Jurnal Riset Veteriner Indonesia
Journal of the Indonesian Veterinary Research
P-ISSN: 2614-0187, E-ISSN:2615-2835
Volume 6 No. 1 (January 2022), pp. 5-14
journal.unhas.ac.id/index.php/jrvi/

Mapping of the Coronavirus Circulating in Asia Based on Sequence of Gene Spike and Membrane Protein Used MEGA-X Aplication

Muhammad Fikri Raditya Jalil[1], Anggini Putri Husada[1],
Nur Azizah Dariming[2] dan Fedri Rell3*

[1] Study Program of Veterinary Medicine, Faculty of Medicine, Hasanuddin University, Jl. Perintis Kemerdekaan Kampus Tamalanrea, Km. 10 Makassar, Indonesia 90245
[2]Medical Education Program, Faculty of Medicine, Hasanuddin University, Jl. Perintis Kemerdekaan Kampus Tamalanrea, Km. 10 Makassar, Indonesia 90245
[3]Laboratory of Microbiology, Study Program of Veterinary Medicine, Faculty of Medicine, Hasanuddin University, Jl. Perintis Kemerdekaan Kampus Tamalanrea, Km. 10 Makassar, Indonesia 90245

*Corresponding author: Fedri Rell (fedrirell@unhas.ac.id)

Abstract

Coronavirus are viruses that can be transmitted to human and animals. Severe acute respiratory syndrome, middle east respiratory syndrome, and Coronavirus disease 2019 are disease can be caused by several subtypes of coronavirus. The aims of this study were to mapping of the coronavirus circulating in Asia based on sequence of gene spike and membrane protein virus. Totally of 67 coronavirus spike protein and membrane gene sequence were accessed via GenBank® (www.ncbi.nlm.nih.gov/genbank/) matched with the ClustalW Method MEGA-X. The result of the study are 20 groups of coronavirus were found based on spike protein gene sequences and 27 groups of coronavirus were found based on membrane protein gene sequences which were different with the first group of coronavirus found in Wuhan. Therefore, it can be concluded that the coronavirus circulate in several Asian countries had been mutate on gene spike and membrane protein.

Keywords: *Asia, Coronavirus, MEGA-X, Membrane Protein, Spike Protein*

Introduction

Coronavirus is a virus that mutates easily. The genetic material of the virus is composed of a positive polarized single-stranded RNA (ssRNA) nucleic acid and an enveloped capsid with lenght of the genome ± 30,000 nucleotides (nt). RNA viruses have a fairly high mutation rate and higher than DNA viruses. RNA viruses are not doing the proofreading process during the nucleic acid replication process so that there is no improvement in the acid-base composition and causes mutations. The existence of mutations results in changes in viral genetic information that can increase viral malignancy and the emergence of new subtypes of coronavirus (Flint *et al*, 2015).

The process of transmission of the coronavirus also plays an important role in changing the genetics of the virus. Transmission of coronavirus can occur between humans and humans, animals and animals, and animals to humans or vice versa which results in viral genetic recombination either by antigenic shift or antigenic drift. Several animals such as bats, camels, and pangolins have been reported as sources of coronavirus that occurs in humans. Transmission of the virus occurs through droplets, aerosols, and faeces-oral (Flint *et al*, 2015).

The spike protein is one of the 4 major coronavirus antigen proteins that play a role in the main structure of gene writing (Yuliana, 2020). Spike plays a role in the attachment and entry of the virus into the host cell for replication (Huang *et al.*, 2020). In addition, the membrane or commonly referred to as the M protein is one of the 4 main proteins determining the pathogenicity of the coronavirus. The membrane is the most abundant viral protein in the virion particle which plays a role in giving shape to the virus. The membrane on the coronavirus has a role as a coronavirus assembly (Sunday *et al.*, 2021).

Coronavirus mutations can occur for three reasons, namely mutations due to changes in the number of nitrogenous bases, mutations due to changes in the type of nitrogenous bases and changes in the location of the sequence of nitrogenous bases in nucleic acids. Mutations due to changes in the number of nitrogenous bases can occur by addition, deletion, insertion, and duplication. Mutations due to changes in nitrogen base types can occur due to transitions or transversions (Sanjuan and Pilar, 2016).

The mutated coronavirus can be very dangerous. Addition is one of the causes of mutations due to changes in the number of nitrogenous bases by adding nitrogenous bases. Deletions can occur due to the reduction of nitrogen bases. Insertion can occur when a nitrogenous base is inserted into the chain of a nucleic acid nitrogenous base. Duplication can occur due to doubling in the sequence of nitrogenous bases. Transition is the replacement of a nitrogen base pair with another nitrogen base pair of the same type. Transversion is the replacement of a nitrogen base pair with another nitrogen base pair that is not similar (Sanjuan *et al.*, 2020).

Molecular-based coronavirus mapping is very important to know the genetic changes of the virus. The spread of the coronavirus that has occurred in many countries allows changes to the viral protein gene sequences. Molecular-based mapping was carried out using the MEGA-X application (Rell *et al.*, 2021). MEGA-X application is an application to analyze the kinship relationship of a virus according to the viral protein gene sequence. Viral protein gene sequences as a source of genetic change information can be accessed through GeneBank. The viral protein gene sequences that have been analyzed can be mapped into kinship groups based on the source of isolation from humans in a country or from animals as well as the comparison between animals and humans.

Material and Method

Sequence data of the coronavirus gene spike and membrane protein from Asia were accessed from Genebank (https://www.ncbi.nlm.nih.gov/genbank/). A total of 67 protein spike and membrane coronavirus gene sequences were analyzed using MEGA-X. The nucleotide sequences that have been collected are then matched with the ClustalW Method from MEGA-X. Phylogenetic analysis was carried out using the Phylogeny Construct test Neighbour-Joining Tree Method from MEGA-X (Kumar *et al.*, 2018). This research was conducted on 01 June - 31 August 2021 virtually.

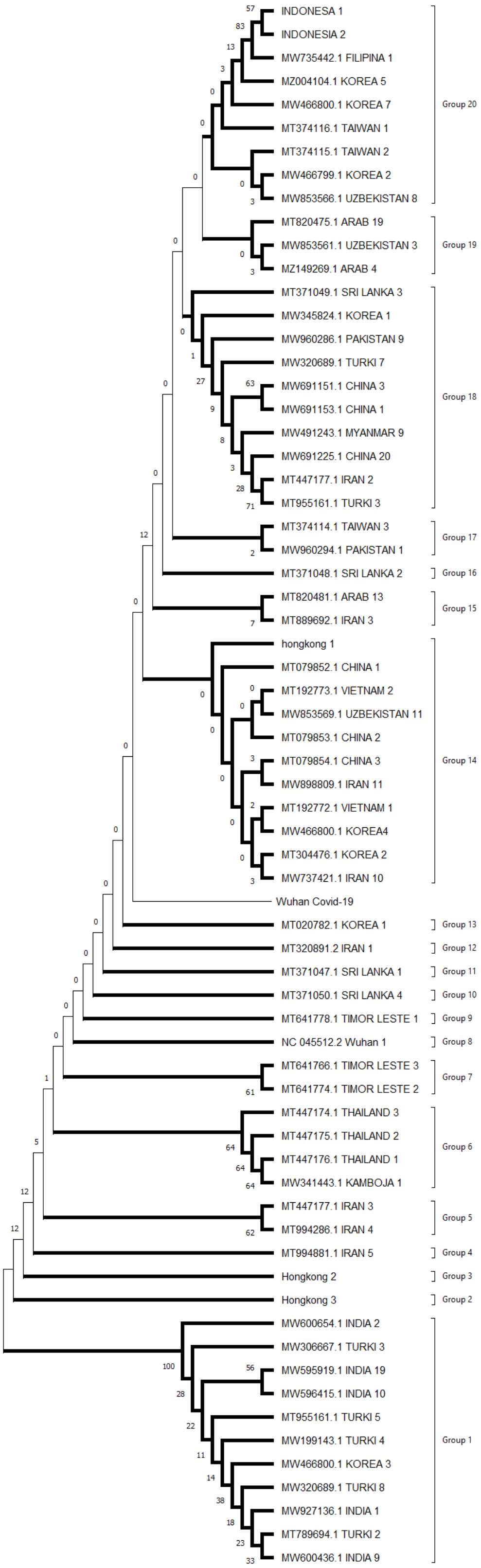
Result and Discussion

Reconstruction of the phylogenetic tree using neighbour join tree model MEGA-X on the spike protein gene sequences of coronavirus in Asia formed 20 groups. The group of coronavirus spike protein gene sequences can be divided into 10 large groups and 10 small groups. Large groups such as groups 1, 5, 6, 7, 14, 15, 17, 18, 19, 20 consisted of several countries that had relatively close kinship of coronavirus protein gene sequences in one group, while small groups consisted of countries that have a relative distance of the coronavirus protein gene sequence from each other so that they form their own group because they have significant differences (Subari *et al.,* 2021). Changes in viral genetics will affect functional properties and may alter infectivity, disease severity or interactions with the host (Harvey *et al.,* 2021).

The spike protein gene sequence of coronavirus that was circulating in India has a phylogenetic branch adjacent to Turkey and Korea. This indicates that there is a fairly close relationship between the coronavirus spike protein gene sequences circulating in India, Turkey and Korea because the relatively close coronavirus spike protein gene sequences have many of the same sequences in one group.

Iran has coronavirus spike protein gene sequence that forms its own group and indicates that the coronavirus spike protein gene sequence circulating in Iran has a protein gene sequence that is quite different from other countries in Asia but still has sufficient kinship with other countries which have adjacent phylogenetic branches. In other words, the spike coronavirus circulating in Iran is still closely related to the spike protein gene sequence circulating in India, Turkey, and Korea.

The coronavirus spike protein gene sequence isolated in Wuhan has a fairly distant branch from the India, Turkey, Korea and Iran group, which indicates that the coronavirus spike protein gene sequence circulating in Turkey, India, Iran and Korea is closely related to the protein gene sequence. spike coronavirus isolated in Wuhan. The distant kinship between the groups occurred due to differences in the spike protein gene sequence due to changes that occurred from time to time. Higher mutation rates lead to higher genetic diversity, except in special cases, it is not possible to infer mutation rates directly from observed population mutation frequencies. Knowledge of the processes underlying viral mutation rates has implications for understanding and managing drug resistance, immune flight, vaccination, pathogenesis, and the emergence of new diseases (Sanjuan and Pilar, 2016).



Picture 1. Spike Coronavirus Protein Phylogenetic Tree

Thailand has the same phylogenetic branch as Cambodia, and it can be interpreted that the gene sequences of the coronavirus spike protein circulating in the two countries are quite close. The clustering that occurs in the phylogenetic branch occurs when there are spike protein gene sequences that have great similarities or can be said to have close kinship. Another branch adjacent to the Thailand and Cambodia branches is the State of Timor Leste.

Timor Leste formed its own phylogenetic branch due to significant differences in the gene sequences of the coronavirus spike protein circulating in that country. Although the spike protein gene sequence circulating in Timor Leste forms its own branch, the spike protein gene sequence still has a fairly close relationship with Thailand and Cambodia because it has a fairly close phylogenetic branch. Both Thailand and Cambodia have spike protein gene sequences that are quite distantly related to the coronavirus spike protein gene sequence circulating in Wuhan. Meanwhile, the State of Timor Leste has a fairly close relationship with the coronavirus spike sequence isolated in Wuhan. This can be explained by the differences in the protein gene sequences of Thailand and Cambodia which are quite close to the State of Timor Leste but have far enough differences from the spike protein gene sequences of the coronavirus isolated in Wuhan. Meanwhile, the coronavirus spike protein gene sequence circulating in Timor Leste has a closer relationship with the coronavirus circulating in Wuhan because it has adjacent branches.

The relatively distant relationship between the coronavirus spike protein gene sequences circulating in Thailand and Cambodia and the coronavirus spike protein gene sequences isolated in Wuhan indicates that there are differences in gene sequences due to changes in the coronavirus spike protein gene sequences isolated in Wuhan. Changes in viral spikes can also alter viral tropism, including new hosts and enhance viral pathogenesis (Ortega *et al.,* 2020).

The spike protein gene sequences of coronavirus originating from China are closely related to the coronavirus spike protein gene sequences due to the large number of sequences that have similar similarities to Vietnam, Uzbekistan, Iran and Korea. Some of the coronavirus spike protein gene sequences in Group 14 such as Iran and Korea are closely related to the coronavirus spike protein gene sequences isolated in Wuhan. Meanwhile, other countries have relatively far kinship from the coronavirus spike protein gene sequence isolated in Wuhan.

Saudi Arabia which still has a fairly close kinship with several countries in the phylogenetic branch, the State of Saudi Arabia has a fairly distant relationship with the spike coronavirus gene isolated in Wuhan. This indicates that there are differences due to changes in the coronavirus spike protein gene sequence isolated in Wuhan and in several countries in Asia. As a COVID-19 therapeutic agent, spike protein-specific neutralizing antibodies can interact with the SARS-CoV-2 S protein to inhibit the virus from invading the human body. Monoclonal antibodies have been approved for emergency use. It has been reported that viral mutants can reduce the effectiveness of neutralizing antibodies (Jia and Wenping, 2021).

The spike protein gene sequences of coronavirus originating from Sri Lanka, Korea, China, Pakistan, Turkey, Myanmar, and Iran have phylogenetic branches that are close to each other and indicate that they are closely related and have almost the same sequence. Pakistan and Taiwan also have a relatively close kinship of spike coronavirus with these countries which can be seen in the phylogenetic branches that are close to each other even though they form new branches but still have almost the same gene sequences. The coronavirus spike protein gene sequences originating from these countries have a fairly distant spike gene relative from the coronavirus spike gene isolated in Wuhan. This indicates that there are differences in the sequence of the coronavirus spike protein gene sequences that can occur through viral mutations. Virus mutations can cause damage to some body functions that are more severe than previous viruses (Dawood, 2020).

The spike protein gene sequences of coronavirus from Indonesia, the Philippines, Korea, Taiwan, and Uzbekistan have a fairly close relationship between one gene and another. This means that the gene sequences of the spike coronavirus protein circulating in Indonesia still have similar sequences to countries from the same group but have quite far kinship in countries from different groups. The spike protein gene sequences for coronavirus originating from countries that are in group 20 are quite distantly related to the spike protein gene sequences isolated in Wuhan. This indicates that the sequence of the coronavirus membrane protein gene sequence has changed due to the coronavirus mutation so that it forms a group that has a fairly distant relationship from the coronavirus isolated in Wuhan. Transmission of the coronavirus can occur between humans and humans, animals and animals, and animals to humans or vice versa which results in viral genetic recombination either by antigenic shift or antigenic drift, causing changes in the spike protein gene sequences that circulating in Asia (Flint *et al.*, 2015).

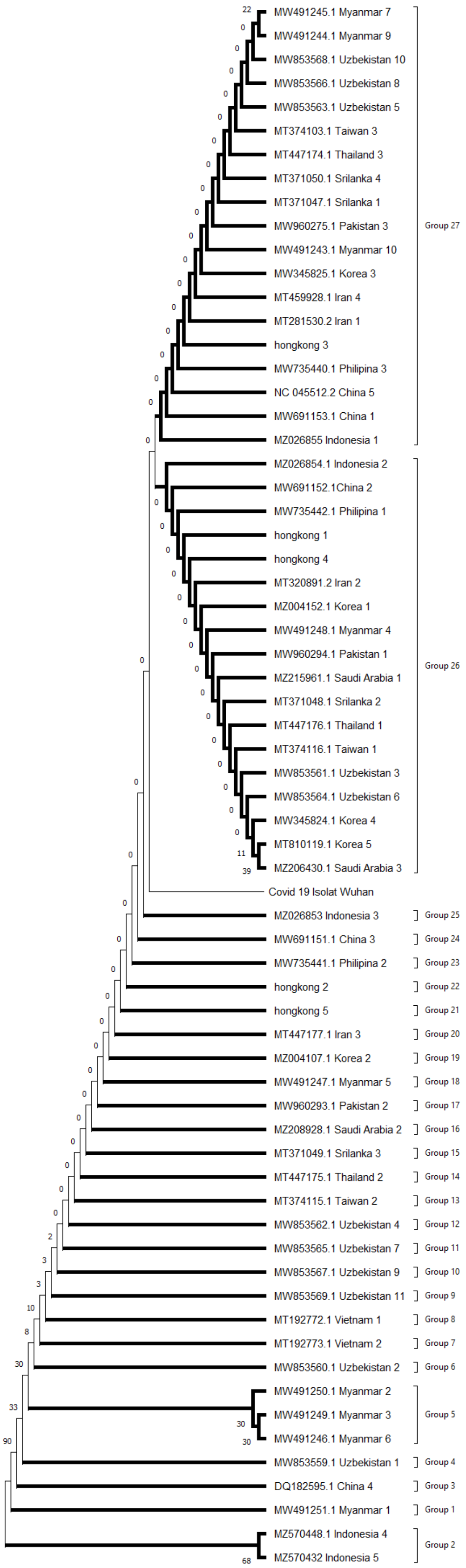
Reconstruction of the phylogenetic tree using the neighbour joining tree model MEGA-X on the coronavirus membrane protein gene sequences in Asia formed 27 groups. The group of coronavirus membrane protein gene sequences can be divided into 4 large groups and 23 small groups. Large groups such as groups 2, 5, 26 and 27 consist of several countries that have relatively close kinship of coronavirus protein gene sequences in one group, while small groups consist of countries that have relatively distant coronavirus protein gene sequence kinship between with each other so that they form their own groups because they have significant differences (Subari et al., 2021). The mutation rate of RNA viruses is very high, up to a million times higher than that of the host and this high rate is correlated with modulation of virulence and evolutionary ability, traits considered beneficial for viral adaptation (Pachetti *et al.*, 2021).

Membrane protein gene sequences of coronavirus originating from Indonesia still have a fairly close relationship with membrane protein gene sequences of coronavirus because of many sequences that have the same similarities. Membrane protein gene sequences of coronavirus originating from Myanmar have phylogenetic branches that are close to each other and indicate that there is a fairly close relationship and have the same sequences. However, the sequence is quite far from membrane protein gene sequences of coronavirus isolated in Wuhan. This indicates that there are differences due to changes in the gene sequences of the coronavirus membrane protein isolated in Wuhan with those in Indonesia and Myanmar. The occurrence of branch differences between coronavirus membrane protein gene sequences from one another can occur due to changes in protein gene sequences that cause fairly distant relationships (Subari *et al.*, 2021). Quite distant kinship can occur due to mutations in the coronavirus membrane protein gene sequence in one country to another (Sanjuan and Pilar, 2016).

Gene sequences of membrane proteins coronavirus from several sequences including Indonesia, China, the Philippines, Hong Kong, Iran, Korea, Myanmar, Pakistan, Saudi Arabia, Sri Lanka, Thailnd, Taiwan and Uzbekistan who still have kinship gene sequences of membrane proteins coronavirus that is close enough for the number of sequences that have the same resemblance. Saudi Arabia and Korea membrane protein gene sequences are closely related to the coronavirus membrane protein gene sequence isolated in Wuhan. Meanwhile, other countries have quite distant kinship from the coronavirus membrane protein gene sequence isolated in Wuhan. This indicates that there are differences due to changes in the gene sequences of the coronavirus membrane protein isolated in Wuhan and in several countries in Asia. Through this branching, it can also be understood that the gene sequences of the coronavirus membrane protein circulating in Indonesia still have similar sequences to other countries in the group. This fairly distant relationship can be caused by evolutionary processes or mutations that cause changes in the coronavirus membrane protein gene sequence, causing branch differences (Subari et al., 2021).

Differences in protein gene sequences in both spike proteins and membrane proteins in coronaviruses can be caused by gene mutations in the coronavirus nucleic acid chain. The mutation rate of RNA viruses is very high, up to one million times higher than that of the host and this high rate is correlated with modulation of virulence and evolutionary ability, traits considered beneficial for viral adaptation (Pachetti *et al.*, 2021). Changes in viral genetics will affect functional properties and can change infectivity, disease severity or interactions with the host (Harvey *et al.*, 2021).

The gene sequences of the coronavirus membrane protein originating from Myanmar, Uzbekistan, Taiwan, Thailand, Sri Lanka, Pakistan, Korea, Iran, Indonesia, the Philippines, China, and Hong Kong are closely related to each other. This means that the gene sequences of the coronavirus membrane protein circulating in that country still have similar sequences to countries belonging to the same group but are quite distantly related in different countries. The coronavirus membrane protein gene sequences originating from the above countries are quite distantly related to the membrane protein gene sequences isolated in Wuhan. This indicates that the sequence of the coronavirus membrane protein gene has changed due to mutations in the coronavirus, thus forming a group that is relatively distant from the coronavirus isolated in Wuhan. Mutations can affect the gene sequences of coronavirus membrane proteins circulating in a country, causing relatively distant relationships between coronaviruses (Subari and Pilar, 2016).



Picture 2. Coronavirus Membrane Protein Phylogenetic Tree

Mutations can occur due to transmission coronavirus can occur either between human and human, animal, and animal or animal-to-human or otherwise that result in genetic recombination of virus either leap antigenic (antigenic shift) and drift antigenic (antigenic drift) causing changes in circulating spike protein gene sequences in Asia (Flint et al., 2015). Virus mutations can cause damage to some body functions that are more severe than previous viruses and (Dawood., 2020).

Changes in spike protein gene sequences and viral membranes can change viral tropism, including new hosts and increase viral pathogenesis (Ortega et al., 2020). The same or similar changes can alter the shape of spike protein and damage or even destroy the NAb binding site. Therefore, by extrapolation, vaccine efficacy can be compromised. These "escape mutations" usually arise when viruses are placed under selective stress by antibodies that limit but do not eliminate viral replication (Moore and Paul, 2021).

Phylogenetic analysis based on spike and membrane protein gene sequences will add the viral data circulating in Asia so that IT can be used in prevention and control of coronavirus infections such as vaccine production and community mobilization restrictions (Alosaimi et al., 2021).

Conclusion

Based on the results of the analysis conducted on 67 samples of spike protein and membrane coronavirus gene sequences circulating in Asia, it was found that several new groups of coronaviruses were formed in the phylogenetic tree analysis using the MEGA-X application. By analysing the spike protein gene sequences of coronavirus, it could be found that there were 20 new groups formed and 27 new groups by analysing the membrane gene of coronavirus, which each of them was different with coronavirus protein gene sequences isolated in Wuhan. The group differences that occur prove that there has been a change in the gene sequence of coronavirus protein circulating in Wuhan City and several countries in Asia.

Acknowledgement

The authors would like to thank the Kementerian Riset, Teknologi dan Pendidikan Tinggi who has founded this research. The authors also would like to thank the Team of the POKJA PKM Hasanuddin University and reviewer for their guidance and advice for the goodness of this research.

Reference

Alosaimi, B., Asif N., Majed F. A., Lilian O., Maaweya E. H., Ahmad S. A., Athari A. dan Mushira. 2021. Structural Mapping of Mutations in Spike, RdRp and Orf3a Genes of SARS-CoV-2 in Influenza Like Illness (ILI) Patients. Viruses. 13 (136): 1-16.

Banoun, Helene. 2021. Evolution of SARS-CoV-2: Review of Mutations, Role of the Host Immune System. *Nephron*. 145(1): 392–403.

Dawood, A.A. 2020. Mutated COVID-19 may foretell a great risk for mankind in the future. *New Microbes and New Infections.* 35(1): 1-8.

Flint, J., Vincent R. R., Glenn F. R., Anna M. S. dan Lynn W. E. 2015. *Principles of Virology*. Edisi ke-4, ASM Press. Washington DC. USA.

Harvey, William T., Alessandro M. C., Ben Jackson, Ravindra K. G., Emma C. T., Ewan M. H., Catherine L., Richard R., Andrew R., COVID-19 Genomics UK (COG-UK) Consortium, Sharon J. P., dan David L. R. 2021. SARS-CoV-2 Variants, *Spike* Mutations and Immune Escape. *Nature Reviews Microbiology*. 19(1): 409-424.

Huang, Y., Chan Y., Xin-Feng X., Wei X dan Shu-Wen L. 2020. Structural and Functional Properties of SARS-CoV-2 *Spike* Protein: Potential Antivirus Drug Development for COVID-19. *Acta Pharmacologica Sinica*. 41(3): 1141-1149.

Jia, Zaixing dan Wenping Gong. 2021. Will Mutations in the *Spike* Protein of SARS-CoV-2 Lead to the Failure of COVID-19 Vaccines?. *Journal of Korean Medical Science.* 36(18): 1-11.

Kumar, S., Glen S., Michael L., Christina K. dan Koichiro T. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology Evolution*. 35(6): 1547-1549.

Minggu, R. B., Janette M. R. dan Grace L. A. T. 2021. Struktur Genom Severe Acute Respiratory Syndrome *Coronavirus*-2 (SARS-CoV-2). *Jurnal Biomedik*. 2(13): 233-240.

Moore, John P. dan Paul A. Offit. 2020. SARS-CoV-2 Vaccines and the Growing Threat of Viral Variants. *JAMA Networks*. 325(9): 821-822.

Ortega, Joseph Thomas, Maria Luisa Serrano, Flor Helene Pujol, dan Hector Rafael Rangel. 2020. Role Of Changes In Sars-Cov-2 *Spike* Protein In The Interaction With The Human Ace2 Receptor: An In Silico Analysis. *EXCLI Journal*. 19(1): 410-417.

Pachetti, Maria, Francesca Benedetti, Fabiola Giudici, Elisabetta Mauro, Paola Storici, Claudio Masciovecchio, Silvia Angeletti, Massimo Ciccozzi, Robert C. Gallo, Davide Zella, dan Rudy Ippodrino. 2020. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Transtitional Medicine*. 179(18): 1-9.

Rell, F., Anak A. A. A. M. A. dan I Gusti N. K. M. 2021. Analisis Filogeni Virus Newcastle Disease Isolat Bali Tahun 2013 Sampai 2014 Berdasarkan Sekuen Daerah Pemotongan Protein Fusion. Buletin Veteriner Udayana. 13(1): 67-74.

Sanjuan R., Miguel R. N., Nicola C., Louis M. M. dan Robert B. 2020. Viral Mutation Rates. Journal of Virology. 84(19): 9733-974