



# The dynamics of circulating SARS-CoV-2 lineages in Bogor and surrounding areas reflect variant shifting during the first and second waves of COVID-19 in Indonesia

Anggia Prasetyoputri<sup>1,\*</sup>, Anik B. Dharmayanthi<sup>2</sup>, Syam B. Iryanto<sup>3</sup>, Ade Andriani<sup>1</sup>, Isa Nuryana<sup>1</sup>, Andri Wardiana<sup>1</sup>, Asep M. Ridwanuloh<sup>1</sup>, Sri Swasthikawati<sup>1</sup>, Hariyatun Hariyatun<sup>1</sup>, Herjuno A. Nugroho<sup>2</sup>, Idris Idris<sup>2</sup>, Indriawati Indriawati<sup>1</sup>, Zahra Noviana<sup>2</sup>, Listiana Oktavia<sup>4</sup>, Yuliawati Yuliawati<sup>1</sup>, Masrukhin Masrukhin<sup>2</sup>, Erwin F. Hasrianda<sup>2</sup>, Linda Sukmarini<sup>1</sup>, Fahrurrozi Fahrurrozi<sup>1</sup>, Nova Dilla Yanthi<sup>1</sup>, Alfi T. Fathurahman<sup>1</sup>, Ari S. Wulandari<sup>1</sup>, Ruby Setiawan<sup>2</sup>, Syaiful Rizal<sup>2</sup>, Ahmad Fathoni<sup>1</sup>, Wien Kusharyoto<sup>1</sup>, Puspita Lisdiyanti<sup>1</sup>, Ratih A. Ningrum<sup>1</sup> and Sugiyono Saputra<sup>2,\*</sup>

<sup>1</sup> Research Center for Biotechnology, National Research and Innovation Agency (BRIN), Bogor, West Java, Indonesia

<sup>2</sup> Research Center for Biology, National Research and Innovation Agency (BRIN), Bogor, West Java, Indonesia

<sup>3</sup> Research Center for Informatics, National Research and Innovation Agency (BRIN), Bogor, West Java, Indonesia

<sup>4</sup> Research Center for Chemistry, National Research and Innovation Agency (BRIN), Bogor, West Java, Indonesia

\* These authors contributed equally to this work.

## ABSTRACT

**Background.** Indonesia is one of the Southeast Asian countries with high case numbers of COVID-19 with up to 4.2 million confirmed cases by 29 October 2021. Understanding the genome of SARS-CoV-2 is crucial for delivering public health intervention as certain variants may have different attributes that can potentially affect their transmissibility, as well as the performance of diagnostics, vaccines, and therapeutics.

**Objectives.** We aimed to investigate the dynamics of circulating SARS-CoV-2 variants over a 15-month period in Bogor and its surrounding areas in correlation with the first and second wave of COVID-19 in Indonesia.

**Methods.** Nasopharyngeal and oropharyngeal swab samples collected from suspected patients from Bogor, Jakarta and Tangerang were confirmed for SARS-CoV-2 infection with RT-PCR. RNA samples of those confirmed patients were subjected to whole genome sequencing using the ARTIC Network protocol and sequencer platform from Oxford Nanopore Technologies (ONT).

**Results.** We successfully identified 16 lineages and six clades out of 202 samples (male  $n = 116$ , female  $n = 86$ ). Genome analysis revealed that Indonesian lineage B.1.466.2 dominated during the first wave ( $n = 48$ , 23.8%) while Delta variants (AY.23, AY.24, AY.39, AY.42, AY.43 dan AY.79) were dominant during the second wave ( $n = 53$ ,

Submitted 11 November 2021

Accepted 26 February 2022

Published 22 March 2022

Corresponding author

Sugiyono Saputra,  
sugiyono.saputra@brin.go.id

Academic editor

Joël Mossong

Additional Information and  
Declarations can be found on  
page 13

DOI 10.7717/peerj.13132

© Copyright

2022 Prasetyoputri et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

26.2%) following the highest number of confirmed cases in Indonesia. In the spike protein gene, S\_D614G and S\_P681R changes were dominant in both B.1.466.2 and Delta variants, while N439K was only observed in B.1.466.2 ( $n = 44$ ) and B.1.470 ( $n = 1$ ). Additionally, the S\_T19R, S\_E156G, S\_F157del, S\_R158del, S\_L452R, S\_T478K, S\_D950N and S\_V1264L changes were only detected in Delta variants, consistent with those changes being characteristic of Delta variants in general.

**Conclusions.** We demonstrated a shift in SARS-CoV-2 variants from the first wave of COVID-19 to Delta variants in the second wave, during which the number of confirmed cases surpassed those in the first wave of COVID-19 pandemic. Higher proportion of unique mutations detected in Delta variants compared to the first wave variants indicated potential mutational effects on viral transmissibility that correlated with a higher incidence of confirmed cases. Genomic surveillance of circulating variants, especially those with higher transmissibility, should be continuously conducted to rapidly inform decision making and support outbreak preparedness, prevention, and public health response.

**Subjects** Molecular Biology, Virology, Epidemiology, Infectious Diseases, COVID-19

**Keywords** SARS-CoV-2, COVID-19, Indonesian lineages, Bogor, Variant shifting, Nanopore, Genomic surveillance, Delta variants

## INTRODUCTION

First detected in Wuhan, China in December 2019, a novel coronavirus named as the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is the causative agent of Coronavirus Disease 2019 (COVID-19). COVID-19 cases rapidly increased to reach an epidemic scale, and was therefore announced as a Public Health Emergency of International Concern (PHEIC) on January 31st, 2020 and subsequently characterized as a pandemic on March 11th, 2020 ([WHO, 2020](#)). This status indicated that multiple countries may pose risks and an international response is required ([Li et al., 2020b](#)). According to the World Health Organization (WHO) ([WHO, 2021b](#)), a total of 246 million COVID-19 cases have been detected worldwide (as of 25 October 2021) with approximately 2.1% mortality rate. Since the first case was reported in March 2020, Indonesia had reached 1 million cases by late January 2021 with a mortality rate of 2.7% ([BNPB, 2021](#)). In July 2021, Indonesia experienced the second wave, which reached the highest peak since the beginning of the pandemic. More than 56,000 cases in one day were then recorded, with a cumulative number of 4.2 million cases by end of October 2021 ([COVID-19 S, 2021](#)). The two provinces with the highest cases were DKI Jakarta and West Java, which accounted for 20% and 16% of national cases, respectively ([COVID-19 S, 2021](#)).

Genetic characterization of SARS-CoV-2 is critical to enable monitoring of viral evolution and its epidemiology. Information on its circulating variants is useful to determine potential markers of transmissibility, altered antigenicity or severity of disease. The role of sequence data availability will become increasingly important as SARS-CoV-2 vaccines and antivirals become available, as it informs on circulating variants that can be

inhibited by the vaccines and the possible emergence of antiviral resistance (*Nasir et al., 2020; Takenouchi et al., 2021*).

In October 2021, there are more than 4.7 million SARS-CoV-2 published sequences worldwide through the Global Initiative on Sharing All Influenza Data (GISAID) and those have been classified into nine large clades (*GISAID, 2021*). There were five major lineages (A, B, B.1, B.1.1, and B.1.177) that have been proposed by Phylogenetic Assignment of Named Global Outbreak LINEages (PANGOLIN) (*Rambaut et al., 2020*) but it has since been expanded into numerous lineages circulating globally. Furthermore, WHO has labeled several variants that pose an increased risk to global public health with Variants of Interest (VOIs) and Variants of Concern (VOCs). In this study, we aimed to identify circulating variants of SARS-CoV-2 in Bogor, West Java and surrounding areas as those cities also experienced massive waves of COVID-19 cases during the pandemic.

## MATERIALS & METHODS

### Ethical clearance/statement

The study protocol was reviewed and approved by the Health Research Ethics Committee, University of Indonesia and Cipto Mangunkusumo Hospital (HREC-FMUI/CMH) (20-10-1321\_EXP). All oropharyngeal and nasopharyngeal swab samples used in this study are those accompanied by written informed consents signed by the patients, agreeing to donate their samples for research purposes.

### Sample collection, RNA extraction and qRT-PCR

The study was carried out at Biosafety Level-3 and Biosafety Level-2 Laboratory, Cibinong Science Center, National Research and Innovation Agency (BRIN; formerly known as Indonesian Institute of Sciences or LIPI). Samples were oropharyngeal and nasopharyngeal swabs obtained from various clinics, hospitals, and diagnostic laboratories in Bogor (Bogor City and Bogor Regency), Jakarta and Tangerang (Tangerang and Tangerang Selatan). Bogor and Tangerang have a close proximity to Jakarta and a great proportion of its residents commute to work in Jakarta, the capital city of Indonesia. Nasopharyngeal and oropharyngeal swab samples were collected between May 2020 and August 2021, which were sent to BSL-3 and BSL-2 Lab for qRT-PCR confirmation of SARS-CoV-2 infection. Viral RNA was extracted using Viral Nucleic Acid Extraction (Geneaid Biotech Ltd., Taiwan) according to the manufacturer's instructions. Amplification analysis was conducted using CE IVD real time PCR kit (Real Q 2019-nCoV Detection Kit Biosewoom, Korea) with envelope (E) and RNA dependent RNA polymerase (RdRp) as targeted genes and human RNase P as internal control. The kit cut-offs were set at cycle threshold (Ct)  $\leq 38$  (E and RdRp) and  $\leq 35$  (RNase P) with limit of detection (LOD) sensitivity of 6.25 copies/ $\mu$ L for sputum and 3.125 copies/ $\mu$ L for swab in CFX 96 BioRad RT-PCR instrument. The qRT-PCR conditions were as follows: 50 °C for 30 min (1 cycle), 95 °C for 15 min (1 cycle), 95 °C for 15 s and 62 °C for 45 s (40 cycles).

### Library preparation

SARS-CoV-2 viral RNA from swab specimens confirmed by qRT-PCR with Ct values of 11-30 were selected ([File S1](#)). Following a decrease of COVID-19 cases after August 2021,

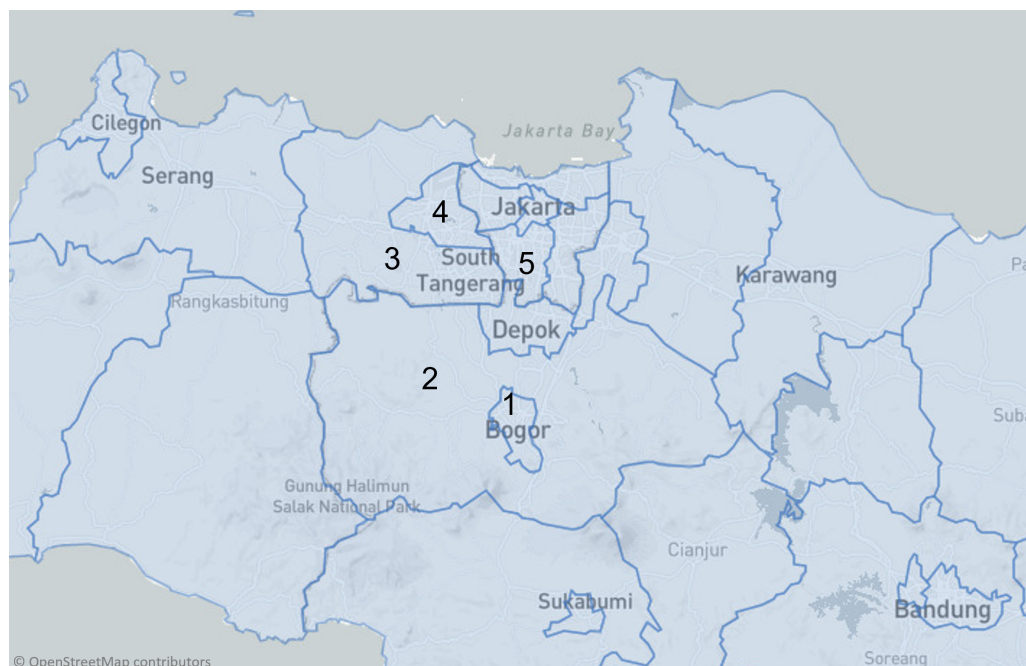
many confirmed positive samples were not qualified to be subjected for whole genome sequencing due to higher Ct values. Samples with Ct values <15 were diluted 10 times with nuclease-free water. Library preparation was performed according to the ARTIC nCoV-2019 sequencing protocol V2 ([Quick, 2020](#)) and then V3 after protocol update was released. Briefly, the complementary DNA (cDNA) was synthesized using SuperScript IV Reverse Transcriptase and either random hexamers (NEB) (V2) or LunaScript® RT SuperMix Kit (V3). cDNA was amplified by touch-down PCR (the annealing temperature is decreased by 0.1 °C after every cycle) using nCoV-2019/V3 primers ([Itokawa et al., 2020](#)) (IDT) to generate overlapping 400 base pair (bp) amplicons covering all the ~30 Kbp SARS-CoV-2 genomes. A modification had been applied in V3 protocol by adding a primer pair that covers the #74 amplicon (0.3 µl, final concentration 2.5 pmol) in pool B reaction. Subsequently, crude PCR amplicons were cleaned up using AMPure XP beads (Beckman) in a 1:1 ratio. The purified amplicons were then quantified using the Qubit High Sensitivity double strand DNA (dsDNA) assay kit and measured on a Qubit 2.0 instrument (ThermoFisher Scientific). A total input DNA of 50 ng per sample was applied to a one-pot native barcoding reaction and subsequently was subjected to an end prep reaction using NEBNext Ultra II End Prep kit (NEB), followed by native barcoding using NB Expansion Kit 1-24 (Oxford Nanopore Technologies, ONT). Up to 23 samples (including a negative control) were barcoded and sequenced in any single run. All barcoded amplicons were then pooled and subjected to AMPure XP beads purification and Qubit quantification. Adapter ligation was then performed using AMII adapter mix (ONT) and T4 DNA ligase (NEB). Following another AMPure XP beads clean up and Qubit quantification, the final library was put on ice until use. Approximately 15 ng of final library were loaded onto an R9.4.1 flow cell mounted on a MinION Mk1b/Mk1c or PromethION sequencer (ONT).

### Sequencing and bioinformatics analysis

Sequencing run was performed using MinKNOW (v20.06.4) with fast basecalling mode and the sequencing output was monitored in real time using RAMPART (v2.1.0) software package. The run was terminated when the coverage of sequences achieved >99% compared to the Wuhan-Hu-1 reference genome ([MN908947.3](#)). A complete bioinformatics ARTIC v1.1.0 ([Loman et al., 2020](#)) was used to generate consensus genome sequences of SARS-CoV-2 with default parameters. Guppy (v 4.0.14, ONT) with a high-accuracy mode was used to perform basecalling to obtain fastq files and classified the basecalled reads based on both barcodes. The reads were then aligned to the Wuhan-Hu-1 reference genome (<https://www.ncbi.nlm.nih.gov/nuccore/MN908947.3>) using minimap2 (v 2.10-r761) ([Li, 2018](#)) and trimmed to remove the primer sequences from the end-read alignments. The filtered reads were then used to call the variant and build a consensus SARS-CoV-2 sequence using Medaka (v 1.0.3) workflows (ONT) and bcftools (v 1.10.2) ([Li, 2011](#)), respectively.

### Lineage, variant calling and phylogenetic analysis

Lineages of 202 SARS-CoV-2 genomes were identified using PANGOLIN v3.1.16 with designated pangoleARN version v1.2.105 ([O'Toole et al., 2020](#); [O'Toole et al., 2021](#)). The phylogenomic tree (Maximum Likelihood; GTR+F+I evolutionary model) was build



**Figure 1** A map of locations from which samples were obtained in Bogor and surrounding areas. Samples were obtained from laboratories located in Bogor (1 and 2), Tangerang (3 and 4) and Jakarta (5), and subsequently subjected to qRT-PCR confirmation and whole genome sequencing at the Biosafety Laboratory Level 2 and 3 BRIN, Cibinong Science Center-Botanical Garden.

Full-size DOI: 10.7717/peerj.13132/fig-1

using Augur utility ([github.com/Nextstrain/augur](https://github.com/Nextstrain/augur)) from the Nextstrain bioinformatic pipeline (*Hadfield et al., 2018*), employing MAFFT (*Katoh & Standley, 2016*) and IQ-TREE (*Nguyen et al., 2015*). The ultrafast bootstrap (UFBoot) (*Minh, Nguyen & Von Haeseler, 2013*) was performed and evaluated with 1,000 bootstrap replications. Unless mentioned otherwise, all plots were constructed using R v 4.1.2 (*Team RC, 2021*) as uploaded to GitHub (<https://github.com/zahranoviana/VenomCov-Figures/tree/main>).

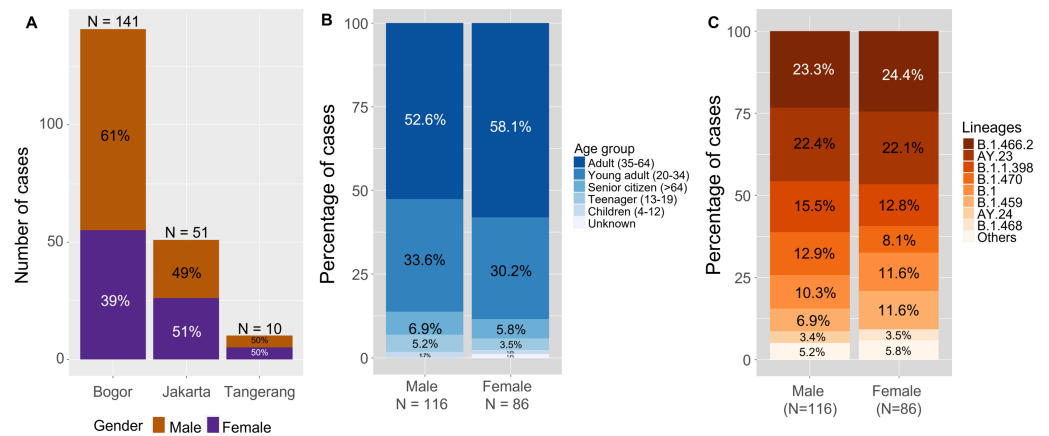
## RESULTS

### Patient demographics

Clinical samples ( $n = 202$ ) were collected between May 2020 and August 2021 from various hospitals, diagnostic laboratories, and healthcare facilities in Bogor and its surrounding areas (Bogor  $n = 141$ , Jakarta  $n = 51$  and Tangerang  $n = 10$ ) as shown in [Figs. 1](#) and [2A](#). Patients were male (57%) and female (43%) with the dominant age group being adults (35–64 years old; 52.6% and 58.1%), followed by young adults (20–34 years old; 33.6% and 30.2%). ([Figs. 2A–2B](#)).

### Variations of lineages

The SARS-CoV-2 genome was amplified using nCoV-2019/V3 primers resulting in 400 bp amplicons that covers 28,999–29,782 bp. All viral genomes have been published in GISAID ([File S1](#)). Using PANGOLIN phylogenetic classification (*Rambaut et al., 2020*),



**Figure 2 Summary of the patients' profiles.** Composition of males and females are shown across sampling locations (A), age groups (B) and lineage distribution (C) from the 202 SARS-CoV-2 genomes sequenced. Others in male lineages (5.2%, 6/116) include B.56 ( $n = 2$ ), B.1.36.19, AY.39, AY.42 and AY.79; while others in female lineages (5.8%, 5/86) include B.1.1, B.1.1.53, B.1.1.10, AY.43 and unclassified. The samples were mainly obtained from originating laboratory in Bogor (69.8%). Over 50% of samples in both males and females were from patients aged 35–64 years. A higher proportion of lineages B.1.1.398 and B.1.470 were observed in male patients compared to females, whereas lineages B.1 and B.1.459 were more prevalent in females compared to males.

Full-size DOI: [10.7717/peerj.13132/fig-2](https://doi.org/10.7717/peerj.13132/fig-2)

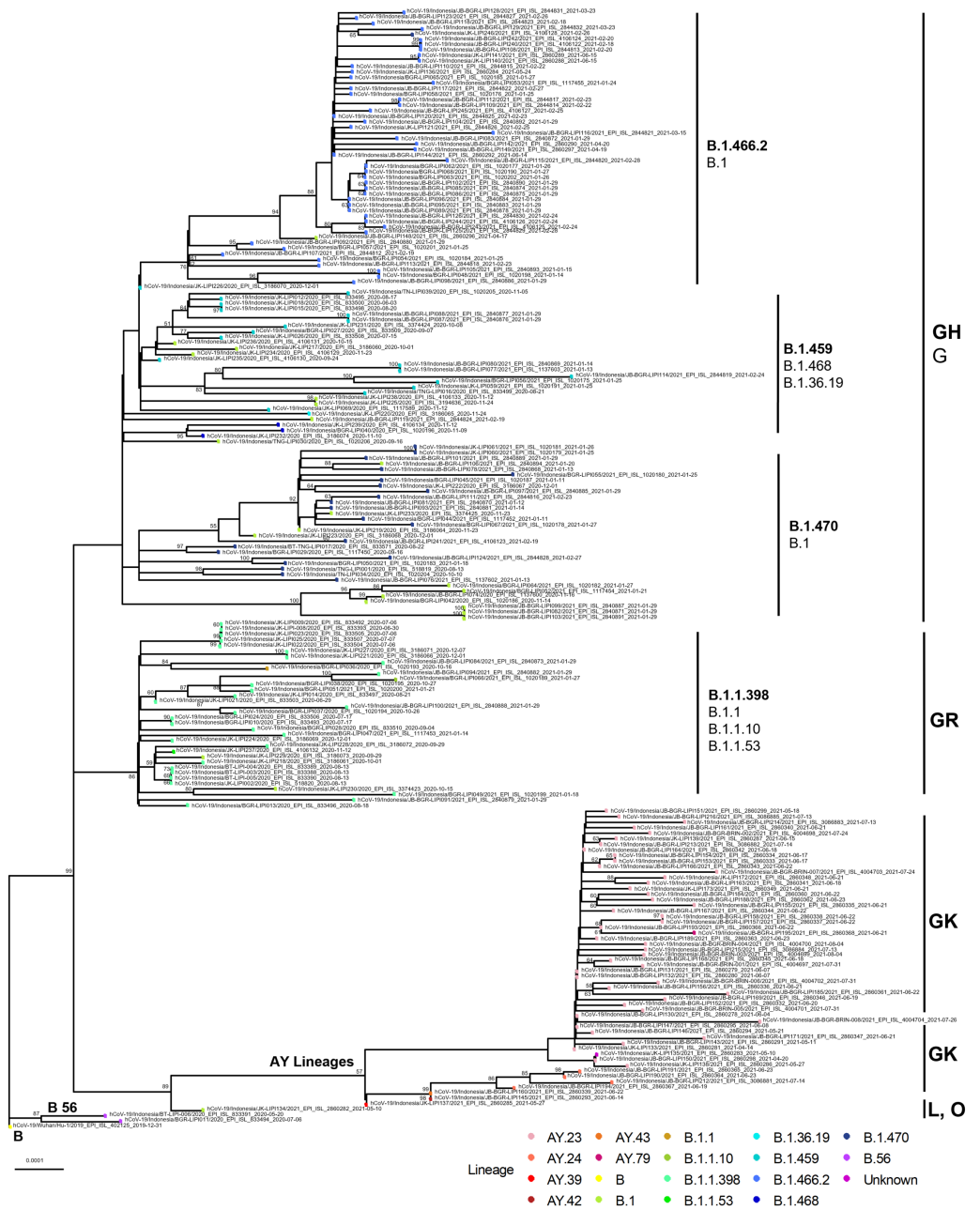
the 201 sequenced genomes were determined to belong to 17 lineages. One virus (hCoV-19/Indonesia/JK-LIPI135/2021) was categorized as unclassified based on PANGOLIN but clustered together with Delta variants in clade GK by GISAID (stretches of NNNs 3.91% of overall sequences). We found that the majority of genomes of samples collected belonged to Delta variants or B.1.617.2 descendant ( $n = 49$ , 24.3%); Indonesian lineage B.1.466.2 ( $n = 48$ , 23.8%), B.1.1.398 ( $n = 29$ , 14.4%), B.1.470 ( $n = 22$ , 10.9%), B.1.459 ( $n = 18$ , 9.9%); European lineage B.1 ( $n = 22$ , 10.9%). In this study, the Delta variants include the lineage from AY.23 ( $n = 41$ , 20.3%), AY.24 ( $n = 4$ , 2%) as well as one virus of AY.39, AY.42, AY.43 and AY.79 lineage. The phylogenetic tree of all viruses is presented in Fig. 3.

### Variations of clades

According to the GISAID clade nomenclature, the samples in this study could be classified into GH ( $n = 105$ , 52%), GK ( $n = 54$ , 26.7%), GR ( $n = 32$ , 15.8%), G ( $n = 6$ , 3%), O ( $n = 3$ , 3%) and L ( $n = 2$ , 1%). The L clade is an early marker belonging to the MN908947.3 reference sequence and designated for B.56 in this study, while GR and GH clades were previously known as the G clade marked with an S\_D614G change that mainly consists of several lineages in the first wave. The GK clade consists of lineages from Delta variants that were dominant in the second wave.

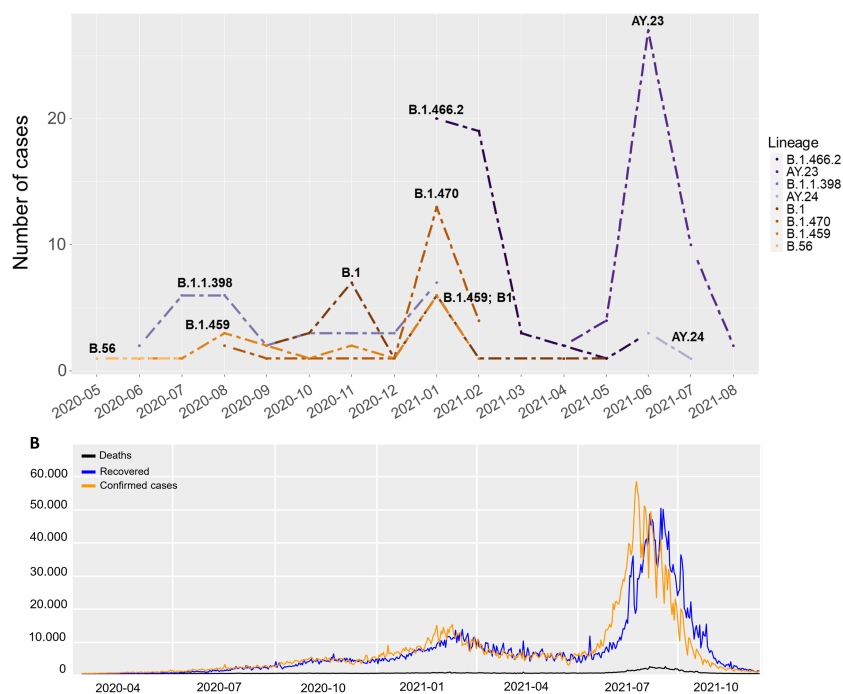
### Distribution of lineages

Overall, the lineage distribution did not show a great variation across gender. Exceptions were noted for B.1.1.398 and B.1.470 that was detected in a higher proportion in males (15.5% and 12.9%) compared to females (12.8% and 8.1%), while B.1 and B.1.459 were respectively found higher in females (11.6% and 11.6%) compared to males (10.3% and



**Figure 3** Phylogenetic tree from 202 viruses, representing 17 lineages and six GISAID clades. The clustering of certain lineages is indicative of the extent of genomic variations within the genomes. Three major clades were observed, including GH, GR and GK, while the minor clades were L, O and G. Clades GH and GR include several Indonesian lineages, while clade GK consists of Delta variants (AY lineages). Phylogenetic tree was built using Augur utility from the Nexstrain bioinformatic pipeline by employing MAFFT, IQ-TREE and 1000 bootstrap replications (UFBoot).

Full-size DOI: 10.7717/peerj.13132/fig-3



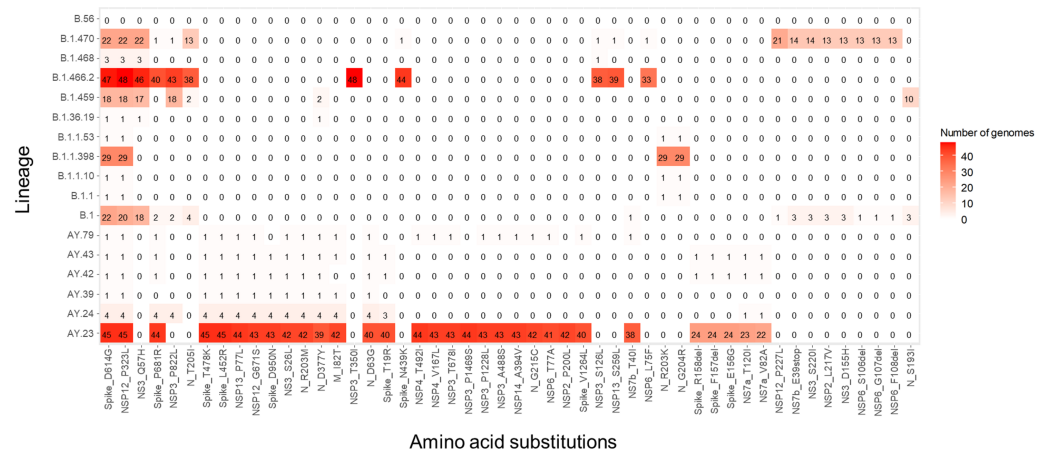
**Figure 4** Lineage distribution of SARS-CoV-2 genomes compared to the national statistics. Side-by-side comparison of lineage distribution is shown across dates of sampling (A) and the official statistics of deaths, recovered and new confirmed cases of COVID-19 (B) at national level (<https://covid19.go.id/peta-sebaran>), demonstrates two peaks at similar timelines, representing the first and second wave of pandemic. Full-size [DOI: 10.7717/peerj.13132/fig-4](https://doi.org/10.7717/peerj.13132/fig-4)

6.9%) (Fig. 2C). We found that the majority of genomes during the first wave (date of collection before 16 May 2021,  $n = 147$ ) was dominated by B.1.466.2 ( $n = 44$ , 29.9%). We used 15 May 2021 as the cut-off date between the two waves because the lowest number of new confirmed cases since the beginning of pandemic was recorded on that day, before the number of cases started to increase in the following day. On the other hand, Delta variants were dominant amongst all identified genomes sequenced during the second wave (date of collection 16 May 2021 onwards,  $n = 53$ ). Those samples mainly consisted of AY.23 ( $n = 41$ , 77.4%) and AY.24 ( $n = 4$ , 7.5%) which were also described as predominantly Indonesia ([https://cov-lineages.org/lineage\\_list.html](https://cov-lineages.org/lineage_list.html)). The lineage distribution during the first and the second waves of COVID-19 is shown in Fig. 4A, compared to the official statistics of deaths, recovered and new confirmed cases of COVID-19 in Indonesia (Fig. 4B).

### Key amino acid changes

Notably, a total 199 out of 202 viruses (99.9%) that we sequenced carried the S\_D614G and NSP12\_P323L non-synonymous mutations. The three genomes that were not found to carry those mutations included two belonging to the lineage B.56 and one from the lineage B.1.416.2, which showed a high N content in the region encoding the spike protein (hCoV-19/Indonesia/JP-BGR-LIPI242/2021, stretches of NNNs 20.66% of overall sequence). Unique changes were found in Indonesian lineages belonging to B.1.1.398 (N\_G204R and N\_R203K), B.1.470 (NSP\_12\_P227L) and B.1.466.2 (S\_N439K and NSP6\_L75F) lineages.





**Figure 5** Amino acid substitutions among lineages identified in this study. Only the top 50 substitutions with the highest number of incidences are shown. Several unique substitutions were observed in Delta variants and other variants circulating during the first wave, with the Delta variants having the majority of amino acid substitutions. The highest incidence of substitutions includes S<sub>D614G</sub> ( $n = 199$ ) and NSP12\_P323L ( $n = 197$ ). Two viruses from B.56 lineages were found to carry the least number of mutations ( $n = 4$ ) and no S<sub>D614G</sub> substitution. One virus (hCoV-19/Indonesia/JK-LIP135/2021) not shown in this figure was found to carry multiple amino acid changes, including S<sub>D614G</sub>, S<sub>P681R</sub>, S<sub>T478K</sub>, S<sub>L452R</sub>, NSP6\_T77A, NSP3\_P1228L, NSP12\_P323L, NSP12\_G671S, and NSP14\_A394V.

Full-size [DOI: 10.7717/peerj.13132/fig-5](https://doi.org/10.7717/peerj.13132/fig-5)

A summary of amino acid substitutions is presented in Fig. 5. We also observed that Delta variants carried various amino acid changes that differ from the first wave, which included changes within the spike protein, namely S<sub>T19R</sub>, S<sub>G142D</sub>, S<sub>E156G</sub>, S<sub>F157del</sub>, S<sub>R158del</sub>, S<sub>L452R</sub>, S<sub>T478K</sub>, and S<sub>D950N</sub>. Additionally, amino acid changes were also found in 22 non-spike protein-encoding genes, namely M, N, NS7a, NS7b, NSP12, NSP13, NSP14, NSP2, NSP3, NSP4 and NSP6S.

## DISCUSSION

Whole genome sequencing of SARS-CoV-2 isolates for genomic surveillance has been instrumental in understanding the virus evolution and the dynamics of transmission, as well as monitoring the emergence of mutations (O'Toole *et al.*, 2021; Hadfield *et al.*, 2018; Katoh & Standley, 2016). Such information is critical in aiding public health intervention, the development of vaccines and antivirals, including determining the efficacy of vaccines. This study, which sought to genetically characterize a regional collection of clinical samples from COVID-19 patients using Nanopore sequencer, led to two main findings. First, we observed a shift in the lineages of circulating SARS-CoV-2 in Bogor and surrounding areas, where the Indonesian lineages were dominant during the first wave of pandemic while the second wave was dominated by the Delta variants. Second, we observed unique amino acid substitutions in Delta variants that were different from those found in the first wave, indicating genotypic attributes that may have accounted for higher transmissibility during the second wave of COVID-19, coinciding with the highest record of confirmed cases in provincial and national level at the time.

Like in any other parts of the world, a shift of SARS-CoV-2 variants in a certain geographical area is part of a continuous evolution of the virus to adapt to their environment. Subsequently, new variants may emerge over time as viruses acquire mutations within their genomes, which may improve their fitness and transmissibility ([Harvey et al., 2021](#)). Introduction from other areas or imported cases *via* travelers are also likely to contribute to a high proportion of total COVID-19 cases in many countries ([Russell et al., 2021](#)) and subsequently taking over other existing variants, as shown by this study for certain areas in Indonesia. Depending on which genomic changes are more advantageous, competition between different variants may exist and thus certain variants could outnumber others. Additionally, improved host immunity could also affect the proportion of circulating variants. Our study indicated that a variant shift occurred during the two waves in Indonesia, as reflected in the changes in variant dominance. Characteristics of the Delta variants that are more favorable for their survival in the host and transmission could have greatly influenced the dynamics of the different circulating variants between the two waves, resulting in the Delta variants' dominance in the second wave.

During the first wave of COVID-19 pandemic, up to 15 May 2021, four lineages were predominant in Indonesia, including B.1.466.2, B.1.1.398, B.1.470, and B.1.459. Lineage B.1.466.2 was found in a higher proportion compared to other lineages, accounting for 40% of total viruses published by Indonesian Consortium of SARS-CoV-2 Genomic Surveillance in GISAID ( $n = 3,732$ ) ([Initiative G, 2021](#)) during the first wave of COVID-19. According to the variant graph at Regeneron COVID-19 dashboard ([Regeneron, 2021](#)), B.1.466.2 was slowly increasing since November 2020 and reaching the highest peak at the end of March 2021 with 72% of total genomes submitted to GISAID. As such, WHO has labeled this lineage as Variant Under Monitoring on 28 April 2021 ([WHO, 2021a](#)) due to some indication that it potentially has higher transmissibility and poses a future public health risk. During this first wave, confirmed new cases at national level peaked up to 14,518 on 30 January 2021 and then slowly declined to the lowest confirmed cases (2,385 cases) recorded on 15 May 2021 ([COVID-19 S, 2021](#)).

The presence of the S\_N439K and S\_P681R substitutions in the Indonesian lineages are important to note, as these mutations have been shown to confer certain advantages for the variants. The S\_N439K mutation has been associated with a similar level of fitness and virulence to the wild-type virus, whilst potentially conferring a stronger binding affinity to human ACE2 receptor and allowing immune escape from antibody-mediated immunity ([Thomson et al., 2021](#)). A study employing molecular dynamic simulation further confirmed those attributes and also provided potential mechanisms by which the stronger hACE2 binding affinity and the reduction in efficacy of neutralizing antibodies occur ([Zhou et al., 2021](#)). A distinctive feature of the Delta variant, the S\_P681R mutation within the S protein has been correlated to higher fusogenicity, which in turn led to viruses that exhibit enhanced pathogenicity *in vivo* and greater resistance to neutralizing antibodies compared to the parental virus without this mutation ([Saito et al., 2021](#)). Similarly, increased infection due to enhanced cleavage of the S protein has also been associated with the presence of S\_P681R mutation, leading to improved viral replication efficiency of Delta variant

compared to the Alpha variant ([Liu et al., 2021](#)). However, functional assays showed that the S\_P681R substitution on its own is not enough to promote enhanced infectivity and transmissibility associated with variants bearing this mutation ([Lubinski et al., 2021](#)). Another substitution at the same position, S\_P681H, has also been shown to enhance cleavage of S protein, but did not necessarily result in improved viral entry ([Lubinski et al., 2022](#)). Collectively, these studies highlight the importance of these mutations and their potential roles in facilitating enhanced transmission of the Delta variant, which could explain the eventual dominance of the Delta variant in the second wave.

A change in proportion of dominating variants was apparent as the second wave occurred. The Delta variants, including their ancestor B.1.617.2 and AY lineages, began to surpass the Indonesian lineages as they reached 53% by the end of May 2021 and remained as dominant variants towards the end of August—reaching up to 85% of the total genomes from Indonesia submitted to GISAID (June–August 2021,  $n = 4,189$ ) ([GISAID Initiative, 2021](#)). These variants accounted for 91% of total viruses submitted to GISAID with collection date from June to August 2021. According to the officials, a nation-wide record during this second wave amounted to 56,757 new cases on 15 July 2021, followed by a rapid decline to only under 2,000 cases within two months ([COVID-19 S, 2021](#)). The dynamics of COVID-19 cases at the provincial level, West Java ([COVID-19 PIK, 2021](#)), showed a similar pattern as described at the national level ([COVID-19 S, 2021](#)) for both the first and the second waves of pandemic.

Variations in key mutations also followed the shift in dominant variants. For Delta variants, we observed 12 key amino acid changes in spike protein that were mostly similar to other studies, including S\_T19R, S\_T95I, S\_G142D, S\_E156-, S\_F157-, S\_R158G, S\_K417N, S\_L452R, S\_T478K, S\_D614G, S\_P681R, and S\_D950N ([CDC, 2021](#)). There is evidence that some of these genetic changes are linked to increased transmissibility ([Allen et al., 2021](#)) and reduced effectiveness of vaccines and monoclonal antibodies ([FDA TUSFaDA, 2021](#); [Deng et al., 2021](#)). On the other hand, substitutions in the region encoding the spike protein in Indonesian lineages mainly consisted of three key mutations, namely S\_D614G, S\_N439K and S\_P681R. Despite the VUM label assigned by WHO, there were no further evidence of either higher transmissibility nor reduced effectiveness of vaccines and therapy linked to the Indonesian lineages.

The most prevalent amino acid change found in our study was the S\_D614G, as 199 out of 202 (99%) genomes were found to carry it. This is consistent with the prevalence of this mutation in SARS-CoV-2 genomes in Indonesia (93.1%) and worldwide (97.6%) ([Groves, Rowland-Jones & Angyal, 2021](#)). Variants carrying the S\_D614G mutation have been suggested to benefit from increased binding affinity to its target ACE2 ([Ozono et al., 2021](#)). Although not associated with worse disease severity, the S\_D614G change has been implicated in enhanced transmission and higher viral loads ([Korber et al., 2020](#); [Plante et al., 2020](#); [Volz et al., 2021](#)). The mutation highly correlates with loss of smell (anosmia) ([Von Bartheld, Hagen & Butowt, 2021](#)). Fortunately, a number of experimental and *in silico* analyses also suggested that the presence of this mutation may not affect the efficacy of vaccines ([McAuley et al., 2020](#)). The D614G variant is considered to be G clade by GISAID and B.1 clade by the PANGOLIN ([Korber et al., 2020](#)).

Presenting the analysis of genomes obtained using the Oxford Nanopore Technologies (ONT) platform, this study adds to the body of literature that demonstrates the feasibility of Nanopore sequencing for genomic surveillance of SARS-CoV-2, including for the detection of variations within the genome. Studies using a similar platform from ONT showed that nucleotide substitutions (*Bull et al., 2020; Li et al., 2020a*), as well as deletions within the SARS-CoV-2 genomes (*Moore et al., 2020*) have also been successfully characterized in samples obtained in the United Kingdom, Brazil and South Africa. Nanopore sequencing also led to the identification of the B.1.1.251 lineage in Brazil in January 2021 (*Dos Santos et al., 2021*) and eight different lineages of SARS-CoV-2 in South Africa (*Engelbrecht et al., 2021*). Other advantages that include comparatively lower costs and shorter turnaround time without compromising yield and quality of resulting genomes are attractive features for limited resource settings. Further reduction in turnaround time and costs can also be achieved, as exemplified by a combination use of primer sets that generate 1,200 bp amplicons and the ONT Rapid Barcoding Kit (*Freed et al., 2020*). Moreover, protocols amenable to high throughput to accelerate sequencing efforts are also available (*Baker et al., 2021; Pater et al., 2021*). The ongoing improvement in efficiency of cost and time is likely to increase the adoption of Nanopore sequencing of SARS-CoV-2 genomes in many more countries, thereby contributing to a better worldwide genomic surveillance of SARS-CoV-2.

Our study demonstrated a variant shift between two pandemic waves in Indonesia, accompanied by variations in key amino acid changes in genomes of circulating SARS-CoV-2 viruses in Bogor and surrounding areas. The relatively small sampling coverage in a subset of geographic areas may not be a sufficient representation of Indonesia, which would be one of the limitations of this study. However, the regions we selected represent areas with a considerable load of human traffic that potentially influence the transmission dynamics of SARS-CoV-2. More importantly, the results of our current monitoring showcased a similar pattern at provincial and national levels. The current prioritization strategy for sample selection in this study was mainly based on low Ct value to maximise the likelihood of obtaining whole genomes, as at the time we had difficulties obtaining sequencing libraries from samples with Ct values of 30 or above. Further studies would benefit from sample selection based on current local outbreaks, or concerned cases such as reinfection, sporadic cases, as well as those based on severity of illness and cases of vaccinated persons. These considerations are important for a more thorough genomic surveillance in Indonesia, especially in large cities and their surrounding areas, such as Bogor, Jakarta and Tangerang. Nevertheless, we successfully obtained information on Indonesian clade and lineage classifications, as well as the amino acid changes carried by those variants, which would contribute to the much-needed data on viral evolution and transmission in Indonesia.

## CONCLUSIONS

The second wave of the pandemic in Indonesia witnessed the Delta variants becoming the dominant SARS-CoV-2 variants in Bogor and surrounding areas, taking over the dominance of Indonesian variants during the first wave of pandemic and, reflecting a shift in circulating variants at the national level. Distinct mutations exhibited by the Delta variants compared to other previously identified variants were potentially the reason behind the Delta variants being the cause of the highest average of confirmed COVID-19 cases in the second wave. This was also corroborated by evidence of higher transmissibility of Delta variants that are responsible for massive outbreaks in many countries. In order to prioritise global monitoring and research, and ultimately to inform the ongoing response to the COVID-19 pandemic, whole genome sequencing remains an indispensable tool. Sequencing enables detection of common variants and various amino acid changes in the SARS-CoV-2 genome, contributing to important information on circulating SARS-CoV-2 variants in Bogor and its surrounding areas.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the COVID-19 diagnostic testing team of BSL-3 and BSL-2 Laboratory, National Research and Innovation Agency (BRIN)-Indonesia; Dr. Andika Chandra Putra and Dr. Jamal Zaini from RSUP Persahabatan, Jakarta; and originating laboratories and submitting laboratories as part of Indonesian Consortium of SARS-CoV-2 Genomic Surveillance where genetic sequence data are shared *via* the GISAID Initiative. We are very grateful for capacity building in the beginning of this project from colleagues at the University of Nottingham: Dr. Susanti Susanti, Dr. Inswasti Cahyani, Prof. Mohammad Ilyas and Prof. Matthew Loose; as part of NICCRAT (Nottingham- Indonesia Collaboration for Clinical Research and Training) partnership.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was supported by the DIPA IPH LIPI 2020 Research Grant (B-5114/IPH/HK.01.03/I/2020) and the RISPRO-LPDP Funding Program for COVID-19 (07/FI/P-KCOVID-19.2B3/X/2020) under the project name “Surveilans Genom SARS-CoV-2” (VenomCoV) LIPI/BRIN. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

DIPA IPH LIPI 2020 Research Grant: B-5114/IPH/HK.01.03/I/2020.

RISPRO-LPDP Funding Program for COVID-19: 07/FI/P-KCOVID-19.2B3/X/2020.

“Surveilans Genom SARS-CoV-2” (VenomCoV) LIPI/BRIN.

### Competing Interests

The authors declare there are no competing interests.

## Author Contributions

- Anggia Prasetyoputri and Anik B. Dharmayanthi conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Syam B. Iryanto conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Ade Andriani, Isa Nuryana and Andri Wardiana conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Asep M. Ridwanuloh, Sri Swasthikawati, Hariyatun Hariyatun, Herjuno A. Nugroho, Idris Idris, Indriawati Indriawati, Listiana Oktavia, Yuliawati Yuliawati, Masrukhin Masrukhin, Erwin F. Hasrianda, Linda Sukmarini, Fahrurrozi Fahrurrozi, Nova Dilla Yanthi, Alfi T. Fathurahman, Ari S. Wulandari and Syaiful Rizal performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Zahra Noviana analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Ruby Setiawan performed the experiments, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Ahmad Fathoni, Wien Kusharyoto, Puspita Lisdiyanti and Ratih A. Ningrum conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Sugiyono Saputra conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

## Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The study protocol was reviewed and approved by the Health Research Ethics Committee, University of Indonesia and Cipto Mangunkusumo Hospital (HREC-FMUI/CMH) (20-10-1321\_EXP).

## DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

All 202 complete genomes of SARS-CoV-2 are available at GISAID: EPI\_ISL\_1020175-187, 189-191, 193-196, 198-202, 204-206 ( $n = 28$ ); EPI\_ISL\_1117450, 452-455, 589, 600, 602-603 ( $n = 9$ ); EPI\_ISL\_2840868-894 ( $n = 27$ ); EPI\_ISL\_2844812-832 ( $n = 21$ ); EPI\_ISL\_2860278-299 ( $n = 22$ ); EPI\_ISL\_2860332-349, 360-368 ( $n = 27$ ); EPI\_ISL\_3086881-885 ( $n = 5$ ); EPI\_ISL\_3186060-061, 064-074 ( $n = 13$ ); EPI\_ISL\_3194636 ( $n = 1$ ); EPI\_ISL\_3374423-425 ( $n = 3$ ); EPI\_ISL\_4004697-704 ( $n = 8$ ); EPI\_ISL\_4106122-134 ( $n = 13$ ); EPI\_ISL\_518819-820 ( $n = 2$ ); EPI\_ISL\_833388-500 ( $n = 14$ ); EPI\_ISL\_833503-510 ( $n = 8$ ); and EPI\_ISL\_833571 ( $n = 1$ ).

## Data Availability

The following information was supplied regarding data availability:

The qRT-PCR results, demographic patients, lineages, clades, and mutations of 202 samples and the raw data for phylogenetic analysis and all whole-genome sequences are available in the [Supplementary Files](#).

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.13132#supplemental-information>.

## REFERENCES

- Allen H, Vusirikala A, Flannagan J, Twohig KA, Zaidi A, Chudasama D, Lamagni T, Groves N, Turner C, Rawlinson C, Lopez-Bernal J, Harris R, Charlett A, Dabrera G, Kall M, the COVID-19 Genomics UK (COG-UK Consortium). 2021. Household transmission of COVID-19 cases associated with SARS-CoV-2 delta variant (B.1.617.2): national case-control study. *The Lancet Regional Health—Europe* 12:100252.
- Baker DJ, Aydin A, Le-Viet T, Kay GL, Rudder S, Martins LDO, Tedim AP, Kolyva A, Diaz M, Alikhan N-F, Meadows L, Bell A, Gutierrez AV, Trotter AJ, Thomson NM, Gilroy R, Griffith L, Adriaenssens EM, Stanley R, Charles IG, Elumogo N, Wain J, Prakash R, Meader E, Mather AE, Webber MA, Dervisevic S, Page AJ, OGrady J. 2021. CoronaHiT: High throughput sequencing of SARS-CoV-2 genomes. *bioRxiv* 2020.06.24.162156.
- BNPB. 2021. Situasi COVID-19 di Indonesia. Available at <https://experience.arcgis.com/experience/57237ebe9c5b4b1caa1b93e79c920338> (accessed on 13 May 2021).
- Bull RA, Adikari TN, Ferguson JM, Hammond JM, Stevanovski I, Beukers AG, Naing Z, Yeang M, Verich A, Gamaarachchi H, Kim KW, Luciani F, Stelzer-Braid S, Eden J-S, Rawlinson WD, Van Hal SJ, Deveson IW. 2020. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nature Communications* 11(1):6272 DOI 10.1038/s41467-020-20075-6.
- CDC. 2021. SARS-CoV-2 variant classifications and definitions. Available at <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html> (accessed on October 2021).
- COVID-19 PIK. 2021. Dashboard Statistik Kasus Covid-19 Provinsi Jawa Barat. Available at <https://pikobar.jabarprov.go.id/data> (accessed on October 2021).
- COVID-19 S. 2021. Peta Sebaran COVID-19. Available at <https://covid19.go.id/peta-sebaran-covid19> (accessed on October 2021).
- Deng X, Garcia-Knight MA, Khalid MM, Servellita V, Wang C, Morris MK, Sotomayor-Gonzalez A, Glasner DR, Reyes KR, Gliwa AS, Reddy NP, Martin CSS, Federman S, Cheng J, Balcerak J, Taylor J, Streithorst JA, Miller S, Kumar GR, Sreekumar B, Chen P-Y, Schulze-Gahmen U, Taha TY, Hayashi J, Simoneau CR, McMahan S, Lidsky PV, Xiao Y, Hemarajata P, Green NM, Espinosa A, Kath C, Haw M, Bell J,

- Hacker JK, Hanson C, Wadford DA, Anaya C, Ferguson D, Lareau LF, Frankino PA, Shivram H, Wyman SK, Ott M, Andino R, Chiu CY. 2021. Transmission, infectivity, and antibody neutralization of an emerging SARS-CoV-2 variant in California carrying a L452R spike protein mutation. *medRxiv*.
- Dos Santos CA, Bezerra GVB, De Azevedo Marinho ARRA, Sena LOC, Alves JC, De Souza MSF, Goes MADO, Teixeira DCP, Silva PCR, Siqueira MAMT, Martins-Filho PR. 2021. First Report of SARS-CoV-2 B.1.1.251 Lineage in Brazil. *Journal of Travel Medicine* 28(4):taab033 DOI 10.1093/jtm/taab033.
- Engelbrecht S, Delaney K, Kleinhans B, Wilkinson E, Tegally H, Stander T, Van Zyl G, Preiser W, Oliveira T. 2021. Multiple early introductions of SARS-CoV-2 to Cape Town, South Africa. *Viruses* 13(3):526 DOI 10.3390/v13030526.
- (FDA) TUSFaDA. 2021. Fact sheet for health care providers emergency use authorization (EUA) of BAMLANIVIMAB and ETESEVIMAB.
- Freed NE, Vlková M, Faisal MB, Silander OK. 2020. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biology Methods and Protocols* 5(1):bpaa014 DOI 10.1093/biomethods/bpaa014.
- GISAID. 2021. EpiCoV. Available at <https://www.epicov.org/epi3/frontend> (accessed on 31 October 2021).
- GISAID Initiative. 2021. Epicov. Available at <https://www.epicov.org/epi3/frontend#27a497> (accessed on October 2021).
- Groves DC, Rowland-Jones SL, Angyal A. 2021. The D614G mutations in the SARS-CoV-2 spike protein: Implications for viral infectivity, disease severity and vaccine design. *Biochemical and Biophysical Research Communications* 538:104–107 DOI 10.1016/j.bbrc.2020.10.109.
- Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34(23):4121–4123 DOI 10.1093/bioinformatics/bty407.
- Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A, Peacock SJ, Robertson DL. 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology* 19(7):409–424 DOI 10.1038/s41579-021-00573-0.
- Itokawa K, Sekizuka T, Hashino M, Tanaka R, Kuroda M. 2020. Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PLOS ONE* 15(9):e0239403 DOI 10.1371/journal.pone.0239403.
- Katoh K, Standley DM. 2016. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* 32(13):1933–1942 DOI 10.1093/bioinformatics/btw108.
- Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, Silva TI, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC. 2020. Tracking changes



- in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**(4):812–827.e19 DOI [10.1016/j.cell.2020.06.043](https://doi.org/10.1016/j.cell.2020.06.043).
- Li H. 2011.** A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**(21):2987–2993 DOI [10.1093/bioinformatics/btr509](https://doi.org/10.1093/bioinformatics/btr509).
- Li H. 2018.** Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**(18):3094–3100 DOI [10.1093/bioinformatics/bty191](https://doi.org/10.1093/bioinformatics/bty191).
- Li X, Geng M, Peng Y, Meng L, Lu S. 2020b.** Molecular immune pathogenesis and diagnosis of COVID-19. *Journal of Pharmaceutical Analysis* **10**(2):102–108 DOI [10.1016/j.jpha.2020.03.001](https://doi.org/10.1016/j.jpha.2020.03.001).
- Li J, Wang H, Mao L, Yu H, Yu X, Sun Z, Qian X, Cheng S, Chen S, Chen J, Pan J, Shi J, Wang X. 2020a.** Rapid genomic characterization of SARS-CoV-2 viruses from clinical specimens using nanopore sequencing. *Scientific Reports* **10**(1):17492 DOI [10.1038/s41598-020-74656-y](https://doi.org/10.1038/s41598-020-74656-y).
- Liu Y, Liu J, Johnson BA, Xia H, Ku Z, Schindewolf C, Widen SG, An Z, Weaver SC, Menachery VD, Xie X, Shi P-Y. 2021.** Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. *bioRxiv*.
- Loman N, Rambaut A, Quick J, Rowe W, Simpson J, Beaulaurier J, Wilkinson W, Florek K. 2020.** ARTIC nanopore protocol for nCoV2019 novel coronavirus. Available at <https://github.com/artic-network/artic-ncov2019> 2021.
- Lubinski B, Fernandes MHV, Frazier L, Tang T, Daniel S, Diel DG, Jaimes JA, Whittaker GR. 2022.** Functional evaluation of the P681H mutation on the proteolytic activation of the SARS-CoV-2 variant B.1.1.7 (Alpha) spike. *iScience* **25**(1):103589 DOI [10.1016/j.isci.2021.103589](https://doi.org/10.1016/j.isci.2021.103589).
- Lubinski B, Frazier LE, Phan MVT, Bugembe DL, Tang T, Daniel S, Cotten M, Jaimes JA, Whittaker GR. 2021.** Spike protein cleavage-activation mediated by the SARS-CoV-2 P681R mutation: a case-study from its first appearance in variant of interest (VOI) A.23.1 identified in Uganda. *bioRxiv*.
- McAuley AJ, Kuiper MJ, Durr PA, Bruce MP, Barr J, Todd S, Au GG, Blasdel K, Tachedjian M, Lowther S, Marsh GA, Edwards S, Poole T, Layton R, Riddell S-J, Drew TW, Druce JD, Smith TRF, Broderick KE, Vasan SS. 2020.** Experimental and in silico evidence suggests vaccines are unlikely to be affected by D614G mutation in SARS-CoV-2 spike protein. *npj Vaccines* **5**(1):96 DOI [10.1038/s41541-020-00246-8](https://doi.org/10.1038/s41541-020-00246-8).
- Minh BQ, Nguyen MA, Von Haeseler A. 2013.** Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* **30**(5):1188–1195 DOI [10.1093/molbev/mst024](https://doi.org/10.1093/molbev/mst024).
- Moore SC, Penrice-Randal R, Alruwaili M, Randle N, Armstrong S, Hartley C, Haldenby S, Dong X, Alrezaihi A, Almsaud M, Bentley E, Clark J, Garcia-Dorival I, Gilmore P, Han X, Jones B, Luu L, Sharma P, Shawli G, Sun Y, Zhao Q, Pullan ST, Carter DP, Bewley K, Dunning J, Zhou E, Solomon T, Beadsworth M, Cruise J, Crook DW, Matthews DA, Davidson AD, Mahmood Z, Aljabr W, Druce J,**

- Vipond R, Ng L, Renia L, Openshaw PJM, Baillie J, Carroll MW, Stewart J, Darby A, Semple M, Turtle L, Hiscox JA. 2020. Amplicon-based detection and sequencing of SARS-CoV-2 in nasopharyngeal swabs from patients with COVID-19 and identification of deletions in the viral genome that encode proteins involved in interferon antagonism. *Viruses* 12(10):1164 DOI 10.3390/v12101164.
- Nasir JA, Kozak RA, Aftanas P, Raphenya AR, Smith KM, Maguire F, Maan H, Alruwaili M, Banerjee A, Mbareche H, Alcock BP, Knox NC, Mossman K, Wang B, Hiscox JA, McArthur AG, Mubareka S. 2020. A comparison of whole genome sequencing of SARS-CoV-2 using amplicon-based sequencing, random hexamers, and bait capture. *Viruses* 12(8):895 DOI 10.3390/v12080895.
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32(1):268–274 DOI 10.1093/molbev/msu300.
- O’Toole A, Scher E, Rambaut A, Van Heus P, McCrone JT, Poon A, Colquhoun R, Underwood A, Maguire F, Hinrichs A, Watson M, Taylor B, Kruczkiewicz P, Nguyen SH, Roemer C, Hill V, Innes MJ, Jackson B, Maricato V, Sjödin A, Grüning B, Fairley D, Menzel P, Hoesven E. 2020. Software package for assigning SARS-CoV-2 genome sequences to global lineages. Available at <https://github.com/cov-lineages/pangolin>.
- O’Toole A, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, Colquhoun R, Ruis C, Abu-Dahab K, Taylor B, Yeats C, Plessis L, Maloney D, Medd N, Attwood SW, Aanensen DM, Holmes EC, Pybus OG, Rambaut A. 2021. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evolution* 7(2):veab064.
- Ozono S, Zhang Y, Ode H, Sano K, Tan TS, Imai K, Miyoshi K, Kishigami S, Ueno T, Iwatani Y, Suzuki T, Tokunaga K. 2021. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nature Communications* 12(1):848 DOI 10.1038/s41467-021-21118-2.
- Pater AA, Bosmeny MS, Parasrampur M, Eddington SB, Ovington KN, White AA, Metz PE, Sylvain RJ, Chilamkurthy R, Yinusa AO, Benzinger SW, Hebert MM, Gagnon KT. 2021. High throughput nanopore sequencing of SARS-CoV-2 viral genomes from patient samples. *bioRxiv* 2021.02.09.430478.
- Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, Zhang X, Muruato AE, Zou J, Fontes-Garfias CR, Mirchandani D, Scharton D, Bilello JP, Ku Z, An Z, Kalveram B, Freiberg AN, Menachery VD, Xie X, Plante KS, Weaver SC, Shi P-Y. 2020. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592(7852):116–121.
- Quick J. 2020. nCoV-2019 sequencing protocol v2 (GunIt) V.2. protocolsio.
- Rambaut A, Holmes EC, O’Toole Á, Hill V, McCrone JT, Ruis C, Plessis L, Pybus OG. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology* 5(11):1403–1407 DOI 10.1038/s41564-020-0770-5.

- Regeneron.** 2021. Regeneron COVID-19 Dashboard.pdf. Available at [https://covid19dashboard.regeneron.com/?tab=Variant\\_Graphs&subTab=Top\\_Variants\\_Over\\_Time\\_\(Select\\_Location\)](https://covid19dashboard.regeneron.com/?tab=Variant_Graphs&subTab=Top_Variants_Over_Time_(Select_Location)) (accessed on October 2021).
- Russell TW, Wu JT, Clifford S, Edmunds WJ, Kucharski AJ, Jit M.** 2021. Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study. *The Lancet Public Health* 6(1):e12-e20  
DOI 10.1016/S2468-2667(20)30263-2.
- Saito A, Irie T, Suzuki R, Maemura T, Nasser H, Uriu K, Kosugi Y, Shirakawa K, Sadamasu K, Kimura I, Ito J, Wu J, Iwatsuki-Horimoto K, Ito M, Yamayoshi S, Loeber S, Tsuda M, Wang L, Ozono S, Butlertanaka EP, Tanaka YL, Shimizu R, Shimizu K, Yoshimatsu K, Kawabata R, Sakaguchi T, Tokunaga K, Yoshida I, Asakura H, Nagashima M, Kazuma Y, Nomura R, Horisawa Y, Yoshimura K, Takaori-Kondo A, Imai M, Tanaka S, Nakagawa S, Ikeda T, Fukuhara T, Kawaoka Y, Sato K.** 2021. Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature* 602(7896):300–306.
- Takenouchi T, Iwasaki YW, Harada S, Ishizu H, Uwamino Y, Uno S, Osada A, Abe K, Hasegawa N, Murata M, Takebayashi T, Fukunaga K, Saya H, Kitagawa Y, Amagai M, Siomi H, Kosaki K.** 2021. Clinical utility of SARS-CoV-2 whole genome sequencing in deciphering source of infection. *Journal of Hospital Infection* 107:40–44 DOI 10.1016/j.jhin.2020.10.014.
- Team RC.** 2021. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Thomson EC, Rosen LE, Shepherd JG, Spreafico R, Da Silva Filipe A, Wojcechowskyj JA, Davis C, Piccoli L, Pascall DJ, Dillen J, Lytras S, Czudnochowski N, Shah R, Meury M, Jesudason N, De Marco A, Li K, Bassi J, O’Toole A, Pinto D, Colquhoun RM, Culap K, Jackson B, Zatta F, Rambaut A, Jaconi S, Sreenu VB, Nix J, Zhang I, Jarrett RF, Glass WG, Beltramello M, Nomikou K, Pizzuto M, Tong L, Cameroni E, Croll TI, Johnson N, Iulio JD, Wickenhagen A, Ceschi A, Harbison AM, Mair D, Ferrari P, Smollett K, Sallusto F, Carmichael S, Garzoni C, Nichols J, Galli M, Hughes J, Riva A, Ho A, Schiuma M, Semple MG, Openshaw PJM, Fadda E, Baillie JK, Chodera JD, Rihn SJ, Lycett SJ, Virgin HW, Telenti A, Corti D, Robertson DL, Snell G.** 2021. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* 184(5):1171–1187 e20  
DOI 10.1016/j.cell.2021.01.037.
- Volz E, Hill V, McCrone JT, Price A, Jorgensen D, Á O’Toole, Southgate J, Johnson R, Jackson B, Nascimento FF, Rey SM, Nicholls SM, Colquhoun RM, Filipe ADS, Shepherd J, Pascall DJ, Shah R, Jesudason N, Li K, Jarrett R, Pacchiarini N, Bull M, Geidelberg L, Siveroni I, Goodfellow I, Loman NJ, Pybus OG, Robertson DL, Thomson EC, Rambaut A, Connor TR.** 2021. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* 184(1):64–75.e11  
DOI 10.1016/j.cell.2020.11.020.

- Von Bartheld CS, Hagen MM, Butowt R. 2021.** The D614G virus mutation enhances anosmia in COVID-19 patients: evidence from a systematic review and meta-analysis of studies from South Asia. *ACS Chemical Neuroscience* **12(19)**:3535–3549 DOI [10.1021/acscemneuro.1c00542](https://doi.org/10.1021/acscemneuro.1c00542).
- WHO. 2020.** Archived: WHO Timeline—COVID-19. Available at <https://www.who.int/news/item/27-04-2020-who-timeline---covid-19> (accessed on 13 May 2021).
- WHO. 2021a.** Tracking SARS-CoV-2 variants. Available at <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>.
- WHO. 2021b.** WHO Coronavirus (COVID-19) dashboard. Available at <https://covid19.who.int/> (accessed on 24 December 2021).
- Zhou W, Xu C, Wang P, Luo M, Xu Z, Cheng R, Jin X, Guo Y, Xue G, Juan L, Anashkina AA, Nie H, Jiang Q. 2021.** N439K variant in spike protein alter the infection efficiency and antigenicity of SARS-CoV-2 based on molecular dynamics simulation. *Frontiers in Cell and Developmental Biology* **9**:697035 DOI [10.3389/fcell.2021.697035](https://doi.org/10.3389/fcell.2021.697035).