ISSN 2449-8955 European Journal of Biological Research *Research Article*

DOI: http://dx.doi.org/10.5281/zenodo.7607992

Analyses of Omicron genomes from India reveal BA.2 as a more transmissible variant

Ashwin Atkulwar [1], Aakif Rehman [2], Yashal Imaan [3], Mumtaz Baig [4]*

[1] Department of Zoology, Amolakchand Mahavidyalaya, Godhani Road, Yavatmal-445001, India

[2] Pace Junior Science College, Dadar, Mumbai, 400028, India

[3] Mohawk College, 135 Fennell Avenue West Hamilton, L9C OE5, ON Canada

[4] Department of Zoology, Laboratory of Molecular and Conservation Genetics, Govt. Vidarbha Institute of Science and Humanities, VMV Road, Amravati 444604, India

* Corresponding author e-mail: mumtaz@lmcg.in

Received: 23 September 2022; **Revised submission:** 30 December 2022; **Accepted:** 1 February 2023

https://jbrodka.com/index.php/ejbr

ABSTRACT: In the current study, the phylodynamics and phylogenomics of Omicron variants are being examined to provide insight into their evolution. We analyzed 564 genomes deposited to the GISAID database from various states of India. A Pangolin COVID-19 Lineage Assigner tool was used to assign lineages to all retrieved genomes. Maximum likelihood (MLE) tree construction and Reduced Median Joining (RM) network were performed. For phylodynamic analysis, the basic reproduction number (R0) was estimated. A Maximum likelihood tree (MLE) confirms the separation of genomes into two distinct clades, BA. 1. and BA. 2. A very high reproduction number (R0) of 2.445 was estimated for the lineage BA.2. Telangana has the highest R0 value in the country, indicating a high prevalence of the BA.2 lineage. The construction of the Reduced Median (RM) network reveals an evolution of some autochthonous haplogroups and haplotypes, which further supports the rapid evolution of Omicron as opposed to its previous variants. Phylogenomic analyses using maximum likelihood (ML) and RM also reveal the likelihood of the emergence of sub-sublineages and novel haplogroups respectively. Due to the recombinant nature and high transmissibility of the Omicron virus, we suggest continuous and more widespread genome sequencing in all states of India to track the evolution of SARS-CoV-2.

Keywords: Omicron; Phylogenomics; SARS-CoV-2; India; Phylodynamics; Reduced Median Network.

1. INTRODUCTION

Virus genomes display a high diversity and can adapt to a variety of hosts with a wide range of tissue tropism and thus increasing the chances of the emergence of novel variants [1]. Through mutation, viruses continuously evolve into new variants. According to their severity, these variants may be classified under variants being monitored, variants of interest, variants of concern, and variants of high consequence. Following the report of the first case on December 31, 2019, WHO officially declared an outbreak of n-CoV-2 as a global pandemic on March 11, 2020. Alpha, Beta, Gamma, Delta, and Omicron are the five subvariants of n-CoV-2 that have been identified so far. The first cases of Omicron (O) belonging to the B.1.1.529 lineage were reported

from Botswana and South Africa on 11th and 14th November 2021, respectively [2]. The variant of concern, Omicron (B.1.1.529), is circulating across the globe with different sub-lineages including BA.1 (B.1.1.529.1), BA.2 (B.1.1.529.2) and BA.3 (B.1.1.529.3). A report by the WHO confirms cases of Omicron in 149 countries, but current data suggest that this variant is detected in approximately 177 countries, so international travel is the most likely way for this variant to spread globally. The Omicron gains attention due to the number of mutations in the spike protein gene, which influences its viral transmissibility, replication, and binding of antibodies, and its dramatic increase in South Africa [3]. Based on baseline studies, it was confirmed that the new variant was able to significantly escape immunity developed from previous infections and vaccinations [4, 5].

According to the WHO report and evidence from the studies, the risk associated with Omicron remains very high. Omicron shows major advantages over the Delta variant in terms of rapid community transmission, causing a dramatic increase in the number of cases in communities. Despite the lower risk of severe symptoms and deaths, the Omicron infection rate in many countries, including India, is very high. This results in higher hospitalization rates and increased pressure on healthcare facilities. There were 91,781 genomes submitted to the GIASID database from around the world, of which India shared 4283. Genome sequencing and annotation of novel SARS-CoV-2 variants from different countries provide baseline data for investigating the molecular epidemiology of this continuously evolving virus [6]. With molecular epidemiology, it is possible to monitor the ongoing pandemic in relation to vaccine development. To determine the underlying phylogeographic and phylodynamic pattern of this Omicron variant in the Indian context, we analysed, 564 genomes deposited from different states of India. This study examined (i) different sub-variants circulating in different states (ii) their evolution within and outside states (iii) estimation of transmission with respect to reproduction number (R0) (iv) efforts of various states in genome sequencing in tracking the variants.

2. MATERIALS AND METHODS

Using the GISAID database, we obtained Omicron (B.1.1.529) variants whole genome sequences from different Indian states deposited until 29th January 2022. We analyzed only complete genomes with high coverage, while gaps and incomplete genomes with missing information were discarded. Overall, 564 genomes were aligned using MAFT version 7 [7], and haplotype numbers were estimated using DnaSP v624 [8]. Prior to alignment, all 564 genomes were assigned lineages using Pangolin COVID-19 Lineage Assigner (https://pangolin.cog-uk.io/) [9]. We compiled a final dataset consisting of 297 haplotypes was compiled after excluding seven genomes with missing bases. A multi-type birth-death model was used in BEAST v2.6.6 [10], in which an HKY nucleotide substitution model with a gamma category count of 4, with a strict clock rate of 2.4E-4 was applied. In MCMC analysis, parameters were sampled every 1000 generations over 10 million generations. The Become Infectious Rate and basic reproduction number (R0) were set to 52.0, assuming a 7-day recovery period. Tracer v1.7.0 was used to determine parameters such as, effective sample size, 95% highest posterior density intervals, reproduction number (R0) and Become Infectious Rate [11]. Network program version 10 was used to construct the Reduced Median Network (RM) from 290 genomes from all major Indian states [12]. PhyML3.0 online version was used to construct a maximum likelihood tree using the same number of genomes [13].

3. RESULTS

3.1. Haplotype estimation, lineage assignment and reproduction number (R0) estimation

After aligning 564 genomes, 297 haplotypes were estimated and used in downstream analyses. Using Pangolin COVID-19 Lineage Assigner, 198 genomes were classified as BA.1, 92 genomes as BA.2 and none as BA.3. For the BA.2 lineage, a reproduction number (R0) of 2.445 was found (HPD=1.525-3.761) (Figure 1). Reproduction number (R0) for lineage BA.1 was found to be 1.700 (HPD=1.143-2.460). Similarly, R0 values of Omicron variants from various states in India range from 0.959 to 1.624 (Table 1 and Figure 1). The highest R0 was found in Telangana, which had the highest prevalence of BA.2 lineage in India, while the lowest reproduction rate (R0) was observed in Uttar Pradesh (UP) (Table 1).

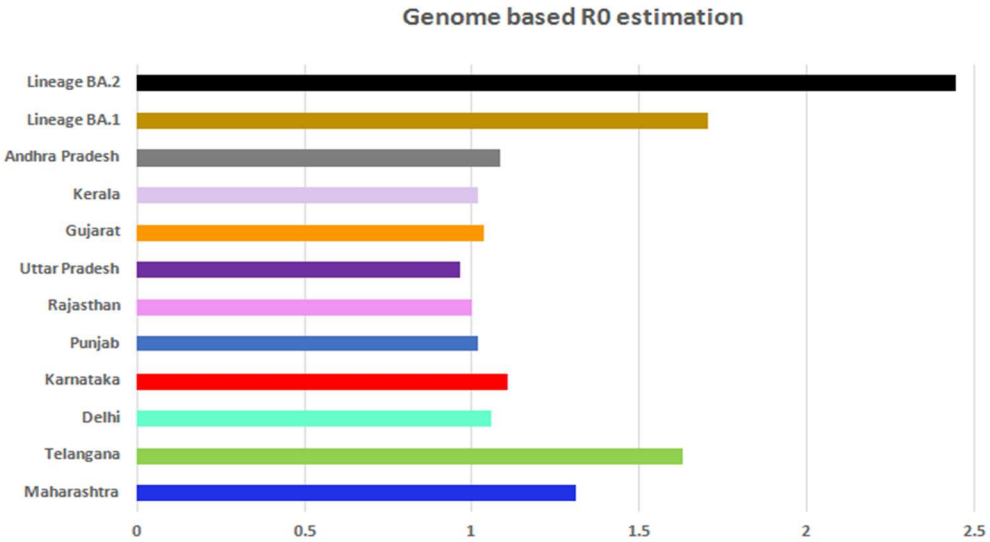

Figure 1. The basic reproduction number (R0), calculated for the various states of India. The graph also demonstrates the comparative analysis of R0's of BA.1 and BA.2 omicron lineages.

Table 1. Comparative R0s of the omicron variants from different states of India with 95% HPD intervals.

Sr. No.	State	R0	95 % HPD Interval
1	Uttar Pradesh	0.959	0.171 – 1.526
2	Kerala	1.013	0.146 – 1.525
3	Andhra Pradesh	1.08	1.010 – 1.184
4	Telangana	1.765	1.226 – 2.695
5	Delhi	1.052	1.012 – 1.110
6	Maharashtra	1.306	0.996 – 1.734
7	Gujarat	1.033	1.000 – 1.081
8	Karnataka	1.103	1.002 – 1.275
9	Punjab	1.018	0.994 – 1.046
10	Rajasthan	0.995	0.931 – 1.058

3.2. Phylogenomic analysis using maximum likelihood (MLE) tree construction

Maximum likelihood-based phylogenetic tree construction using 564 genomes indicates the clustering of an entire dataset into two clades representing two distinct sub-lineages, BA.1 and BA.2 (Figure 2). In addition, this clustering was not found to be very tight and showed the emergence of some sub-sub-lineages visible in the form of short and long branches radiating out within these clades.

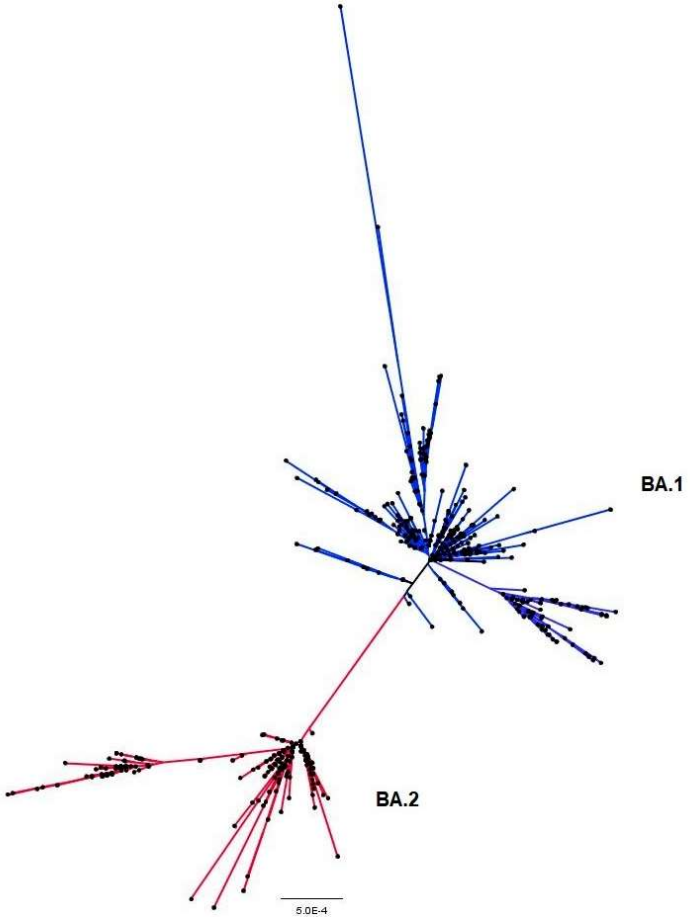

Figure 2. The Maximum likelihood tree constructed online by PhyML shows the clustering of 564 omicron genomes in two major clades. Note, the looseness of the cluster, which is characterized by short and long branches emerging as possible sub-sub lineages.

3.3. Reduced Median (RM) Network construction

Reduced Median (RM) constructed from 297 haplotypes, identified eight major haplogroups, O1 to O8 (Figure 3). Furthermore, these major haplogroups and their associated singleton haplotypes (shown by single color according to state) show interesting relatedness. A singleton haplotype as well as major and minor haplogroups are linked by lineages defined by mutation numbers ranging between 1 to 9 (Figure 3). RM also reveals a total of over 30 median vectors in the network represented by black dots, which signify either unsampled, missing, or extinct taxa. The eight major haplogroups (O1-O8) are formed by the shared haplotypes between genomes of viruses distributed across a maximum of nine states to a minimum of five states. Also, there are equal numbers of minor haplogroups that results from the sharing of haplotypes between four and two states (not marked). A variant haplogroup marked as OA, represents a unique ancestral position in the network because of its direct connection to the first Omicron genome discovered in South Africa (Figure 3). Additionally, haplogroups O5 and O8 in the reconstructed RM are of particular interest since they show a broader distribution and acquired their haplotypes by sharing them across nine and eight states, respectively. Moreover, the expansion of these two haplogroups (O5 and O8) is more star-like, indicating the emergence of more autochthonous haplogroups and haplotypes.

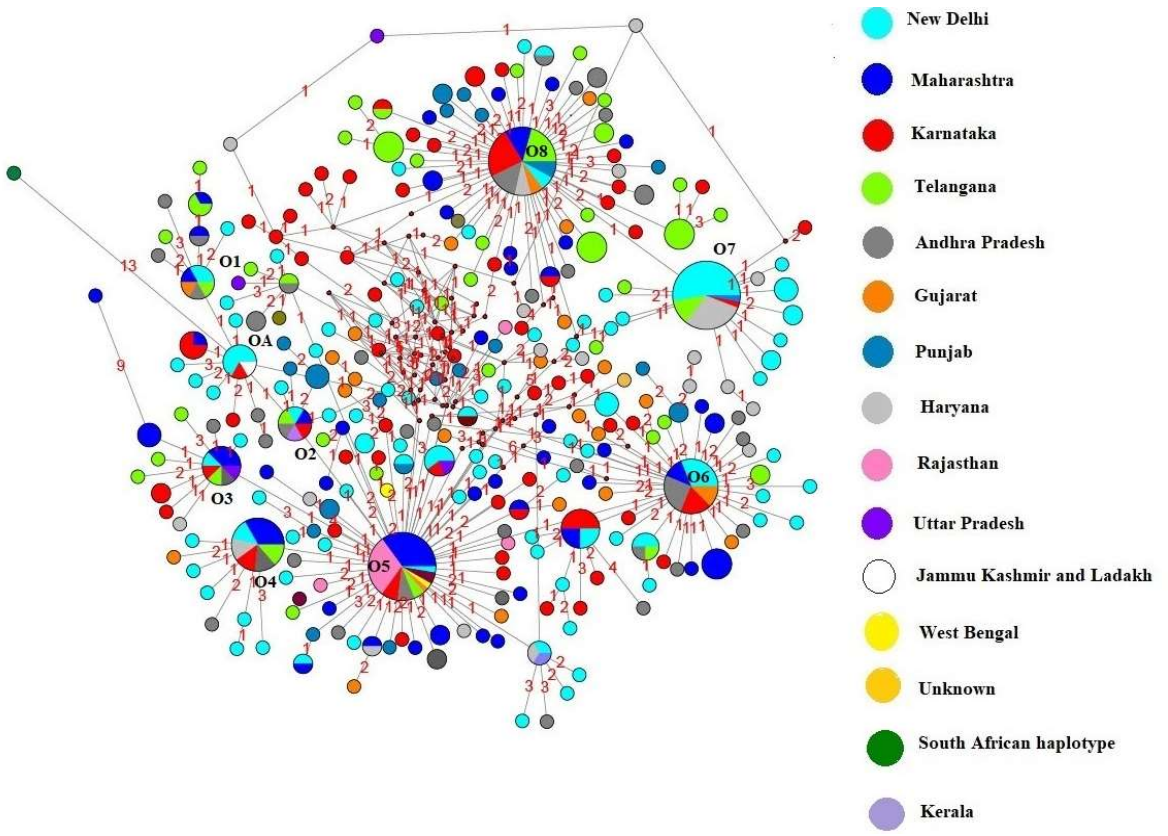

Figure 3. Reduced Median (RM) network constructed out of 297 omicron haplotypes representing various states of India. The major haplogroups inferred in the network are labelled O1-O8, while OA represents an ancestral haplogroup with connection to the early omicron cases detected in South Africa. The size of the circles corresponds to the number of shared haplotypes; Links with red numbers indicate number of mutated nucleotides. The black dots represent median vectors.

4. DISCUSSION

The higher reproduction number (R0) of 2.445 for the BA.2 lineage is in line with Lyngse et al. [14] finding that BA.2 lineage is more transmissible in Denmark. According to Lyngse et al., secondary infections in the vaccinated and unvaccinated Danish population confirm Omicron's immune-evasive properties and higher transmission. Moreover, the Omicron variant exhibited higher recombination rates, facilitating rapid evolution [15]. The technical advisory group on SARS-CoV-2 virus evolution recommends public health authorities to monitor the BA.2 sub-lineage as a distinct sub-lineage of Omicron. In addition, the presence of short and long branches within the BA.1 and BA2 clades inferred by the maximum likelihood tree indicates that the SARS-CoV-2 virus might be capable of recombination and evolving into new lineages (Fig. 2). In addition to this, the density of the population, and vaccination status of the hosts, would also further determine the evolution of sub-sub lineages within the sub-lineages into novel variants of concern. A study published by Loconsole et al. [16], reveals an outbreak of SARS-CoV-2 Omicron infection and successful immune escape in 15 booster-vaccinated healthcare workers in Italy. While the study by Liu et al. [15] demonstrated chimerism in Omicron variants, because of genomic recombination between two SARS-CoV-2 lineages involving the spike protein gene. Consequently, in this course of Omicron evolution, the emergence of new sub-lineages and variants in the future cannot be ruled out. According to several previous studies, BA.2 lineage is the most transmissible among Omicron variants. Because a significant number of novel mutations in the SARS-CoV-2 receptor binding domain (RBD) increases the binding affinity. It has also been confirmed that BA2 has a larger effective population size than BA.1 as well as different antigenicity between BA.1 and BA.2 [17]. In addition,

BA.2 spikes had higher binding affinities than BA.1 spikes, suggesting that BA.2 is more pathogenic than BA.1 [18]. Due to their ability to efficiently efficient assemble, enter and escape neutralization by the antibodies, Omicron variants are more transmissible than previous ones [19, 20]. Median Joining (MJ) and Reduced Median (RM) networks enable geneticists to study genetic patterns in the context of species or taxa-relatedness. Despite its usefulness, it should be used with caution in organisms that are prone to recombination and horizontal gene transfer [16]. The rapid evolution of the Omicron variant as observed in the maximum likelihood tree is also supported by the RM network constructed using the same dataset. The higher number of major haplogroups including sub-haplogroups and associated singleton haplogroups further confirms the rapid evolutionary rate of recombination in both lineages (Fig. 3). Haplogroups are generally regarded as relics of shared ancestry [21, 22]. Haplotype sharing has been used in contact tracing accurately at the beginning of the SARS-CoV-2 pandemic [23]. In India, information obtained from such networks can be used for contact tracing, filling gaps, and community transmission. A significant difference was observed between this RM and the first phylogenetic network published by Forster et al. 2020 at the onset of the epidemic and with our own previously published MJ network [24]. Compared with our previously published MJ, RM exhibits a more distinct and faster evolution of variants/haplogroups (colored circles) and the accumulation of mutations across connecting lineages [25]. We suggest that several novel haplogroups as well as sub-sublineages inferred by our phylogenetic and RM network construction methods may evolve into more transmissible and infectious variants. As a result, we recommend strict tracking of SARS-CoV-2 lineages through regular genome sequencing across India with the aim of developing more effective vaccines in the future.

5. CONCLUSION

The first-ever Omicron genome-based study representing various states of India has revealed two lineages, BA.1 and BA.2. The BA.2 exhibits high reproduction number (R0) and virulence. The study also shows the evolution of sub-sublineages and autochthonous haplogroups. Considering these findings, we suggest continuous genome sequencing, and full vaccination of the Indian population to counter the spread and evolution of the SARS-CoV-2 virus in a highly populous landscape.

Authors' contributions: MB conceived the idea; AA, AR and YI performed analyses; MB and AA wrote the paper. All authors discussed the results and contributed to the final manuscript.

Conflict of interest: The authors declare no potential conflict of interest.

Funding: This research was not funded by any funding agency.

Acknowledgments: We gratefully acknowledge all authors and submitting laboratories of the genomes from GISAID.

REFERENCES

1. Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of corona virus cross-species transmission. J Virol. 2010; 84(7): 3134-3146.
2. World Health Organization Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern; https://www.who.int/news-room/statements/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern; 2021.
3. Spreeuwenberg P, Kroneman M, Paget J. Reassessing the Global Mortality Burden of the 1918 Influenza Pandemic. Am J Epidemiol. 2018; 187(12): 2561-2567.
4. Biondi-Zoccai G, Landoni G, Carnevale R, Cavarretta E, Sciarretta S, Frati G. SARS-CoV-2 and COVID-19: facing the pandemic together as citizens and cardiovascular practitioners. Minerva Cardioangiol. 2020; 68(2): 61-64.

5. Woo P C, Huang Y, Lau S K, Yuen KY. Corona virus genomics and bioinformatics analysis. Viruses. 2010; 2(8): 1804-1820.
6. Zhang YZ, Holmes EC. A genomic perspective on the origin and emergence of SARS-CoV-2. Cell. 2020; 181(2): 223-227.
7. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2010; 4: 772-780.
8. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio C, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large datasets. Mol Biol Evol. 2017; 34: 3299-3302.
9. Toole AO, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019; 15(4): e1006650.
10. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019; 15(4): e1006650.
11. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst Biol. 2018: syy032.
12. Bandelt HJ, Forster P, Roehl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999; 16: 37-48.
13. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol. 2010; 59(3): 307-321.
14. Lyngse FP, Kirkeby CT, Denwood M, Christiansen LE, Molbak K, Moller CH, et al. Transmission of SARS-CoV-2 Omicron VOC subvariants BA.1 and BA.2: Evidence from Danish Households. medRxiv. 2022: 2022.01.28.22270044.
15. Liu IS, Guo Y, Guo Y, Liu L, Chan JF, Huang Y, et al. Antibody evasion properties of SARS-CoV-2 Omicron sublineages. Nature. 2022; 604(7906): 553-556.
16. Loconsole D, Bisceglia L, Centrone F, Sallustio A, Accogli M, Dalfino L, et al. Autochthonous Outbreak of SARS-CoV-2 Omicron Variant in Booster-Vaccinated (3 Doses) Healthcare Workers in Southern Italy: Just the Tip of the Iceberg? Vaccines. 2022; 10(2): 283.
17. Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. J Med Virol. 2022; 94(4): 1641-1649.
18. Yamasoba D, Kimura I, Nasser H, Morioka Y, Nao N, Ito J, et al. Genotype to Phenotype Japan (G2P-Japan) Consortium, Sawa H, Saito A, Irie T, Tanaka S, Matsuno K, Fukuhara T, Ikeda T, Sato K. Virological characteristics of the SARS-CoV-2 Omicron BA.2 spike. Cell. 2022; 185(12): 2103-2115.e19.
19. Syed AM, Ciling A, Taha TY, Chen IP, Khalid MM, Sreekumar B, et al. Omicron mutations enhance infectivity and reduce antibody neutralization of SARS-CoV-2 virus-like particles. Proc Natl Acad Sci USA. 2022; 119(31): e2200592119.
20. Yu J, Collier AY, Rowe M, Mardas F, Ventura JD, Wan H, et al. Neutralization of the SARS-CoV-2 Omicron BA.1 and BA.2 Variants. N Engl J Med. 2022; 386(16): 1579-1580.
21. Makarenkov V, Mazoure B, Rabusseau G, Legendre P. Horizontal gene transfer and recombination analysis of SARS-CoV-2 genes helps discover its close relatives and shed light on its origin. BMC Ecol Evol. 2021; 21(1): 5.
22. Frantz Laurent AF, Bradley DG, Larson G, Orlando L. Animal domestication in the era of ancient genomics. Nat Rev Genet. 2020; 21(8): 449-460.
23. Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, et al. The genetic legacy of the Mongols. Am J Hum Genet. 2003; 72(3): 717-721.

24. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc Natl Acad Sci. USA. 2020; 117: 9241-9243.
25. Farah S, Atkulwar A, Praharaj M R, Khan R, Gandham R, Baig M. Phylogenomics and phylodynamics of SARS-CoV-2 genomes retrieved from India. Future Virol. 2020; 15: 747-753.