

Analysis of SARS-CoV-2 nucleocapsid protein sequence variations in ASEAN countries

Mochammad Rajasa Mukti Negara,¹ Ita Krissanti,^{2,3} Gita Widya Pradini³



pISSN: 0853-1773 • eISSN: 2252-8083
<https://doi.org/10.13181/mji.oa.215304>
Med J Indones. 2021;30:89–95

Received: January 23, 2021
Accepted: April 14, 2021
Published online: June 18, 2021

Authors' affiliations:

¹Faculty of Medicine, Universitas Padjadjaran, Sumedang, Indonesia,
²Veterinary Medicine Program, Faculty of Medicine, Universitas Padjadjaran, Sumedang, Indonesia, ³Department of Biomedical Science, Faculty of Medicine, Universitas Padjadjaran, Sumedang, Indonesia

Corresponding author:

Gita Widya Pradini
 Department of Biomedical Science,
 Faculty of Medicine, Universitas Padjadjaran, Jalan Raya Bandung Sumedang km 21, Jatinangor, Sumedang 40161, West Java, Indonesia
 Tel/Fax: +62-22-7795594/
 +62-22-7795595
 E-mail: gita.widya.pradini@unpad.ac.id

ABSTRACT

BACKGROUND Nucleocapsid (N) protein is one of four structural proteins of SARS-CoV-2 which is known to be more conserved than spike protein and is highly immunogenic. This study aimed to analyze the variation of the SARS-CoV-2 N protein sequences in ASEAN countries, including Indonesia.

METHODS Complete sequences of SARS-CoV-2 N protein from each ASEAN country were obtained from Global Initiative on Sharing All Influenza Data (GISAID), while the reference sequence was obtained from GenBank. All sequences collected from December 2019 to March 2021 were grouped to the clade according to GISAID, and two representative isolates were chosen from each clade for the analysis. The sequences were aligned by MUSCLE, and phylogenetic trees were built using MEGA-X software based on the nucleotide and translated AA sequences.

RESULTS 98 isolates of complete N protein genes from ASEAN countries were analyzed. The nucleotides of all isolates were 97.5% conserved. Of 31 nucleotide changes, 22 led to amino acid (AA) substitutions; thus, the AA sequences were 94.5% conserved. The phylogenetic tree of nucleotide and AA sequences shows similar branches. Nucleotide variations in clade O (C28311T); clade GR (28881–28883 GGG>AAC); and clade GRY (28881–28883 GGG>AAC and C28977T) lead to specific branches corresponding to the clade within both trees.

CONCLUSIONS The N protein sequences of SARS-CoV-2 across ASEAN countries are highly conserved. Most isolates were closely related to the reference sequence originating from China, except the isolates representing clade O, GR, and GRY which formed specific branches in the phylogenetic tree.

KEYWORDS nucleocapsid proteins, phylogeny, SARS-CoV-2, sequence alignment

Coronaviruses are a large virus family that cause infectious diseases in animals and humans.¹ Coronavirus has been recorded to cause pandemics several times, including severe acute respiratory syndrome (SARS) caused by SARS coronavirus (SARS-CoV) in 2002 and Middle East respiratory syndrome (MERS) caused by MERS coronavirus (MERS-CoV) in 2012.² It is an enveloped virus with a spherical virions structure and diameter of 70–160 nm. It has a single-stranded, non-segmented, positive-sense RNA

genome. Its nucleocapsid has a helical shape and diameter of 11–13 nm.³

SARS-CoV-2 is the seventh discovery of coronavirus that causes infectious disease in humans and is classified into *Betacoronavirus*.¹ It has a 79% similarity to SARS-CoV and 50% to MERS-CoV.⁴ Having 29,903 nucleotides in the genome makes it the second-largest RNA virus. This viral genome has 11 open reading frames (ORFs) encoding 27 proteins. The first ORF (ORF 1/ab) covers two-thirds of the viral genome and encodes 16

nonstructural proteins, while the other one-third of the genome encodes four structural proteins and six additional proteins (accessory).^{5,6}

SARS-CoV-2 has a structural protein named nucleocapsid (N) protein. This protein has 1,259 nucleotides located from nucleotides 28,274 to 29,533 and is encoded by 419 amino acids (AAs).^{7,8} It has two domains, which are the N-terminal domain (46–176 AA) and C-terminal domain (247–364 AA). Between these two domains, there is a linkage region (182–247 AA) containing the serine/arginine-rich domain (184–204 AA). These three domains have electrostatic interactions with the RNA genome.^{9,10}

The N protein of SARS-CoV-2 has a vital role during the transcription and replication of viral RNA. It forms a ribonucleoprotein helix during RNA genome packaging and regulates viral RNA synthesis during replication, transcription, and modulation of infected cells' metabolism.¹¹ It is widely expressed during infection as the interferon antagonists.¹⁰ It is also a highly immunogenic protein. The anti-N protein antibodies (IgG, IgA, and IgM) were detected in confirmed coronavirus disease 2019 (COVID-19) patients' sera.⁹ This region is conserved, with only three variations found (S194L, K249I, and P344S).¹² Hence, this protein has been used as a protein target for SARS-CoV-2 diagnostic and future vaccine.^{10,13} There are limited data analyzing the sequence variation of SARS-CoV-2 N protein along with the translated AAs in ASEAN countries. This study aimed to analyze the variation of the SARS-CoV-2 N protein sequences in ASEAN countries, including Indonesia. The phylogenetic tree was also generated to describe phylogenetic relatedness among N sequence isolates in ASEAN countries.

METHODS

This study analyzed the SARS-CoV-2 N protein gene sequences from GenBank and Global Initiative on Sharing All Influenza Data (GISAID) from December 2019 to March 2021. The SARS-CoV-2 N protein reference sequence (GenBank accession number: NC_045512.2) was downloaded from GenBank (www.ncbi.nlm.nih.gov/genbank). SARS-CoV-2 N protein gene sequences from ASEAN countries were downloaded from the GISAID website under the EpiCoV™ platform (www.gisaid.org). Data collection from GISAID were adjusted to the countries; two

of each GISAID clade were taken as representative isolates. The collected sequences were aligned by MUSCLE using the MEGA-X software version 10.1.8.¹⁴

Multiple sequence alignment of nucleotide was done to analyze nucleotide substitution mutations and their impact on translated AA (synonymous and non-synonymous mutation) compare with the reference sequence. Non-synonymous mutation, which led to AA substitution, was then subsequently checked for the AA polarity changes based on the AA classification.¹⁵ The phylogenetic trees of nucleotide sequences were built with MEGA-X software using the maximum likelihood method and the Tamura-Nei model with a 1,000 replications of the bootstrap test. The translated protein sequences used the maximum likelihood method and the Jones-Taylor-Thornton model with 1,000 replications of the bootstrap test to construct the AA phylogenetic tree. Analysis of variations in the multiple sequence alignment was carried out on nucleotide and translated protein sequences. Nucleotide and AA phylogenetic trees were built to determine the relationship of the N gene of SARS-CoV-2 from ASEAN countries.

RESULTS

Ninety eight isolates represented the SARS-CoV-2 N protein complete coding sequences (each sequence was 1,259 nucleotides in length) were collected (Table 1). No isolate was found in the clade GV from ASEAN countries, and there were no data from Laos in the GISAID database (data were collected up to March 2021). Among all isolates, there were 31 nucleotide changes (2.4%) and 97.5% of sequences were being conserved. All of the clade O isolates had substitution mutation at C28311T. Isolates from clade S had substitution mutation at G28378C. All isolates in the GR clade had substitution mutations at position 28881 (GGG to AAC). These mutations caused two AA substitutions (R203K and G204R). All isolates from clade GRY had four substitutions, making them branched out from the GR clade. The substitutions were at position 28280–28282 (GAT to CTA) and C28977T (Table 1). Other isolates which do not undergo any nucleotide changes can be found in Table 2.

The nucleotide sequences were then translated into AAs with appropriate codon and displayed the similarities and positions between all clades (Figure 1). Of all AA sequences, 94.5% were conserved with 22

Table 1. The nucleotide substitution mutations and corresponding AA substitution within nucleocapsid (N) protein gene

Nucleotide substitution mutations	AA substitution		AA polarity changes*	Isolates (country name GISAID accession ID clade designation)
Non-synonymous mutation				
G28280C, A28281T, T28282A	D3L	Asp3Leu	NP to P	Indonesia EPI_ISL_1169046 GRY, Indonesia EPI_ISL_1239143 GRY, Cambodia EPI_ISL_1098603 GRY, Cambodia EPI_ISL_1098602 GRY, Malaysia EPI_ISL_934424 GRY, Thailand EPI_ISL_1346626 GRY, Thailand EPI_ISL_877227 GRY
C28409T	P46S	Pro46Ser	NP to P	Thailand EPI_ISL_877227 GRY
C28311T	P13L	Pro13Leu	None (NP)	Brunei Darussalam EPI_ISL_435677 O, Brunei Darussalam EPI_ISL_443187 O, Philippines EPI_ISL_434555 O, Philippines EPI_ISL_434558 O, Indonesia EPI_ISL_560991 O, Malaysia EPI_ISL_459953 O, Malaysia EPI_ISL_528743 O, Myanmar EPI_ISL_512844 O, Singapore EPI_ISL_549004 O, Thailand EPI_ISL_455910 O
G28878A	S202N	Ser202Asn	None (P)	Indonesia EPI_ISL_1169052 S, Indonesia EPI_ISL_1257844 S, Cambodia EPI_ISL_918367 S
G28580T	D103Y	Asp103Tyr	None (P)	Singapore EPI_ISL_548994 GR
C28695T	T141I	Thr14Ile	P to NP	Malaysia EPI_ISL_501211 V
A28704G	D144G	Asp144Gly	P to NP	Singapore EPI_ISL_536442 GH
G28812T	S180I	Ser180Ile	P to NP	Vietnam EPI_ISL_511896 GH
G28842T	S190I	Ser190Ile	P to NP	Malaysia EPI_ISL_501206 V
G28851T	S193I	Ser193Ile	P to NP	Indonesia EPI_ISL_1117592 GH
C28854T	S194L	Ser194Leu	P to NP	Singapore EPI_ISL_536442 GH, Thailand EPI_ISL_447906 G, Indonesia EPI_ISL_1257824 G, Malaysia EPI_ISL_861722 GH, Malaysia EPI_ISL_877232 G, Malaysia EPI_ISL_877233 G, Thailand EPI_ISL_882796 GH, Thailand EPI_ISL_882797 GH
A28860T	N196I	Asn196Ile	P to NP	Malaysia EPI_ISL_861722 GH
G28881A, G28882A, G28883C	R203K G204R	Arg203Lys Gly204Arg	None (P) NP to P	Indonesia EPI_ISL_528747 GR, Indonesia EPI_ISL_538509 GR, Indonesia EPI_ISL_1300660 GR, Indonesia EPI_ISL_1284048 GR, Malaysia EPI_ISL_490101 GR, Malaysia EPI_ISL_501176 GR, Singapore EPI_ISL_548994 GR, Singapore EPI_ISL_549003 GR, Thailand EPI_ISL_447015 GR, Thailand EPI_ISL_455915 GR, Vietnam EPI_ISL_416431 GR, Vietnam EPI_ISL_511898 GR, Cambodia EPI_ISL_1098603 GRY, Cambodia EPI_ISL_1098602 GRY, Indonesia EPI_ISL_1169046 GRY, Indonesia EPI_ISL_1239143 GRY, Malaysia EPI_ISL_934424 GRY, Thailand EPI_ISL_1346626 GRY, Thailand EPI_ISL_877227 GRY
C28887T	T205I	Thr205Ile	P to NP	Indonesia EPI_ISL_1257841 GH, Indonesia EPI_ISL_1341793 G, Malaysia EPI_ISL_1186028 GH
G28908T	G212V	Gly212Val	None (NP)	Malaysia EPI_ISL_490101 GR
C28932T	A220V	Ala220Val	None (NP)	Malaysia EPI_ISL_728158 GV, Thailand EPI_ISL_708804 GV, Thailand EPI_ISL_708821 GV
C28977T	S235F	Ser235Phe	P to NP	Thailand EPI_ISL_1346626 GRY, Thailand EPI_ISL_877227 GRY, Malaysia EPI_ISL_934424 GRY, Cambodia EPI_ISL_1098602 GRY, Cambodia EPI_ISL_1098603 GRY, Indonesia EPI_ISL_1169046 GRY, Indonesia EPI_ISL_1239143 GRY
G29000T	G243C	Gly243Cys	None (P)	Malaysia EPI_ISL_728158 GV
G29384A	D371N	Asp371Asn	None (P)	Thailand EPI_ISL_458024 G
G29402T	D377Y	Asp377Tyr	None (P)	Indonesia EPI_ISL_1284114 GR
C29415T	A381V	Ala381Val	None (NP)	Indonesia EPI_ISL_568694 L

Table continued on next page

Table 1. (continued)

Nucleotide substitution mutations	AA substitution		AA polarity changes*	Isolates (country name GISAID accession ID clade designation)
Synonymous mutation				
A29086T	NA	NA	NA	Malaysia EPI_ISL_528743 O
C29218T	NA	NA	NA	Malaysia EPI_ISL_528743 O
A28615G	NA	NA	NA	Thailand EPI_ISL_437611 V
C29509T	NA	NA	NA	Thailand EPI_ISL_882796 GH, Thailand EPI_ISL_882797 GH
G28378C	NA	NA	NA	Indonesia EPI_ISL_1169052 S, Indonesia EPI_ISL_1257844 S, Cambodia EPI_ISL_918367 S
T28861C	NA	NA	NA	Indonesia EPI_ISL_1117592 GH

AA=amino acid; GISAID=Global Initiative on Sharing All Influenza Data; NA=no AA changes; NP=non-polar AA; P=polar AA
 *The polarity of AAs was based on the AA classification¹⁵

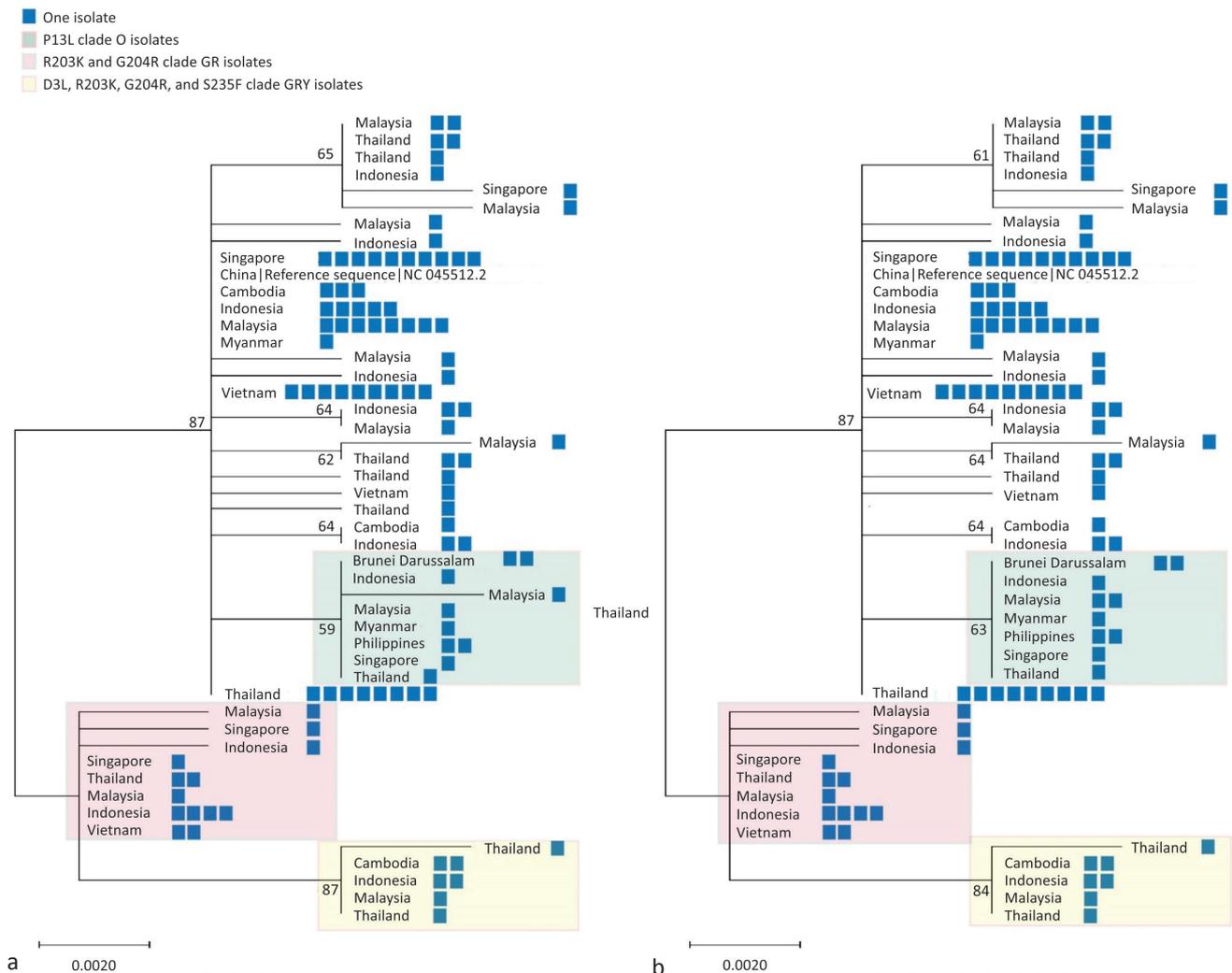


Figure 1. Phylogenetic trees of nucleotide (a) and translated protein sequences (b)

sites of AA substitution (non-synonymous mutation) (Table 1).

Twelve out of 22 non-synonymous mutations resulted in polarity changes of the corresponding AAs (Table 1). Of 12 non-synonymous mutations, 9

were changed from polar AA to non-polar AA, and 3 substitutions changed from non-polar AA to polar AA. Ten isolates in the clade O had the P13L (substitution of proline to leucine at position 13), which both were non-polar AAs.

Table 2. Conserved N sequence isolates

Country (n)	GISAID accession ID	Clade
Cambodia (3)	EPI_ISL_411902	L
	EPI_ISL_1096347	GH
	EPI_ISL_1096348	GH
Indonesia (5)	EPI_ISL_435281	L
	EPI_ISL_435283	O
	EPI_ISL_437188	GH
	EPI_ISL_568689	GH
	EPI_ISL_574619	G
Malaysia (7)	EPI_ISL_416829	L
	EPI_ISL_455313	GH
	EPI_ISL_455790	S
	EPI_ISL_490015	G
	EPI_ISL_490048	L
	EPI_ISL_490103	GH
	EPI_ISL_501222	G
Myanmar (1)	EPI_ISL_434709	G
Singapore (10)	EPI_ISL_420104	V
	EPI_ISL_420111	L
	EPI_ISL_428828	G
	EPI_ISL_443231	S
	EPI_ISL_462433	O
	EPI_ISL_538458	S
	EPI_ISL_548998	L
	EPI_ISL_548999	G
	EPI_ISL_549005	GH
	EPI_ISL_574513	V
Thailand (8)	EPI_ISL_403962	L
	EPI_ISL_437612	S
	EPI_ISL_437622	O
	EPI_ISL_447011	L
	EPI_ISL_447017	GH
	EPI_ISL_447026	S
	EPI_ISL_455943	GH
	EPI_ISL_512862	V
Vietnam (10)	EPI_ISL_416427	S
	EPI_ISL_416429	S
	EPI_ISL_418269	I
	EPI_ISL_435314	GH
	EPI_ISL_435316	V
	EPI_ISL_498177	G
	EPI_ISL_511893	L
	EPI_ISL_511894	G
	EPI_ISL_511897	V
	EPI_ISL_511898	GR

N=nucleocapsid; GISAID=Global Initiative on Sharing All Influenza Data. All the isolates were conserved in comparison to the reference sequence (GenBank accession number: NC_045512.2)

Based on the phylogenetic tree of the N gene nucleotide sequence (Figure 1a), all isolates from clade G, clade GH, clade L, clade O, clade S, and clade V, together with the reference sequence, did not form specific branches. All isolates originated from clade O formed one specific branch due to substitution mutations at position 28311. Moreover, isolate Malaysia | EPI_ISL_528743 | O had additional nucleotides changes at position 29086 and 29218 (Table 1), making additional branch arose from clade O. All isolates in clade GR formed a separate branch (Figure 1a) due to nucleotide substitutions at position 28881 to 28883 (Table 1). The entire isolates of clade GRY were branched out from the clade GR (Figure 1a), thus formed one branch due to nucleotide substitutions at position 28280 to 28282 (GAT to CTA) and position 28977 (C to T) (Table 1). The AA phylogenetic tree formed similar branches to the phylogenetic tree of nucleotide sequences (Figure 1b). The entire sequences in clade O formed a branch due to the P13L AA substitution (Table 1). Meanwhile, all isolates within the GR clade formed a branch due to AA substitutions in R203K and G204R (Table 1). The isolates of the GRY clade was branched out from the GR clade (Figure 1b) because of AA substitutions of D3L and S235F (Table 1).

DISCUSSION

As the COVID-19 pandemic continues, there is a fast, ongoing transmission of SARS-CoV-2 across the world.¹⁶ This study analyzed 98 sequences of the N protein gene of SARS-CoV-2 from ASEAN countries. The association of 10 countries in Southeast Asia (ASEAN countries) has enabled the ease of mobility within the internal regions. Thus, it promotes the virus to spread across the countries. Since human migration is a significant factor in viral evolution, investigating the possible mutations of the circulating SARS-CoV-2 in ASEAN countries is crucial.¹³ The phenomena of human migration naturally drive the virus to adapt to various host immune system and geographic condition with several mechanisms, including mutations, deletions, and/or recombinations.¹⁶

Overall, the nucleotides and AAs sequences of SARS-CoV-2 N protein (1,259 nucleotides and 419 AAs) across ASEAN countries are highly conserved (97.5% of conserved nucleotides and 94.5% of conserved AAs). These findings support the previous report on

the N protein gene as a more conserved and stable gene.¹⁰ Meanwhile, a study of 61,485 N protein gene sequences across the world by Rahman et al¹⁷ found that the mutations of the N protein gene sequences occurred more frequently, with 75.66% of AA positions underwent evolutionary changes in the SARS-CoV-2 nucleocapsid. However, as we specifically analyze the SARS-CoV-2 N protein sequences within ASEAN countries, our finding might raise the question of whether any specific region in which N protein is more conserved compared with other countries. Although the mutation frequency on N gene sequences is high, Rahman et al¹⁷ found no substitution of AAs compared with the reference in 29 countries and/or regions. These contrary findings highlight a challenge in designing SARS-CoV-2 N protein for vaccine candidates and diagnostic tools. The nucleotide variations within N sequences of concern were the substitutions at position 28881 (GGG to AAC), which were found in clade GR and GRY. Additionally, nucleotide substitution C28887T was found in some clade G and GH isolates. All of these variations were included in the target for forwarding primer of the N gene from China Centers for Disease Control and Prevention.¹⁸ A study has found mismatches in the primer binding region because of nucleotide variations, leading to a false-negative result of N gene detection by polymerase chain reaction assay.¹⁹

The phylogenetic tree of the SARS-CoV-2 N protein across ASEAN countries showed a similar branching pattern between the nucleotide and AA tree. The phylogenetic tree of AA sequences showed a more even branching pattern, suggesting variations within the nucleotide sequences mostly did not impact AA substitutions. Thus, the sequences of the N protein gene analyzed in this study are conserved. There were three clades that consistently grouped as one clade: clade O (n = 10), clade GR (n = 13), and clade GRY (n = 7) (Figure 1a), whereas the remaining isolates (n = 68) consisted of clades G, GH, L, S, and V along with the reference sequence or “wild type” did not form specific branches according to the original clade. This pattern suggests the N gene sequence insufficiently represents the nucleotide polymorphisms that composed the grouping of each clade. The O clade forms its branch characterized by AA substitution P13L, while the GR clade exhibits two AA substitutions R203K and G204R. The R203K and G204R mutations were known to be implicated in destabilized and decreased general

structural flexibility of the N protein.¹⁹ Clade GRY was separated from the GR clade by other AA substitutions D3L and S235F.

Multiple sequence alignment of AA sequences found 12 non-synonymous mutations which contribute to AA's polarity changes (Table 1). The R-groups of non-polar AAs have either aliphatic or aromatic groups, making them to have hydrophobic features. Hydrophobic AAs tend to repel water, making them structurally buried in the hydrophobic core of the protein and less exposed to the surface, thus less antigenic.²⁰ AA changes in the antibody-antigen interactions play an important role in the maturation of antibody affinity responses and antigenic variations. AA substitutions in the antibody-antigen interaction interface are important in a biological context. In particular, antigen changes can affect the entire antigen interaction with the antibody, providing an effective means of antigenic variation.²¹ Moreover, antigenic variation within N protein could affect the performance of the antigen-based rapid diagnostic test since this protein is often used as the target analyte.²² Thus, any site with changes in the AA polarity warrants further attention to study its impact on the antigenicity of the corresponding site.

This study has limitations since there were only 98 sequences from ASEAN countries. A more robust study with larger samples that cover many regions would better describe the N gene variations. Furthermore, a study describing the structural prediction of the N protein is suggested to confirm the impact of AA variations on antigenicity.

In conclusion, the N protein sequence of SARS-CoV-2 across ASEAN countries showed high similarity and indicated the N protein gene conservation within ASEAN countries. Based on the nucleotide and AA phylogenetic tree, most isolates were closely related to the reference sequence (China).

Conflict of Interest

The authors affirm no conflict of interest in this study.

Acknowledgment

We gratefully acknowledge the authors, originating and submitting laboratories who contributed to SARS-CoV-2 sequence data in the GISAID's EpiCoV™ database. All individuals who submitted isolates analyzed in this study may be searched using GISAID's accession ID via the GISAID website, www.gisaid.org under the EpiCoV™ platform. We also thank Professor Herman Susanto for supporting this publication.

Funding Sources

None.

REFERENCES

1. Wang Z, Qiang W, Ke H. A handbook of 2019-nCoV pneumonia control and prevention. Hubei Sci Technol Press. 2020;1–108.
2. Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, et al. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *Int J Epidemiol.* 2020;49(3):717–26.
3. Suprobowati OD, Kurniati I. *Virology Medical Laboratory Technology (TLM) Teaching Materials* [Internet]. Jakarta: Ministry of Health of the Republic of Indonesia; 2018. [cited 2020 Jun 24]. Available from: http://perpus.poltekeskupang.ac.id/index.php?p=show_detail&id=3150&keywords=TLM.Indonesian.
4. Harapan H, Itoh N, Yufika A, Winardi W, Keam S, Te H, et al. Coronavirus disease 2019 (COVID-19): a literature review. *J Infect Public Health.* 2020;13(5):667–73.
5. Helmy YA, Fawzy M, Elswad A, Sobieh A, Kenney SP, Shehata AA. The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. *J Clin Med.* 2020;9(4):1225.
6. Kopecky-Bromberg SA, Martínez-Sobrido L, Frieman M, Baric RA, Palese P. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. *J Virol.* 2007;81(2):548–57.
7. Surjit M, Liu B, Kumar P, Chow VT, Lal SK. The nucleocapsid protein of the SARS coronavirus is capable of self-association through a C-terminal 209 amino acid interaction domain. *Biochem Biophys Res Commun.* 2004;317(4):1030–6.
8. Shahhosseini N, Wong G, Kobinger GP, Chinikar S. SARS-CoV-2 spillover transmission due to recombination event. *Gene Rep.* 2021;23:101045.
9. Zeng W, Liu G, Ma H, Zhao D, Yang Y, Liu M, et al. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochem Biophys Res Commun.* 2020;527(3):618–23.
10. Dutta NK, Mazumdar K, Gordy JT. The nucleocapsid protein of SARS-CoV-2: a target for vaccine development. *J Virol.* 2020;94(13):e00647–20.
11. Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol.* 2020;92(4):418–23.
12. Kang S, Yang M, Hong Z, Zhang L, Huang Z, Chen X, et al. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm Sin B.* 2020;10(7):1228–38.
13. Yong SK, Su PC, Yang YS. Molecular targets for the testing of COVID-19. *Biotechnol J.* 2020;15(6):e2000152.
14. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35(6):1547–9.
15. Azad S. Amino acids: its types and uses. *Int J Clin Diagnostic Pathol.* 2018;1(1):13–6.
16. Islam MR, Hoque MN, Rahman MS, Alam ASMRU, Akther M, Puspo JA, et al. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Sci Rep.* 2020;10:1404.
17. Rahman MS, Islam MR, Alam ASMRU, Islam I, Hoque MN, Akter S, et al. Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. *J Med Virol.* 2021;93(4):2177–95.
18. Arena F, Pollini S, Rossolini GM, Margaglione M. Summary of the available molecular methods for detection of SARS-CoV-2 during the ongoing pandemic. *Int J Mol Sci.* 2021;22(3):1298.
19. Khan KA, Cheung P. Presence of mismatches between diagnostic PCR assays and coronavirus SARS-CoV-2 genome. *R Soc Open Sci.* 2020;7(6):200636.
20. Guan Q, Sadykov M, Mfarrej S, Hala S, Naeem R, Nugmanova R, et al. A genetic barcode of SARS-CoV-2 for monitoring global distribution of different clades during the COVID-19 pandemic. *Int J Infect Dis.* 2020;100:216–23.
21. Colman PM. Effects of amino acid sequence changes on antibody-antigen interactions. *Res Immunol.* 1994;145(1):33–6.
22. World Health Organization. Antigen-detection in the diagnosis of SARS-CoV-2 infection using rapid immunoassays: interim guidance, 11 September 2020. World Health Organization; 2020. Available from: <https://apps.who.int/iris/handle/10665/334253>.