Analysis of the SARS-CoV-2 envelope (E), nucleocapsid (N), and non-structural protein12 (nsp12) genes from COVID-19 patients in West Java

Azzania Fibriani*1, Irin Annisa Evitayani1, Gusti Ayu Prani Pradani2, Rebecca Stephanie2, Ema Rahmawati3, Ryan Bayusantika Ristandi3, Cut Nur Cinthia Alamanda3, Rifky Waluyajati Rachman3, Rini Robiani3, Isak Solihin3

1Department of Biology, School of Life Sciences and Technology, Bandung Institute of Technology, Bandung 40132, Indonesia
2Department of Microbiology, School of Life Sciences and Technology, Bandung Institute of Technology, Bandung 40132, Indonesia
3Laboratorium Kesehatan Provinsi Jawa Barat, Bandung 40161, Indonesia

Fibriani A, Evitayani IA, Pradani GAP, Stephanie R, Rahmawati E, Ristandi RB, Alamanda CNC, Rachman RW, Robiani R, Solihin I. 2021. Analysis of the SARS-CoV-2 envelope (E), nucleocapsid (N), and non-structural proteins12 (nsp12) gene from COVID-19 patients in West Java. Journal of Microbial Systematics and Biotechnology 3(1), 9-20

Abstract
According to World Health Organization, as of January 2021, Indonesia is the only Southeast Asian country in which COVID-19 is still occurring in community transmission. West Java is one of the provinces holding the highest positive cases number. With the envelope (E), nucleocapsid (N), and non-structural protein 12 (nsp12) being the target genes of SARS-CoV-2 diagnostic kits and several antiviral drugs, the study of genetic variations has become relevant and greatly important. Out of 267 oro-nasopharyngeal swab specimens that were previously confirmed positive for COVID-19 in qPCR diagnostic test in Laboratorium Kesehatan Provinsi Jawa Barat, ten samples with acceptable qualities were selected and three samples were sequenced using Sanger sequencing. Nonsynonymous mutations were observed in the envelope gene (L21F) and in the nucleocapsid genes (R203K, G204R, A211S, and S193I). Phylogenetic analysis showed that samples were clustered with other sequences carrying identical mutations, but clustered non-discriminatively with all sequences when carrying no mutation. No pattern in geographical areas and clades, except for R203K-G204R for being a marker for the GR clade. Protein structure analysis showed that mutations observed did not change the hydrophobicity and the secondary structure of the nucleocapsid, while stability change (ΔΔG) showed that all mutations, aside from the R203K-G204R, have neutral effect on the protein stability. Therefore, it can be concluded that mutations observed in this experiment did not impart preference to disperse in certain geographical areas or cause any significant structural change in the protein.

Keywords: envelope, nsp12, nucleocapsid, SARS-CoV-2, Sanger sequencing
Introduction

The SARS-CoV-2 causing the COVID-19 pandemic is a positive single-stranded RNA virus belonging to Nidovirales order, Cornidovirinae suborder, Coronaviridae family, Orthocoronavirinae subfamily, Betacoronavirus genus, and Sarbecovirus subgenus. Other Coronaviruses that also belong to the Betacoronavirus genus are the SARS-CoV that caused SARS (severe acute respiratory syndrome) and the MERS-CoV that caused MERS (Middle East respiratory syndrome). Other human coronaviruses, known to cause common colds in humans, are hCoV OC43 (Organ Culture 43) and HKU1 (Hong Kong University 1) from the Betacoronavirus genus, and 229E (specimen code 229E) and NL63 (Netherland 63) from the Alphacoronavirus genus (Liu et al., 2020).

Since its first emergence on December 2019 in Wuhan, China, COVID-19 has infected 88 million individuals, as of January 2021, and has taken more than 1.9 million lives worldwide (World Health Organization, 2021). Considering the high number of positive cases and death toll caused by this disease, the development of SARS-CoV-2 diagnostic kits in parallel with the development of antiviral drugs becomes critical to reduce the spread of this virus, the efficacy and efficiency of which the presence of mutations in the viral genome should not hinder (Kilic et al., 2020). This is especially relevant in countries that have not yet tackled the issue regarding the uncontrollable spread of this virus-one of them is Indonesia, which is currently the only Southeast Asian country that has not yet escaped the community transmission category in WHO observation (World Health Organization, 2021). Moreover, out of its 34 provinces, West Java is among the provinces with the highest number of positive cases and death tolls, along with the national capital Jakarta and East Java.

Despite being a central part in the viral replication, nsp12 is a region with the highest number of mutations in the viral genome-globally, in Indonesia nationally, or in West Java in specific (Pachetti et al., 2020; Internal Laboratorium Kesehatan Provinsi Jawa Barat, 2020). Not only nsp12, the nucleocapsid is also observed to hold the third-highest number of mutations throughout the viral genome, coming only after the ORF1AB and spike region (Koyama et al., 2020). This raised the importance of studying the genetic variations within these genes, considering nsp12 being the target of the only COVID-19 drug with emergency use authorization from the Food and Drugs Administration (Robson et al., 2020). Changes in the secondary and tertiary structures of the nsp12 and nucleocapsid due to genetic variations have also been speculated to alter the functions of these proteins (Rahman et al., 2020), and even as far as interfering the interactions of these proteins with the designed drugs (Kumar et al., 2020; Tung & Limtung, 2020).

The envelope, on the other side, is not observed to house as many mutations (Dilucca et al., 2020), but the exclusion of this particular gene from mutation studies will prove to be unwise, considering the pivotal role this protein plays in the virion assembly, induction of budding in the replication process, and the formation of the viroporin structure (Schoeman & Fielding, 2019). The envelope is also known to be targeted by other antiviral drugs for coronaviruses, such as amantadine and rimantadine.

Therefore, the study of genetic variations and mutations of these genes and its implication on the protein structure and phylogenetic relationship to other SARS-CoV-2 sequences, particularly in West Java, Indonesia, become relevant.

Materials and methods

Primer design

Conserved regions within target genes (nsp12, N, and E gene) were obtained by aligning reference and random sequences using ClustalW algorithm. Reference sequences include full SARS-CoV-2 genome (NCBI RefSeq Accession ID NC_045512.2), all available
Indonesian SARS-CoV-2 sequences (as of July 2020), random full SARS-CoV-2 genome worldwide, and random sequences of SARS-CoV, MERS-CoV and human coronavirus (hCoV 229E, NL63, OC43, HKU1). Primer sets were generated using SnapGene 1.1.3, and the final sets were then selected based on the intrinsic properties of each primer evaluated by OligoAnalyzer Tool.

**Data collection**

Samples were oro-nasopharyngeal swab specimens in viral transport media isolated from COVID-19 patients in Laboratorium Kesehatan Provinsi Jawa Barat (Labkes) from September 1st to 15th of 2020. Inclusion criteria of samples were: (a) test positive for COVID-19 when detected with commercial diagnostic qPCR kits in all target genes, (b) give Cq values below 25 in all target genes, and (c) be contained in viral transport media with acceptable conditions (showing no signs of drying, change in turbidity, and are available in sufficient volume). This study was conducted in accordance to the approval of Padjajaran University Research Ethics Committee (No: 320/UN6.KEP/EC/2021)

**Isolation and sequencing of target genes**

RNA extraction was performed with QIAamp Viral RNA Mini Extraction Kit (QIAGEN), followed by reverse transcription by Tetro cDNA Synthesis Kit (Bioline USA) using random hexamer. Amplification was performed using GoTaq Green Master Mix (Promega). Samples were purified agarose incisions containing only the desired electrophoretic bands. Sequencing of target genes was performed using Sanger sequencing (with Applied Biosystems SeqStudio Genetic Analyzer), using primer sets previously used in amplification.

**In-silico phylogenetic, protein structure, and protein stability analysis**

Sequencing reads were trimmed and assembled into contiguous sequences and variations of sequencing reads from the reference sequence were used to establish the phylogenetic analysis of target genes. Phylogenetic trees were then generated using the Jukes–Cantor genetic distance model and Neighbor-Joining clustering method. Sequences included in the trees are those of the analysed samples, SARS-CoV-2 reference sequence (NCBI RefSeq Accession ID NC_045512.2), random SARS-CoV-2 sequences worldwide (representing each WHO region), and all Indonesian SARS-CoV-2 sequences available in GISAID database. Phylogenetic trees were expected to illustrate the genetic distance between samples and other SARS-CoV-2 sequences circulating worldwide.

Secondary protein structure analysis was performed by measuring hydrophobicity with Kyte-Doolittle scale and predicting secondary structure with RaptorX Property method. Using nsp12 (ID PDB 6M71), N (ID PDB P0DTC9), and E (ID PDB 5X29) proteins modelled by i-TASSER (University of Michigan) as references, mutant protein stability analysis was performed using foldX plugin for YASARA (European Molecular Biology Laboratory and Center for Genomic Regulation) by measuring the Gibbs free energy of wildtype and mutant proteins.

**Results and Discussion**

**Data collection**

Between September 1st and 15th of 2020, 267 samples were collected from oro-nasopharyngeal swab specimens that previously tested positive in SARS-CoV-2 qPCR diagnostic kit in Labkes Jawa Barat. As a result, 196 samples showed amplification in all target genes used in qPCR (ORF1AB, N and E), with 41 samples giving Cq below 25, and 14
samples observed to still be in acceptable conditions. 10 samples were then selected, as shown in Table 1 below.

**Table 1.** Profile of samples used in this experiment. Samples in grey are not included in the experiment.

| Sample | Sample Code | Sex | Age (years) | City | Specimen Collection Date | Cq: | human RNase P (internal control) |
|--------|-------------|-----|-------------|------|--------------------------|-----|---------------------------------|
|        |             |     |             |      |                          |     | ORF1ab-RdRp                     |
| A      | R76116      | F   | 38          | Kuningan | 01/09/2020               | 15.80 | 14.52 17.62 19.42              |
| B      | R82914      | M   | 66          | Sukabumi | 04/09/2020               | 20.63 | 19.32 21.52 21.04              |
| C      | R83053      | F   | 34          | Sumedang | 04/09/2020               | 20.40 | 19.91 21.41 17.81              |
| D      | L14840      | F   | 59          | Sleman   | 04/09/2020               | 16.29 | 15.44 16.06 17.25              |
| E      | L14839      | M   | 63          | Sleman   | 04/09/2020               | 13.54 | 12.20 13.14 30.38              |
| F      | R76088      | M   | 5           | Kuningan | 01/09/2020               | 17.49 | 15.53 18.17 26.46              |
| G      | R74919      | F   | 34          | Bandung  | 31/08/2020               | 15.65 | 16.14 15.52 29.22              |
| H      | R76079      | M   | 37          | Kuningan | 01/09/2020               | 20.83 | 18.64 21.52 21.79              |
| I      | R76269      | F   | 16          | Sukabumi | 01/09/2020               | 18.66 | 17.66 18.26 23.15              |
| J      | R75479      | F   | 25          | Bandung  | 31/08/2020               | 19.12 | 14.54 18.07 16.98              |
| K      | R75489      | M   | 31          | Bandung  | 31/08/2020               | 10.86 | 7.91 10.12 19.74               |
| L      | R82899      | M   | 66          | Sukabumi | 04/09/2020               | 19.70 | 18.98 20.57 21.37              |
| M      | R74721      | F   | 53          | Kuningan | 31/08/2020               | 19.11 | 17.49 18.67 19.38              |
| N      | R75284      | M   | 38          | Bandung  | 31/08/2020               | 17.25 | 16.71 16.94 30.17              |

**Primer design**

Primer sets were designed by aligning Indonesian SARS-CoV-2 sequences, and random global SARS-CoV-2, MERS-CoV, SARS-CoV, hCoV 229E, hCoV NL63, hCoV OC43, and hCoV HKU1 sequences. After selecting against primer candidates with unsatisfactory properties, the following sets of primer were obtained, as shown below in Table 2.

**Table 2.** List of primers

| Name of Primers | Sequences                                      |
|-----------------|------------------------------------------------|
| E Forward       | 5'-GACGACTACTAGCGTGCCCTTTG-3'                 |
| E Reverse       | 5'-GGTAAATAGTACCGTTGGAATCTGCG-3'              |
| N Forward 1     | 5'-GGGTAGTCTTGTAGTCGTTGTC-3'                  |
| N Forward 2     | 5'-GACGTTGTCCAGAAACAAAAAC-3'                  |
| N Reverse 1     | 5'-GCCGTGATGTTGACGACTG-3'                     |
| N Reverse 2     | 5'-CGGTGAACCAAAGACGCAG-3'                     |
| NSP12 Forward 1 | 5'-CTAGCTAGTGGGGACAACC-3'                     |
| NSP12 Forward 2 | 5'-GCACCGTGCTTTGGGATACTG-3'                   |
| NSP12 Forward 3 | 5'-GCATCTCAAGGTCTAGTGCGTAAC-3'                |
| NSP12 Reverse 1 | 5'-CTGTCAGACGACACTTCTACGTAC-3'               |
| NSP12 Reverse 2 | 5'-GTAGCTCTCTCTAGTGGCGGC-3'                   |
| NSP12 Reverse 3 | 5'-GCTTGCCGTACACGTTACC-3'                    |
Amplification of target genes

Amplification of target genes was performed using the outermost primer sets for the nucleocapsid and envelope genes (E_F and E_R, N_R1 and N_F1), while amplification of nsp12 used two sets of primers (nsp12_F1 and nsp12_R2, nsp12_F2 and nsp12_R1) to accommodate the full coverage of the significantly longer gene. Not all samples gave clear and defined electrophoretic band, as shown in Figures 1 and 2. Samples D, G, and I (L14840, R74919, and R76269 respectively) were then selected for giving amplification in all target genes.

Figure 1. Amplification of the RdRp-A and RdRp-B fragment. Electrophoresis was performed using 1.8% agarose gel and DNA HyperLadder™ 1kb DNA ladder (shown as Ld). Negative control for cDNA synthesis is shown as C(-), and negative control for amplification is shown as P(-). (a.) Amplification of the RdRp-A fragment is expected to yield an amplicon of 2083 bp. (b.) Amplification of the RdRp-B fragment is expected to yield an amplicon of 2139 bp.

Figure 2. Amplification of the nucleocapsid and envelope gene. Electrophoresis was performed using 1.8% agarose gel and DNA HyperLadder™ 1kb DNA ladder (shown as Ld). Negative control for cDNA synthesis is shown as C(-), and negative control for amplification is shown as P(-). (a.) Amplification of the nucleocapsid gene is expected to yield an amplicon of 1352 bp. (b.) Amplification of the envelope gene is expected to yield an amplicon of 354 bp.
**In-silico phylogenetic, protein structure, and protein stability analysis**

Trimming and assembly of sequencing read yielded several genetic variations in the three samples when compared to the reference sequence, summarized below in Table 3. However, the attempt to sequence the full length of nsp12 did not provide satisfactory results in all samples and be omitted from further analysis. R76269 was also observed to give insufficient reading quality for the N gene, and so had to be omitted from further analysis regarding the nucleocapsid.

**Table 3.** Several genetic variations were observed in E and N gene in the three designated samples.

| Target gene | Sample D (L14840) | Sample G (R74919) | Sample I (R76269) |
|-------------|-------------------|-------------------|-------------------|
| E           | R203K, G204R, A211S | S193I             | L21F             |
| N           |                   |                   |                   |

One nonsynonymous mutation, L21F, was observed in the envelope region in R76269 (sample I) where a change of leucine into phenylalanine occurred on the 21st amino acid. As of January 26th 2021, L21F is observed with a frequency of 0.06% worldwide, documented in only 30 countries worldwide, and with only 6 documented cases in Indonesia out of a total of 196 sequences. Due to its low frequency, none of the randomly sampled sequences in the phylogenetic tree carried the L21F, resulting in R76269 clustering non-discriminatively with all the sequences in the tree. To resolve this, 20 additional sequences known to carry this mutation were added in order to provide a more meaningful phylogenetic tree. R76269 was then observed to cluster together with the L21F mutants, as seen in Figure 3.

Three nonsynonymous mutations were observed in the nucleocapsid region in L14840 (sample D): R203K, G204R, and A211S; while R74919 (sample G) was observed to only carry one nonsynonymous mutation, S193I. R203K and G204R coexisted as a result of three consecutive mutations involved in coding the 203rd and 204th amino acid while A211S and S193I occurred due to a point mutation in a single nucleotide. A product of three sequential point mutations, R203K and G204R are observed to almost always coincide and occur with such a high frequency worldwide in their coexisting form that R203K-G204R became the marker gene for the GR SARS-CoV-2 clade (Mercatelli & Giorgi, 2020). Indeed, Figure 3 showed that all sequences that clustered together with L14840 in the phylogenetic tree belong to the GR clade, except one sequence that belongs to the G clade, a parent of GR clade, possibly due to the ambiguous nucleotide in the marker position. Currently, R203K-G204R can be found in all continents, as shown in Figure 4, and is recorded in 22.9% of all sequences worldwide. S193I, with a frequency of 0.30%, is documented in 1576 sequences worldwide, 8 of them being Indonesian sequences that evidently clustered together in the phylogenetic tree. A211S, on the other hand, is documented with a frequency 0.04%, and is recorded in only 255 sequences worldwide, with only 2 cases found in Indonesia.
Figure 3. Phylogenetic tree of the nucleocapsid and envelope protein. Shown in red is the SARS-CoV reference sequence (NC_004718.3) acting as the outgroup, in orange non-human SARSr-CoV reference sequence (batCoV RaTG13; EPI_ISL_402131), and in green SARS-CoV-2 reference sequence (NC_045512.2). Black stars denote the sequenced samples. (a.) Shown in teal are sequences carrying S193I and blue carrying R203K-G204R in the nucleocapsid. (b.) Shown in magenta are sequences carrying L21F in the envelope.
Figure 4. Global distribution of mutations occurring in the envelope and nucleocapsid protein. Shown in faint blue are countries with low number of occurrences of this mutation, and red with high number of occurrences. The mutation is not documented to occur in countries in white. (a.) L21F, (b.) S193I, (c.) R203K-G204R, and (d.) A211S.

In addition to the four nonsynonymous mutations in the nucleocapsid, one ambiguous peak was observed in R74919 (Sample G) on the 277th nucleotide (relative to the reading frame), bearing equal amount of guanine and adenine. Although it is still unclear whether the ambiguous peak is a result of a polymerase mistake or the existence of two different populations, this ambiguous peak will nevertheless code for arginine, indicating that this phenomenon will result in synonymous mutation.

It should be noted that these phylogenetic trees demonstrate that none of these mutations, apart from the R203K-G204R, have any affinity to appear in certain clades or geographical areas as these mutations are not located in the position of any of the markers and do not show any causal relationship with heightened transmissibility.

Protein structure and stability analysis of SARS-CoV-2 E and N gene

L21F is located in the transmembrane domain of the envelope protein, but as leucine and phenylalanine are both hydrophobic amino acids, L21F is not presumed to contribute a significant change in the overall protein structure. Indeed, further analysis using Kyte-Doolittle hydrophobicity scale showed that amino acids surrounding the mutation site are hydrophobic when carrying a leucine or a phenylalanine on the 21st position. L21F is located in a helix-dominated region (Wang et al., 2016), and secondary structure modelling using RaptorX Property method in Table 4 showed that a change from leucine to phenylalanine did not change the secondary structure of the protein around the mutation site. Furthermore, protein stability analysis using foldX showed that L21F increased the overall protein stability of envelope protein by 0.29 kcal/mole, indicating that this mutation theoretically does not make any significant change on the protein thermodynamically. In addition to these results, this particular mutation is also observed not to disturb the residues known to heavily drive the oligomerization of the viroporin, Asn15 and Val25 (Schoeman & Fielding, 2019), and so it can be safely surmised that L21F does not impart a significant change regarding the ability of the envelope to form the viroporin (Nieto-Torres et al., 2014). With this in mind, it should...
also stand to reason that L21F will not interfere with the interaction with envelope-targeting drugs.

**Table 4.** Secondary structure comparison stability change between wildtype envelope and nucleocapsid protein and their mutants. Secondary structures shown are helices (h in red), and coils (c in grey), while mutation sites are shown in bold. Stability change is shown as ΔGmutant – ΔGwildtype. Also shown is the stability change of proteins from samples carrying more than one mutation site.

| Protein          | Envelope | Nucleocapsid |
|------------------|----------|--------------|
| Mutation         | L21F     | S193I        |
| Sample           | R76269   | R74919       |
| Sample           | L14840   | A211S        |

| Secondary structure comparison | Wildtype | Mutants |
|---------------------------------|----------|---------|
| Amino acid position             | h        | h       | h       |
| Secondary structure             |          |         |         |
| Amino acid position             |          |         |         |
| Secondary structure             |          |         |         |
| Stability (ΔG) (kcal/mol)       | 60.52    | 248.55  |
| Stability change (ΔΔG) (kcal/mol)| 60.81   | 248.13  |
| ΔG > 0.5 = non-favourable       | 0.29     | -0.42   |
| -0.5 < ΔG < 0.5 = neutral       |          |         |
| ΔG < -0.5 = favourable          | 1.46     | 1.5     |

All of the mutations observed in the nucleocapsid are located in the linker domain (Zeng et al., 2020), with S193I and R203K-G204R located in a serine-arginine-rich region, known for its function in phosphorylation, and A211S located adjacent to this region (Maitra et al., 2020). All mutations involved changed the polarity of amino acids on the mutation sites. However, R203K-G204R is not only known not to disturb the overall performance of this particular serine-arginine-rich region (Cortey et al., 2020), it is also hypothesized to contribute for heightened stability of this region as measurements of free energy of this motive in isolation when carrying the R203K-G204R mutation showed a more favourable structure (Rahman et al., 2020). This may or may not account for the wide spread of this mutation, as studies surrounding this conjecture are yet to be carried out. Similarly, A211S is also hypothesized to possibly increase the conformational flexibility of this region (Garvin et al., 2020). Due to its limited spread, the understanding around A211S is not yet complete.

Linker proteins are known to consist of hydrophilic amino acids, with coiled or bent secondary structure to facilitate its function to interact with other molecules (George & Heringa, 2002). Indeed, Kyte-Doolittle hydrophobicity scale showed that the mutation sites in the wildtype and mutant proteins are both located in a hydrophilic region and secondary structure modelling RaptorX Property method showed that both wildtype and mutant proteins gave the same coil-dominated secondary structure. Protein stability analysis, however, showed that S193I and A211S do not give any thermodynamic advantage for the protein.
stability. Counterintuitively, R203K-G204R is observed to be the mutation that lessen the thermodynamic stability of the nucleocapsid protein most out of the other mutations, despite being the most widely spread worldwide. It is predicted that the mutations observed in the nucleocapsid in this experiment do not impart a significant change to the protein, and that correlation between the wide spread of some of the mutations might be due to factors that have not been explored in this study. Due to its dynamic form, studies regarding the effect of mutation in the nucleocapsid, specifically the interaction with antiviral drugs, are still scarce. Only one mention of possible alteration in this context was found in the 156th amino acid (Azad, 2021), but no study has proven the effect of R203K-G204R on the ability of antiviral drugs to bind on the nucleocapsid, although this variation is widely spread across the globe.

**Conclusion**

One of three samples of the Indonesian SARS-CoV-2 that was successfully sequenced showed L21F mutation in enveloped protein, while two other samples showed R203K, G204R, A211S, and S193I mutation in nucleocapsid protein. However, all of the mutations did not change the secondary and the tertiary structure of E and N SARS-CoV-2 Protein. All samples are phylogenetically close to the SARS-CoV-2 Wuhan reference sequences and cluster together with the similar SARS-CoV-2 mutants from other countries.

It should also be noted that this study did not include any experiments that could illuminate any causal relationship between the mutations and the viral viability, pathogenicity, and the overall transmissibility of the disease, and so further research to cover this scope should prove to be interesting.

**Conflict of interest**

The authors state no conflict of interest from this manuscript.

**Acknowledgment**

This research was supported by the Laboratorium Kesehatan Provinsi Jawa Barat, Indonesia.

**Author contributions**

ER, RBR, CNCA, RWR, RR, IS supplied the oro-nasopharyngeal swab specimens and laboratory facilities used in this experiment. AF, IAE, GAPP, RS performed all in-silico analysis, the isolation of target genes, and wrote the manuscript. All authors designed the experiments, and analyzed the data.

**References**

Azad GK. 2021. The molecular assessment of SARS-CoV-2 Nucleocapsid Phosphoprotein variants among Indian isolates. Heliyon. 7, e06167. DOI: 10.1016/j.heliyon.2021.e06167.

Cortey M et al. 2020. N SARS-CoV-2 amino acid substitutions widely spread in the human population are mainly located in highly conserved segments of the structural proteins. o Title. bioRxiv. DOI: https://doi.org/10.1101/2020.05.16.099499.

Diluca M, Forcelloni S, Georgakilas AG, Giansanti A, Pavlopoulou A. 2020. Codon Usage and Phenotypic Divergences of SARS-CoV-2 Genes. Viruses. 12, 498. DOI: 10.3390/v12050498.

Garvin MR et al. 2020. Potentially adaptive SARS-CoV-2 mutations discovered with novel spatiotemporal and explainable AI models. Genome Biol. 21, 1–26. DOI: 10.1186/s13059-020-02191-0.

George RA, Heringa J. 2002. An analysis of protein domain linkers: Their classification and
role in protein folding. Protein Eng. DOI: 10.1093/protein/15.11.871.
Internal Balai Laboratorium Kesehatan Provinsi Jawa Barat. 2020. Analisis Genom SARS-CoV-2 Indonesia.
Kilic T, Weissleder R, Lee H. 2020. Molecular and Immunological Diagnostic Tests of COVID-19: Current Status and Challenges. iScience. 23, 101406. DOI: 10.1016/j.isci.2020.101406.
Koyama T, Platt D, Parida L. 2020. Variant analysis of SARS-cov-2 genomes. Bull. World Health Organ. 98, 495–504. DOI: 10.2471/BLT.20.253591.
Kumar A et al. 2020. Characterization of Nucleocapsid (N) Protein from Novel Coronavirus SARS-CoV-2. Preprints. DOI: 10.20944/preprints202005.0413.v1.
Liu DX, Liang JQ, Fung TS. 2020. Human Coronavirus-229E, -OC43, -NL63, and -HKU1. In: Reference Module in Life Sciences. Elsevier. DOI: 10.1016/b978-0-12-809633-8.21501-x.
Maitra A et al. 2020. Mutations in SARS-CoV-2 viral RNA identified in Eastern India: Possible implications for the ongoing outbreak in India and impact on viral structure and host susceptibility. J. Biosci. 45. DOI: 10.1007/s12038-020-00046-1.
Mercatelli D, Giorgi FM. 2020. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Front. Microbiol. 11, 1800. DOI: 10.3389/fmicb.2020.01800.
Nieto-Torres JL et al. 2014. Severe Acute Respiratory Syndrome Coronavirus Envelope Protein Ion Channel Activity Promotes Virus Fitness and Pathogenesis Denison, MR, editor. PLoS Pathog. 10, e1004077. DOI: 10.1371/journal.ppat.1004077.
Pachetti M et al. 2020. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J. Transl. Med. 18, 179. DOI: 10.1186/s12967-020-02344-6.
Rahman MS et al. 2020. Evolutionary dynamics of SARS-CoV-2 nucleocapsid (N) protein and its consequences. J. Med. Virol. DOI: 10.1002/jmv.26626.
Robson F et al. 2020. Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. Mol. Cell. 79, 710–727. DOI: https://doi.org/10.1016/j.molcel.2020.07.027.
Schoeman D, Fielding BC. 2019. Coronavirus envelope protein: Current knowledge. Virol. J. 16, 1–22. DOI: 10.1186/s12985-019-1182-0.
Tung HYL, Lintung P. 2020. Mutations in the phosphorylation sites of SARS-CoV-2 encoded nucleocapsid protein and structure model of sequestration by protein 14-3-3. Biochem. Biophys. Res. Commun. 532, 134–138. DOI: 10.1016/j.bbrc.2020.08.024.
Wang S, Ma J, Xu J. 2016. AUCpreD: proteome-level protein disorder prediction by AUC-maximized deep convolutional neural fields. Bioinformatics. 32, i672–i679. DOI: 10.1093/bioinformatics/btw446.
World Health Organization. 2021. WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int/.
Zeng W et al. 2020. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. Biochem. Biophys. Res. Commun. 527, 618–623. DOI: 10.1016/j.bbrc.2020.04.136.