

# Sequence Conservation Analysis and Gene Relationships of Nucleocapsid (N) Gene in Orthocoronavirinae Subfamily

Husna Nugrahapraja<sup>1\*,2</sup>, Adi Nugraha<sup>1</sup>, and Alidza Fauzi<sup>1</sup>

<sup>1</sup>School of Life Sciences and Technology, Institut Teknologi Bandung, Bandung, 40132, West Java, Indonesia

<sup>2</sup>University Center of Excellence for Nutraceuticals, Bioscience and Biotechnology Research Center, Institut Teknologi Bandung, Bandung, 40132, West Java, Indonesia

**Abstract.** Coronavirus (CoV) is a virus that causes respiratory and gastrointestinal diseases in animals and humans. It belongs to the Orthocoronavirinae. The nucleocapsid protein (N) plays multiple roles in virus assembly, RNA transcription, and interaction with host cells. This study aimed to analyse the N protein by identifying conserved residues and exploring the gene and protein relationships within the Orthocoronavirinae. Therefore the results of this study are expected to help identify conserved regions of N protein in SARS-CoV-2 which can be used as probes for the virus identification process and can be used as target areas in vaccine development. We used 159 N gene and protein sequences, including 64 from Alpha, 51 from Beta-, 11 from Delta-, and 20 from Gammacoronavirus genera of the Orthocoronavirinae. Three sequences from Tobaniviridae were used as outgroups. Multiple sequence alignment (MSA) and phylogenetic tree analysis were performed using the neighbour-joining and Maximum Likelihood. The MSA results revealed several conserved residues, ranging from 18 to 41, were located in the N-terminal and C-terminal domains, the linker region, Nuclear Localization Signal (NLS), Nuclear Export Signal (NES) motifs, and Packing Signal (PS) binding sites. The phylogenetic tree analysis indicated that Gammacoronavirus and Deltacoronavirus were closely related to Betacoronavirus, while Alfacoronavirus showed the most distant relationship. Furthermore, the study identified 23 conserved residues involved in RNA binding, including amino acids such as Ser89, Val111, Pro112, Gly124, Tyr125, Phe150, Tyr151, Gly154, Thr155, Gly156, Trp180, Val181, Gly409, Arg411, Asn419, Gly421, and Pro443. These residues interacted with phosphate groups, nitrogenous bases, and pentose sugars and exhibited non-specific interactions with RNA. In summary, this study investigated the N protein in the Orthocoronavirinae subfamily, providing insights into its function, structure, and evolutionary relationships.

**Keywords:** Coronavirus, Motives, Phylogenetics, Mutations, Conserved Residues

## 1 Introduction

Coronaviruses are a group of viruses with RNA and membrane genetic material. Coronaviruses have a wide distribution, and are distributed among mammals and birds. Coronavirus is a member of the Orthocoronavirinae subfamily in the Coronaviridae family. The characteristics of the coronavirus are its large positive sense single-stranded genome which is around 26.4-31.7 kilobases, polyadenylated and has a stamp at the 5' end, has a viral membrane, and a spike protein shaped like a beater [1].

Coronavirus infection in animals was discovered since the early 20th century. One of the earliest known coronaviruses was the transmissible gastroenteritis virus (TGEV) which infected pigs in the early 20th century. TGEV is causing epidemics and pig deaths in the United States, especially because the mortality rate is up to 100%. This causes economic losses to the pig industry [2, 3]. After being discovered in pigs, in 1930 a coronavirus was discovered in birds, namely the infectious bronchitis virus

(IBV) which causes infectious bronchitis in birds [4]. Until now, IBV is spread throughout the world and affects the ability of poultry to produce meat and eggs [5]. This causes economic losses, especially in countries that have intensive poultry industries [6].

TGEV and IBV are two examples of the many coronaviruses that cause disease in mammals and birds. In mammals, the coronavirus also infects mice, cats, dogs, cattle, beluga sharks, porcupines, camels, horses and civets [7 – 15]. In poultry, the coronavirus infects poultry, turkeys, and some birds, namely the nightingale, thrush, bondol, white-eye, finches, magpie robin, night heron, wigeon, and common morhen [16, 17]. Several coronaviruses in mammals and poultry have caused massive economic losses, threatening zoonoses, especially in several viruses that are spread throughout the world (IBV, MHV, FIPV, CCoV, BCoV, FCoV, MERS-CoV, PHEV and PDCoV), which have caused epidemics (TGEV, PHEV, PEDV, PDCoV, SeACoV, TCoV, and CCoV), and those with hosts with close human interactions (MHV, LRNV, FIPV, FCoV, CCoV, CRCoV, MERS-CoV, FRSCV, FRECV, and ECoV). [2,

\* Corresponding author: [nugrahapraja@sith.itb.ac.id](mailto:nugrahapraja@sith.itb.ac.id)

5, 9, 10, 13, 15, 18-27]. MERS-CoV is a zoonotic product of camels which causes epidemics with a case fatality rate of 28-35%. In addition, studies by Lednicky and Vlasova revealed successive infections of PDCoV and CRCoV in humans [28, 29] although, it is necessary to re-examine whether these viruses cause disease in infected patients, and can be transmitted to other humans or not. Therefore, coronaviruses in animals can have a negative impact on the economy, health, and even human survival.

The earliest coronaviruses found to cause infection in humans were HCoV-229E and HCoV-OC43 in 1966 and 1967 respectively. Both of them together with HCoV-NL63 which was discovered in 2004 caused mild respiratory infections. HCoV-229E, HCoV-OC43, and HCoV NL63 had case fatality rates of 25%, 9.1%, and 12.5%, respectively [30, 31]. HCoV-HKU1 was found in a patient in Hong Kong in 2004 with a low case fatality rate, usually death occurs because the patient has other serious illnesses or his immune system is very weak or not functioning properly [32].

This coronavirus is associated with several respiratory diseases, from the mildest is common fever to the most severe are pneumonia and bronchitis. SARS-CoV and MERS-CoV caused epidemics, while SARS-CoV-2 caused pandemics. All three cause acute respiratory system syndrome (SARS) and respectively have case fatality rates of 11%, 28-35%, and 7.3% [33-35]. Of the seven coronaviruses that cause disease in humans, HCoV-229E and HCoV-NL63 belong to the Alphacoronavirus genus, while the others belong to the Betacoronavirus genus. In addition, HCoV-OC43 and HCoV-HKU1 have natural hosts in rodents, while the other five coronaviruses have natural hosts in bats. In humans, the SARS-CoV-2 pandemic with a small case fatality rate can affect economic stability [33].

Allegations regarding the re-infection of the coronavirus in the future both to humans and other animals became clearer when there were cases handled by Lednicky and Vlasova regarding PDCoV and CRCoV infections to humans [28-29] and epidemics SeACoV in pigs in 2017 [25].

Interspecies transmission or cross-species transmission is the transmission of an infectious pathogen from one species to another. When a pathogen has been contracted from another species, the pathogen can cause disease in that species. Even these pathogens can be transmitted to their own species and cause epidemics or pandemics. Two transitional stages are necessary for the emergence of interspecies transmission, namely human contact with infectious agents, and interspecies transmission of these agents. In addition, there are two transitional stages that are important for a pathogen to cause a pandemic but do not occur in many pathogens that have occurred zoonoses, namely human-to-human transmission that supports, and genetic adaptation to the host [36].

Interspecies transmission of coronaviruses is supported by their long-term presence in nature, their rapid mutagenesis, high diversity, the evolution of coronaviruses within the host [37], as well as human interactions with several coronavirus hosts [38]. Therefore, as a long-term solution to coronavirus infection, the characterization of the genetic and

biological components of the coronavirus becomes very important.

The coronavirus genome is single-stranded RNA, positive sense, and is about 26.4-31.7 kilobases (International Committee on Taxonomy of Viruses, 2012). The coronavirus genome encodes four structural proteins, namely spike protein (S), envelope protein (E), matrix protein (M), and nucleocapsid protein (N). In addition, the coronavirus genome also encodes 16 nonstructural proteins (nsp1-16) that form the replicase-transcriptase complex (RTC) and accessory proteins [39].

The nucleocapsid (N) protein is a structural protein measuring 43–46 kDa. The N protein plays a role in the packaging of the RNA genome because it forms a ribonucleoprotein, the efficiency of transcription and processing of the viral RNA genome through its interaction with nsp3, the assembly of viruses through its interaction with the M protein, and influences host cells and host cell cellular mechanisms by blocking the G1/S phase transition [33, 40, 41]. Important characteristics of N protein that can be used as vaccine candidates, a good inhibitor target is highly immunogenic, expressed in large quantities during infection, and can induce protective immunity against SARS-CoV and SARS-CoV-2 [42].

Comprehensive characterization of the protein N coronavirus in terms of sequence, phylogenetics, and implications for its structure and function can provide insight into potential treatment targets, epitopes for vaccines, inhibitor targets, etc. that can be used long-term for coronavirus infection in humans and animals. In the future. Therefore, this study aims to identify which residues are the most conserved from the N protein sequences in the Orthocoronavirinae subfamily, analyze the effect of these residues on the function and structure of the domain, motif or sequence region in the N protein, and analyze the kinship. Subfamily Orthocoronavirinae using N protein and gene sequences.

## 2 Materials and Methods

### 2.1 Classification and Sequence Alignment of Current Orthocoronavirinae

Data retrieval from the National Center for Biotechnology Information (NCBI) website is based on only the N gene sequence manually. A total of 156 cDNA sequences of the N gene were taken from members of the four genera in the Orthocoronavirinae subfamily. Sampling was carried out for each complete sequence, annotated on the N gene, and differed in terms of host and country of origin of the sample. Based on data collection of samples taken, 64 sequences out of 156 sequences came from the genus Alphacoronavirus, 61 sequences came from Betacoronavirus, 11 sequences came from the genus Deltacoronavirus, and 20 sequences came from the genus Gammacoronavirus.

Next translation of the N gene sequence was carried out using EMBOSS Transeq on the EMBL-EBI website. After the translation is done, check again whether the reading frame is correct with Jalview 2.11.0.

Furthermore addition of outgroups is carried out only for compilation of sequences that will be used for the construction of phylogenetic trees, while for compilation of sequences used for the identification of sustainable amino acid residues it is not added. As many as 3 outgroup individuals, namely bovine torovirus from the Tobaniviridae family. The Tobaniviridae family is in the same order as the Orthocoronavirinae subfamily, namely Nidovirales.

Additionally multiple Sequence Alignment of gene and protein N sequences was carried out both in sequence compilation to identify sustainable amino acid residues and in constructing phylogenetic trees using ClustalOmega on the EMBL-EBI website.

## 2.2 Construction and Visualization of Phylogenetic Tree

The construction of the phylogenetic tree was carried out by compiling cDNA and protein sequences with the addition of outgroups. The construction of phylogenetic trees was carried out using MrBayes and MEGA X. The method used in MrBayes is Bayesian, while the methods used in MEGA X are Neighbor Joining (NJ) and Maximum Likelihood (ML). The writer choose MEGA X because MEGA X is a versatile tool that covers a broader spectrum of molecular biology tasks, including sequence alignment and basic phylogenetic analysis. Moreover it is known for its user-friendly interface. On the other hand, Mr. Bayes is a specialized software specifically designed for Bayesian phylogenetic inference, offering a high level of flexibility and accuracy [43].

The parameters used in MEGA X for building NJ trees are gamma rates, including transitions and transversions, and 1000 times bootstrap with default configuration. The difference between the model for cDNA and protein sequences is that the model for cDNA sequences uses Maximum Composite Likelihood (MCL), while the model for protein sequences is the Jones-Taylor-Thornton (JTT) model. In building ML trees, the parameters used are gamma distributed rates with invariant sites (G+I), use of all sites, the ML heuristic method is Nearest Neighbor-Interchange (NNI), and bootstrap 1000 times. The ML tree of cDNA sequences was constructed using the GTR model and estimation of the pairwise distance matrix using MCL, while for protein sequences using the JTT model and estimation [43].

Label color modification based on genus, appearance of bootstraps, and leaf sorting on phylogenetic trees were used using iTOL

## 2.3 Identification of Conserved Amino Acid Residues

Identification of conserved amino acid residues was carried out using Jalview 2.11.0 on the MSA-produced protein sequences. The sequence identity thresholds used were 80%, 90%, 95%, 97.5%, and 100%. After that, using WebLogo 2.8.2, a visualization of the sustainability of each amino acid residue in the protein sequence from MSA was made.

## 2.4 Analysis and Comparison between Phylogenetic Tress with Reference

Lastly from Figure 1 analysis of the conserved residue data was carried out by identifying the domains, motifs, and regions present in the N protein sequence that have conserved residues. After that, estimation of the location of the domains, motifs, and regions contained in the N protein sequence was carried out using the results of MSA and the characterization of the N protein in several coronaviruses. Some of the coronaviruses whose characterization results are used as references are MERS-CoV, HCoV-229E, IBV, MHV, NL63, OC43, PEDV, and HKU1. Then, a literature study was carried out regarding the effect of these conserved amino acid residues on the structure and/or function of a domain, motif, and region contained in the N protein sequence. Comparison of phylogenetic trees was carried out between phylogenetic trees that had been made and with reference trees that had been made by Tabibzadeh [43].

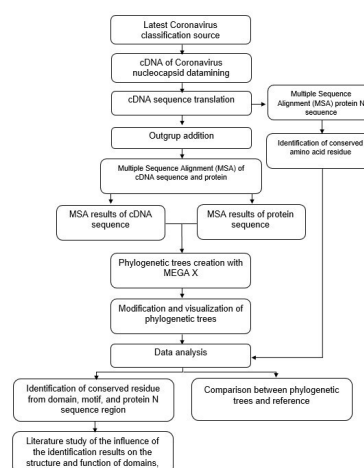


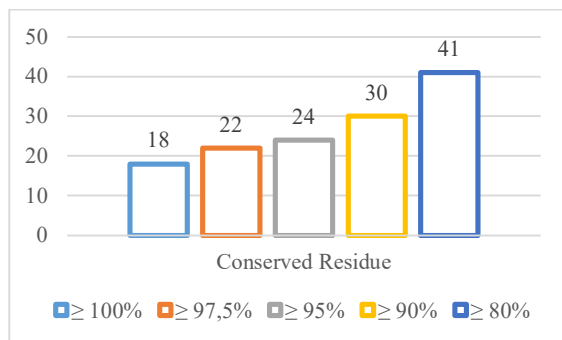
Fig. 1. Research Workflow

## 3 Results and Discussion

### 3.1 Predicted Conserved Amino Acid Residues

Preservation analysis is one of the most widely used methods in the prediction of functionally important residues in protein sequences [44]. From Figure 2 of the N protein sequence length of around 400 residues in the analyzed coronavirus, residues that exceed the 80% sequence identity threshold have 41 residues or around 10%. In addition, 18 residues with 100% sequence identity were found in the sample sequences tested. By knowing that the N protein has a low sequence identity but has the same modular organizational structure, this can prove that there are conserved residues used by N protein to maintain its function.

When compared with the MSA performed by Laude & Masters, it can be seen that the residues that were sustainable in Laude & Masters [45] but not preserved in the results of the MSA performed were two arginine residues, leucine, histidine, two alanine residues, aspartic acid, and valine.



**Fig. 2.** Conserved residues and sustainability thresholds based on MSA results

### 3.2 Implication of Conserved Residues on Function and Structure

As we can see in Figure 3. The visualization of the MSA in the N-terminal domain, serine residues in the 89th site and arginine in the 200th site in the visualization of the MSA results are Ser64 and Arg164 residues in HCoV-OC43 which play a role in direct interaction with the 2'-hydroxyl group in pentose sugar RNA. The Tyr residue at the 151st site is also the Tyr126 residue in HCoV-OC43 which plays a role in its interaction with nitrogenous bases [46].

In addition, this residue along with the arginine residue at the 133th site is a Tyr94 and Arg76 residue in IBV which play a role in binding RNA. The aromatic properties of the tyrosine residues and the alkaline properties of the arginine residues play a role in binding RNA by creating a large surface that is in contact with the viral gRNA [47].

Although not confirmed, we can see that in the Table 1. Tyr residues at sites 125, 151, 152, and tryptophan at site 180 which correspond to residues Tyr87, Tyr112, Tyr113, and Trp133 in SARS-CoV are on the same  $\beta$ -sheet surface and play a role in packaging RNA and is a residue that plays a role in forming hydrophobic pockets in HCoV-OC43 which can orient nitrogenous bases on the protein surface, rather than selecting a protein-RNA sequence [46, 48]. The  $\beta$ -hairpin structures are similar in structure but vary in electrostatic surface and topology, which may indicate a specific adaptive function. As we can see from the Table 2. This structure motif plays a functionally important role in the N-NTD for binding RNA [46] and neutralizing the phosphate group [47], in this structure there is an arginine residue at the 133th site and glycine at the 139th site. The  $\beta$ -sheet core consists of the secondary structure  $\beta 1\beta 2\beta 5\beta 6\beta 7$ . In this structure there are conserved residues that occupy  $\beta 5$  and  $\beta 6$ . The role of the  $\beta$ -sheet core structure is to "hold" the RNA by neutralizing the phosphate groups of the RNA, and the aromatic amino acid residues in this section interact with the base portion of the RNA [47, 49] In the border area between NTD and LKR, namely at sites 213 to 245 on MSA results, Schuster (2020) said that this area is the result of recombination that also occurs in several

coronaviruses, namely Pangolin-CoV MP789, Bat-CoV RaTG13, and bat-SL-CoVZXC21.

**Table 1.** Conserved residues in the domain in the N protein based on their location on the MSA result site

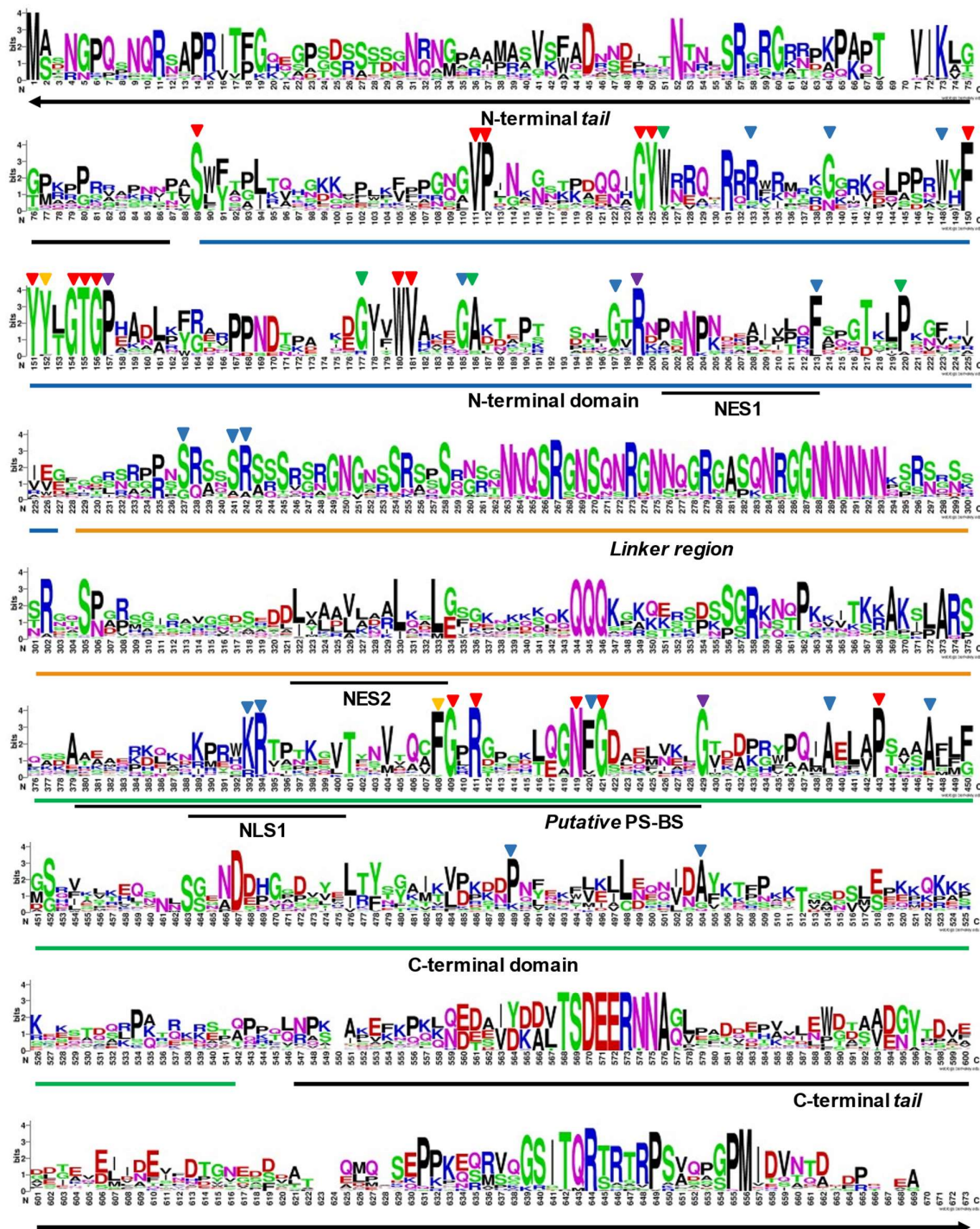
| Domain/Motif/Region     | Residue                                                                                                                                                                                                       |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| NTD (N-terminal domain) | SER89, VAL111, PRO112, GLY124, TYR125, TRP126, ARG133, GLY139, TRP148, PHE150, TYR151, TYR152, GLY154, THR155, GLY156, PRO157, GLY177, TRP180, VAL181, GLY185, ALA186, GLY197, ARG199, PHE213, PRO220, SER237 |
| LKR (linker region)     | SER241, ARG242, LEU330                                                                                                                                                                                        |
| CTD (C-terminal domain) | LYS393, ARG394, PHE408, GLN409, ARG411, ASN419, PHE420, GLY421, GLY429, ALA439, PRO443, ALA447, PRO489, ALA504                                                                                                |

**Table 2.** Conserved residues in the motifs and regions in protein N based on their location on the MSA result site

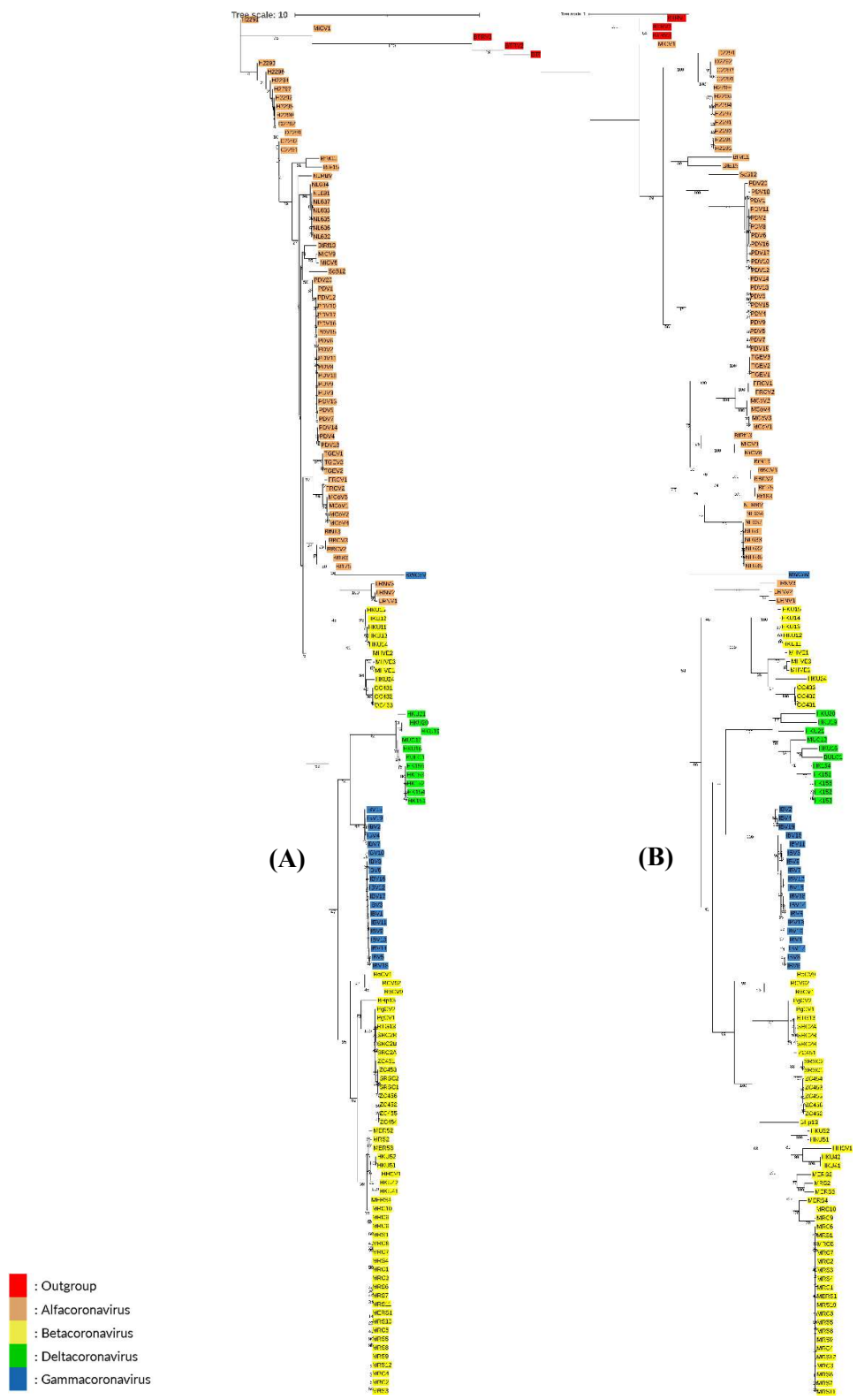
| Motif/Region                                 | Residue                                |
|----------------------------------------------|----------------------------------------|
| NLS1 (Nuclear Localization Signal 1)         | LYS393, ARG394                         |
| NES1 (Nuclear Export Signal 1)               | PHE213                                 |
| NES2                                         | LEU330                                 |
| Putative PS-BS (Packing Signal-Binding Site) | LYS393, ARG394, PHE408, GLN409, ARG411 |

### 3.3 Gene and Protein N Phylogenetic Tree Analysis

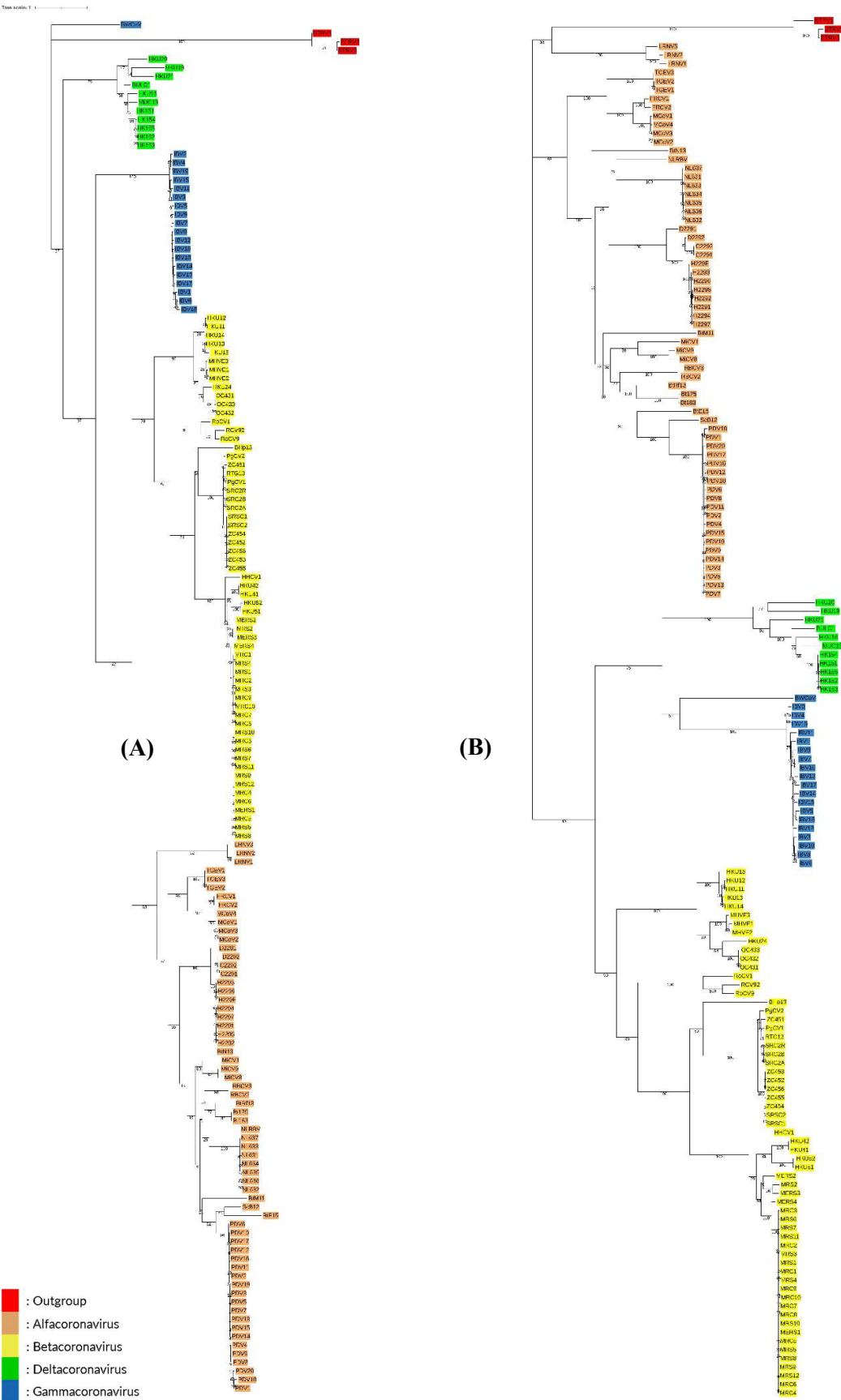
As we can see in the Figure 4 the phylogenetic tree obtained from the cDNA data, that the two trees have high similarity. This can happen because the resulting NJ tree can produce the correct topology [50] and the use of the Maximum Composite Likelihood (MCL) model on the Neighbor Joining tree is robust in consistency with the Maximum Likelihood Estimation (MLE), also in terms of efficiency and overall the computation. When compared with the Tabibzadeh tree [43], which uses cDNA data using the Neighbor Joining method, you can see the similarities to the two trees from the cDNA data generated. The similarity is that the Betacoronavirus kinship is close to Gammacoronavirus, and Alphacoronavirus is the genus that is most distantly related. Meanwhile, the difference is in the tree topology which can be caused by the number of different samples, outgroups and Deltacoronavirus samples which are not present in the Tabibzadeh tree [43] but are present in the created tree



**Fig. 3.** Visualization of conserved residues in MSA results with WebLogo. Conserved residues are indicated by the triangle above the residue illustration. The color of the triangle above the residue indicates that the residue has exceeded the residue identity threshold a) blue: 80%, b) green: 90%, c) orange: 95%, d) purple: 97.5%, and e) red: 100% Motifs and domains that have conserved residues are mentioned in the visualization. The site numbering on the MSA results is below the residuals.



**Fig. 4.** The phylogenetic tree obtained from (A) the Neighbor Joining method and (B) the Maximum Likelihood method from cDNA data.



**Fig. 5.** The phylogenetic tree obtained from (A) the Neighbor Joining method and (B) the Maximum Likelihood method from protein data.

In the phylogenetic tree obtained from the protein data, it can be seen from Figure 5 that the two trees tend to be dissimilar. The similarities between the two trees are only found in their kinship structure, where Gammacoronavirus is closely related to Deltacoronavirus, then both are close to Betacoronavirus, and the most distantly related is Alfahcoronavirus. However, the difference with other trees is that the genera closest to the outgroup, the cDNA tree and the ML tree from the protein data show that the closest to the outgroup is Alfahcoronavirus, not Deltacoronavirus as in the NJ tree obtained from protein data. Errors in the Neighbor Joining method are caused significantly by zero-length branches in the tree [51]. In the generated NJ tree, a large number of zero-length branches can be seen, especially in sequences with the same virus species.

The ML tree obtained from the protein data is similar to both the Tabibzadeh tree [43] and the tree constructed with cDNA data in terms of the genus arrangement in the tree. This is because the resulting tree from the ML method has the advantage of having low variance compared to other methods, being robust against violations of assumptions in the evolutionary model, being able to outperform the performance of the parsimony or distance methods even though the sequences used are very short, evaluating different tree topologies, using all the information on the sequence. Compared to the distance method, and is better for calculating branch length [52].

## 4 Conclusions

It is concluded that the most conserved residues in the nucleocapsid protein of the coronavirus are Ser89, Val111, Pro112, Gly124, Tyr125, Phe150, Tyr151, Gly154, Thr155, Gly156, Trp180, Val181, Gly409, Arg411, Asn419, Gly421, and Pro443. Second, 2. Ser89 and Arg200 play a role in their interaction with the 2'-hydroxyl group pentose sugar. Tyr151 plays a role in its interaction with nitrogenous bases. Tyr125, Tyr151, and Trp180 play a role in orienting bases on the protein surface and RNA packaging. Phe150, Tyr151, and Trp180 and Val181 can "grasp" RNA by neutralizing the phosphate group of RNA. Arg411 plays a role in binding negatively charged oligonucleotides based on non-specific interactions. Gly409, Arg411, Asn419, Gly421 play a role in binding packing signal (PS). Pro443 plays a role in hydrophobic interactions. Finally, the genera Deltacoronavirus and Gammacoronavirus are the most closely related, followed by the genus Betacoronavirus, while the genus Alfahcoronavirus is the most distant. The results of this study are expected to help identify conserved regions of N protein in SARS-CoV-2 which can be used as probes for the virus identification process and can be used as target areas in vaccine development.

## Acknowledgements

The authors thank the School of Life Sciences and Technology for partially funding this research under the Biology Study Program.

## Conflicts of interest statement

The authors declared that they have no conflicts of interest in this study.

## References

1. International Committee on Taxonomy of Viruses, 2012. *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, Oxford: Elsevier.
2. Spickler, A. R., Roth, J. A., Galyon, J. & Brown, G., 2016. *Transboundary and Emerging Diseases of Animals*. 1st ed. Ames: Iowa State University.
3. Xia Xia, L., Yang, Y., Wang, J., Jing, Y., & Yang, Q. (2018). Impact of TGEV infection on the pig small intestine. *Viol. J.*, 15(1), 1-7.
4. Jackwood, M. W., 2012. Review of Infectious Bronchitis Virus Around the World. *Avian Disease*, Volume 56, 634-641.
5. The Pirbright Institute, n.d. *Infectious bronchitis virus*. [Online] Available at: <https://www.pirbright.ac.uk/viruses/infectious-bronchitis-virus> [Accessed 28 August 2021].
6. Cavanagh, D., 2007. Coronavirus avian infectious bronchitis virus. *Veterinary Research*, **38**, 281-297.
7. Wang, W., Lin, X. D., Guo, W. P., Zhou, R. H., Wang, M. R., Wang, C. Q., ... & Zhang, Y. Z. 2015. Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. *Virology*, **474**, 19-27.
8. Bailey, O. T., Pappenheimer, A. M., Cheever, F. S., & Daniels, J. B. 1949. A murine virus (JHM) causing disseminated encephalomyelitis with extensive destruction of myelin: II. Pathology. *J.Exp.Med.*, 90(3), 195-212.
9. Pratelli, A., 2006. Genetic evolution of canine coronavirus and recent advances in prophylaxis. *Veterinary Research*, **37**(2), 191-200.
10. Centre for Agriculture and Bioscience International, 2019. *bovine coronavirus infection*. [Online] Available at: <https://www.cabi.org/isc/datasheet/9170> [Accessed 28 August 2021].
11. Mihindukulasuriya, K. A., Wu, G., St. Leger, J., Nordhausen, R. W., & Wang, D. 2008. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J.Virol.*, **82**(10), 5084-5088.
12. Corman, V. M., Kallies, R., Philipps, H., Göpner, G., Müller, M. A., Eckerle, I., & Drexler, J. F., 2014. Characterization of a Novel Betacoronavirus Related to Middle East Respiratory Syndrome Coronavirus in European Hedgehogs. *Viol.J.*, **88**(1), 717-724.
13. Müller, M. A., Corman, V. M., Jores, J., Meyer, B., Younan, M., Liljander, A., & Drosten, C., 2014. MERS coronavirus neutralizing antibodies in camels, Eastern Africa, 1983–1997. *Emerg. Infect. Dis.*, **20**(12), 2093–2095.
14. Pusterla, N., Vin, R., Leutenegger, C. M., Mittel, L. D., & Divers, T. J., 2018. Enteric coronavirus infection in adult horses. *Vet. J.*, **231**, 13-18.

15. Murray, J., Kiupel, M. & Maes, R. K., 2010. Ferret Coronavirus-Associated Diseases. *Vet Clin North Am Exot Anim Pract.*, **13**(3), 543-560.
16. Woo, P. C., Lau, S. K., Lam, C. S., Lai, K. K., Huang, Y. I., Lee, P., & Yuen, K. Y., 2009. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. *Virology*, **83**(2), 908-917.
17. Woo, P. C., Lau, S. K., Lam, C. S., Lau, C. C., Tsang, A. K., Lau, J. H., & Yuen, K. Y., 2012. Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavi. *Virology*, **86**(7), 3995-4008.
18. Huang, D. S., Emancipator, S. N., Fletcher, D. R., Lamm, M. E., & Mazanec, M. B., 1996. Hepatic pathology resulting from mouse hepatitis virus S infection in severe combined immunodeficiency mice. *Lab. Anim. Sci.*, **46**(2), 167-173.
19. Holzworth, J., 1963. Some important disorders of cats. *The Cornell Veterinarian*, Volume 53, 157-160.
20. Taharaguchi, S., Soma, T. & Hara, M., 2012. Prevalence of Feline Coronavirus Antibodies in Japanese Domestic Cats during the Past Decade. *J. Vet. Med. Sci.*, **74**(10), 1355-1358.
21. Mora-Díaz, J. C., Piñeyro, P. E., Houston, E., Zimmerman, J., & Giménez-Lirola, L. G., 2019. Porcine Hemagglutinating Encephalomyelitis Virus: A Review. *Front. vet. sci.*, **6**(53)
22. Boley, P. A., Alhamo, M. A., Lossie, G., Yadav, K. K., Vasquez-Lee, M., Saif, L. J., & Kenney, S. P., 2020. Porcine Deltacoronavirus Infection and Transmission in Poultry, United States. *Emerg. Infect. Dis.*, **26**(2), 255-265.
23. Snelson, H., 2014. *Comprehensive Discussion of PEDV, Slide 3, American Association of Swine Veterinarians presentation to the 2014 Pork Management Conference.* [Online] Available at: <http://www.slideshare.net/trufflemedia/dr-harry-snelson-pedv-lessons-learned> [Accessed 28 August 2021].
24. Pérez-Rivera, C., Ramírez-Mendoza, H., Mendoza-Elvira, S., Segura-Velázquez, R., & Sánchez-Betancourt, J. I., 2019. First report and phylogenetic analysis of porcine deltacoronavirus in Mexico. *Transbound. Emerg Dis*, **66**(4), 1436-1441.
25. Yang, Y.-L., Yu, J.-Q. & Huang, Y.-W., 2020. Swine enteric alphacoronavirus (swine acute diarrhea syndrome coronavirus): An update three years after its discovery. *Virus Res*, 285.
26. Peterson, E. & Hymas, T., 1951. Antibiotics in the Treatment of an Unfamiliar Turkey Disease. *Poult. Sci.*, **30**(3), 466-468.
27. Priestnall, S. L., Brownlie, J., Dubovi, E. J. & Erles, K., 2006. Serological prevalence of canine respiratory coronavirus. *Vet. Microbiol.*, **115**(1), 43-53.
28. Lednický, J. A., Tagliamonte, M. S., White, S. K., Elbadry, M. A., Alam, M. M., Stephenson, C. J., & Morris Jr, J. G., 2021. Emergence of porcine delta-coronavirus pathogenic infections among children in Haiti through independent zoonoses and convergent evolution. *medRxiv. Medicine*, **90**(3), 195-212.
29. Vlasova, A. N., Diaz, A., Dantie, D., Xiu, L., Toh, T. H., Lee, J. S. Y., ... & Gray, G. C., 2021. Novel Canine Coronavirus Isolated from a Hospitalized Patient With Pneumonia in East Malaysia. *Clin. Infect. Dis.*, Issue ciab456.
30. Krvavac, A., Patel, T. P., Karle, E. M., Epstein, N. B., Reznikov, E. A., Gates, L. G., & Holliday, Z. M., 2020. Increased Incidence, Morbidity, and Mortality in Human Coronavirus NL63 Associated with ACE Inhibitor Therapy and Implication in SARS-CoV-2 (COVID-19). *Missouri Medicine*, **117**(4), 346-354.
31. Choi, W. I., Kim, I. B., Park, S. J., Ha, E. H., & Lee, C. W., 2021. Comparison of the clinical characteristics and mortality of adults infected with human coronaviruses 229E and OC43. *Scientific Reports*, **11**(4499).
32. Kanwar, A., Selvaraju, S. & Esper, F., 2017. Human Coronavirus-HKU1 Infection Among Adults in Cleveland, Ohio. *Open Forum Infect. Dis.*, **4**(2).
33. Artika, I. M., Dewantari, A. K. & Wiyatno, A., 2020. Molecular biology of coronaviruses: current knowledge. *Heliyon*, **6**(8), e04743.
34. Chang, C. K., Hou, M. H., Chang, C. F., Hsiao, C. D., & Huang, T. H., 2014. The SARS coronavirus nucleocapsid protein – Forms and functions. *Antiviral Research*, **103**, 39-50.
35. Ritchie, H., 2021. *Mortality Risk of COVID-19.* [Online] Available at: <https://ourworldindata.org/mortality-risk-covid> [Accessed 5 August 2021].
36. Childs, J. E., Richt, J. A. & Mackenzie, J. S., 2007. Introduction: Conceptualizing and Partitioning the Emergence Process of Zoonotic Viruses from Wildlife to Humans. In: S. Rallison & A. Clauss, eds. *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission.* Heidelberg: Springer, 1-31.
37. Zhou, Z., Qiu, Y. & Ge, X., 2021. The taxonomy, host range and pathogenicity of coronaviruses and other viruses in the Nidovirales order. *Animal Diseases*, **1**(5).
38. Parrish, C. R., Holmes, E. C., Morens, D. M., Park, E. C., Burke, D. S., Calisher, C. H., & Daszak, P., 2008. Cross-Species Virus Transmission and the Emergence of New Epidemic Diseases. *Microbiol. Rev.*, **72**(3), 457-470.
39. Hartenian, E., Nandakumar, D., Lari, A., Ly, M., Tucker, J. M., & Glaunsinger, B. A., 2020. The molecular virology of coronaviruses. *J. Biol. Chem.*, **295**(37), 12910–12934.
40. Lu, S., Ye, Q., Singh, D., Cao, Y., Diedrich, J. K., Yates III, J. R., & Corbett, K. D., 2021. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat. Commun.*, **12**(502).
41. Gao, T., Gao, Y., Liu, X., Nie, Z., Sun, H., Lin, K., & Wang, S., 2021. Identification and functional analysis of the SARS-COV-2 nucleocapsid protein. *BMC Microbiology*, **12**(58).

42. Surjit, M. & Lal, S. K., 2008. The SARS-CoV nucleocapsid protein: a protein with multifarious activities. *Infect. Genet. Evol.*, **8**, 397-405.
43. Tabibzadeh, A., Esghaei, M., Soltani, S., Yousefi, P., Taherizadeh, M., Safarnezhad Tameshkel, F., ... & Karbalaie Niya, M. H., 2020. Evolutionary study of COVID-19, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as an emerging coronavirus: Phylogenetic analysis and literature review. *Vet. Med. Sci.*, **7**(2), 1-13.
44. Capra, J. A. & Singh, M., 2007. Predicting functionally important residues from sequence conservation. *Bioinformatics*, **23**(15), 1875-1882.
45. Laude, H. & Masters, P. S., 1995. The Coronavirus Nucleocapsid Protein. In: S. G. Siddell, ed. *The Coronaviridae*. New York: Plenum Press, 141-163.
46. Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., & Chen, S., 2020. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B.*, **10**(7), 1226-1238.
47. McBride, R., Zyl, M. v. & Fielding, B. C., 2014. The Coronavirus Nucleocapsid Is a Multifunctional Protein. *Viruses*, **6**(8), 2991-3018.
48. Chan-Yeung, M. & Xu, R.-H., 2003. SARS: epidemiology. *Respirology*, **8**(Supplementary 1), S9-S14.
49. Schuster, N., 2020. Using the nucleocapsid protein to investigate the relationship between SARS-CoV-2 and closely related bat and pangolin coronaviruses. *bioRxiv*, 1-27.
50. Saitou, N. & Nei, M., 1987. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, **4**(4), 406-425.
51. Kumar, S. & Gadagkar, S. R., 2000. Efficiency of the Neighbor-Joining Method in Reconstructing Deep and Shallow Evolutionary Relationships in Large Phylogenies. *J. Mol. Evol.*, **51**, 544-533.
52. Will, K., 2012. *Principles of Phylogenetics - University of California, Berkeley*. [Online] Available at: [http://ib.berkeley.edu/courses/ib200a/lect/ib200a\\_lect11\\_Will\\_likelihood.pdf](http://ib.berkeley.edu/courses/ib200a/lect/ib200a_lect11_Will_likelihood.pdf) [Accessed August 2021].