Tran GV, Le TV, Dwyer DE, van Doorn HR; OUCRU COVID-19 Research Group. Genetic diversity of SARS-CoV-2 and clinical, epidemiological characteristics of COVID-19 patients in Hanoi, Vietnam. PLoS One. 2020 Nov 17;15(11):e0242537.
7. Hou YJ, Chiba S, Halfmann P, Ehre C, Kuroda M, Dinnon KH 3rd, Leist SR, Schäfer A, Nakajima N, Takahashi K, Lee RE, Mascenik TM, Graham R, Edwards CE, Tse LV, Okuda K, Markmann AJ, Bartelt L, de Silva A, Margolis DM, Boucher RC, Randell SH, Suzuki T, Gralinski LE, Kawaoka Y, Baric RS. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. Science. 2020 Nov 12:eabe8499.
8. Beeching NJ, Fletcher TE, Fowler R. BMJ best practice. Coronavirus Disease 2019 (COVID-19). https://bestpractice.bmj.com/topics/en-us/3000168/prognosis Accessed on October 10, 2020.

9. https://ugene.net/wiki/display/UUOUM30/About+UGENE Accessed on August 20, 2020.

10. GISAID. https://platform.gisaid.org/epi3/cfrontend#8dc5e

11. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol. 2018;35:1547-1549.

12. Mercatelli D, Giorgi FM. Geographic and genomic distribution of SARS-CoV-2 Mutations. Front Microbiol 2020, 11, 1800, doi: 10.3389/fmicb.2020.01800

13. Wruck W, Adjaye J. Transmission of SARS-COV-2 from China to Europe and West Africa: a detailed phylogenetic analysis. bioRxiv preprint. doi: https://doi.org/10.1101/2020.10.02.323519.

14. Dearlove B, Lewitus E, Bai H, Li Y, Reeves DB, Joyce MG, Scott PT, Amare MF, Vasan S, Michael NL, Modjarrad K, Rolland M. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. Proc Natl Acad Sci USA 2020, 202008281, doi: 10.1073/pnas.2008281117

15. Wagner C, Roychoudhury P, Frazar C, Lee J, Muller N, Moncla L, Hadfield J, Hodcroft E, Pelle B, Richardson M, Behrens C, Huang M, Mathias P, Pepper G, Shrestha L, Xie H, Addetia A, Nguyen T, Rachleff V, Gautom R, Melly G, Hiatt B, Dykema P, Adler A, Brandsetter E, Han P, Fay K, Ilcsin M, Lacombe K, Sibley T, Tryong M, Wolf C, Cowgill K, Schrag S, Duchin J, Boeckh M, Englund J, Famulare M, Lutz B, Rieder M, Thompson M, Neher R, Baird G, Starita L, Chu H, Shendure J, Lindquist S, Nickerson D, Greninger A, Jerome K, Bedford T. Comparing viral load and clinical outcomes in Washington State across D614G mutation in spike protein of SARS-CoV-2. 2020; https://github.com/blab/ncov-D614G

16. Lorenz-Redondo R, Nam HH, Roberts SC, Simons LM, Jennings LJ, Qi C, Achenbach CJ, Hauser AR, Ison MG, Hulqust JF, Ozer EA. A unique clade of SARS-CoV-2 viruses is associated with lower viral loads in patient upper airways. Preprint. medRxiv 2020, 2020.05.19.20107144, doi: 10.1101/2020.05.19.20107144

17. Zou L, Dai L, Zhang Y, Fu W, Gao Y, Zhang Z, Zhang Z. Clinical characteristics and risk factors for disease severity and death in patients with Coronavirus Disease 2019 in Wuhan, China. Front Med (Lausanne). 2020;7:532.

18. Brufsky A. Distinct viral clades of SARS-CoV-2: implications for modeling of viral spread. J. Med. Virol 2020, doi: 10.1002/jmv.25902.
Figure Titles and Legends

Figure 1. Clade distribution of SARS-CoV-2 genomes in Indonesia until the submission date of September 2020, showing that 65% contained the D614G mutation. All G clades (G, GH and GR) carried the D614G mutation.

Figure 2. Phylogenetic analysis of SARS-CoV-2 genomes from Indonesia and different countries. A phylogenetic tree was constructed from 29,400 nt length of the open reading frame (ORF) of SARS-CoV-2 using the maximum likelihood statistical method with 1,000 bootstrap replications and best-fitting DNA substitution model (GTR+G+I) for the dataset. Virus samples from Yogyakarta and Central Java are indicated in red, while those from other regions in Indonesia are indicated in blue. The tree is drawn to scale (0.0001), with branch lengths measured in the number of substitutions per site.

Table List

Table 1. Characteristics of four patients with COVID-19 and SARS-CoV-2 virus samples from Yogyakarta and Central Java.

Table 2. The genome-wide SNPs of four SARS-CoV-2 virus samples from Yogyakarta and Central Java (positions referred to the reference sequence: NC_045512.2).

Table 3. The amino acid variations of four SARS-CoV-2 virus samples from Yogyakarta and Central Java (positions referred to the reference sequence: NC_045512.2).
Figure 1

Clade distribution of SARS-CoV-2 genomes in Indonesia until the submission date of September 2020, showing that 65% contained the D614G mutation. All G clades (G, GH and GR) carried the D614G mutation.
Figure 2

Phylogenetic analysis of SARS-CoV-2 genomes from Indonesia and different countries.

A phylogenetic tree was constructed from 29,400 nt length of the open reading frame (ORF) of SARS-CoV-2 using the maximum likelihood statistical method with 1,000 bootstrap replications and best-fitting DNA substitution model (GTR+G+I) for the dataset. Virus samples from Yogyakarta and Central Java are indicated in red, while those from other regions in Indonesia are indicated in blue. The tree is drawn to scale (0.0001), with branch lengths measured in the number of substitutions per site.
Manuscript to be reviewed
Table 1 (on next page)

Characteristics of four patients with COVID-19 and SARS-CoV-2 virus samples from Yogyakarta and Central Java.

* Name of SARS-CoV-2 proteins (bold) are followed by related genes (italic) and amino acid mutation indicated in bracket.

CT, cycle threshold;

Ref. sequence: hCoV-19/Wuhan/Hu-1/2019 (NC_045512.2)
Table 1. Characteristics of four patients with COVID-19 and SARS-CoV-2 virus samples from Yogyakarta and Central Java.

| Patient No | Sex  | Age (yo) | COVID-19 severity | C\textsubscript{T} value | Virus name (GI\textsubbox{S}AID Accession ID) | Average Coverage | Collection Date | Lineage (GI\textsubbox{S}AID clade) | Amino acid mutations\textsuperscript{*} (no. mutation and position of proteins-encoded genes) |
|------------|------|----------|-------------------|------------------------|---------------------------------------------|------------------|----------------|-------------------------------|-----------------------------------------------------------------------------------|
| 1          | Female | 83       | Moderate          | 16.9                   | hCoV19/Indonesia/YO-UGM-781481/2020 (EPI\_ISL\_516829) | 3748x           | 10/08/2020 | GH                           | 9: NSP\textsubbox{3}-ORF1ab (P679S), NSP12-ORF1ab (P323L, A656S), NSP13-ORF1ab (M576I), Spike-S (D614G), NS3-ORF3a (A54V, Q57H, A99S), NP-N (Q160R) |
| 2          | Male   | 77       | Moderate          | 19.7                   | hCoV19/Indonesia/YO-UGM-202449/2020 (EPI\_ISL\_516800) | 22088x          | 22/06/2020 | GH                           | 4: NSP\textsubbox{3}-ORF1ab (P822L), NSP12-ORF1ab (P323L), Spike-S (D614G), NS3-ORF3a (Q57H) |
| 3          | Female | 55       | Moderate          | 24.7                   | hCov19/Indonesia/JT-UGM-202538/2020 (EPI\_ISL\_525492) | 347x            | 26/06/2020 | GH                           | 5: NSP\textsubbox{3}-ORF1ab (P822L), NSP12-ORF1ab (P323L), Spike-S (D614G), NS3-ORF3a (Q57H), NS7a-ORF7a (H73Y) |
| 4          | Male   | 30       | Mild              | 27.9                   | hCoV19/Indonesia/YO-UGM-200927/2020 (EPI\_ISL\_516806) | 102x            | 16/05/2020 | L                            | 1: NSP\textsubbox{5}-ORF1ab (M49I)                                                  |

\textsuperscript{*} Name of SARS-CoV-2 proteins (bold) are followed by related genes (italic) and amino acid mutation indicated in bracket.

CT, cycle threshold;

Ref. sequence: hCoV-19/Wuhan/Hu-1/2019 (NC\_045512.2)
**Table 2** (on next page)

The genome-wide SNPs of four SARS-CoV-2 virus samples from Yogyakarta and Central Java (positions referred to the reference sequence: NC_045512.2).

Name of SARS-CoV-2 proteins (bold) are followed by related genes (italic) and nucleotide variations are colour-shaded

SNPs are shown according to nucleotide positions starting from 5'-UTR (untranslated region) of SARS-CoV-2 genome
Table 2. The genome-wide SNPs of four SARS-CoV-2 virus samples from Yogyakarta and Central Java (positions referred to the reference sequence: NC_045512.2)

| No | Virus ID                                                                 | 5' UTR       | NSP3-ORF1ab | NSP5-ORF1ab   | NSP12-ORF1ab | NSP13-ORF1ab | NSP14-ORF1ab | Spike-S | NS3-ORF3a | Matri x-M | NP-N |
|----|--------------------------------------------------------------------------|--------------|-------------|---------------|--------------|--------------|--------------|----------|-----------|-----------|------|
| 1  | hCoV19/Wuhan/Hu-1/2019 (NC_045512.2)                                     | A C C T C C  | G C G C C C  | G G C G C A C | G G C A C G G | G G C A C G G | G G C A C G G |         |           |           |      |
| 2  | hCoV19/Indonesia/YO-UGM-781481/2020 (EPI ISL_516829)                      | A T T C T C  | G C G T T T  | T T C C T C   | G C T G T C   | G C T G T C   | G C T G T C   |         |           |           |      |
| 3  | hCoV19/Indonesia/YO-UGM-202449/2020 (EPI ISL_516800)                      | A T T T C T  | G T G C T C  | G G T T G C   | G T G C T G   | G T G C T G   | G T G C T G   |         |           |           |      |
| 4  | hCoV19/Indonesia/JT-UGM-202538/2020 (EPI ISL_525492)                      | G T T T C T  | G T T C T C  | G G T T G C   | G T G C T G   | G T G C T G   | G T G C T G   |         |           |           |      |
| 5  | hCoV19/Indonesia/YO-UGM-200927/2020 (EPI ISL_516806)                      | A C C T C C  | G C C C C G  | G C C A C G   | G C C A C G   | G C C A C G   | G C C A C G   |         |           |           |      |

Name of SARS-CoV-2 proteins (bold) are followed by related genes (italic) and nucleotide variations are colour-shaded. SNPs are shown according to nucleotide positions starting from 5' UTR (untranslated region) of SARS-CoV-2 genome.
**Table 3** (on next page)

The amino acid variations of four SARS-CoV-2 virus samples from Yogyakarta and Central Java (positions referred to the reference sequence: NC_045512.2).

Name of SARS-CoV-2 proteins (bold) are followed by related genes (italic) and amino acid variations are colour-shaded

Amino acid variations are shown starting from translated region of each protein.
**Table 3.** The amino acid variations of four SARS-CoV-2 virus samples from Yogyakarta and Central Java (positions referred to the reference sequence: NC_045512.2).

| No | Virus ID                                                                 | NSP3-ORF1ab | NSP5-ORF1ab | NSP12-ORF1ab | NSP13-ORF1ab | Spike-S | NS3-ORF3a | Matrix-M | NS7a-ORF7a | NP-N |
|----|--------------------------------------------------------------------------|-------------|-------------|--------------|--------------|---------|-----------|----------|------------|------|
| 1  | hCoV-19/Wuhan/Hu-1/2019 (NC_045512.2)                                     | F           | D           | P            | M            | 1       | D         | A        | Y          | L    |
|    |                                                                          | 1           | 0           | 7            | 7            | 2       | 4         | 9        | 1          | 5    |
| 2  | hCoV19/Indonesia/YO-UGM-781481/2020 (EPI_ISL_516829)                      | F           | D           | S            | P            | M       | N         | L        | D          | L    |
|    |                                                                          | 1           | 0           | 7            | 7            | 2       | 4         | 8        | 2          | 3    |
| 3  | hCoV19/Indonesia/YO-UGM-202449/2020 (EPI_ISL_516800)                      | F           | D           | P            | L            | M       | N         | L        | D          | D    |
|    |                                                                          | 1           | 0           | 7            | 7            | 2       | 4         | 3        | 2          | 3    |
| 4  | hCoV19/Indonesia/JT-UGM-202538/2020 (EPI_ISL_525492)                      | F           | D           | P            | L            | M       | N         | L        | D          | D    |
|    |                                                                          | 1           | 0           | 7            | 7            | 2       | 4         | 8        | 2          | 3    |
| 5  | hCoV19/Indonesia/YO-UGM-200927/2020 (EPI_ISL_516806)                      | F           | D           | P            | I            | N       | L         | D        | P          | A    |
|    |                                                                          | 1           | 0           | 7            | 7            | 2       | 4         | 3        | 2          | 1    |

Amino acid variations are shown starting from translated region of each protein.

Name of SARS-CoV-2 proteins (bold) are followed by related genes (italic) and amino acid variations are colour-shaded.