Jurnal Teknologi Laboratorium
Vol.9, No.1, *Special Edition* 2020, pp. 13 – 20
ISSN 2580-0191 (Online), ISSN 2338 – 5634(Print)
DOI: 10.29238/teknolabjournal.v9i1.221
Journal homepage: https://www.teknolabjournal.com/index.php/Jtl/index

Original Research

Immunobioinformatics analysis and phylogenetic tree construction of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Indonesia: spike glycoprotein gene

Arif Nur Muhammad Ansori[1,2a], Viol Dhea Kharisma[2,3b], Yulanda Antonius[4c], Martia Rani Tacharina[5d]*, Fedik Abdul Rantam[5e]

[1] Doctoral Program in Veterinary Science, Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia
[2] Computational Virology and Complexity Science, Division of Molecular Biology and Genetics, Generasi Biologi Indonesia Foundation, Gresik, Indonesia
[3] Master Program in Biology, Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Malang, Indonesia
[4] Department of Biology, Faculty of Biotechnology, University of Surabaya, Surabaya, Indonesia
[5] Department of Veterinary Microbiology, Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia

[a]E-mail address : arif.nma-17@fkh.unair.ac.id
[b]E-mail address : violdhea@gmail.com
[c]E-mail address : yulandaantonius@staff.ubaya.ac.id
[d]E-mail address : martia.rt@fkh.unair.ac.id
[e]E-mail address : fedik-a-r@fkh.unair.ac.id

HIGHLIGHTS

- The basis of designing an epitope-based vaccine against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

ARTICLE INFO

Article history:

Received Date: April 28th, 2020
Revised Date: May 27th, 2020
Accepted Date: June 07th, 2020

Keywords:

COVID-19
Epitope-based vaccine
Immunobioinformatics
Phylogenetic tree
SARS-CoV-2

A B S T R A C T

The outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes coronavirus disease 2019 (COVID-19), has spread worldwide and as a result, the World Health Organization (WHO) declared it a pandemic. At present, there are no approved vaccines against SARS-CoV-2. Therefore, the aim of this study was to predict epitope-based vaccines using bioinformatics approaches and phylogenetic tree construction of SARS-CoV-2 against the backdrop of the COVID-19 pandemic. In this study, we employed 27 isolates of SARS-CoV-2 spike glycoprotein genes retrieved from GenBank® (National Center for Biotechnology Information, USA) and the GISAID EpiCoV™ Database (Germany). We analyzed the candidate epitopes using the Immune Epitope Database and Analysis Resource. Furthermore, we performed protective antigen prediction with VaxiJen 2.0. Data for B-cell epitope prediction, protective antigen prediction, and the underlying phylogenetic tree of SARS-CoV-2 were obtained in this research. Therefore, these data could be used to design an epitope-based vaccine against SARS-CoV-2. However, advanced study is recommended for confirmation (*in vitro* and *in vivo*).

*Corresponding Author:
Martia Rani Tacharina
Department of Veterinary Microbiology, Faculty of Veterinary Medicine, Universitas Airlangga
Kampus C Universitas Airlangga, Surabaya, 60115, Indonesia
E-mail: martia.rt@fkh.unair.ac.id
Phone: (+6231)5992785 ext. 5993016

1. INTRODUCTION

On 30 December 2019, the Chinese government reported an outbreak of pneumonia disease in Wuhan.[1] Furthermore, the causative agent identified by the International Committee on Taxonomy of Viruses (ICTV) was named SARS-CoV-2.[2] This new virus has rapidly spread across China and other countries.[3,4,5] In addition, the WHO announced a new name for the disease caused by SARS-CoV-2 coronavirus disease 2019 (COVID-19).[6] At present, there are three coronaviruses that cause disease in humans: severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2.[7]

Coronaviruses belong to the *Coronaviridae* family of the order *Nidovirales*. The coronavirus family consists of alpha, beta, gamma, and delta coronaviruses. The term "corona" reflects the crown-like spikes on the surface of the virus. They are roughly 65–125 nm in diameter with a single-stranded RNA of approximately 26–32 kbs in length. The structural proteins are encoded by four structural genes, including the envelope (E), membrane (M), nucleocapsid (N), and spike (S) genes; orf1ab is the largest gene in SARS-CoV-2.[8] Interestingly, the spike glycoprotein plays an important role in binding to the receptors of the host cell and is the major target for neutralizing antibodies.[7,9]

Vaccines have been demonstrated to decrease the morbidity and mortality levels of many infectious diseases.[10] Therefore, the development of an effective vaccine against SARS-CoV-2 infection is urgently required. Progress in molecular biology and biotechnology is driving the construction of novel concepts in vaccinology. Synthetic recombinant proteins containing epitopes could be produced efficiently with modern biotechnology methods.[11] Hence, an epitope-based vaccine is suggested as a novel possibility for effective vaccines against SARS-CoV-2. In this study, we applied bioinformatics analysis to obtain data for B-cell epitope prediction, protective antigen prediction, and molecular phylogenetic tree construction of SARS-CoV-2 isolates.

2. MATERIALS AND METHOD

2.1 SARS-CoV-2 isolates

The isolates of SARS-CoV-2 were retrieved from GenBank® (National Center of Biotechnology Information, USA) and the GISAID EpiCoV™ Database (Table 1).

Table 1. SARS-CoV-2 isolates retrieved from GenBank®.

Accession ID	Origin	Host	Isolation Source	Database
MT240479.1	Pakistan: Gilgit	*Homo sapiens*	Throat swab	GenBank®
MT188341.1	USA: Minnesota (MN)	*Homo sapiens*	Nasopharyngeal or oropharyngeal swab	GenBank®
MN908947.3	China: Wuhan	*Homo sapiens*	Unknown	GenBank®
LC529905.1	Japan	*Homo sapiens*	Unknown	GenBank®
MT039890.1	South Korea	*Homo sapiens*	Unknown	GenBank®
MT066156.1	Italy	*Homo sapiens*	Sputum	GenBank®
MT072688.1	Nepal	*Homo sapiens*	Oropharyngeal swab	GenBank®
MN985325.1	USA: Washington (WA)	*Homo sapiens*	Oropharyngeal swab	GenBank®
MN988713.1	USA: Illinois (IL)	*Homo sapiens*	Sputum	GenBank®
MT253701.1	China: Zhejiang, Hangzhou	*Homo sapiens*	Sputum	GenBank®
MT007544.1	Australia: Victoria	*Homo sapiens*	Unknown	GenBank®
MT192759.1	Taiwan	*Homo sapiens*	Sputum	GenBank®
MT039888.1	USA: Massachusetts (MA)	*Homo sapiens*	Oropharyngeal swab	GenBank®
MT106054.1	USA: Texas (TX)	*Homo sapiens*	Sputum	GenBank®

Accession ID	Origin	Host	Isolation Source	Database
MT233519.1	Spain: Valencia	*Homo sapiens*	Nasopharyngeal exudate	GenBank®
MT126808.1	Brazil	*Homo sapiens*	Nasopharyngeal swab	GenBank®
MN938384.1	China: Shenzhen	*Homo sapiens*	Nasopharyngeal swab	GenBank®
MT012098.1	India: Kerala State	*Homo sapiens*	Throat swab	GenBank®
MT093571.1	Sweden	*Homo sapiens*	Unknown	GenBank®
MT135041.1	China: Beijing	*Homo sapiens*	Unknown	GenBank®
MT121215.1	China: Shanghai	*Homo sapiens*	Throat swab	GenBank®
EPI_ISL_435281	Indonesia: Jakarta	*Homo sapiens*	Nasopharyngeal and Oro-pharyngeal swab	GISAID EpiCoV™
EPI_ISL_435282	Indonesia: Jakarta	*Homo sapiens*	Nasopharyngeal and Oro-pharyngeal swab	GISAID EpiCoV™
EPI_ISL_435283	Indonesia: Jakarta	*Homo sapiens*	Nasopharyngeal swab	GISAID EpiCoV™

2.2 Nucleotide sequence preparation

SARS-CoV-2 nucleotide sequences (spike glycoprotein gene) from all isolates were retrieved from GenBank® and the GISAID EpiCoV™ Database. Multiple sequence alignment of nucleotide sequences was performed using Molecular Evolutionary Genetics Analysis X (MEGA X).[10]

2.3 Prediction of B-cell epitopes and protective antigens

A B-cell epitope, which is known as a B-cell antigenic determinant, is a specific antigen region with high affinity for B-cell lymphocytes. This interaction induces B-cells to produce an antigen-specific antibody and memory cells.[12] To predict the B-cell epitope, analysis was conducted using the IEDB online webserver with default thresholds (0.350) and VaxiJen v2.0.[10]

2.4 Molecular phylogenetic analysis

Phylogenetic modeling and tree visualization were achieved by applying MEGA X with the Maximum Likelihood method.[10] The phylogenetic tree was validated by running the analysis on 1000 bootstrapped input datasets[13] and cross-referencing it against the Tamura-Nei substitution model.[14]

3. RESULTS AND DISCUSSION

A novel coronavirus, SARS-CoV-2, spread across the world very rapidly. Human-to-human transmission has been confirmed, and the number of global cases has been increasing at a fast pace.[9] To date, there are more than 2 million people infected with SARS-CoV-2 worldwide based on the online interactive dashboard hosted by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. This interactive web-based dashboard visualizes and tracks reported cases of COVID-19 in real time.[15]

In addition, the spike glycoprotein has recently been regarded as a highly expectant antigen formulation for the construction of a SARS-CoV-2 vaccine. There are two reasons for this: (1) it is involved in surface exposure, directly recognized by the host immune system, and (2) it mediates the interaction with the host cells by binding to the receptor, ACE2.[16] Muthumani *et al*. reported that a DNA vaccine encoding MERS-CoV spike glycoprotein was immunogenic in rhesus macaques, camels, and mice.[17] Moreover, Pallesen *et al*. demonstrated a higher titer of neutralizing antibodies in mice immunized by the recombinant prefusion, MERS-CoV spike glycoprotein.[18] In addition, Zhang *et al*. described various types of SARS-CoV-2 vaccines based on a number of different developmental processes, such as whole-cell killed and live-attenuated vaccines, subunit vaccines, mRNA vaccines, DNA vaccines, live vector vaccines, and synthetic peptides.[16]

Epitopes from all samples of SARS-CoV-2 isolates were predicted using the IEDB online webserver to determine the potential for B-cell recognition with an accuracy of approximately

75%. This prediction was based on the combination of Hidden Markov Model (HMM) statistical methods and trend scale.[10] Furthermore, the peptides were predicted using VaxiJen v2.0 to determine the characteristics of immunogenicity or protective antigens. Therefore, they can be distinguished as either non-antigens or antigens. The prediction has an accuracy of approximately 70-89%. The performance of this server was developed based on the physicochemical properties of the target protein without alignment.[19]

Table 2. Prediction of B-cell epitopes and protective antigens of SARS-CoV-2

Predicted Peptides	Position	Length	Protective Antigens
RTQLPPAYTNS	21-31	11	0.8710 (Probable ANTIGEN)
SGTNGTKRFDN	71-81	11	0.5906 (Probable ANTIGEN)
LTPGDSSSGWTAG	249-261	13	0.4950 (Probable ANTIGEN)
VRQIAPGQTGKIAD	407-420	14	1.2606 (Probable ANTIGEN)
YQAGSTPCNGV	473-483	11	0.0881 (Probable NON-ANTIGEN)
YGFQPTNGVGYQ	495-506	12	0.7136 (Probable ANTIGEN)
RDIADTTDAVRDPQ	567-580	14	0.4400 (Probable ANTIGEN)
QTQTNSPRRARSV	675-687	13	0.1763 (Probable NON-ANTIGEN)
ILPDPSKPSKRS	805-816	12	0.5322 (Probable ANTIGEN)
VYDPLQPELDSF	1137-1148	12	0.0903 (Probable NON-ANTIGEN)
KNHTSPDVDLG	1157-1167	11	1.4039 (Probable ANTIGEN)
FDEDDSEPVL	1256-1265	10	0.3154 (Probable NON-ANTIGEN)

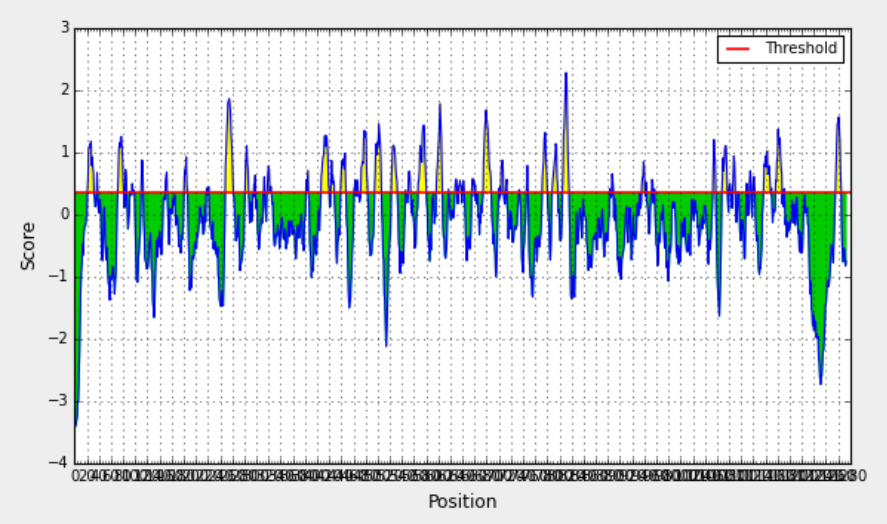

Figure 1. Prediction of B-cell epitopes from the amino acids of spike glycoprotein of SARS-CoV-2. B-cell epitope prediction was performed using the IEDB online webserver. The yellow region was positive, whereas the green region was a negative prediction of B-cell epitopes.

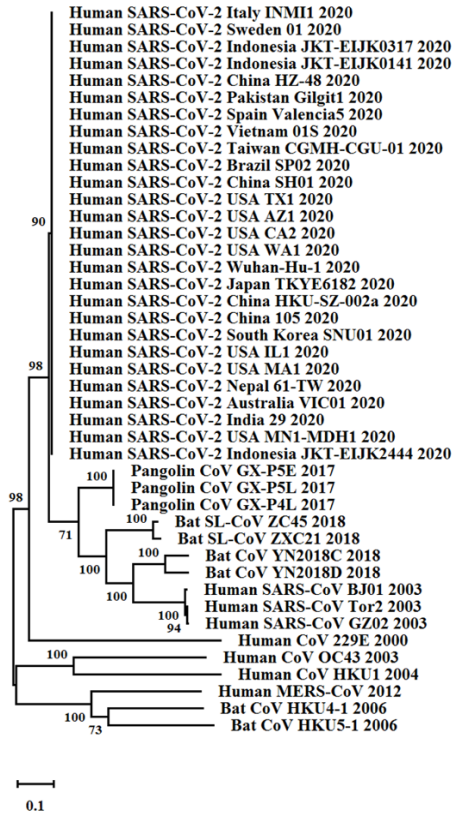

Figure 2. Phylogenetic tree of the first three SARS-CoV-2 in Indonesia and other coronaviruses.

In this study, we obtained 12 predicted peptides and 8 other peptides considered as protective antigens (Table 2 and Figure 1). B-cell epitope prediction is a methodology to predict protein regions that could be recognized as an epitope response to B-cells. An epitope is part of an antigen molecule that binds to the antibodies. This area is essential for designing certain types of vaccines or specific antibodies. The design of a seed vaccine requires an epitope that can determine the active side of the antigen that binds to the antibodies.[12]

Recently, there have been no epitope-based vaccines seen on the market.[20] However, over the past decade, research and development to reveal epitope-based vaccines has garnered much interest within the vaccine industry. Progress has been made utilizing various methods, such as cell culture techniques, recombinant DNA technology, and immunoinformatics.[11] Epitope-based vaccines play an essential part in the present volume of research with several advantages over conventional vaccines, including robust safety and stability, high specificity, and better manufacturing and retention. As a result, epitope-based vaccines have become an increasingly popular field of vaccinology. Several studies have identified the potency of epitope-based antigens that efficiently generate high immunity and protection against various pathogens.[10,12]

Molecular phylogenetic analysis is used to address both applied and fundamental issues of virus research, including phylogeography, diagnostics, evolution, origin, epidemiology, taxonomy, and forensics. It may provide an evolutionary view of the variety of any character that can be measured for a cluster of viruses.[21] The molecular phylogenetic analysis in this study exhibited the genetic relationship between SARS-CoV-2 isolates from various countries and other coronaviruses originating from humans, bats, and pangolins (Figure 2). Interestingly, we identified the newly submitted three Indonesian SARS-CoV-2 isolates (provided by Eijkman Institute for Molecular Biology, Ministry of Research and Technology/National Agency for Research and Innovation of the Republic of Indonesia) from the GISAID EpiCoV™ Database. We revealed that there were not many differences in the spike glycoprotein gene between viruses isolated from various countries and those from Indonesia. On the other hand, Andersen *et al*. stated that SARS-CoV-2 is clearly not a laboratory construct or manipulated virus.[22]

In addition, Lam *et al*. demonstrated that coronaviruses are present in many wild mammals in Asia.[1] Currently, it is crucial to investigate the likelihood of intermediate hosts of SARS-CoV-2 to contain COVID-19 spread. Pangolin CoV is 91.02% identical to SARS-CoV-2 at the whole-genome level.[23] Previously, Zhou *et al*. stated that SARS-CoV-2 shares a 96% whole genome with a bat CoV from *Rhinolophus affinis* (BatCoV RaTG13) collected from Yunnan Province, China.[24] Based on this study, we recommend that further surveillance studies be carried out on bats and pangolins in the natural environment, especially in China and Southeast Asia, in order to understand the risk of zoonotic transmission in the near future.

Overall, vaccination is an effective method for controlling diseases and regulating human and animal health. Specifically, vaccines are agents that enhance the adaptive immune response and can reduce the effects of infections and diseases. This is accomplished through the immune system recognizing the vaccine as a foreign object, then destroying it, and placing it in memory. Moreover, the novel concept of reverse vaccinology has revolutionized the study of vaccine development. The ability to obtain a whole-genome sequence from a virulent organism has led to the *in silico* analysis of the most protective antigens before conducting confirmation experiments (*in vitro* and *in vivo*). Several advantages, such as low cost, speed, along with the success of the bioinformatics approach, depend on the accuracy of predictions, which are supported by many tools.

4. CONCLUSION

This study supplied data for B-cell epitope prediction, protective antigen prediction, and the phylogenetic tree of SARS-CoV-2. In summary, this study could serve as the basis to design an epitope-based vaccine against SARS-CoV-2. However, an advanced study is suggested for confirmation (*in vitro* and *in vivo*).

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

ACKNOWLEDGEMENT

We thank Editage for editing the manuscript.

FUNDING INFORMATION

PMDSU Scholarship Batch III from the Directorate General of Higher Education, Ministry of Education and Culture of the Republic of Indonesia.

REFERENCES

1. Lam TTY, Shum MHH, Zhu HC, et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature*. 2020. doi:10.1038/s41586-020-2169-0.
2. Gorbalenya AE, Baker SC, Baric RS, et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*. 2020;5(4):536-544. doi:10.1038/s41564-020-0695-z.
3. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet*. 2020;395(10223):470-473. doi:10.1016/S0140-6736(20)30185-9.
4. Holshue ML, DeBolt C, Lindquist S, et al. First case of 2019 novel coronavirus in the United States. *N Engl J Med*. 2020;382(10):929-936. doi:10.1056/NEJMoa2001191.
5. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395(10223):507-513. doi:10.1016/S0140-6736(20)30211-7.
6. Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents*. 2020;55(3):105924. doi:10.1016/j.ijantimicag.2020.105924.
7. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function,

and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*. 2020;181(2):281-292.e6. doi:10.1016/j.cell.2020.02.058.
8. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *J Adv Res*. 2020;24:91-98. doi:10.1016/j.jare.2020.03.005.
9. Phan T. Genetic diversity and evolution of SARS-CoV-2. *Infect Genet Evol J*. 2020;81(2020). doi:https://doi.org/10.1016/j.meegid.2020.104260.
10. Ansori ANM, Kharisma VD. Characterization of Newcastle Disease Virus in Southeast Asia and East Asia: Fusion Protein Gene. *EKSAKTA J Sci Data Anal*. 2020;1(1):14-20. doi:10.20885/eksakta.vol1.iss1.art3.
11. Callaway E. The race for coronavirus vaccines. *Nature*. 2020;580(7805):576-577. doi:10.1038/d41586-020-01221-y.
12. Adianingsih OR, Kharisma VD. Study of B cell epitope conserved region of the Zika virus envelope glycoprotein to develop multi-strain vaccine. *J Appl Pharm Sci*. 2019;9(1):98-103. doi:10.7324/JAPS.2019.90114.
13. Guzzi PH, Mercatelli D, Ceraolo C, Giorgi FM. Master regulator analysis of the SARS-CoV-2/human interactome. *J Clin Med*. 2020;9(4):982. doi:10.3390/jcm9040982.
14. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30(12):2725-2729. doi:10.1093/molbev/mst197.
15. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. 2020;20(5):533-534. doi:10.1016/S1473-3099(20)30120-1.
16. Zhang J, Zeng H, Gu J, Li H, Zheng L, Zou Q. Progress and prospects on vaccine development against sars-cov-2. *Vaccines*. 2020;8(2):1-12. doi:10.3390/vaccines8020153.
17. Muthumani K, Falzarano D, Reuschel EL, et al. A synthetic consensus anti–spike protein DNA vaccine induces protective immunity against Middle East respiratory syndrome coronavirus in nonhuman primates. *Sci Transl Med*. 2015;7(301):301ra132-301ra132. doi:10.1126/scitranslmed.aac7462.
18. Pallesen J, Wang N, Corbett KS, et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc Natl Acad Sci U S A*. 2017;114(35):E7348-E7357. doi:10.1073/pnas.1707304114.
19. Zahroh H, Ma'rup A, Tambunan USF, Parikesit AA. Immunoinformatics approach in designing epitopebased vaccine against meningitis-inducing bacteria (Streptococcus pneumoniae,Neisseria meningitidis,and Haemophilus influenzae type b). *Drug Target Insights*. 2016;10:19-29. doi:10.4137/DTI.S38458.
20. Fosgerau K, Hoffmann T. Peptide therapeutics: Current status and future directions. *Drug Discov Today*. 2015;20(1):122-128. doi:10.1016/j.drudis.2014.10.003.
21. Gorbalenya AE. Phylogeny Viruses. In: *Reference Module in Biomedical Sciences*. Main St, Frisco: Elsevier Inc.; 2017:1-7. doi:10.1016/B978-0-12-801238-3.95723-41.
22. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med*. 2020;26(4):450-452. doi:10.1038/s41591-020-0820-9.
23. Zhang T, Wu Q, Zhang Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr Biol*. 2020;30(7):1346-1351. doi:10.1016/j.cub.2020.03.022.
24. Zhou P, Yang X Lou, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-273. doi:10.1038/s41586-020-2012-7.

SHORT BIOGRAPHY

Arif Nur Muhammad Ansori, M.Sc. has been received the PMDSU Scholarship Batch III (2017) for his M.Sc. and Ph.D. program and is currently a Ph.D. candidate in veterinary science at Universitas Airlangga, Indonesia. His research area focuses on the field of vaccinology and virology.

Viol Dhea Kharisma, M.Sc. is student in biology at Universitas Brawijaya, Indonesia. His research area focuses on the field of computational virology and complexity science.

Yulanda Antonius, M.Sc. is a lecturer at University of Surabaya, Indonesia. Her research area focuses on the field of stem cells and bioinformatics.

Martia Rani Tacharina, DVM, M.Sc. is a lecturer at Universitas Airlangga, Indonesia. Her research area focuses on the field of virology and immunology.

Prof. Dr. Fedik Abdul Rantam, DVM is a professor at Universitas Airlangga, Indonesia. His research area focuses on the field of stem cells and virology.