The Impact of SARS-CoV-2 Multiple Spike Protein Mutations on COVID-19 Outcomes

Gunadi (✉ drgunadi@ugm.ac.id)

Universitas Gadjah Mada/Dr. Sardjito Hospital

Mohamad Saifudin Hakim

Universitas Gadjah Mada

Hendra Wibawa

Disease Investigation Center, Wates, Yogyakarta, Ministry of Agriculture

Marcellus

Universitas Gadjah Mada

Ika Trisnawati

Universitas Gadjah Mada/Dr. Sardjito Hospital

Endah Supriyati

Universitas Gadjah Mada

Afiahayati

Universitas Gadjah Mada

Riat El Khair

Universitas Gadjah Mada/Dr. Sardjito Hospital

Kristy Iskandar

Universitas Gadjah Mada/ UGM Academic Hospital

Siswanto

Universitas Gadjah Mada/ UGM Academic Hospital

Irene

Balai Besar Teknik Kesehatan Lingkungan dan Pengendalian Penyakit

Nungki Anggorowati

Universitas Gadjah Mada/Dr. Sardjito Hospital

Edwin Widyanto Daniwijaya

Universitas Gadjah Mada/ UGM Academic Hospital

Dwi Aris Agung Nugrahaningsih

Universitas Gadjah Mada

Yunika Puspadewi

Universitas Gadjah Mada/Dr. Sardjito Hospital

Susan Simanjaya

Universitas Gadjah Mada

Dyah Ayu Puspitarani

Universitas Gadjah Mada

Hana Fauzyyah Hanifin

Universitas Gadjah Mada

Alvina Alexandra Setiawan

Universitas Gadjah Mada

Irene Tania

Universitas Gadjah Mada

Cita Shafira Amalia

Universitas Gadjah Mada

I Putu Aditio Artayasa

Universitas Gadjah Mada

Haries Rachman

Universitas Gadjah Mada

Herdianto Mulyawan

Disease Investigation Center, Wates, Yogyakarta, Ministry of Agriculture

Nur Rahmi Ananda

Universitas Gadjah Mada/Dr. Sardjito Hospital

Eggi Arguni

Universitas Gadjah Mada/Dr. Sardjito Hospital

Titik Nuryastuti

Universitas Gadjah Mada

Tri Wibawa

Universitas Gadjah Mada

Research Article

Keywords: COVID-19 outcomes, comorbidity, genomic surveillance, multiple spike protein mutations, phylogenetic analysis, SARS-CoV-2 variants, whole genome sequencing

Posted Date: July 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-700897/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Recent studies focusing on the association of SARS-CoV-2 variants of concern (VOC) on COVID-19 outcomes have been reported. However, studies of the impact of multiple mutations within the spike (S) protein of SARS-CoV-2 on COVID-19 illness are limited. This study determined the association between multiple mutations within the S protein, prognosis factors, and the disease outcomes of SARS-CoV-2 infection.

Methods: We included 51 COVID-19 patients from Yogyakarta and Central Java, Indonesia. Whole genome sequences of SARS-CoV-2 were determined by the Illumina MiSeq next-generation sequencer, followed by the phylogenetic analysis of 170 full-genomes of SARS-CoV-2 from different regions. We analyzed characteristics of COVID-19 patients and multiple mutations in association with different outcomes.

Results: Among 51 patients, the clinical manifestations of COVID-19 were as follows: without any symptoms (13.7%), mild (47%), moderate (19.6%), severe (4%), critical (2%), and died (13.7%). The age of hospitalized patients (53.4 ± 18 years) was higher than non-hospitalized patients (34.6 ± 19) ($p=0.001$). A significant association between diabetes, hypertension, and anticoagulant and the hospitalization of patients was noted with p -value of 0.039 (OR=4.47 [95% CI=1.07-18.58]), 0.001 (OR=17 [95% CI=2-144]), and 0.02 (OR=27.97 [95% CI=1.54-507.13]), respectively; whereas a strong association between patients' age, diabetes, anticoagulant, and steroid with the mortality of patients was revealed with p -value of 0.016 (OR=8.44 [95% CI=1.5-47.49]), 0.019 (OR=8.5 [95% CI=1.43-50.66]), 0.001 (46.8 [95% CI=4.63-472.77]), and 0.009 (OR=15.75 [95% CI=2-123.86]), respectively. All viruses contained the D614G variant, except one case. Accordingly, the samples were classified as the following clade: L (2%), GH (84.3%), GR (11.7%), and O (2%). Besides the D614G, the most common variants in the S protein were L5F (18.8%), V213A (18.8%), and S689R (8.3%). There was no significant association between multiple S protein variants with either hospitalization or mortality of COVID-19 ($p=0.11$ and 0.69, respectively). Multivariate analysis showed that hypertension and anticoagulant were the strong factors affecting the hospitalization and mortality of patients with COVID-19 with a p -value of 0.009 (OR=17.06 [95% CI=2.02-144.36]) and 0.001 (OR=46.8 [95% CI=4.63-472.77]), respectively. Interestingly, the multiple S protein variants almost reached a significant level affecting the hospitalization of patients ($p=0.07$). Phylogenetic analysis showed that although most of the viruses from this study belonged to clade GH, none were detected as the variant of concern (VOC) and the variant of interest (VOI) of SARS-CoV-2.

Conclusions: Here, we show for the first time the association between SARS-CoV-2 mutations within the S protein besides the VOC with the COVID-19 outcomes. Our findings suggest that multiple mutations in the S protein might affect the severity of COVID-19. Our study further suggests the importance of genomic surveillance to monitor SARS-CoV-2 variants, particularly those that might influence the outcomes of COVID-19 patients.

Introduction

After one year of the COVID-19 pandemic, SARS-CoV-2 has infected approximately 185 million people and causes 4 million deaths worldwide [1,2]. Indonesia has documented 2,379,397 COVID-19 cases and 62,908 deaths on July 7, 2021, and become the highest cases country in the South-East Asian region [3].

The outcome of SARS-CoV-2 infection is determined by multiple factors, including the viral and host genetics and age and comorbidities [4,5]. It is hypothesized that the host genetic factors might influence the outcome of SARS-CoV-2 infection. Three genes encoding the angiotensin-converting enzyme 2 (ACE2), the human leukocyte antigen (HLA), toll-like receptor, and complement pathway are suggested to be the primary determinant of COVID-19 outcomes [6]. For viral genetic factors, a previous study indicated that variations within the ORF1ab (4715L) and S protein (614G) had a significant positive correlation with fatality rates of COVID-19 [7]. The viral mutation may affect the presentation to MHC-I and MHC-II and consequently determine the magnitude of cellular immune responses [8]. The emergence of variants of concern (VOC) has attracted public health authorities to assess its impact on clinical presentation and severity. Indeed, the currently known VOCs (alpha, beta, gamma and delta) have been associated with a possible increased risk of hospitalization and disease severity [9].

SARS-CoV-2 has continuously and rapidly spread worldwide, providing a high opportunity for mutation events, especially on the S protein. However, the studies of the impact of multiple mutations within the spike (S) protein of SARS-CoV-2 on COVID-19 illness are limited. We determined the association between the SARS-CoV-2 mutations within the S protein identified from non-VOC isolates and the COVID-19 outcomes.

Material And Methods

RNA extraction and whole genome sequencing

RNA was extracted from all samples of COVID-19 patients from Yogyakarta and Central Java provinces using QiAMP Viral RNA mini kit (Qiagen, Hilden, Germany), followed by real-time polymerase chain reaction (RT-PCR) using Real-Q 2019-nCoV Detection Kit (BioSewoom, Seoul, South Korea) with LightCycler® 480 Instrument II (Roche Diagnostics, Mannheim, Germany).

The double-stranded cDNA was synthesized using Maxima H Minus Double-Stranded cDNA Synthesis (Thermo Fisher Scientific, MA, United States), followed by purification of cDNA using a GeneJET PCR Purification Kit (Thermo Fisher Scientific, MA, United States) and library preparations using the Nextera DNA Flex for Enrichment using Respiratory Virus Oligos Panel. Next-generation sequencing (NGS) was performed to sequence the whole-genome of SARS-CoV-2 using the Illumina MiSeq instrument (Illumina, San Diego, CA, United States) with Illumina MiSeq reagents v3 150 cycles (2 x 75 cycles).

The assembly of our sample genomes was mapped into the reference genome from Wuhan, China (hCoV-19/Wuhan/Hu-1/2019, GenBank accession number: NC_045512.2) using Burrow-Wheeler Aligner (BWA) algorithm embedded in UGENE v. 1.30 [10]. Single nucleotide polymorphisms (SNPs) were identified using the number of high confidence base calls (consensus sequence variations of the

assembly) that disagree with the reference bases for the genome position of interest, followed by exporting all SNPs a vcf file and visualizing them in MS Excel.

Phylogenetic study

For the phylogenetic study, we utilized a dataset of 170 available SARS-CoV-2 genomes from our region and other countries that were retrieved from GISAID (Acknowledgment Table is provided in Supplementary Table 1), followed by multiple nucleotide sequence alignment using the MAFFT program (<https://mafft.cbrc.jp/alignment/server/>). Neighbour Joining statistical method with 1,000 bootstrap replications was used to construct a phylogenetic tree from 29,563 nt length of the open reading frame (ORF) of SARS-CoV-2 virus genome [11,12]. The Kimura 2-parameter method and the gamma distribution with estimated shape parameter (α) for the dataset were utilized to compute the evolutionary distances and model the rate variation among sites, respectively [13]. DAMBE version 7 [14] was used to calculate the estimation of the α gamma distribution, while MEGA version 10 (MEGA X) [15] was utilized for all other analyses.

Patients and prognostic factors

This study was a retrospective study. We included all patients with COVID-19 from Yogyakarta and Central Java provinces, Indonesia, who sent their samples for whole genome sequencing into our institution from June to October 2020. The exclusion criteria were incomplete medical records.

Various clinical manifestations of COVID-19 have been noted, including asymptomatic until pneumonia with varying degrees. The degree of pneumonia of COVID-19 was classified according to the WHO classifications: 1) mild, without evidence of hypoxia or pneumonia; 2) moderate, pneumonia but not severe; 3) severe, pneumonia plus one of the following signs: respiratory rate > 30 breaths/minute (or based on age for children), severe respiratory distress, or $\text{SpO}_2 < 90\%$ in room air; and 4) critical, Acute Respiratory Distress Syndrome (ARDS), sepsis, or septic shock, or other complications [16–18].

We determined several prognostic factors to be associated with the outcomes of patients with COVID-19 according to previous reports, including age, sex, comorbidity, smoking and treatment [5,17]. We classified the outcomes of patients into two groups: hospitalized vs non-hospitalized and survived vs died.

The Medical and Health Research Ethics Committee of the Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/Dr. Sardjito Hospital approved this work (KE/FK/0563/EC/2020). All participants or guardians signed written informed consent for participating in this study.

Statistical analysis

The data were presented as frequency (percentage) and mean $\pm$ SD. The association between prognostic factors and outcomes was analyzed using Chi-square or Fisher exact tests with 95% confidence interval (CI), followed by a multivariate logistic regression test. The association was considered significant if the p -values of < 0.05 . The IBM Statistical Package for the Social Sciences (SPSS) version 21 (Chicago, USA) was used to perform statistical analysis.

Results

Association between prognostic factors and hospitalization of patients with COVID-19.

Among 51 patients, the clinical manifestations of COVID-19 were as follows: without any symptoms (13.7%), mild (47%), moderate (19.6%), severe (4%), critical (2%), and died (13.7%). The age of hospitalized patients (53.4 ± 18 years) was higher than non-hospitalized patients (34.6 ± 19) ($p = 0.001$) (Table 1).

Table 1
Association between prognostic factors and hospitalization of patients with COVID-19.

Characteristics	All (n = 51)	Hospitalized (n = 29) (N, %; mean $\pm$ SD)	Non-hospitalized (n = 22) (N, %; mean $\pm$ SD)	p-value	OR (95% CI)
RT-PCR Ct value		20.3 $\pm$ 4.2	18.9 $\pm$ 3.9	0.26	
Age (years)		53.4 $\pm$ 18	34.6 $\pm$ 19	0.001*	3.81 (0.72–20.16)
■ $\geq$ 65	10	8 (27.6)	2 (9.1)	0.12	
■ < 65	41	21 (72.4)	20 (90.9)		
Sex					
■ Male	30	19 (65.5)	11 (50)	0.27	1.9 (0.61–5.9)
■ Female	21	10 (34.5)	11 (50)		
Comorbidity					
■ Obesity	3	3 (10.3)	0	0.25	5.94 (0.29–121.31)
■ Diabetes	15	12 (41.4)	3 (13.6)	0.039*	4.47 (1.07–18.58)
■ Hypertension	14	13 (44.8)	1 (4.5)	0.001*	17 (2–144)
■ Cardiovascular disease	9	8 (27.6)	1 (4.5)	0.06	8 (0.92–69.72)
■ Chronic kidney disease	2	2 (6.9)	0	0.37	4.09 (0.19–89.65)
Smoking	4	1 (3.4)	3 (13.6)	0.21	0.23 (0.02–2.34)
Therapy					
■ ACEI/ARB	4	4 (13.8)	0	0.17	7.94 (0.33–66.14)
■ Anticoagulant	11	11 (37.9)	0	0.02*	27.97 (1.54–507.13)
■ Steroid	5	5 (17.2)	0	0.12	10.1 (0.53–193.23)
*, significant ($p < 0.05$); ACEI, angiotensin-converting enzyme inhibitors; ARB angiotensin receptor blocker; CI, confidence interval; OR, odds ratio					

Association between prognostic factors and mortality of patients with COVID-19

A significant association between diabetes, hypertension, and anticoagulant and the hospitalization of patients was found with p -value of 0.039 (OR = 4.47 [95% CI = 1.07–18.58]), 0.001 (OR = 17 [95% CI = 2–144]), and 0.02 (OR = 27.97 [95% CI = 1.54–507.13]), respectively (Table 1). A strong association between patients' age, diabetes, anticoagulant, and steroid and the mortality of patients was revealed with p -value of 0.016 (OR = 8.44 [95% CI = 1.5–47.49]), 0.019 (OR = 8.5 [95% CI = 1.43–50.66]), 0.001 (46.8 [95% CI = 4.63–472.77]), and 0.009 (OR = 15.75 [95% CI = 2–123.86]), respectively (Table 2).

Table 2
Association between prognostic factors and mortality of patients with COVID-19.

Characteristics	All (n = 51)	Died (n = 7) (N, %; mean $\pm$ SD)	Survived (n = 44) (N, %; mean $\pm$ SD)	p-value	OR (95% CI)
RT-PCR Ct value		18.7 $\pm$ 5.0	19.9 $\pm$ 3.9	0.57	
Age (years)		66.8 $\pm$ 14	41.8 $\pm$ 19	0.002*	8.44 (1.5–47.49)
■ $\geq$ 65	10	4 (57.1)	6 (13.6)	0.016*	
■ < 65	41	3 (42.9)	38 (86.4)		
Sex				0.69	1.9 (0.33–10.88)
■ Male	30	5 (71.4)	25 (56.8)		
■ Female	21	2 (28.6)	19 (43.2)		
Comorbidity					
■ Obesity	3	1 (14.3)	2 (4.5)	0.34	3.5 (0.27–44.75)
■ Diabetes	15	5 (71.4)	10 (22.7)	0.019*	8.5 (1.43–50.66)
■ Hypertension	14	4 (57.1)	10 (22.7)	0.07	4.53 (0.87–23.72)
■ Cardiovascular disease	9	3 (42.9)	6 (13.6)	0.08	4.75 (0.84–26.71)
■ Chronic kidney disease	2	1 (14.3)	1 (2.3)	0.18	7.17 (0.39–130.31)
Smoking	4	0	4 (9.1)	0.74	0.6 (0.03–12.34)
Therapy					
■ ACEI/ARB	4	2 (28.6)	2 (4.5)	0.05	8.4 (0.96–73.43)
■ Anticoagulant	11	6 (85.7)	5 (11.4)	0.001*	46.8 (4.63–472.77)
■ Steroid	5	3 (42.9)	2 (4.5)	0.009*	15.75 (2–123.86)
*, significant ($p < 0.05$); ACEI, angiotensin-converting enzyme inhibitors; ARB angiotensin receptor blocker; CI, confidence interval; OR, odds ratio					

Molecular and phylogenetic analysis

All viruses contained the D614G variant, except one isolate. Accordingly, the samples were classified as the following clade: L (2%), GH (84.3%), GR (11.7%), and O (2%). Besides the D614G mutation, the most common mutation in the S protein was L5F (18.8%), V213A (18.8%), and S689R (8.3%) (Table 3).

Phylogenetic analysis showed that although most virus samples belonged to the clade GH and followed by the clade GR, none of these was detected as the variant of concern (VOC) and the variant of interest (VOI) of SARS-CoV-2 (Fig. 1). While two viruses of clade L and O formed separate clusters to the GH and GR clades.

Association between multiple S protein variants with COVID-19 patients' outcomes and prognostic factors

There was no significant association between multiple spike protein with either hospitalization or mortality of COVID-19 patients ($p = 0.11$ and 0.69 , respectively) (Table 4). Moreover, none of the prognostic factors was associated with multiple spike protein mutations ($p > 0.05$) (Table 4).

Table 4
Association between multiple spike protein mutations with outcomes of patients with COVID-19 and prognostic factors

Variables	S protein mutation		OR (95% CI)	p-value
	Multiple (N, %; mean $\pm$ SD)	None/Single (N, %; mean $\pm$ SD)		
Hospitalized			0.36 (0.11–1.25)	0.11
■ Yes	16 (48.5)	13 (72.2)		
■ No	17 (51.5)	5 (27.8)		
Survival			1.43 (0.25–8.23)	0.69
■ Died	5 (15.2)	2 (11.1)		
■ Live	28 (84.8)	16 (88.9)		
RT-PCR Ct value	19.2 $\pm$ 3.7	20.6 $\pm$ 4.7		0.26
Age (years)			1.35 (0.30-6.0)	0.70
■ $\geq$ 65	7 (21.2)	3 (16.7)		
■ < 65	26 (78.8)	15 (83.3)		
Sex	19 (57.6)	11 (61.1)	0.86 (0.27–2.79)	0.81
■ Male	14 (42.4)	7 (38.9)		
■ Female				
Comorbidity				
■ Obesity	2 (6.1)	1 (5.6)	1.10 (0.09-13.0)	0.94
■ Diabetes	11 (33.3)	4 (22.2)	1.75 (0.46–6.59)	0.41
■ Hypertension	6 (18.2)	8 (44.4)	0.28 (0.08-1.0)	0.05
■ Cardiovascular disease	6 (18.2)	3 (16.7)	1.11 (0.24–5.09)	0.89
■ Chronic kidney disease	2 (6.1)	0	2.94 (0.14–64.55)	0.49
Smoking	2 (6.1)	2 (11.1)	0.51 (0.07–4.01)	0.53

Variables	S protein mutation		OR (95% CI)	<i>p</i> -value
Therapy				
■ ACEI/ARB	3 (9.1)	1 (5.6)	1.7 (0.16–17.65)	0.66
■ Anticoagulant	7 (21.2)	4 (22.2)	0.94 (0.23–3.78)	0.93
■ Steroid	4 (12.1)	1 (5.6)	2.34 (0.24–22.73)	0.46
*, significant; ACEI, angiotensin-converting enzyme inhibitors; ARB, angiotensin receptor blocker; CI, confidence interval; OR, odds ratio				

Multivariate analysis

Multivariate analysis showed that hypertension and anticoagulant were the substantial factors affecting the hospitalization and mortality of patients with COVID-19 with a *p*-value of 0.009 (OR = 17.06 [95% CI = 2.02-144.36]) and 0.001 (OR = 46.8 (95% CI = 4.63-472.77), respectively. Interestingly, the multiple S protein mutations almost reached a significant level affecting the hospitalization of patients (*p* = 0.07) with the OR of 4.64 (95% CI = 0.87–24.68) (Table 5).

Table 5
Multivariate analysis of the association between prognostic factors and outcomes of patients with COVID-19

Prognostic Factor	Hospitalized		Mortality	
	OR (95% CI)	p-value	OR (95% CI)	p-value
Multiple S protein mutations	4.64 (0.87–24.68)	0.07	0.91 (0.04–22.85)	0.96
Age (≥ 65 years)	0.10 (0.004–3.07)	0.19	4.56 (0.01–2267.77)	0.63
Sex (male)	2.5 (0.5–12.6)	0.24	3.45 (0.01–941.16)	0.67
Comorbidity				
■ Obesity	-	1	0.05 (0.0001–23.23)	0.33
■ Diabetes	2.74 (0.33–22.76)	0.35	14.27 (0.16–1286.04)	0.25
■ Hypertension	17.06 (2.02–144.36)	0.009	3.44 (0.03–386.31)	0.61
■ Cardiovascular disease	5.52 (0.18–168.92)	0.33	1.31 (0.01–340.36)	0.92
■ Chronic kidney disease	-	1	3.68 (0.04–353.38)	0.58
Smoking	6.72 (0.23–197.53)	0.27	-	1
Therapy				
■ ACEI/ARB	-	1	13.69 (0.02–11919.51)	0.45
■ Anticoagulant	-	1	46.8 (4.63–472.77)	0.001
■ Steroid	-	1	43.96 (0.05–41926.76)	0.28
*, significant ($p < 0.05$); ACEI, angiotensin-converting enzyme inhibitors; ARB, angiotensin receptor blocker; CI, confidence interval; OR, odds ratio				

Discussion

Our study is able to show the effect of hypertension and the use of anticoagulants on the severity of COVID-19 patients from the Indonesian population. The patients with hypertension have a ~ 17-fold higher risk of hospitalization than those without hypertension, in line with previous reports [17,19]. The effect of hypertension on COVID-19 severity is controversial [20]. The pathogenesis of hypertension affecting the COVID-19 severity is complex [20]. The impact of hypertension on the severity of COVID-19 is significant when accompanied by cardiovascular diseases, including myocardial injury [21]. However, our study did not show an association between the use of angiotensin-converting enzyme inhibitors/angiotensin receptor blockers (ACEI/ARB) and COVID-19 severity. Similar to hypertension, the effect of ACEI/ARB on COVID-19 severity is still inconclusive [20]. The S protein of SARS-CoV-2 binds to the ACE2 receptor to enter the human cells, suggesting that the use of ACEI/ARB might worsen the

prognosis of COVID-19 [22]. However, current reports showed that ACEI/IRB was not associated with the poorer outcomes of COVID-19 [23,24].

Our study also demonstrated the association between the use of anticoagulants and COVID-19 mortality with an increased risk of approximately ~ 47-fold. SARS-CoV-2 often induces a pro-coagulative state due to several mechanisms, including endothelial dysfunction, cytokine storm, and complement hyperactivation [25]. While a recent study showed that the use of anticoagulants decreased the mortality of patients with COVID-19, it was not the case with our findings. These differences probably because we grouped the hospitalized and non-hospitalized into one group, classified into anticoagulant versus non-anticoagulant groups. Of note, we have only a limited sample size. These limitations should be considered during the interpretation of our findings. Further study with larger sample size is necessary to clarify and confirm our study.

Most previous reports focused on the impact of VOC on the COVID-19 outcomes, including B.1.1.7 (alpha), B.1.351 (beta), P1 (gamma), and the most recent VOC, B.1.617.2 (delta) [26–29, 30]. Indonesia has reported the identification of alpha, beta, and delta variants since January 2021 [31]. In this present study, we have not found the VOC, and VOI strains in our samples of the sample collection occurred during the period from June to October 2020 or before the first detection of VOC (B.1.1.7 lineage) in Indonesia in January 2021. Currently, the delta variant is identified as the most frequent VOC [31]. However, the actual frequency of the circulating VOCs in Indonesia might be biased due to our limited whole-genome sequencing capacities.

Interestingly, we revealed that patients with multiple S protein mutations might have a ~ 5-fold higher possibility of being hospitalized than those with none or a single S protein mutation. However, the association between mutation and clinical outcome of COVID-19 is inconclusive. A study in Uruguay found that mutation in structural and non-structural protein was not associated with COVID-19 fatalities [32]. Another recent study analyzed the association between viral genomic variants and COVID-19 outcomes. They showed that 17 variants had a two-fold higher risk of severe COVID-19, while 67 variants were associated with less severe COVID-19 [33]. This is in line with another study from France and the US, suggesting that different viral variants may result in different infection severity and risk of hospitalization [34,35]. Since SARS-CoV-2 is an RNA virus, its dynamic evolution is expected to influence its biological characteristic [36], including its virulence and pathogenicity [33]. Interestingly, as an RNA virus, the critical aspect of the SARS-CoV-2 life cycle is not implied by the protein sequence [37]. Indeed, one study showed the importance of synonymous substitutions on the selection of SARS-CoV-2 [38].

All our samples, except one, contained the D614G variant. It has already been established that the D614G mutation was not associated with the COVID-19 illness [39,40]. Indeed, almost all viruses circulating globally consist of the D614G mutation [31]. We also observed other S protein mutations in our samples, including L5F, V213A, and S68SR. None of the mutations lies on the receptor-binding domain (RBD) of the S protein.

Interestingly, a previous report showed that one variant in non-RBD S protein, V1176F, might lead to RBD-ACE2 binding changes and was associated with a high mortality rate of COVID-19 [41]. Moreover, a recent study revealed that variants within the different proteins of SARS-CoV-2 have associated with different patients' outcomes [42]. Further *in vitro* study is essential to clarify whether multiple non-RBD S protein variants associate with the COVID-19 severity in our patients.

There are several limitations of our study. First, we have only a limited sample size that may result in bias in our analysis. Second, the S protein continuously evolves and may result in new mutation(s) that significantly affect virulence and disease pathogenesis. Third, we only analyzed mutations located within the S protein. Mutations in other structural and non-structural proteins may significantly influence the COVID-19; however, they are not investigated in our study.

Conclusions

Here, we show for the first time the association between the SARS-CoV-2 mutations within the S protein besides the VOC and the COVID-19 outcomes, revealing that multiple S protein mutations might affect the severity of COVID-19, in addition to hypertension and the use of anticoagulants. Our study further suggests the importance of genomic surveillance to monitor the SARS-CoV-2 variants, particularly those that might influence the outcomes of COVID-19 patients.

Abbreviations

CI, confidence interval; OR, odds ratio; S protein, spike protein; SNPs, single nucleotide polymorphisms; VOC, variant of concern; VOI, variant of interest

Declarations

Ethics approval and consent to participate

This study was approved by the Medical and Health Research Ethics Committee, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/Dr Sardjito Hospital, Yogyakarta, Indonesia (KE/FK/0563/EC/2020) and written informed consent was obtained. The research has been performed following the Declaration of Helsinki.

Consent to publish

All participants or guardians signed written informed consent for participating in this study.

Availability of data and material

All data generated or analyzed during this study are included in the submission. The sequence and metadata are shared through GISAID (www.gisaid.org).

Competing interests

The authors declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding

The Ministry of Education, Culture, Research and Technology, Indonesia, funded our study. The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

Authors' contributions

G, KI, and NA conceived the study. G drafted the manuscript, and MSH, HW, and TW critically revised the manuscript for important intellectual content. M, SS, DAP, HFH, AAS, ITa, CSA, IPAA, HR, and HM performed the library preparation and NGS. G, MSH, M, IT, REK, I, A, S, EWD, ES, DAAN, YP, NRA, EA, TN and TW collected the data; and G, MSH, HW, M, and A analyzed the data. All authors have read and approved the manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

We thank the Collaborator Members of the Yogyakarta-Central Java COVID-19 study group: Eko Budiono, Heni Retnowulan, Sumardi, Bambang Sigit Riyanto, Munawar Gani, Satria Maulana, Ira Puspitawati, Osman Sianipar (Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada [FK-KMK UGM]/RSUP Dr. Sardjito), Bagoes Poermadjaja (Balai Besar Besar Veteriner Wates, Yogyakarta), Indaryati and Havid Setyawan (Balai Besar Teknik Kesehatan Lingkungan dan Pengendalian Penyakit, Yogyakarta), Ludhang Pradipta Rizki and Sri Fatmawati (FK-KMK UGM), Safitriani and Muhammad Taufiq Soekarno (PT. Pandu Biosains). We gratefully acknowledge the authors, the originating and submitting laboratories for their sequence and metadata shared through GISAID. We also thank to Besar Besar Veteriner Wates (Disease Investigation Center Wates) where the virus samples were sequenced using the NGS Illumina MiSeq instrument. All submitters of data may be contacted directly via www.gisaid.org. The Acknowledgments Table for GISAID is reported as Supplementary Table 1.

References

1. World Health Organization. <https://www.who.int/news/item/29-06-2020-covidtimeline> Accessed on July 2, 2021.
2. Phelan AL, Katz R, Gostin LO. The novel coronavirus originating in Wuhan, China: challenges for global health governance. JAMA. 2020 323, 709–710, doi:10.1001/jama.2020.1097
3. World Health Organization. <https://covid19.who.int/table> Accessed on July 2, 2021.

4. Brodin P. Immune determinants of COVID-19 disease presentation and severity. *Nat Med*. 2021;27(1):28-33. doi:10.1038/s41591-020-01202-8
5. Awortwe C, Cascorbi I. Meta-analysis on outcome-worsening comorbidities of COVID-19 and related potential drug-drug interactions. *Pharmacol Res*. 2020;161:105250. doi:10.1016/j.phrs.2020.105250
6. Debnath M, Banerjee M, Berk M. Genetic gateways to COVID-19 infection: Implications for risk, severity, and outcomes. *The FASEB Journal*. 2020;34(7):8787-8795. doi:10.1096/fj.202001115r
7. Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J Hum Genet*. 2020;65(12):1075-1082. doi:10.1038/s10038-020-0808-9
8. de Sousa E, Ligeiro D, Lérias J et al. Mortality in COVID-19 disease patients: Correlating the association of major histocompatibility complex (MHC) with severe acute respiratory syndrome 2 (SARS-CoV-2) variants. *International Journal of Infectious Diseases*. 2020;98:454-459. doi:10.1016/j.ijid.2020.07.016
9. World Health Organization. <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—8-june-2021> Accessed on July 2, 2021.
10. About UGENE - Unipro UGENE Online User Manual v. 1.30 - WIKI [Internet]. Ugene.net. 2020 [cited 22 December 2020]. Available from: <https://ugene.net/wiki/display/UUOUM30/About+UGENE>.
11. Saitou N. and Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. 1997;4:406-425.
12. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*. 1985;39:783-791.
13. Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*. 1980;16:111-120.
14. Xia X. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. *Mol Biol Evol*. 2018;35:1550–1552.
15. Kumar S. Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol*. 2018;35:1547-1549
16. Gunadi, Wibawa H, Marcellus, Hakim MS, Daniwijaya EW, Rizki LP, et al. Fulllength genome characterization and phylogenetic analysis of SARS-CoV-2 virus strains from Yogyakarta and Central Java, Indonesia. *PeerJ*. 2020;8:e10575 doi: 10.7717/peerj.10575
17. Beeching NJ, Fletcher TE, Fowler R. BMJ best practice. Coronavirus Disease 2019 (COVID-19). <https://bestpractice.bmj.com/topics/en-us/3000168/prognosis> Accessed on May 23, 2020.
18. Gunadi, Wibawa H, Hakim MS, Marcellus, Trisnawati I, Khair RE, et al. Molecular epidemiology of SARS-CoV-2 isolated from COVID-19 family clusters. *BMC Med Genomics*. 2021;14(1):144. doi: 10.1186/s12920-021-00990-3.
19. Alrashed AA, Khan TM, Alhusseini NK, Asdaq SMB, Enani M, Alosaimi B, et al. Severity of COVID-19 infection in ACEI/ARB users in specialty hospitals: A retrospective cohort study. *J Infect Public*

- Health. 2021;14(6):726-733.
20. Clark CE, McDonagh STJ, McManus RJ, Martin U. COVID-19 and hypertension: risks and management. A scientific statement on behalf of the British and Irish Hypertension Society. *J Hum Hypertens*. 2021;35(4):304-307.
 21. Guo T, Fan Y, Chen M, Wu X, Zhang L, He T, et al. Cardiovascular implications of fatal outcomes of patients with coronavirus disease 2019 (COVID-19). *JAMA Cardiol*. 2020;5:811–18.
 22. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. *JAMA*. 2020;323:1824–36.
 23. Mancia G, Rea F, Ludergnani M, Apolone G, Corrao G. Renin–angiotensin–aldosterone system blockers and the risk of covid-19. *N Engl J Med*. 2020;382:2431–40.
 24. Reynolds HR, Adhikari S, Pulgarin C, Troxel AB, Iturrate E, Johnson SB, et al. Renin–angiotensin–aldosterone system inhibitors and risk of covid-19. *N Engl J Med*. 2020;382:2441–8.
 25. Carfora V, Spiniello G, Ricciolino R, Di Mauro M, Migliaccio MG, Mottola FF, et al. Anticoagulant treatment in COVID-19: a narrative review. *J Thromb Thrombolysis*. 2021;51(3):642-648.
 26. Singh J, Samal J, Kumar V, Sharma J, Agrawal U, Ehtesham NZ, Sundar D, Rahman SA, Hira S, Hasnain SE. Structure-Function Analyses of New SARS-CoV-2 Variants B.1.1.7, B.1.351 and B.1.1.28.1: Clinical, Diagnostic, Therapeutic and Public Health Implications. *Viruses*. 2021;13(3):439.
 27. Esper FP, Cheng YW, Adhikari TM, Tu ZJ, Li D, Li EA, Farkas DH, Procop GW, Ko JS, Chan TA, Jehi L, Rubin BP, Li J. Genomic Epidemiology of SARS-CoV-2 Infection During the Initial Pandemic Wave and Association With Disease Severity. *JAMA Netw Open*. 2021;4(4):e217746.
 28. Horby P, Huntley C, Davies N, et al. NERVTAG paper on COVID-19 variant of concern B.1.1.7. London: Department of Health and Social Care, Scientific Advisory Group for Emergencies, January 2021 (<https://www.gov.uk/government/publications/nervtag-paper-on-covid-19-variant-of-concern-b117>. opens in new tab)
 29. Abdool Karim SS, de Oliveira T. New SARS-CoV-2 Variants - Clinical, Public Health, and Vaccine Implications. *N Engl J Med*. 2021;384(19):1866-1868.
 30. World Health Organization. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20210511_Weekly_Epi_Update_39.pdf Accessed on May 23, 2021.
 31. 2021. Pandemic coronavirus causing COVID-19 [Online]. Available at <https://platform.gisaid.org/epi3/cfrontend#8dc5e> (accessed 24 May 2021).
 32. Elizondo V, Harkins GW, Mabvakure B, et al. SARS-CoV-2 genomic characterization and clinical manifestation of the COVID-19 outbreak in Uruguay. *Emerg Microbes Infect*. 2021;10(1):51-65. doi:10.1080/22221751.2020.1863747
 33. Voss JD, Skarzynski M, McAuley EM, Maier EJ, Gibbons T, Fries AC, Chapleau RR. Variants in SARS-CoV-2 Associated with Mild or Severe Outcome. *medRxiv* 2020.12.01.20242149; doi: 10.1101/2020.12.01.20242149

34. Dao TL, Hoang VT, Nguyen NN, et al. Clinical outcomes in COVID-19 patients infected with different SARS-CoV-2 variants in Marseille, France [published online ahead of print, 2021 May 24]. *Clin Microbiol Infect.* 2021;S1198-743X(21)00270-6. doi:10.1016/j.cmi.2021.05.029
35. Nakamichi K, Shen JZ, Lee CS, et al. Hospitalization and mortality associated with SARS-CoV-2 viral clades in COVID-19. *Sci Rep.* 2021;11(1):4802. Published 2021 Feb 26. doi:10.1038/s41598-021-82850-9
36. Armengaud J, Delaunay-Moisan A, Thuret JY, van Anken E, Acosta-Alvear D, Aragón T, et al. The importance of naturally attenuated SARS-CoV-2 in the fight against COVID-19. *Environ Microbiol.* 2020;22(6):1997-2000.
37. Berrio A, Gartner V, Wray GA. Positive selection within the genomes of SARS-CoV-2 and other Coronaviruses independent of impact on protein function. *PeerJ.* 2020;8:e10234.
38. Velazquez-Salinas L, Zarate S, Eberl S, Gladue DP, Novella I, Borca MV. Positive Selection of ORF1ab, ORF3a, and ORF8 Genes Drives the Early Evolutionary Trends of SARS-CoV-2 During the 2020 COVID-19 Pandemic. *Front Microbiol.* 2020;11:550674.
39. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell.* 2020;182(4):812-827.e19.
40. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole Á, et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell.* 2021;184(1):64-75.e11.
41. Farkas C, Mella A, Haigh JJ. Large-scale population analysis of SARS-CoV2 whole genome sequences reveals host-mediated viral evolution with emergence of mutations in the viral Spike protein associated with elevated mortality rates. medRxiv 2020.10.23.20218511; doi: <https://doi.org/10.1101/2020.10.23.20218511>.
42. Nagy Á, Pongor S, Györfy B. Different mutations in SARS-CoV-2 associate with severe and mild outcome. *Int J Antimicrob Agents.* 2021;57(2):106272.

Table 3

Table 3 can be found in the Supplemental Files

Figures

Figure 1

The evolutionary history was inferred using the Neighbor-Joining method [11]. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [12]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method [13] and are in the units of

the number of base substitutions per site. This analysis involved 170 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 29,563 positions in the final dataset. Evolutionary analyses were conducted in MEGA10.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table3Gunadi.xlsx](#)
- [SuppTable1Gunadi.xlsx](#)