Pakistan J. Zool., pp 1-6, 2024. **DOI**: https://dx.doi.org/10.17582/journal.pjz/20221107051149

Phylogenetic Analysis and Structural Prediction of NSP9 and NSP11 of SARS-CoV-2 Variants Across the Globe

Rabia Nawaz[1]*, Mohi-Ud-Din Nasir[1], Muhammad Khizar[1], Haroon Riaz[1], Jalees ul Hassan[2], Ammara Ahad[1], Ghazanfar Ali[1], Zohal Hassan[1], M. Mushtaq[1], M. Shahid[3] and Uqba Mehmood[1]

[1]*Department of Biological Sciences, The Superior University, Lahore*
[2]*Department of Fisheries, University of Veterinary and Animal Sciences, Lahore*
[3]*Center of Excellence in Molecular Biology, University of the Punjab, Lahore*

Article Information
Received 07 October 2022
Revised 05 February 2024
Accepted 27 February 2024
Available online 09 August 2024 (early access)

Authors' Contribution
RN, MDN and MK participated in the experimental design, material collection and wrote the manuscript. HR and JH participated in the bioinformatics analysis. AA, GA and ZH helped in writing manuscript. All authors read and approved the final manuscript.

Key words
SARS-CoV-2, Selected protein NSP9, NSP11, Phylogenetic analysis, Structural prediction, 3d structure

ABSTRACT

In December 2019, a new disease appeared in Wuhan, China, later called Coronavirus disease 2019 (COVID-19). The causative agent was named Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2). It has four structural and sixteen nonstructural proteins. Nonstructural proteins (NSP1 to NSP16) are encoded by virus but are not a part of viral particles. In this study, the In Silico analysis of SARS-CoV-2 selective non-structural proteins, NSP9 (genomic region 4141-4253) and NSP11 (genomic region 4393-4405) for the years of 2020-2022 was performed. Phylogenetic analysis of NSP9 and NSP11 was performed by the Neighbor-Joining method through CLC-Sequence Viewer 8.0 version. The reference sequences of the coronavirus variants of concern Alpha, Kappa, Delta, Omicron, and initial refseq from China were taken to determine the similarities and differences between these sequences among different countries. We conclude that in NSP9 the rate of genetic mutation was far higher as compared to NSP11 which in turn means that NSP11 was more conserved. The structural prediction had been done in order to observe the protein sequences of Alpha, Delta, Omicron, and refseq for NSP9 and NSP11 of SARS-CoV-2 using the Swiss Model. We observed that the behavior of viruses and mutations across the globe are closely related to each other.

INTRODUCTION

The coronavirus disease appeared in Wuhan, China, and rapidly turned into a huge global challenge. Previously two other human viruses from the same family emerged in 2002 and 2012, named severe acute respiratory syndrome (SARS-CoV) and Middle East respiratory syndrome (MERS), respectively. There are four structural proteins, spike, envelope, membrane, and nucleocapsid in SARS-CoV-2. The virus recognizes a specific receptor which is found in many parts of the body such as the heart, lungs, intestine, and kidney. The receptor is known as angiotensin-converting enzyme receptor (ACE2). The membrane protein defines the shape of viral envelope. The virus and our body cells work together to make viral proteins. The nucleocapsid or N protein plays multiple functions. In particular, it provokes the defense mechanism of host cells and allows the replication of viral RNA and the development of new viral particles (Hasöksüz *et al*., 2020).

Nonstructural proteins (NSPs) are encoded by a virus but are not a part of viral particles. There are a total of 16 NSPs (NSP1-NSP16). NSP9 has a length of 113 amino acids and the range is 4141 to 4253 on the genome it is bound with helicase enzyme (Yadav *et al*., 2021) The NSP9-NSP8 complex is essential for the replication of RNA and functions as a viral virulence factor (Raj, 2021). NSPs plays a vital role in primase activity (Reshamwala, *et al*., 2021) The NSP11 is similar to the first segment of NSP12 the range is 4393 to 4405 on the genome and has a length of 13 amino acid. The NSP11 plays a part in the action of the endoribonuclease. It appears that NSP11 participates in the replication process. NSP11 is regarded as a target for the development of medications against coronavirus 2 (Yadav *et al*., 2021).

Due to their increased transmissibility, the potential for significant effects, and/or ability to evade neutralizing antibodies, new SARS-CoV-2 lineages have attracted

* Corresponding author: dr.rabia.nawaz8@gmail.com
0030-9923/2024/0001-0001 $ 9.00/0

great attention globally. The initial SARS-CoV-2 variant of concern (VOC), VOC Alpha (B.1.1.7), has been reported to have a 43–90% greater reproduction rate and a 75% more infectiousness than the preceding strains, and it has now become the predominant variety in the United Kingdom and endemic globally. The rate of B.1.1.7 transmission also increased with age and viral load. After that, the Delta (B.1.617.2) variety spread throughout India, outcompeting earlier lineages like B.1.1.7 (Alpha), and COVID-19 cases shot up quickly in various nations throughout the world. These variations are thought to be more infectious and harmful because they manage to elude neutralizing antibodies (Zhao *et al*., 2022; Shahzaib *et al*., 2023).

We analyzed the non-structural proteins 9 and 11 for phylogenetic analysis to find the evolutionary relationship amongst these proteins between different variants of the virus from neighboring countries. In this analysis, we used multiple sequence alignment to form an evolutionary tree. The protein structural prediction had been done using the Swiss Model, in order to observe any visible conformational difference among different variants. Ramachandran plot shows the statistical distribution of the combinations of the backbone dihedral angles ϕ and ψ.

MATERIALS AND METHODS

Sequence retrieval from NCBI and phylogenetic analysis

152 sequences of NSP9 (4141-4253) and NSP11(4393-4405) for SARS-COV-2 from USA, Germany, India, Japan, France, Bangladesh, Nepal, Pakistan, Turkey, and reference sequences for variants Alpha, Delta, Kappa, Omicron, First China were retrieved from NCBI (https://www.ncbi.nlm.nih.gov/) (Tables I and II). The sequence alignment was done using CLC-sequence viewer 8.0 under the following parameters: gap open cost, 10; gap extension cost, 1.0 and very accurate progressive algorithm. The phylogenetic analysis showed the differences in the form of distance of different sequences which reveals the close evolutionary relationship and shows common ancestry. Phylogenetic tree of 152 sequences of NSP9 and NSP11 of SARS-CoV-2 was constructed using the neighbor-joining method to confirm the accuracy of phylogenetic tree and the circular cladogram layout was selected in CLC-sequencer viewer 8.0.

Structure prediction

The structural prediction to observe the protein sequences of Alpha, Delta, Omicron, and Original China for NSP9 and NSP11 amino acid length for SARS-CoV-2 using the Swiss Model (Table II). The visualization and analysis of selected SARS-CoV-2 variants for protein structure homology modeling were done using Swiss model. Ramachandran plot, which plots a graph that explains the nature and structure of the amino acid sequence, is also useful in validating the 3D structure of protein.

Table I. Accession numbers of selected NSP9 (4141-4253) and NSP11 (4393-4405) of SARS-CoV-2 from USA, Germany, Japan, India, France, Bangladesh, China, Nepal, Pakistan, Australia, and Turkey.

USA	OK096410, OK096405, OK096405, OP341944, OP341943, OP341942, OP341941, OP341940, OP341939, OP34 1937 OP341938, OP341936, OP341935, OP341934, OP341933, OP341932, OP341931, OP341930, OP341929 OP341928, OP341927, OP341926, OK096397, OK096407, OK096415, OK096416, OK096398, OK096400, OK096401, OK096402, OK096403, OK096399, OK096404, OK096406, OK096408, OK096409, OK096411, OK096412, OK096414, MT558706, MT558705, MT558707, MT558702, MT558703
Germany	MT560669, MT560668, MT560670, MT560671, MT560672, MT560674, MT560673, MT560675, MT560677, MT560679, MT560680, MT560682, MT560690, MT560689, MT560694, MT560693, OU975109, OU975111, OU975113, OU975114, OU974925, OU974926, OU974927, OU974929
India	MT560669, MT560668, MT560670, MT560671, MT560672, MT560674, MT560673, MT560675, MT560677, MT560679, MT560680, MT560682, MT560690, MT560689, MT560694, MT560693
Japan	ON444132, ON444134, ON444137, ON444139, ON444141, ON444143, ON444145, ON444146, ON444153, ON444147, BS005267, BS005268, BS005269, BS005270
France	ON444132, ON444134, ON444137, ON444139, ON444141, ON444143, ON444145, ON444146, ON444153, ON444147
Bangladesh	MT566434, MT566435, MT566436, MT566437, MT566438, MT566
China	ON965799, ON965800, ON965801, ON965802, ON965803
Nepal	MZ 157009, MZ157010, MZ157011, MZ157012
Pakistan	MT879619, MT560679, MW400961, MW411960
Australia	ON819429, ON532671, ON532673, ON532672
Turkey	MT560531, MT560530, MT560525

RESULTS

Tables I-II shows accession numbers for all the sequence. CLC-sequencer 8.0 provides statistically improved correct alignment for NSP9 (4141-4253) and NSP11 (4393-4405) of SARS-COV-2. A total of 16 mutations were found in the NSP9 region from 2020, 2021, 2022 while 9 mutations were found in NSP11. The phylogenetic analysis showed the differences in the form of distance of different sequences which reveals the close evolutionary relationship and show common ancestry. Phylogenetic tree of 152 sequences of NSP9 and NSP11 of SARS-CoV-2 was constructed using neighbor joining method in CLC-sequencer viewer 8.0 (Fig. 1).

Structural prediction for NSP9 of SARS-CoV-2

The protein sequence of SARS-CoV-2, non-structural protein 9 of Alpha-variant, Delta-varient, Omicron-varient and first China-variant was extracted from NCBI in FASTA format. The structure prediction of the NSP 9 FASTA sequence was conducted using Swiss-model. Because of the absence of apparent differences in the predicted structures of various SARS-CoV-2 variants, only the structure of the alpha variant is shown in Figure 2A as a reference. The Z-Scores of all predicted structures is given in Table III.

Table II. Accession numbers of reference genomic sequences for phylogenetic tree generation and proteomic sequences for structural prediction of NSP9 and NSP11 of the SARS-CoV-2 Alpha (UK), Delta (India), Omicron (South Africa), First China, and Kappa (USA) were retrieved from NCBI.

	Reference genomic sequences	Proteomic sequences
Alpha (UK)	OW998408	UKQ11044
Delta (India)	OX014251	QUX81271
Omicron (South Africa)	OP093374	UUB67348
Original China	NC045512	YP_009724389
Kappa (USA)	MZ157006	

Ramachandran plot of SARS-CoV-2 NSP9

A protein sequence for NSP9 of SARS-CoV-2 φ (phi) and ψ (psi) torsion angles are plotted in Ramachandra plot shows in Figure 3A.

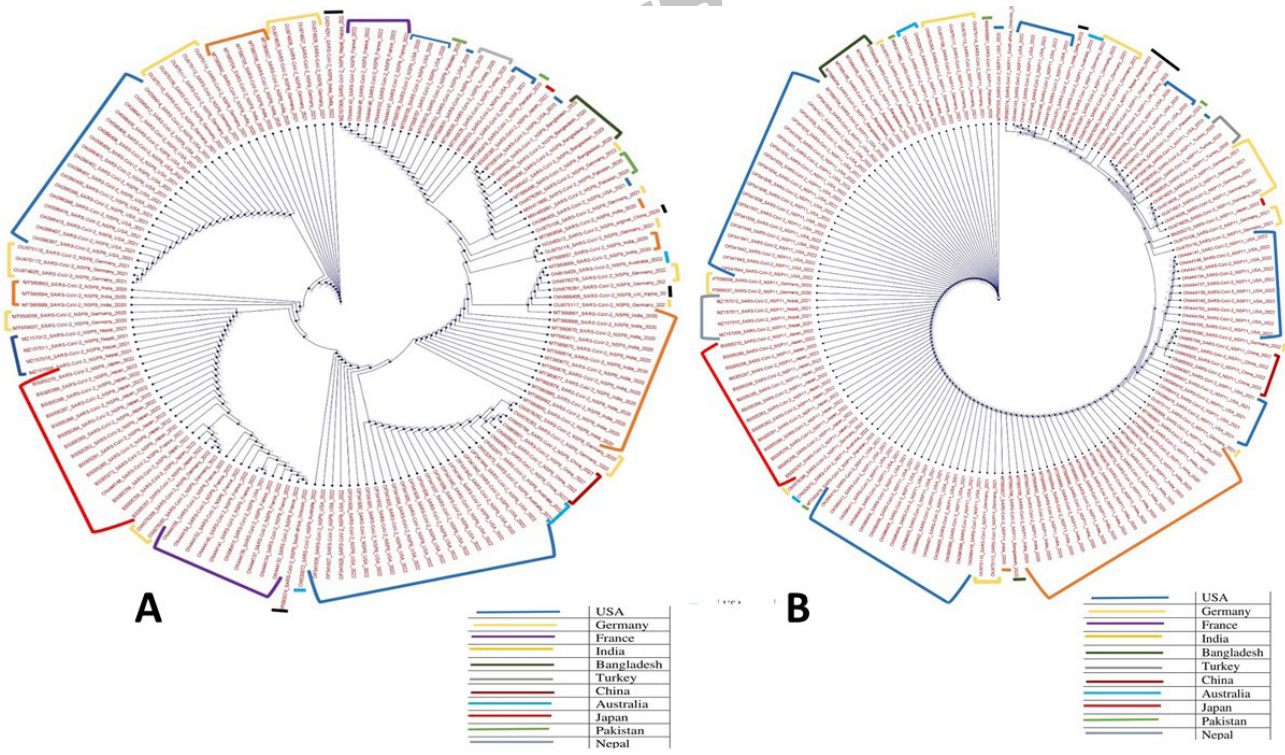

Fig. 1. The phylogenetic tree was constructed by a Neighbor-joining method (Nj), representing 152 NSPs sequences including the five reference sequences. The name of the reference sequence was colored Black and some other countries' sequences are also colored over the cladogram. Colors are mentioned in the Table color key. (A) present phylogenetic tree of NSP 9. (B) present the phylogenetic tree of NSP 11.

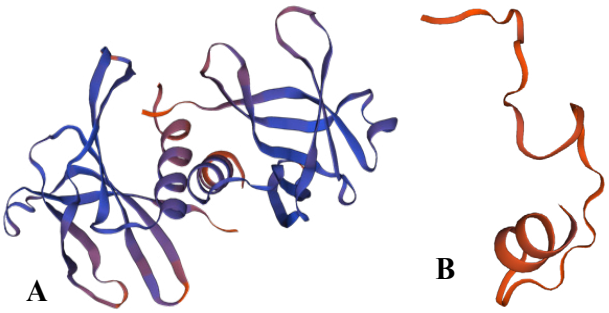

Fig. 2. 3D model of SARS-CoV-2 Alpha (UKQ110445) NSP9 (A) and SARS-CoV-2 Alpha-variant (UKQ11044) NSP11 (B) was built by using Swiss-model.

Structural production for NSP11 of SARS-CoV-2

The protein sequence of SARS-CoV-2, non-structural protein 11 of Alpha-variant, Delta-variant, Omicron-variant and first China-variant was extracted from NCBI in FASTA format. The structural prediction of the NSP11 FASTA sequence was accomplished using Swiss-model. Notably, the structural predictions for various SARS-CoV-2 variants did not reveal any significant differences. Therefore, the structure of alpha variant is shown in Figure 2 as a reference. The Z-Score of predicted SARS-CoV-2 variants is given in Table III.

Table III. QMEAN Z-scores of SARS-CoV-2 NSP9 and SARS-CoV-2 NSPII variants.

S. No.	SARS-CoV-2 variants	QMEAN Z-score				
		QMEAN	Cβ	All Atom	Solva-tion	Tor-sion
NSP9						
1	Alpha variants	-2.22	-0.91	-1.44	-1.70	-1.35
2	Delta variants	-1.68	-1.06	-1.24	-0.68	-1.63
3	Omicron variants	-1.41	1.26	-1.20	-0.45	-1.49
4	China variants	-1.68	-1.06	-1.24	0.68	-1.63
NSPII						
1	Alpha variants	-1.08	-0.63	-1.07	-3.31	0.58
2	Delta variants	-2.64	-0.88	-0.85	-3.85	-1.22
3	Omicron variants	-1.02	-0.38	-0.88	-2.52	0.15
4	China variants	-2.67	-0.39	-0.67	-3.31	-1.74

Ramachandran plot of SARS-CoV-2 NSP11

Protein sequence variant structures for NSP11 of SARS-CoV-2 φ (phi) and ψ (psi) torsion angles are plotted in Ramachandra plot shown in Figure 3B.

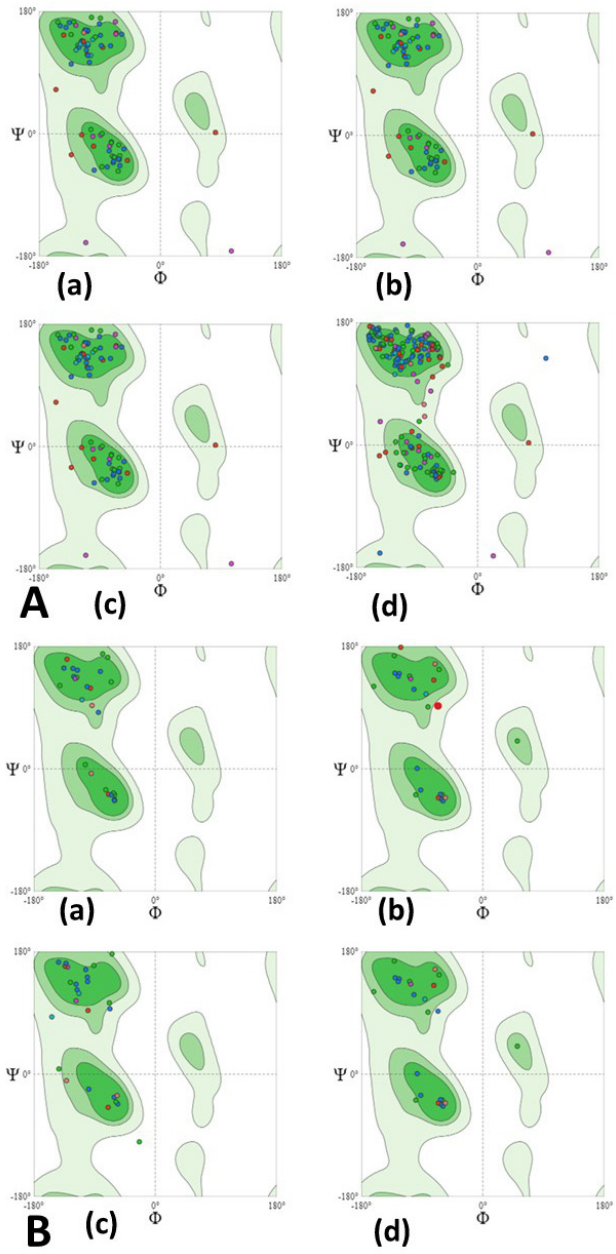

Fig. 3. Ramachandran plot for NSP9 (A) and NSPII (B) of SARS-CoV-2 showing the most favored regions (dark green) show allowed regions but rare (light green) for the torsion angles in alpha helices and beta sheets. (a) presents SARS-CoV-2 Alpha variant ramachandran plot, (b) present SARS-CoV-2 Delta variant ramachandran plot, (c) present SARS-CoV-2 Omicron variant ramachandran plot, and (d) present SARS-CoV-2 0rignal China variant ramachandran plot.

DISCUSSION

The coronavirus disease appeared in Wuhan, China, and rapidly turned into a huge global challenge. The NSP9 plays a vital role in RNA binding subunit in the RNA-synthesizing apparatus in the SARS-CoV-2 replication complex. The NSP11 plays a supportive role in the action of the endoribonuclease. It appears that NSP11 participates in the replication process. NSP11 is regarded as a target for the development of medications against coronavirus 2 (Gadhave *et al*., 2021). Phylogenetic analysis is conducted to find evolutionary relationships within viral genomes from neighboring countries. In this analysis, we used multiple sequence alignment to form an evolutionary tree.

The great majority of SARS-CoV-2 mutations were found to be almost certainly neutral in the beginning of the pandemic (Cagliani *et al*., 2020). In NSP9 the rate of genetic mutation was far higher as compared to NSP11 which in turn means that NSP11 was more conserved. A total of 16 mutations were found in NSP9 region from 2020, 2021, 2022 while 9 mutations were found in NSP11.

To create a neighbor-joining tree that visually depicts the genetic links and distances among mutants dispersed throughout several nations, we chose the VOC Alpha, Omicron, Wild strain, Kappa, and Delta lineages. Due to host selection forces, mutations may differ between countries. The regional differences in the genome sequence are clear, and some mutation sites are constant throughout nations. Sawmya *et al*. (2022) performed the phylogenetic analysis of SAR-CoV-2 strains across the globe. Their study revealed that the genomic sequences of China and Italy were closely related to each other. They also found out that the strain of USA, Germany, Qatar, and Poland were similar. In our study for NSP9 region, three Turkish samples fell in the cluster of US. In this cluster, European sequences were also present. Rehman et.al, found the same thing in novel mutations in SARS-CoV-2 isolates from Turkey (Rehman *et al*., 2020).

We also observed that one of the Pakistani samples was part of the cluster pertaining to other European and Turkey samples cluster. Castillo *et al*. (2020) this study found out that the European sequences fell in Asian cluster. Other two samples of Pakistan fell in the Asian cluster in which majority of the sequences were from Bangladesh and Japan, while one sample was German, and one was American. American samples were part of almost all the clusters which might indicate frequent travel and introduction of the variants back and forth in American territory. Two major clusters were made up of almost all American samples. Further smaller clusters in these American clusters were observed. One was made up of Australian, China, and German samples. In the 2nd one, German and Indian samples were present. American and German samples tended to fall close to each other in several instances.

A major cluster was made up of Japanese samples, in which two samples were from India, two from Germany, three from Nepal and another from France. Ko *et al*. (2021) found the Asian sample fell in European countries of whole genome sequences (WGS) of SARS-CoV-2 from throughout Japan and study was subjected to evolutionary analysis by neighbor joining (NJ) method.

First case of Omicron variant was reported from South Africa. For NSP9, the omicron variant fell into the cluster of US, France, Germany, Japan, and Nepal. Jia *et al*. (2022) and Rehman *et al*. (2022) found the same thing in the study of SARS-CoV-2 Omicron variant sequences.

NSP11 genomic sequences formed one major cluster and the majority of sequences here were from America. The other samples fell in the Asian subcluster in which majority were from India, Japan, Bangladesh, Nepal, China, Bangladesh, Turkey, Pakistan, and Australia. NSP11 was found to be more conserved.

These findings suggest that while some mutations are inherent in the evolution of SARS-CoV-2, other modifications may result from the virus's adaptation to a particular country's social and economic conditions, medical care, sex-age ratio, and environmental factors. The identification of certain substitutions is useful for virus tracing for strains that have established stable Lineage. Further research is needed to determine whether and how these mutations affect the infectivity and immunogenicity of SARS-CoV-2.

On close protein sequences analysis, we found that Alpha and first China variant showed VAL position at 4 b-sheet but delta and omicron showed VAL at 7b-sheet. Some changes were predicted in receptor-binding domain resulting in the reduction of antibody interaction. Although none of the variants have shown any visible change in 3d structure, the local quality estimates and Z-Scores vary for all.

To improve global viral surveillance, we must actively monitor the effects of current genetic changes and keep an eye out for any new ones. To stop widespread local transmission, imported illnesses should be quarantined and watched over. In order to make preventive and control policy more accurate and scientific in the future, normalized epidemic prevention and control measures must be optimized.

CONCLUSION

We conclude that the rate of genetic mutation was far higher in the NSP9 region as compared to NSP11.

American samples were part of almost all the clusters which might indicate frequent travel and introduction of the variants back and forth in American territory. In this study we observed that the behavior of viruses and mutations of different countries are closely related to each other. Observing the genomic behavior of viruses is important in order to find out the nature of the viral mutations, any patterns hidden in those to tackle further complications that can arise in future. Region-wise monitoring of genetic changes can also play a role in development of specified vaccines and help regulatory agencies to plan an appropriate strategy to fight the virus.

Statement of conflict of interest

The authors have decalred no conflict of interest.

REFERENCES

Cagliani, R., Forni, D., Clerici, M. and Sironi, M., 2020. Computational inference of selection underlying the evolution of the novel coronavirus, severe acute respiratory syndrome coronavirus 2. *J. Virol.,* **94**: e00411-00420. https://doi.org/10.1128/JVI.00411-20

Castillo, A.E., Parra, B., Tapia, P., Acevedo, A., Lagos, J., Andrade, W. and Tambley, C., 2020. Phylogenetic analysis of the first four SARS-CoV-2 cases in Chile. *J. med. Virol.,* **92**: 1562-1566. https://doi.org/10.1002/jmv.25797

Gadhave, K., Kumar, P., Kumar, A., Bhardwaj, T., Garg, N. and Giri, R., 2021. Conformational dynamics of 13 amino acids long NSP11 of SARS-CoV-2 under membrane mimetics and different solvent conditions. *Microb. Pathogen.,* **158**: 105041. https://doi.org/10.1016/j.micpath.2021.105041

Hasöksüz, M., Kilic, S. and Saraç, F., 2020. Coronaviruses and sars-cov-2. *Turk. J. med. Sci.,* **50**: 549-556. https://doi.org/10.3906/sag-2004-127

Jia, H., Wang, H., Cao, L., Lai, Z., Cheng, Z., Chen, Q. and Xu, C., 2022. Genetic analysis of a SARS-CoV-2 Omicron variant from a Chinese traveller returning from overseas. *Emerg. Microb. Infect.,* **11**: 306-309. https://doi.org/10.1080/22221751.2022.2025747

Ko, K., Nagashima, S.E.B., Ouoba, S., Akita, T., Sugiyama, A. and Ohge, H., 2021. Molecular characterization and the mutation pattern of SARS-CoV-2 during first and second wave outbreaks in Hiroshima, Japan. *PLoS One,* **16**: e0246383. https://doi.org/10.1371/journal.pone.0246383

Raj, R., 2021. Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. *Biochem. biophys. Rep.,* **25**: 100847. https://doi.org/10.1016/j.bbrep.2020.100847

Rehman, S., Mahmood, T., Aziz, E. and Batool, R., 2020. Identification of novel mutations in SARS-COV-2 isolates from Turkey. *Arch. Virol.,* **165**: 2937-2944. https://doi.org/10.1007/s00705-020-04830-0

Rehman, S., Hayat, F.S., Norin, S., Aziz, A., Rahman, S. and Haq, N., 2022. Epitopes determination for Omicron: The COVID-19 new variant. *Pakistan J. Zool.,* **55**: 641-648. https://dx.doi.org/10.17582/journal.pjz/20220712070711

Reshamwala, S.M., Likhite, V., Degani, M.S., Deb, S.S. and Noronha, S.B., 2021. Mutations in SARS-CoV-2 nsp7 and nsp8 proteins and their predicted impact on replication/transcription complex structure. *J. Med. Virol.,* **93**: 4616-4619. https://doi.org/10.1002/jmv.26791

Sawmya, S., Saha, A., Tasnim, S., Anjum, N., Toufikuzzaman, M., Rafid, A.H.M. and Alam, T., 2022. *Phylogenetic analyses of SARS-CoV-2 strains reveal its link to the spread of COVID-19 across the globe medinfo 2021: One world, one health–global partnership for digital innovation*. IOS Press. pp. 729-733. https://doi.org/10.3233/SHTI220174

Shahzaib, A., Raza, T. and Areej, A., 2023. COVID-19 cytokine storm, co-infections, and secondary infections: Recent information and clinical implications. *Hosts Viruses,* **10**: 1-13. https://dx.doi.org/10.17582/journal.hv/2023/10.1.13

Yadav, R., Chaudhary, J.K., Jain, N., Chaudhary, P.K., Khanra, S., Dhamija, P. and Handu, S., 2021. Role of structural and non-structural proteins and therapeutic targets of SARS-CoV-2 for COVID-19. *Cells*, **10**: 821. https://doi.org/10.3390/cells10040821

Zhao, N., Zhou, N., Fan, H., Ding, J., Xu, X., Dong, X. and Yu, Y., 2022. Mutations and phylogenetic analyses of SARS-CoV-2 among imported COVID-19 from abroad in Nanjing, China. *Front. Microbiol*., **13**: 851323. https://doi.org/10.3389/fmicb.2022.851323