Genome Profiling of SARS-CoV-2 in Indonesia, ASEAN and the Neighbouring East Asian Countries: Features, Challenges and Achievements

Inswasti Cahyani[1], Eko W Putro[2], Asep M Ridwanuloh[2], Satrio Wibowo[2,3], Hariyatun Hariyatun[2], Gita Syahputra[2], Gilang Akbariani[2,3], Ahmad R Utomo[4], Mohammad Ilyas[5], Matthew W. Loose[1], Wien Kusharyoto[2], Susanti Susanti[3,5,6*]

Affiliations:

1. School of Life Sciences, University of Nottingham, Nottingham, United Kingdom
2. Research Centre for Biotechnology, Indonesian Institute of Sciences (LIPI/*Lembaga Ilmu Pengetahuan Indonesia*), Bogor, Indonesia
3. PT. PathGen Diagnostik Teknologi, Cibinong, Bogor, Indonesia
4. Biomedical Postgraduate Program, Universitas Yarsi, Jakarta, Indonesia
5. Molecular Pathology Research Group, School of Medicine, University of Nottingham, United Kingdom
6. Department of Pharmacology and Clinical Pharmacy, Faculty of Pharmacy, Universitas Muhammadiyah, Purwokerto, Indonesia

***Corresponding author**: susanti.susanti@nottingham.ac.uk

Abstract

A year after the World Health Organisation (WHO) declared COVID-19 as a pandemic, much has been learned with respect to SARS-CoV-2 epidemiology, vaccine production and disease treatment. Whole-genome sequencing (WGS) has played a significant role in contributing to our understanding of the epidemiology and biology of this virus. In this paper, we investigate the use of SARS-CoV-2 WGS in Southeast and East Asia and the impact of technological development, access to resources, and demography of individual countries on its uptake. Facilitated by the Nottingham-Indonesia Collaboration for Clinical Research and Training (NICCRAT) initiative, we showcased a bilateral collaboration between the University of Nottingham and the Indonesian Institute of Sciences (LIPI/*Lembaga Ilmu Pengetahuan Indonesia*) to establish WGS of SARS-CoV-2 using Oxford Nanopore Technology® (ONT) in Indonesia. Analyses of SARS-CoV-2 genomes deposited on GISAID from Southeast and East Asian countries reveals the importance of collecting clinical and demographic metadata and the importance of open access and data sharing. Lineage and phylogenetic analyses per 1 June 2021 found that: 1) B.1.466.2 variants were the most predominant in Indonesia, with mutations in the spike protein including D614G at 100%, N439K at 99.1%, and P681R at 69.7% frequency, 2) The variants of concern (VoCs), B.1.1.7 (Alpha), B.1.351 (Beta) and B.1.617.2 (Delta) were first detected in Indonesia in January 2021, 2) B.1.470 was first detected in Indonesia and spread to the neighbouring regions, and 3) The highest rate of virus transmissions between Indonesia and the rest of the world appears to be through interactions with Singapore and Japan, two neighbouring countries with high degree of access and travels to and from Indonesia. Overall, we conclude that WGS of SARS-CoV-2 using ONT platforms fits well with the Indonesian context and can catalyse the increase of sequencing rates in the country.

Keywords

ASEAN, COVID-19, genomic surveillance, GISAID, nanopore, NICCRAT, SARS-CoV-2, variant of concerns, whole-genome sequencing

Introduction

Coronaviruses belong to a large family of RNA viruses that usually cause respiratory illnesses. Most of the time, diseases caused by coronaviruses such as HCoV-229E, OC43, NL63, and HKU1 are mild[1]. Nonetheless, in the past two decades more deadly forms of coronaviruses have emerged including Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV)[2] and the Middle East Respiratory Syndrome Coronavirus (MERS)[3]. Most recently, COVID-19 emerged and was declared a global pandemic by the World Health Organization (WHO) on 11th March 2020[4]. This disease is caused by a novel previously unreported coronavirus strain subsequently named Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2)[5].

COVID-19 was first reported in Wuhan in early December 2019 [6]. In early January 2020, the Chinese government released the first genome sequence of SARS-CoV-2 (*i.e.* labelled as WH-Human_1)[7,8]. By the end of February 2020, the infection had been reported in 51 countries with almost 84,000 cases[9]. Despite being one of the world's most populated countries, with more than 260 million people, and no travel restrictions in place at the time, Indonesia only diagnosed and announced its first case of infection on 2nd March 2020[10]. However, neighbouring countries such as Singapore and Malaysia had reported their first cases of infection on 23rd January and 25th January 2020, respectively. Indonesia's later detection and action might have been influenced to a large degree by some presumed confidence in the society, that the infection would not easily spread in the tropics[11]. There was a widely-held belief that the hot tropical climate and high intensity of UV rays in Indonesia would be sufficient to hinder SARS-CoV-2 survival and spread in the air, much like the first SARS epidemic where cases were low compared to its originating place in China[2,11,12]. Therefore, before March 2020, Indonesia only carried out early testing and screening of individuals with symptoms and travel links[13,14].

In hindsight, the indifference toward COVID-19 when it was still an epidemic was not exclusive to Indonesia, as many governments around the globe shared a similar lack of anticipation and mitigation to prevent COVID-19 epidemic from turning into a pandemic[15]. The Association of Southeast Asian Nations (ASEAN) accommodates free cross-border travels for its members, with Indonesia having the largest area and population size. Indonesia also has bilateral travel agreements that make travel more accessible to other East Asian countries such as Japan and China. Global travel has been a significant contributing factor to the rapid spread of COVID-19 and establishing the pandemic[16]. Therefore, we sought to compare the genomic epidemiology and metadata of SARS-CoV-2 between ASEAN member countries and others in the region.

As of today, more than a year after Indonesia's first COVID-19 case was announced (*i.e.* mid-March 2020), it is thought that the number of cases in Indonesia is still significantly underestimated. Besides the utility of screening and tracing for containing and reducing transmission, there is also a need to pinpoint and follow variants of SARS-CoV-2 that might worsen the pandemic. World Health Organization defines a variant to be a Variant of Interest (VoI) if it has changed phenotypically from the reference genome (or shown

phenotypic implications) and has been identified to cause community transmissions or detected in multiple countries[17]. A VoI can become a Variant of Concern (VoC) if it has been demonstrated to 1) increase transmissibility or detrimental changes in COVID-19 epidemiology; 2) an increase in virulence or change in clinical disease presentation; or 3) decrease in the effectiveness of public health and social measures or available diagnostics, vaccines, therapeutics[17]. The existence and potential emergence of these variants highlight the need for reliable, easy-to-set-up, rapid, and cost-effective genome sequencing technology in Indonesia. We showcase how the Oxford Nanopore Technology® (ONT) platform can be used for WGS of SARS-CoV-2 in the Indonesian context. Moreover, this report discusses the types and transmission patterns of SARS-CoV-2 variants in Indonesia in relation to ASEAN and other Asian countries until 1 June 2021. We highlighted how the knowledge obtained from WGS of SARS-CoV-2 is pivotal in the understanding and tackling of COVID-19 spread in Indonesia and its surrounding regions.

Results

A. The rates of whole genome sequencing in ASEAN and its neighbouring countries correlate to differences in technology development, access to resources and demography.

Whole-genome sequencing of viruses that play a role in an outbreak, epidemic and/or pandemic is emerging as a crucial tool in controlling and mitigating the spread of disease[18-21]. Whole-genome sequencing can inform epidemiological scientists on the nature, behaviour and transmission patterns of viruses[22]. This information can support decision makers in taking appropriate actions. In the era of next-generation sequencing (NGS), WGS for epidemiological monitoring should be straightforward, in principle, with the availability of many tools and relatively low cost of sequencing compared to one or two decades ago. Some examples of WGS that applied these NGS principles in (or after) an outbreak were in the case of the Zika virus in Brazil[18,19], Ebola in West Africa [20] and more recently, a mumps outbreak in Canada [21]. Whole-genome sequencing is also at the heart of global research for vaccines and treatment discoveries of the COVID-19 pandemic and for surveying its VoCs[22].

Unfortunately, WGS, which is using cutting edge sequencing technology, typically means that only developed countries with access and resources to technology are readily able to use it in a pandemic (Figure 1A). Figure 1 shows that despite being the largest continent in the world, Asia has submitted only about 6% of the total SARS-CoV-2 genomes deposited on GISAID per 1 June 2021 (Figure 1A), with Japan submitting the most at just over 44% (Figure 1B-1C). Indonesia contributed to 1.7% of genomes submitted from Asia, relatively behind its close neighbour, the Philippines at 4% (Figure 1B-1C). Indonesia also experienced the highest total number of positive cases in the East and South-East Asian region (Figure 2A). However, the rate of SARS-CoV-2 whole-genome sequencing did not follow the case rate as the number of genomes submitted to GISAID and Indonesia's sequencing rate were relatively low (Figure 1C and 2B, respectively). Some of Indonesia's neighbours, such as Hong Kong, Japan and South Korea tended to have significantly higher sequencing rates (Figure 1C and 2B) most likely because of better access to technology and resources. Better availability of resources also seems to help these countries to push their COVID-19 test rate to parallel their population size (Figure S1). Meanwhile, other countries with similar development status to Indonesia,

such as Thailand, might fare better because of relatively smaller numbers of cases and population size (Figure 2A). In contrast to Indonesia, the Philippines relatively tripled its sequencing rate despite having the second highest number of positive cases in the region (Figure 2).

There has been a lot of debate on the ideal sequencing rate and whether it should be based on the population size or positive cases or both[23]. As of 1 June 2021, Indonesia had sequenced 0.1% of its positive cases (Figure 2B). However, this sequencing rate was calculated on known case number which might have been underestimated due to suboptimal testing rates (Figure S1).

B. Establishment of whole genome sequencing effort in Indonesia and its informative clinical and demographic metadata

1. Establishing WGS of SARS-CoV-2 between NICCRAT and LIPI using ONT platform

In January 2021, the Indonesian Ministry of Health working together with the Ministry of Research and Technology, started a nation-wide consortium for SARS-CoV-2 WGS for pandemic surveillance, bringing together laboratories, research and academic institutions[24]. The consortium was expected to accelerate and scale-up SARS-CoV-2 WGS informing the government on the genomic epidemiology of the pandemic in a timely manner[24]. The creation of nation-wide consortia for SARS-CoV-2 WGS has been shown to work in the United Kingdom with the COG-UK consortium[25]. The UK is currently the second biggest contributor of SARS-CoV-2 genomes on GISAID after the United States of America[26], sequencing about 10% of its positive cases[25]. Therefore, emulating some aspects of this consortium in Indonesia is feasible while considering the local context, such as demography and logistics.

Prior to establishing this consortium, the Indonesian Institute of Sciences (LIPI) had already joined forces with the University of Nottingham in the UK to build capacity to carry out SARS-CoV-2 WGS in Indonesia. This collaboration was enabled through the Nottingham-Indonesia Collaboration in Cancer Research and Training (NICCRAT) initiative, established in 2019[27,28]. This initiative is to foster partnership between the University of Nottingham and several Indonesian academics and research institutions in the field of health and clinical science[27,28]. As COVID-19 became a global pandemic in early 2020, the initiative utilized existing resources and expertise to take part in SARS-CoV-2 WGS (Table 1)[28].

Increasing WGS of SARS-CoV-2 in Indonesia has been challenging. Up-to-standard laboratories are primarily found in the dense west-to-middle parts of Indonesia, such as Java, Sumatera, and Bali. The high-cost and state-of-the-art devices required for NGS are also a limiting factor as well as the expertise and infrastructure required. Oxford Nanopore Technology® (ONT) offers a solution to these issues by providing a portable, scalable, and lower cost of sequencing platform compared to other NGS technologies[29]. Oxford Nanopore Technology® can potentially be the workhorse for WGS in Indonesia. In addition, the ARTIC network[30,31] provided protocols, primer designs and analysis pipelines and tools capable of running on basic compute infrastructure.

We utilised the MinION sequencing platform to sequence the genomes of SARS-CoV-2 (see Method and Table 1), exploiting their speed, portability, relatively low cost, and the existence of well-established protocols. These features are particularly advantageous within Indonesia, where specialized molecular biology laboratories that can handle WGS

are not widely available. This project also benefits through the involvement of the University of Nottingham in the COG-UK consortium[25]. In April 2020, the first skill transfer workshop of SARS-CoV-2 WGS using the ONT platform was carried out by the University of Nottingham team to the Indonesian partner in LIPI. From this, the NICCRAT-LIPI team has pioneered the submission of 12 whole genomes on GISAID, with sample collections dated from September 2020 to January 2021 (Table S1). Following this, WGS of SARS-CoV-2 has been established as a routine program in LIPI.

2. *Comparison of clinical and demographic metadata of deposited genomic sequence in GISAID between Indonesia and its neighbouring countries*

As a part of SARS-CoV-2 WGS, recording sample metadata is a vital aspect to build a clearer picture of genomic epidemiology[32]. However, this is not always easy nor practical to achieve, especially when the institution collecting samples differs to the one sequencing them, such as a hospital and a sequencing research facility. Moreover, countries have different regulations concerning privacy, data access and sharing. In this report, we categorize the metadata accompanying genomes on GISAID as clinical (*e.g.* disease severity, mortality, hospitalization rate, and sample type) or demographic (*e.g.* age, gender, and residency) metadata. Indonesia and its neighbours recorded and dealt with sample metadata in different methods which influence the interpretation of WGS output.

Indonesia performs well in terms of metadata completeness, both in clinical (Table 2, Figure 3A) and demographic metadata (Figure 3B-3C) when compared to other ASEAN and neighbouring countries, with over 70% of the genomes have the corresponding clinical metadata (Table 2). In addition, Indonesia also recorded the sources of the majority of samples (*i.e.* nasal and oropharyngeal swab) used in WGS, which is lacking in many other large datasets (*e.g.* COG-UK) (Figure 3A). On the other hand, many countries, including Japan as the top contributor of SARS-CoV-2 genomes shown in Figure 1B, do not have accompanying clinical and/or demographic metadata submitted to GISAID (Figure 3B-3C). Most likely these metadata are stored in a local/national repository but not reported to GISAID to comply with respective data-sharing regulations.

The majority of the Indonesian samples came from hospitalized patients (Table 2). More than half of these sequenced cases in Indonesia belong to the productive age group of 15–44-year-olds, with the second largest group (29%) being 45-64-year-olds (Figure 3B). This data can merely represent the age distribution of the Indonesian population (Table S2), as it is also reflected in the population age distribution of South Korea (Figure 3B, Table S2). However, it is hard to say for countries such as Hong Kong and China as their unknown categories of this metadata are relatively large (Figure 3B). The same assumption could also be applied to the metadata distribution of gender in Indonesia where 50.5% of the infection came from female patients (Figure 3C), in correlation with the 50:50 gender distribution in the Indonesian population (Table S2). However, a study shows that the male sex of higher age with comorbidity may have more prevalence in requiring intensive care treatment and mortality[33,34].

In conclusion, Indonesia has successfully recorded and shared its genomic metadata on GISAID, despite the sporadic origins of the samples. It is notable that with the different national and international rules governing personal information and data sharing, metadata information should best be made available and accessible to selected relevant parties (*e.g.* health ministry, medical professionals, and epidemiological researchers).

This way, the metadata will put the SARS-CoV-2 WGS output into context and draw as correct conclusions as possible to deal with the COVID-19 pandemic.

C. The distribution of variants by PANGO lineages in Indonesia compared to its neighbours and the rest of the world.

Lineage annotation by Phylogenetic Assignment of Named Global Outbreak Lineages (PANGO Lineages; PANGOLIN) presents a dynamic nomenclature of the SARS-CoV-2[35]. Because it labels the variants that are circulating and active, it can track the transmission of SARS-CoV-2 based on sequence differences while also considering new virus diversity[35]. Therefore, we use the PANGOLIN annotations analysed by GISAID and included in the metadata of the respective genomes (Figure 4 and S2). In the following sections, we also use the WHO's nomenclature of Greek alphabets[36], when generally discussing lineages that have become VoCs.

Rambaut *et.al.* hypothesized that lineage A was the founding lineage[35]. However, currently, the B lineage, in particular B.1.1.7 is the most prevalent variant both in Asia (*i.e.* 19.2%) and the world (*i.e.* 43.1%) (Figure 4A). Lineage B.1.1.7 (the Alpha variant[36]) was first detected in Kent in the United Kingdom as early as September 2020[37] and has now become a VoC[36]. Previous data downloaded in early February 2021 shows that the Alpha variant had not been significantly detected in Asia, and only accounted for 8.3% of the sequenced genomes in the world at that time (Figure S2). The most prevalent variants in early February 2021 in Asia and the world were B.1.1.284 and B.1.177, respectively (Figure S2). However, Indonesia had detected the Alpha variant along with two other VoCs, Beta and Delta, in January 2021 (Table 3).

Figure 4B shows the frequency distribution of Indonesia-detected variants as a proportion of the total number of cases in Indonesia (in pink), ASEAN area (in green), Asia continent (in orange) and the world (in purple). Ten of the B.1.1.7 variant had been detected as of 24 April 2021[38] but it has not significantly influenced variant distribution in Indonesia, in contrast to the world's proportion (Figure 4B and S3, upper panels). The two most predominant variants in Indonesia per 1 June 2021 were B.1.466.2 and B.1.470 at about 25% (456 genomes) and 21% (392 genomes), respectively (Figure 4B, lower panel). Moreover, B.1.470 variant was first detected in Indonesia on 9 April 2020 and then spread to neighbouring countries such as Malaysia, South Korea and Japan[39].

All B.1.466.2 variants in Indonesia carried the D614G spike protein mutation (Table 4). The G614 virus was shown to have an increased capacity to enter ACE2-expressing cells[40] and to enhance its replication in human respiratory cells[41] compared to the original D614 virus. These G614 mutation characteristics were hypothesized to confer a moderate level of infectivity and transmissibility, and could be the reasons it maintained its presence in the globally circulating variants[42,43]. The B.1.466.2 variant also frequently carried two other mutations: 1) N439K at 99.1% frequency and 2) P681R at 69.7% frequency (Table 4). Studies on viruses with these two mutations also showed correlation to an increased infectivity, by way of ACE2 receptor binding, and viral transmissibility[43-46]. Moreover, the P681R mutation was frequently detected in the more-infectious B.1.617.2 (Delta) variant[46]. Therefore, the high prevalence of all three mutations in the Indonesian B.1.466.2 variant might have given it a cumulative fitness advantage over the other variants. Furthermore, the B.1.617.2 (Delta) variant carried more mutations in its spike protein including the two mutations found in the B.1.466.2 variant (Table 4). All the Delta variants had mutations in D614G, P681R, T19R, T478K, and L452R (Table 4).

There was another variant, B.50, with only five genomes recorded in Indonesia per 1 June 2021, designated as an Indonesian lineage on PANGOLIN website[47]. B.50 variant was also detected in Japan and Germany, both at around 11% of the world's total[47]. Hence, this variant is not unique to Indonesia. There is so far no variant detected only in Indonesia and not elsewhere in the world (Figure S3). This might mean that the lineages found in Indonesia: 1) accumulated mutations during local transmissions from imported parental variants, or 2) were exported to other countries, particularly its neighbours. It might also be possible that both scenarios occurred intermittently, as border closure for foreigners and quarantine measures for returning citizens were only put into effect from February 2021[48].

D. Phylogenetic relationship based on PANGOLIN exists between different Indonesian variants and those exported to the neighbouring countries.

In Indonesia, the earliest dated sample for sequencing was collected by *Universitas Airlangga* (Airlangga University) in Surabaya on 12 March 2020 (Figure 5A, Table S3). We used this sequence as the root in the phylogenetic tree analysis (Figure 5B). Nonetheless, the first Indonesian cases (*i.e.* a cluster of two) were detected in Jakarta and officially announced on 2 March 2020[10,13,14]. As these patients immediately underwent quarantine and hospitalization, it was less likely that the first sequencing sample was the result of direct transmission from this cluster, only ten days apart. This implies that COVID-19 might have been transmitted and circulated locally and presumably undetected (as modelled by De Salazar *et.al.*[49]), prior to these first official cases.

In terms of Indonesian variants being associated with those in other countries (*i.e.* country of exposure in the metadata was stated to be Indonesia and/or other mentions of link to Indonesia), a few neighbouring countries are on the list: Singapore, Malaysia, South Korea, Taiwan, Japan, and a bit further away, India (Figure 5A, Figure S4). We term these Indonesian-associated variants as exported variants that occurred in an intermittent manner related to travel and/or contact history with Indonesia (Figure 5A, Figure S4). Since the beginning of March 2020, Indonesian-exported variants could be found in Singapore in parallel to the local transmission of these variants within Indonesia (Figure 5A). Japan also shows a similar pattern to a lesser degree (Figure 5A). This could also suggest that an on-going, multi-way transmission patterns have existed between these countries, facilitated by relatively easy access and travel between them, and considering the high level of economic and industrial collaborations. The fact that these associated variants still exist up to recently in these countries highlights the importance of genomic surveillance to inform border control policies during a pandemic. Meanwhile, based on the downloaded data per 1 June 2021, there was only one imported Indonesian case from Saudi Arabia (*i.e.* EPI_ISL_576387)[50]. However, this seemingly rare, imported case was likely due to a low rate of contact tracing in Indonesia.

All the sequenced genomes of SARS-CoV-2 in Indonesia plus the exported variants were analysed to measure the phylogenetic relationship between them (n = 1830). From the total number of exported variants, Singapore holds the highest number of associations at ~55%, followed by Japan at ~29% (table in Figure 5B). A phylogenetic tree rooted to the earliest Indonesian genome is presented in Figure 5B, with variants from Singapore and Japan annotated in red and blue, respectively, and Indonesia in grey (Figure 5B).

The tree shows a relatively even distribution of Singapore- and Japan-exported variants, branching out from the Indonesian lineages (Figure 5B).

In a more general context, associations between imported/exported lineages in different countries are visualized in Figure 5C (data per 1 June 2021) and Figure S5 (data per 1 February 2021). We filtered 3850 genomes that had different country annotations compared to their countries of exposure, or other location information (Figure 5C)[50]. Worldwide associations show that Singapore was the country having most links to variants in other countries, both in a total number of genomes and a total number of linked countries, followed by Japan (Figure 5C). These numbers are not surprising considering Singapore is a busy travel hub in the Asian region as well as the world [51]. Meanwhile, the diverse foreign transmission links in Japan might have been catalysed by the outbreak onboard the Diamond Princess cruise ship in February 2020. Interestingly, the GISAID data that we downloaded in early February 2021, did not have much information on Singapore variants in terms of their associations to other countries (Figure S5). This contrasts with the newest data (Figure 5C). It is possible that new genomic metadata was added within the four months prior to our last data download (*i.e.* 1 June 2021), and/or back-tracked, or both.

Discussion

The importance of WGS to tracing and understanding pathogens has been demonstrated repeatedly during the ongoing COVID-19 pandemic[29,52,53]. However, establishing distributed networks of cost-effective variable through sequencing in low to middle income countries is challenging. For this purpose, ONT is a robust sequencing platform to do WGS and will fit the Indonesian context well. It has a range of outputs that can be suited to the need and budget of each sequencing facility and crucially, samples can be processed in small or large batches.

One of the difficulties in whole-genome sequencing within a pandemic context is sampling, *i.e*. how to obtain an optimum number of samples representing cases in the population. It is clearly better if sequenced samples come from a wide range of subjects rather than only one group or a subset of individuals[54]. In Indonesia, this has been challenging as samples were mostly non-randomly taken from symptomatic and/or hospitalized patients (Table 2). This will likely affect how the pandemic is observed and may not typify the events occurring in the broader community. More recently, the Indonesian government has been trying to distribute sample collection more evenly by gathering specimens from different provinces across the Indonesian vast archipelago[55]. This is a good progress that needs to be maintained and even improved.

In the case of VoCs, the three major variants of Alpha (B.1.1.7), Beta (B.1.351) and Delta (B.1.617.2) have recently been detected in Indonesia, with the Delta variant quickly rising in the days following our data analyses per 1 June 2021[56]. This pattern is also occurring in the UK where the Delta variant is gaining number against the Alpha variant[57,58]. The emergence of new variants that may become VoIs and VoCs highlights the importance of genomic surveillance of the SARS-CoV-2 to control the COVID-19 pandemic. Furthermore, functional studies of variants with high prevalence in Indonesia, such as B.1.466.2, B.1.470, and B.50 can be pursued to understand their biological significance. However, such effort needs to be approached with caution as the prevalence of these variants will change with time as the pandemic progresses.

Lastly, Indonesia must increase its sequencing rate through the newly established COVID-19 WGS consortium, in parallel with testing (Figure S1). It is also important to allocate resources for asymptomatic case research to study transmission patterns by asymptomatic individuals, which could influence case numbers[59-61]. These aspects of epidemiological research (*i.e.* genomics and diagnostics) will help Indonesia be better prepared and build toward sustainable mitigation of outbreaks. In parallel, the good quality of clinical and demographic metadata recording, despite the difficult terrain faced by many sequencing facilities, should be maintained or even improved to maximize research output.

Methods

Sample collection and WGS

Samples were collected from nasopharyngeal and oropharyngeal swabs of patients in viral transport medium[62-65] or DNA/RNA Shield™ (Zymo Research, USA) abides by the Center for Disease Control's (CDC) guidelines for pathogen inactivation[66]. RNA extraction was done using Viral Nucleic Acid Extraction Kit II (Geneaid Biotech Ltd, Taiwan) according to the manufacturers' instructions. All samples were tested for presence of SARS-CoV-2 with an RT–qPCR assay[62-65,67] using Real-Q 2019-nCoV Detection Kit (Rev.2 (2020.03.25); BioSewoom, Korea) according to the manufacturers' instructions in a CFX96 Touch Real-Time PCR Detection System (BioRad, USA) machine. Samples with a Ct value of 11-30 were chosen for sequencing[31].

Sequencing libraries were prepared based on the ARTIC nCoV-2019 sequencing protocol v2 (GunIt) V.2[31], with minor modifications, using reverse-transcribed cDNA. Primers are obtained from IDT (Version 3)[31]. Libraries were loaded on a MinION device and flow cell. ONT's MinKNOW software was used to run sequencing. The Rampart software was applied to monitor coverage of each barcoded sample in real-time by running fast base-calling.

Bioinformatics processes

A validated pipeline was followed to conduct high accuracy base-calling, demultiplexing, trimming, alignment, and consensus establishment[30]. Medaka workflow was used in generating consensus genome sequences due to the speed and GPU compatibility. A minimum of 20X coverage was set in the pipeline[68]. A concatenated consensus sequences of all barcoded samples which also contained the number of ambiguous nucleotides (Ns) were then further analysed. Assignment of clade membership was done by PANGOLIN (version 27-05-2021)[69]. FASTA files of assembled sequences along with metadata in comma-separated values (.csv) files were prepared for submission to GISAID. The metadata included collection date, location, origin of samples, passage history, and sequencing technology. Further analysis for possible missing metadata, FASTA sequences, a certain information, and frameshift mutations was undertaken following the result of GISAID quality control checking after submission[70]. In total of 12 genomes were submitted to GISAID[50].

Lineage and phylogenetic analyses

Full-genome of SARS-CoV-2 sequences from ASEAN member countries, China, Hong Kong, Japan, South Korea, Taiwan, and the United Kingdom along with their metadata were downloaded from the GISAID database[50] on 4 February and again on 1 June 2021.

The metadata already contains annotations of genomes based on PANGOLIN[35]. These lineages were used to sub-divide the genomes in further analyses.

All Indonesian sequences and those transmitted from or historically linked to Indonesia (*i.e.,* Indonesia as the country of exposure) were selected for phylogenetic analysis; only complete genomes with high coverage were chosen for phylogenetic analysis. Up to 1 June 2021, there were 1830 sequences that were aligned and the 5′ and 3′ ends trimmed using MAFFT v.7.475 with automatic flavour selection[71]. Maximum-likelihood phylogenetic trees were produced from the trimmed sequences using IQ-TREE2[72], employing the GTR+F+R3 model of nucleotide substitution as suggested by the software's model finder[73], with 1000 SH-like approximate likelihood ratio test (SH-aLRT) Ultra-Fast Bootstrap[74]. Output tree was rooted using the earliest Indonesian genome, *i.e.* Indonesia/JI-ITD-136N/2020|EPI_ISL_529961|2020-03-12. The tree was then visualized and annotated with FigTree where the nodes were ordered and rooted[75].

References

1. Liu, D. X., Liang, J. Q. & Fung, T. S. Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (Coronaviridae). in *Encyclopedia of Virology* 428–440 (Elsevier, 2021). doi:10.1016/B978-0-12-809633-8.21501-X

2. Severe Acute Respiratory Syndrome (SARS). Available at: https://www.who.int/health-topics/severe-acute-respiratory-syndrome#tab=tab_1. (Accessed: 8th June 2021)

3. Middle East respiratory syndrome coronavirus (MERS-CoV). Available at: https://www.who.int/health-topics/middle-east-respiratory-syndrome-coronavirus-mers#tab=tab_1. (Accessed: 22nd June 2021)

4. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. Available at: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. (Accessed: 22nd June 2021)

5. Naming the coronavirus disease (COVID-19) and the virus that causes it. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it. (Accessed: 22nd June 2021)

6. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

7. Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, co - Nucleotide - NCBI. Available at: https://www.ncbi.nlm.nih.gov/nuccore/MN908947. (Accessed: 22nd June 2021)

8. Chinese researchers reveal draft genome of virus implicated in Wuhan pneumonia outbreak | Science | AAAS. Available at: https://www.sciencemag.org/news/2020/01/chinese-researchers-reveal-draft-genome-virus-implicated-wuhan-pneumonia-outbreak. (Accessed: 22nd June 2021)

9. (No Title). Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200228-sitrep-39-covid-19.pdf. (Accessed: 22nd June 2021)

10. Kilas Balik Kronologi Munculnya Kasus Pertama Covid-19 di Indonesia Halaman all - Kompas.com. Available at:

https://megapolitan.kompas.com/read/2021/03/02/05300081/kilas-balik-kronologi-munculnya-kasus-pertama-covid-19-di-indonesia?page=all. (Accessed: 2nd June 2021)

11. Benarkah Daya Tahan Tubuh dan Iklim Tropis Jadi Penyebab Indonesia Negatif Virus Corona? Halaman all - Kompas.com. Available at: https://www.kompas.com/tren/read/2020/02/14/100706865/benarkah-daya-tahan-tubuh-dan-iklim-tropis-jadi-penyebab-indonesia-negatif?page=all. (Accessed: 22nd June 2021)

12. A Chronicle on the SARS Epidemic. *Chinese Law Gov.* **36**, 12–15 (2003).

13. Indonesia finally reports two coronavirus cases. Scientists worry it has many more | Science | AAAS. Available at: https://www-sciencemag-org.ezproxy.nottingham.ac.uk/news/2020/03/indonesia-finally-reports-two-coronavirus-cases-scientists-worry-it-has-many-more. (Accessed: 2nd July 2021)

14. Setiawaty, V. *et al.* The identification of first COVID-19 Cluster in Indonesia. *Am. J. Trop. Med. Hyg.* **103**, 2339–2342 (2020).

15. Hassan, I., Mukaigawara, M., King, L., Fernandes, G. & Sridhar, D. Hindsight is 2020? Lessons in global health governance one year into the pandemic. *Nature Medicine* **27**, 396–400 (2021).

16. GLOBALIZATION AND DISASTERS: ISSUES OF PUBLIC HEALTH, STATE CAPACITY AND POLITICAL ACTION on JSTOR. Available at: https://www.jstor.org/stable/24358435?seq=1#metadata_info_tab_contents. (Accessed: 8th June 2021)

17. WHO. Special edition: Proposed working definitions of SARS-CoV-2 Variants of Interest and Variants of Concern. *https://www.who.int/publications/m/item/covid-19-weekly-epidemiological-update* 4 (2021). Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20210225_weekly_epi_update_voc-special-edition.pdf. (Accessed: 8th June 2021)

18. Faria, N. R. *et al.* Mobile real-time surveillance of Zika virus in Brazil. *Genome Med.* **8**, 97 (2016).

19. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1266 (2017).

20. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).

21. Stapleton, P. J. *et al.* Evaluating the use of whole genome sequencing for the investigation of a large mumps outbreak in Ontario, Canada. *Sci. Rep.* **9**, (2019).

22. Cyranoski, D. Alarming COVID variants show vital role of genomic surveillance. *Nature* **589**, 337–338 (2021).

23. Mercer, T. R. & Salit, M. Testing at scale during the COVID-19 pandemic. *Nature Reviews Genetics* 1–12 (2021). doi:10.1038/s41576-021-00360-w

24. Kementerian Kesehatan Republik Indonesia. Available at: https://www.kemkes.go.id/article/view/21010900005/pemerintah-tingkatkan-kapasitas-deteksi-genom-virus-sars-cov-2.html. (Accessed: 25th May 2021)

25. COVID-19 Genomics UK Consortium. Available at: https://www.cogconsortium.uk/. (Accessed: 6th April 2021)

26. GISAID - Submission Tracker Global. Available at: https://www.gisaid.org/test5/submission-tracker-global/. (Accessed: 12th June 2021)

27. NICCRAT-Delivering Impact Partnerships - Asia Business Centre. Available at: https://blogs.nottingham.ac.uk/asiabusinesscentre/2021/05/25/niccrat-delivering-impact-partnerships/. (Accessed: 26th June 2021)

28. Nottingham-Indonesia Collaboration in Cancer Research and Training - Asia Business Centre. Available at: https://blogs.nottingham.ac.uk/asiabusinesscentre/2019/10/18/nottingham-indonesia-collaboration-in-cancer-research-and-training/. (Accessed: 6th April 2021)

29. Bull, R. A. *et al.* Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat. Commun.* **11**, 1–8 (2020).

30. Artic Network. Available at: https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html. (Accessed: 26th May 2021)

31. Quick, J. Josh Quick 2020. nCoV-2019 sequencing protocol v2 (GunIt). *protocols.io* (2020). doi:10.17504/protocols.io.bdp7i5rn

32. Schriml, L. M. *et al.* COVID-19 pandemic reveals the peril of ignoring metadata standards. *Scientific Data* **7**, 1–4 (2020).

33. Peckham, H. *et al.* Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. *Nat. Commun.* **11**, 1–10 (2020).

34. Surendra, H. *et al.* Clinical characteristics and mortality associated with COVID-19 in Jakarta, Indonesia: A hospital-based retrospective cohort study. *Lancet Reg. Heal. Pacific* **9**, 100108 (2021).

35. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).

36. Tracking SARS-CoV-2 variants. Available at: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/. (Accessed: 14th June 2021)

37. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology - Virological. Available at: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563. (Accessed: 14th June 2021)

38. 10 Kasus Varian Baru COVID-19 B.1.1.7 ditemukan di Indonesia | Badan Penelitian dan Pengembangan Kesehatan. Available at: https://www.litbang.kemkes.go.id/10-kasus-varian-baru-covid-19-b-1-1-7-ditemukan-di-indonesia/. (Accessed: 14th June 2021)

39. PANGO lineages B.1.470. Available at: https://cov-lineages.org/lineages/lineage_B.1.470.html. (Accessed: 14th June 2021)

40. Zhang, L. *et al.* SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat. Commun.* **11**, 1–9 (2020).

41. Plante, J. A. *et al.* Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).

42. Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G

Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827.e19 (2020).

43. Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology* **19**, 409–424 (2021).

44. Thomson, E. C. *et al.* Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* **184**, 1171-1187.e20 (2021).

45. Shirakawa, K. *et al.* SARS-CoV-2 spike P681R mutation enhances and accelerates viral fusion 1 2 Akatsuki Saito. *Hesham Nasser* **4**, 16 (2021).

46. Cherian, S. *et al.* Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. *bioRxiv* 2021.04.22.440932 (2021). doi:10.1101/2021.04.22.440932

47. PANGO lineages B.50. Available at: https://cov-lineages.org/lineages/lineage_B.50.html. (Accessed: 14th June 2021)

48. Indonesia Travel Restrictions. Available at: https://kemlu.go.id/losangeles/en/news/11727/indonesia-travel-restrictions. (Accessed: 14th June 2021)

49. De Salazar, P. M., Niehus, R., Taylor, A., Buckee, C. & Lipsitch, M. Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases. *medRxiv Prepr. Serv. Heal. Sci.* 2020.02.04.20020495 (2020). doi:10.1101/2020.02.04.20020495

50. GISAID - Initiative. Available at: https://www.gisaid.org/. (Accessed: 6th April 2021)

51. Why Singapore is a Global Logistics Hub – MICE Singapore. Available at: https://www.visitsingapore.com/mice/en/bulletin-board/why-singapore-is-a-global-logistics-hub/overview/. (Accessed: 21st June 2021)

52. Oude Munnink, B. B. *et al.* Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat. Med.* **26**, 1405–1410 (2020).

53. Rockett, R. J. *et al.* Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. *Nat. Med.* **26**, 1398–1404 (2020).

54. Lancet, T. Genomic sequencing in pandemics. *Lancet* **397**, 445 (2021).

55. Short Report Of Working Group Surveillance Of Covid-19 | Badan Penelitian dan Pengembangan Kesehatan | Province Distribution. Available at: https://www.litbang.kemkes.go.id/short-report-of-working-group-surveillance-of-covid-19/. (Accessed: 14th June 2021)

56. SARS-CoV2 Genomic Surveillance Indonesia | Badan Penelitian dan Pengembangan Kesehatan. Available at: https://www.litbang.kemkes.go.id/sars-cov2-genomic-surveillance-indonesia/. (Accessed: 22nd June 2021)

57. Variants: distribution of cases data, 20 May 2021 - GOV.UK. Available at: https://www.gov.uk/government/publications/covid-19-variants-genomically-confirmed-case-numbers/variants-distribution-of-cases-data. (Accessed: 22nd June 2021)

58. Nextstrain / ncov / global. Available at: https://nextstrain.org/ncov/global. (Accessed: 22nd June 2021)

59. Alene, M. *et al.* Magnitude of asymptomatic COVID-19 cases throughout the course of infection: A systematic review and meta-analysis. *PLoS ONE* **16**,

e0249090 (2021).

60. Pollock, A. M. & Lancaster, J. Asymptomatic transmission of covid-19. *The BMJ* **371**, (2020).

61. Subramanian, R., He, Q. & Pascual, M. Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

62. Bruce, E. A. *et al.* Direct RT-qPCR detection of SARS-CoV-2 RNA from patient nasopharyngeal swabs without an RNA extraction step. *PLoS Biol.* **18**, 1–14 (2020).

63. Lu, X. *et al.* US CDC real-time reverse transcription PCR panel for detection of severe acute respiratory syndrome Coronavirus 2. *Emerg. Infect. Dis.* **26**, 1654–1665 (2020).

64. Corman, V. M. *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* **25**, 2000045 (2020).

65. Won, J. *et al.* Development of a laboratory-safe and low-cost detection protocol for SARS-CoV-2 of the Coronavirus Disease 2019 (COVID-19). *Exp. Neurobiol.* **29**, 107–119 (2020).

66. Guidance on the Inactivation or Removal of Select Agents and Toxins for Future Use | Compliance | Federal Select Agent Program. Available at: https://www.selectagents.gov/compliance/guidance/inactivation/index.htm. (Accessed: 26th May 2021)

67. Jung, Y. J. *et al.* Comparative analysis of primer-probe sets for the laboratory confirmation of SARS-CoV-2. *bioRxiv* 2020.02.25.964775 (2020). doi:10.1101/2020.02.25.964775

68. Kemp, S. A. *et al.* SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **65**, (2021).

69. PANGO lineages. Available at: https://cov-lineages.org/pangolin.html. (Accessed: 26th May 2021)

70. Pater, A. A. *et al.* High Throughput Nanopore Sequencing of SARS-CoV-2 Viral Genomes from Patient Samples. *bioRxiv* 2021.02.09.430478 (2021). doi:10.1101/2021.02.09.430478

71. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

72. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

73. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

74. Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

75. FigTree. Available at: http://tree.bio.ed.ac.uk/software/figtree/. (Accessed: 27th May 2021)

Acknowledgements and Funding

We would like to express our gratitude to the labs and institutions involved in WGS of SARS-CoV-2 and deposited their data on GISAID. Specifically, we thank those in Indonesia and the countries whose data are used for this paper. A list of contributing labs whose data were used in phylogenetic analyses is presented in Supplementary Table S3. We also convey our appreciation to those who had given their consents to their samples being used for SARS-CoV-2 research, and Josh Quick for developing and his guidance of the ARTIC Network's protocol for the ONT platform. We thank Puspita Lisdayanti, head of the Research Centre for Biotechnology, LIPI, for facilitating the collaboration. Lastly, we are grateful for the assistance of COVID-19 detection team in LIPI for sample collection and preparation.

This work was funded by the government of Indonesia through the Indonesian Endowment Fund (LPDP-RISPRO) grant scheme. IC was supported by a Wellcome Trust UK grant. SS was funded by Bowel Cancer Research UK.

Competing Interests

We declare that SW and GA are the employees and SS is the CEO/Founder of PathGen Diagnostik Teknologi. AR, AU, MI and WK are unpaid scientific advisors of the company.

Figure Legends

Figure 1. Profiles of SARS-CoV-2 WGS in Indonesia, its neighbouring countries in Asia, and the world. Metadata was downloaded from GISAID[50] per 1 June 2021.

A) Asia only accounts for about 6% of the total global number of genomes submitted in the world.
B) In Asia, Japan is the country with the highest number of submitted genomes.
C) Distribution map of the number of genomes submitted to GISAID in ASEAN and East Asia region

Figure 2. COVID-19 cases and SARS-CoV-2 sequencing rate

A) Confirmed positive number of COVID-19 cases in ASEAN and East Asia region, where Indonesia currently holds the highest number of cases, followed by the Philippines.
B) The rates of WGS of countries in ASEAN and East Asia region, compared to the number of positive cases in Figure 1A; Hong Kong sequenced almost 24% of its positive cases and led the number in the region.

Figure 3. Metadata profiles between ASEAN, Asian countries, and the UK; Indonesia is relatively thorough in collecting and registering its genomic metadata, comparable to those of South Korea.

A) Proportion of sample sources used for sequencing; Indonesia obtained most of its sequencing samples from swab while the UK did not provide this information to GISAID.
B) Distribution of the age groups of patient samples. Indonesia submitted almost complete metadata of this category to GISAID, as well as South Korea.
C) Distribution of the gender groups of patient samples. Indonesia submitted almost complete metadata of this category to GISAID, as well as South Korea and Taiwan.

Figure 4. The distribution of SARS-CoV-2 variants based on PANGOLIN[69] per data downloaded from GISAID on 1 June 2021

A) In Asia, B.1.1.7 and B.1.1.214 are the most common variants found with a total of more than 35% between them (left panel). B.1.1.7 is also the dominant variant of the submitted genomes in the world (right panel).
B) Distributions of Indonesia-detected variants as proportions in Indonesia, ASEAN, Asia, and the rest of the world. Variants B.1.466.2 and B.1.470 (lower panel) were the two most prevalent variants in Indonesia.

Figure 5. Phylogenetic analyses of the locally transmitted and exported Indonesian genomes (n = 1830) based on data downloaded from GISAID on 1 June 2021

A) Timeline of Indonesian-associated variants by country; exported transmissions were primarily within Southeast and East Asia.
B) Phylogenetic tree rooted to the earliest Indonesian genome shows the Indonesian lineages in grey, spread across the tree. Tree was ordered by nodes, with Singapore and Japan represented by red and blue, respectively. Scale of branch length is given.
C) Circular plot shows associations of variants (*i.e.* exported/imported transmissions) between different countries in the world; transmissions from Indonesia were the highest to Singapore and Japan, both in the number of events and variant types.

Table 1 List of institutions in Indonesia and the numbers of submitted genomes to GISAID (Per 1 June 2021)

INSTITUTION	SEQUENCING PLATFORM				
	Illumina only	Illumina + Sanger	ONT	Sanger only	Sub Total
EIJK	314	38	-	-	352
EIJK-NIHRD	22	-	-	-	22
EIJK-UNWAR	315	-	-	-	315
FKUA	70	-	-	-	70
FKUI	-	-	15	-	15
FKUPNVJ	9	-	-	-	9
GSI	-	-	1	1	2
IRCVS-IAARD	-	-	2	-	2
ITD	126	-	-	-	126
LIPI	-	-	77	-	77
MKS	-	-	17	-	17
MRINUPH	-	-	-	17	17
NIHRD	473	7	-	15	495
RSDS-RCVTD-UNAIR	-	-	10	-	-
SB	38	-	-	-	38
SHSIU	-	-	14	-	14
TFRIC	14	-	2	-	16
TUH	-	-	11	-	11
UGM	66	-	-	-	66
UNS	10	-	-	-	10
UNTAN	5	-	-	-	5
USU	23	-	-	-	23
WJHL-ITB	100	-	-	-	100
TOTAL	**1585**	**45**	**149**	**33**	**1812**

Table 2 Clinical metadata of genomes in Indonesia, Southeast and East Asia (based on data downloaded from GISAID per 1 June 2021)

Country	Live	Deceased	Not Hospitalized/ Asymptomatic	Hospitalized/ Symptomatic	Total No. of Genomes	% Genomes with Metadata
Brunei	0	0	0	0	10	0.0%
Cambodia	278	0	0	0	281	98.9%
China	239	1	2	153	1,220	32.4%
Hong Kong	601	26	0	323	2,837	33.5%
Indonesia	811	25	26	477	1,806	74.1%
Japan	114	16	62	9,114	50,241	18.5%
Malaysia	476	7	3	64	745	73.8%
Myanmar	2	0	0	0	41	4.9%
Philippines	3,593	73	0	11	4,295	85.6%
Singapore	0	0	0	553	2,576	21.5%
South Korea	25	9	0	0	6,216	0.5%
Taiwan	25	0	0	74	193	51.3%
Thailand	8	0	0	3	1,404	0.8%
Timor-Leste	0	0	0	0	19	0.0%
Vietnam	79	0	0	3	158	51.9%

Table 3 The first detection dates of variants of concern (VoCs) in Indonesia per 1 June 2021

Variant of Concern	Number of Genomes	First Detected Case	
		Accession ID	Date
B.1.1.7 (Alpha)	23	hCoV-19/Indonesia/SS-NIHRD-WGS00427/2021	5/01/2021
B.1.351 (Beta)	4	hCoV-19/Indonesia/BA-NIHRD-WGS00725/2021	25/01/2021
B.1.617.2 (Delta)	29	hCoV-19/Indonesia/JK-NIHRD-WGS00007/2021	7/01/2021

Table 4 Most frequent mutations of SARS-CoV-2 spike protein in B.1.466.2, B.1.617.2 (Delta) and all other variants in Indonesia

Amino Acid Change	B.1.466.2	B.1.617.2 (Delta)	All Variants
D614G	100.0%	100.0%	94.2%
N439K	99.1%	0%	26.5%
P681R	69.7%	100.0%	21.0%
T19R	0%	100.0%	0%
T478K	0%	100.0%	0%
L452R	0%	100.0%	0%

Figure 1

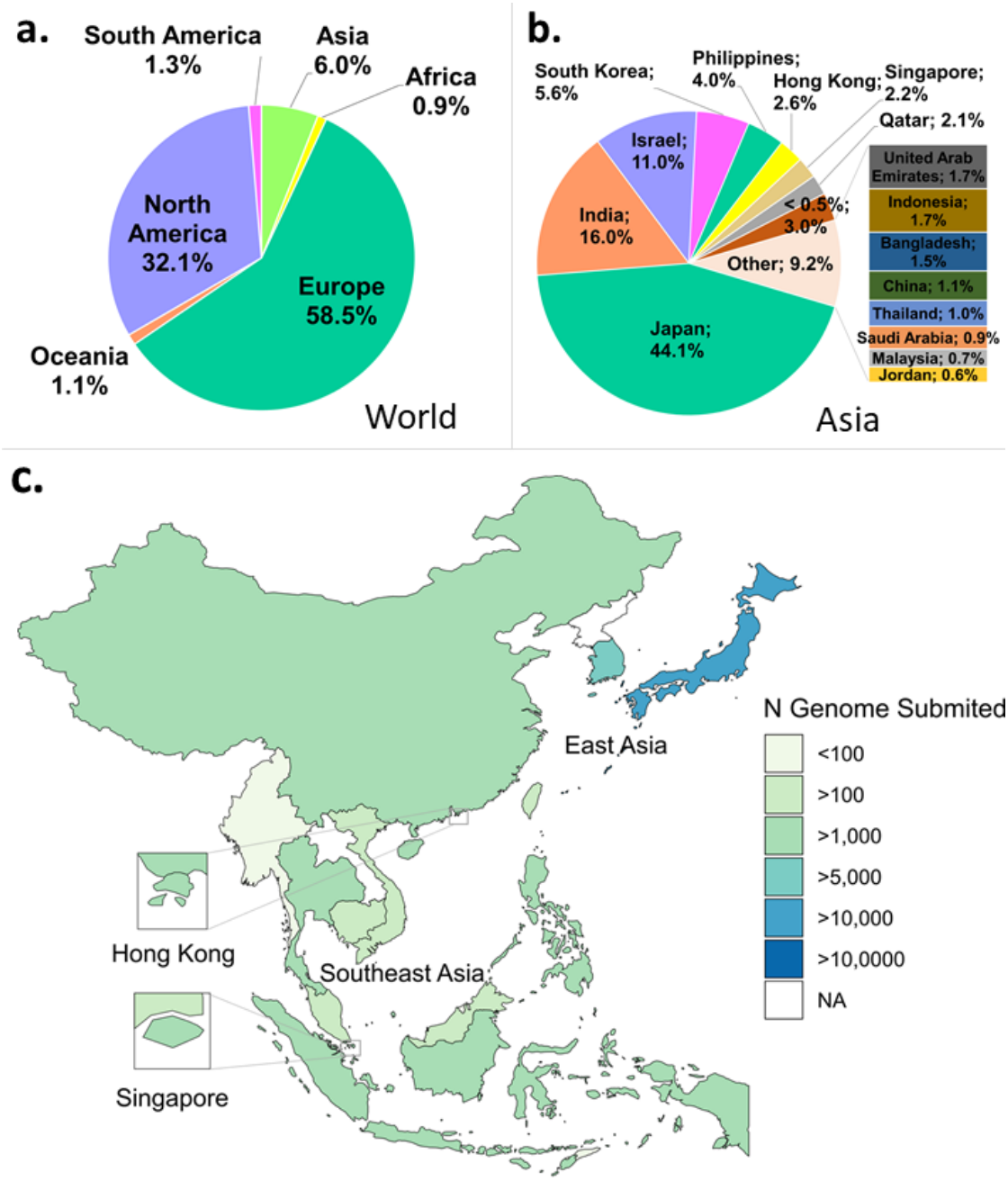

a.
South America 1.3%
Asia 6.0%
Africa 0.9%
North America 32.1%
Europe 58.5%
Oceania 1.1%
World
b.
South Korea; 5.6%
Philippines; 4.0%
Hong Kong; 2.6%
Singapore; 2.2%
Qatar; 2.1%
Israel; 11.0%
India; 16.0%
< 0.5%; 3.0%
Other; 9.2%
Japan; 44.1%
United Arab Emirates; 1.7%
Indonesia; 1.7%
Bangladesh; 1.5%
China; 1.1%
Thailand; 1.0%
Saudi Arabia; 0.9%
Malaysia; 0.7%
Jordan; 0.6%
Asia
c.
East Asia
Hong Kong
Southeast Asia
Singapore
N Genome Submited
<100
>100
>1,000
>5,000
>10,000
>10,0000
NA

Figure 2

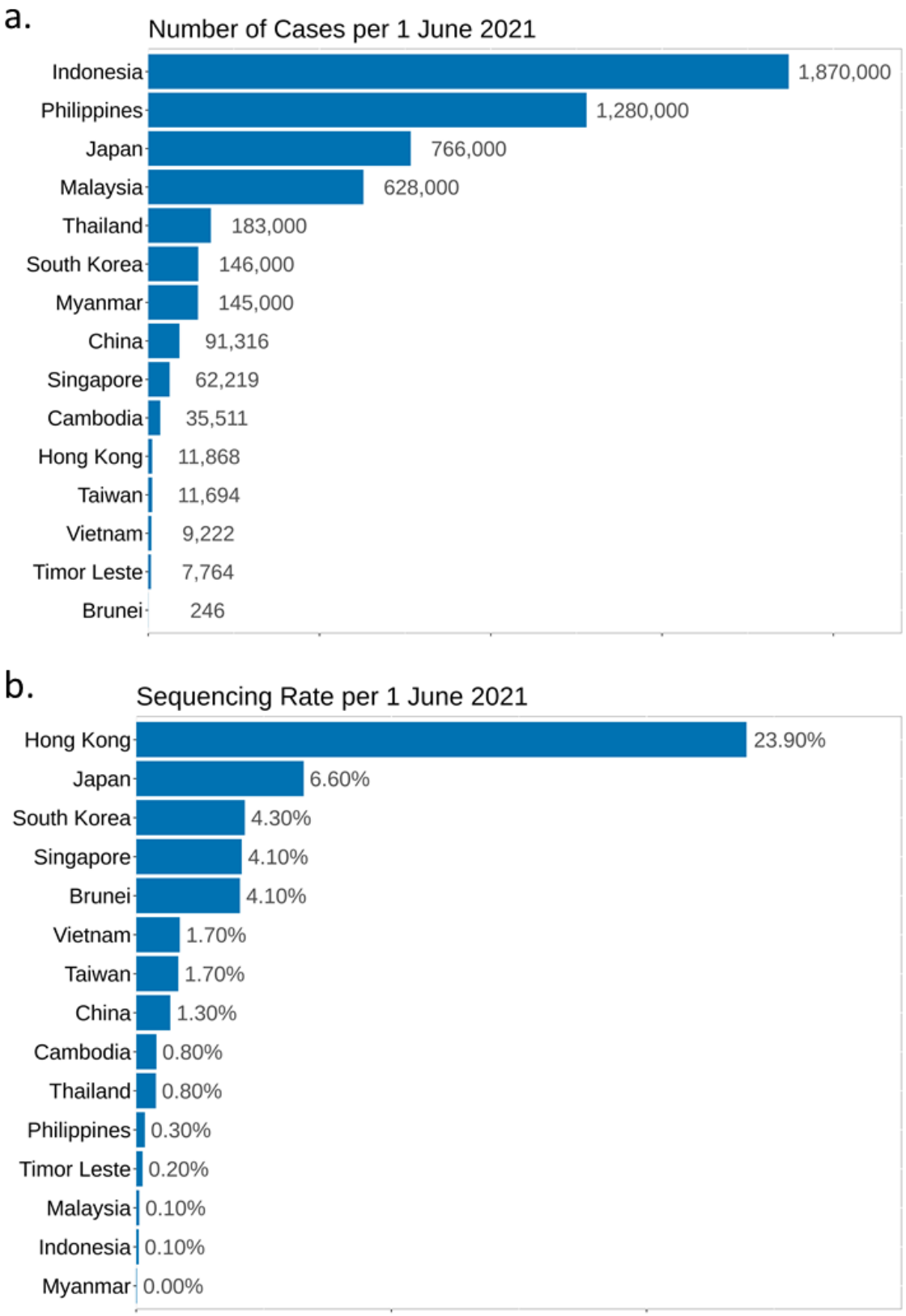

a.
Number of Cases per 1 June 2021
Indonesia 1,870,000
Philippines 1,280,000
Japan 766,000
Malaysia 628,000
Thailand 183,000
South Korea 146,000
Myanmar 145,000
China 91,316
Singapore 62,219
Cambodia 35,511
Hong Kong 11,868
Taiwan 11,694
Vietnam 9,222
Timor Leste 7,764
Brunei 246
b.
Sequencing Rate per 1 June 2021
Hong Kong 23.90%
Japan 6.60%
South Korea 4.30%
Singapore 4.10%
Brunei 4.10%
Vietnam 1.70%
Taiwan 1.70%
China 1.30%
Cambodia 0.80%
Thailand 0.80%
Philippines 0.30%
Timor Leste 0.20%
Malaysia 0.10%
Indonesia 0.10%
Myanmar 0.00%

Figure 3

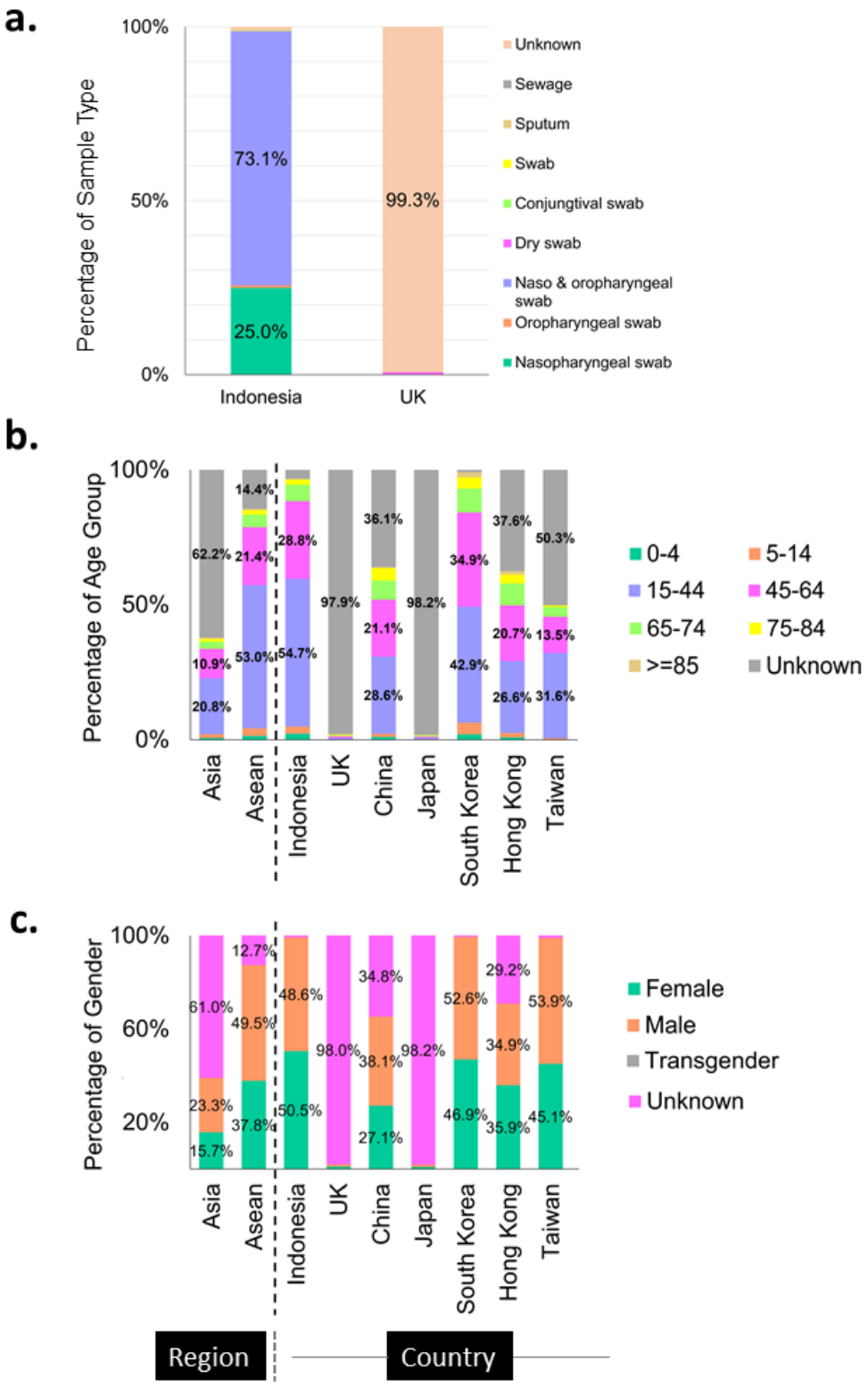

a.
Percentage of Sample Type
100%
50%
0%
73.1%
25.0%
99.3%
Indonesia
UK
Unknown
Sewage
Sputum
Swab
Conjungtival swab
Dry swab
Naso & oropharyngeal swab
Oropharyngeal swab
Nasopharyngeal swab
b.
Percentage of Age Group
100%
50%
0%
Asia
Asean
Indonesia
UK
China
Japan
South Korea
Hong Kong
Taiwan
0-4
5-14
15-44
45-64
65-74
75-84
>=85
Unknown
c.
Percentage of Gender
100%
60%
20%
Female
Male
Transgender
Unknown
Region
Country

Figure 4

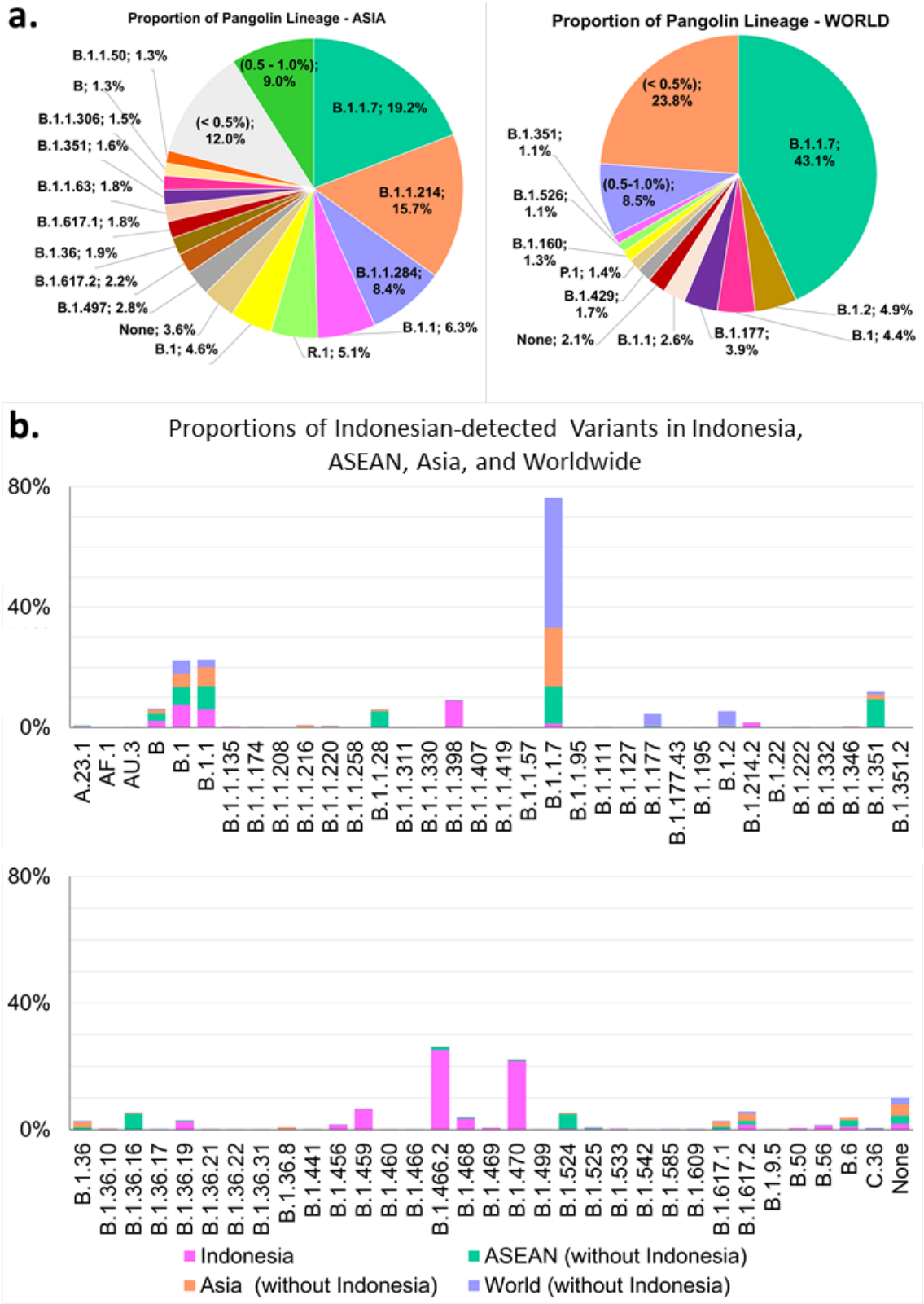
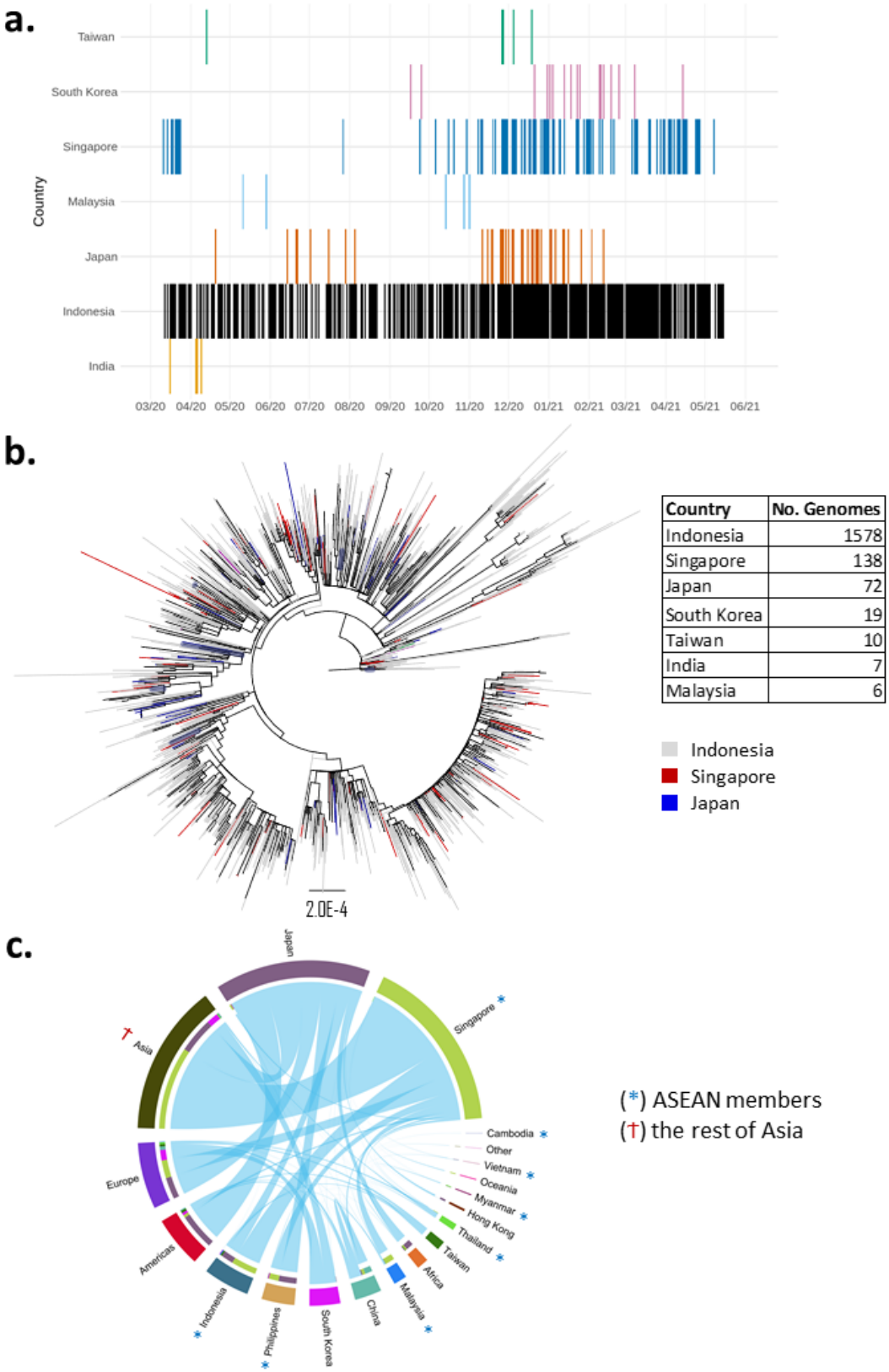

a.
Proportion of Pangolin Lineage - ASIA
B.1.1.7; 19.2%
B.1.1.214; 15.7%
B.1.1.284; 8.4%
B.1.1; 6.3%
R.1; 5.1%
B.1; 4.6%
None; 3.6%
B.1.497; 2.8%
B.1.617.2; 2.2%
B.1.36; 1.9%
B.1.617.1; 1.8%
B.1.1.63; 1.8%
B.1.351; 1.6%
B.1.1.306; 1.5%
B; 1.3%
B.1.1.50; 1.3%
(0.5 - 1.0%); 9.0%
(< 0.5%); 12.0%
Proportion of Pangolin Lineage - WORLD
B.1.1.7; 43.1%
B.1.2; 4.9%
B.1; 4.4%
B.1.177; 3.9%
B.1.1; 2.6%
None; 2.1%
B.1.429; 1.7%
P.1; 1.4%
B.1.160; 1.3%
B.1.526; 1.1%
B.1.351; 1.1%
(0.5-1.0%); 8.5%
(< 0.5%); 23.8%
b.
Proportions of Indonesian-detected Variants in Indonesia, ASEAN, Asia, and Worldwide
80%
40%
0%
A.23.1
AF.1
AU.3
B
B.1
B.1.1
B.1.1.135
B.1.1.174
B.1.1.208
B.1.1.216
B.1.1.220
B.1.1.258
B.1.1.28
B.1.1.311
B.1.1.330
B.1.1.398
B.1.1.407
B.1.1.419
B.1.1.57
B.1.1.7
B.1.1.95
B.1.111
B.1.127
B.1.177
B.1.177.43
B.1.195
B.1.2
B.1.214.2
B.1.22
B.1.222
B.1.332
B.1.346
B.1.351
B.1.351.2
80%
40%
0%
B.1.36
B.1.36.10
B.1.36.16
B.1.36.17
B.1.36.19
B.1.36.21
B.1.36.22
B.1.36.31
B.1.36.8
B.1.441
B.1.456
B.1.459
B.1.460
B.1.466
B.1.466.2
B.1.468
B.1.469
B.1.470
B.1.499
B.1.524
B.1.525
B.1.533
B.1.542
B.1.585
B.1.609
B.1.617.1
B.1.617.2
B.1.9.5
B.50
B.56
B.6
C.36
None
Indonesia
ASEAN (without Indonesia)
Asia (without Indonesia)
World (without Indonesia)

Figure 5



Country	No. Genomes
Indonesia	1578
Singapore	138
Japan	72
South Korea	19
Taiwan	10
India	7
Malaysia	6