Original Paper

Genomic Insights Into the Evolution and Demographic History of the SARS-CoV-2 Omicron Variant: Population Genomics Approach

Kritika M Garg[1,2], PhD; Vinita Lamba[3,4], MSc; Balaji Chattopadhyay[1,3], PhD

[1]Department of Biology, Ashoka University, Sonipat, India

[2]Centre for Interdisciplinary Archaeological Research, Ashoka University, Sonipat, India

[3]Trivedi School of Biosciences, Ashoka University, Sonipat, India

[4]J William Fulbright College of Arts and Sciences, Department of Biological Sciences, University of Arkansas, Fayetteville, AR, United States

Corresponding Author:
Balaji Chattopadhyay, PhD
Trivedi School of Biosciences
Ashoka University
Rajiv Gandhi Education City
Sonipat, 131029
India
Phone: 91 8073119246
Email: balaji.chattopadhyay@ashoka.edu.in

Abstract

Background: A thorough understanding of the patterns of genetic subdivision in a pathogen can provide crucial information that is necessary to prevent disease spread. For SARS-CoV-2, the availability of millions of genomes makes this task analytically challenging, and traditional methods for understanding genetic subdivision often fail.

Objective: The aim of our study was to use population genomics methods to identify the subtle subdivisions and demographic history of the Omicron variant, in addition to those captured by the Pango lineage.

Methods: We used a combination of an evolutionary network approach and multivariate statistical protocols to understand the subdivision and spread of the Omicron variant. We identified subdivisions within the BA.1 and BA.2 lineages and further identified the mutations associated with each cluster. We further characterized the overall genomic diversity of the Omicron variant and assessed the selection pressure for each of the genetic clusters identified.

Results: We observed concordant results, using two different methods to understand genetic subdivision. The overall pattern of subdivision in the Omicron variant was in broad agreement with the Pango lineage definition. Further, 1 cluster of the BA.1 lineage and 3 clusters of the BA.2 lineage revealed statistically significant signatures of selection or demographic expansion (Tajima's $D<-2$), suggesting the role of microevolutionary processes in the spread of the virus.

Conclusions: We provide an easy framework for assessing the genetic structure and demographic history of SARS-CoV-2, which can be particularly useful for understanding the local history of the virus. We identified important mutations that are advantageous to some lineages of Omicron and aid in the transmission of the virus. This is crucial information for policy makers, as preventive measures can be designed to mitigate further spread based on a holistic understanding of the variability of the virus and the evolutionary processes aiding its spread.

(JMIR Bioinform Biotech 2023;4:e40673) doi: 10.2196/40673

KEYWORDS

SARS-CoV-2; Omicron; evolutionary network; population subdivision; genome evolution; COVID-19; microevolution

Introduction

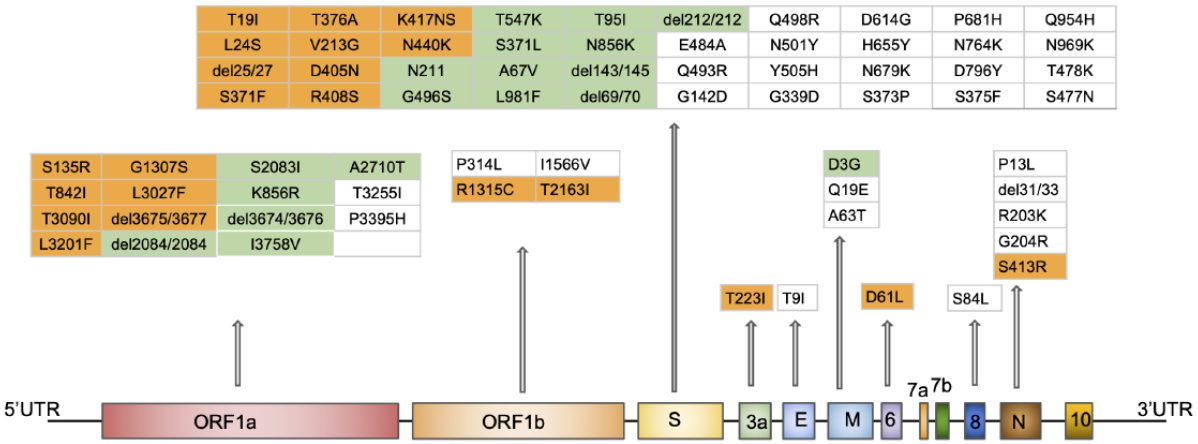
The past 2 decades have witnessed multiple zoonotic coronavirus outbreaks, with the latest being the COVID-19 outbreak, which was caused by SARS-CoV-2. The virus emerged in Wuhan, China, and it quickly spread across the globe, resulting in more than 6.5 million deaths [1-3]. SARS-CoV-2 is a *Betacoronavirus* with a positive, single-stranded RNA genome. The genome is approximately 30 kilobases in length and encodes for 26 proteins (16 nonstructural, 4 structural, and 6 accessory proteins; Figure 1) [4,5].

XSL•FO
RenderX

Extensive genomic surveillance programs were established across the globe to monitor the evolution of the virus. This resulted in an exponential increase in the number of SARS-CoV-2 genomes that has in turn presented a unique set of challenges for data analysis [6-9]. With over 10 million genome sequences already available, new algorithms are being designed to tackle the deluge of data [6-9]. Most available analytical tools are designed to identify the overall evolutionary relationship between various lineages. However, obtaining a finer-level understanding of the diversity and subdivision within a lineage can provide important insights into pathogen evolution, particularly during ongoing pandemics [6]. Pango lineage classification is one such nomenclature method for identifying fine-scale, phylogenetically relevant clusters of SARS-CoV-2 based on the mutations in the spike protein [6].

In this study, we used population genomics methods to understand the subdivision of the Omicron lineage of SARS-CoV-2 as it spread across the globe, in an attempt to elucidate the evolutionary history of the variant. The Omicron variant was first identified within Botswana, Southern Africa, in November 2021, and within a short span of time, it emerged as the main variant driving SARS-CoV-2 infections across the globe, replacing the Delta variant [10,11]. The Omicron variant was also of immediate concern due to the large number of mutations observed in its spike protein (Figure 1). Among the 60 mutations that this variant accumulated when compared to the reference Wuhan sequence, the majority were concentrated in the spike protein (38 mutations in the BA.1 lineage and 31 mutations in the BA.2 lineage; Figure 1) [12,13]. Some of these mutations increased both the transmission ability and the antibody escape of the virus, allowing the Omicron variant to rapidly spread across the globe [11,13]. Given the high infection rate and rapid spread of the virus across the globe, alternative methods for inspecting fine-scale subdivision and transmission are necessary to understand the evolution of the virus and devise any strategy to reduce its spread. Thus, we investigated the subdivision and demographic history of the BA.1 and BA.2 lineages of Omicron, along with identifying mutations that are correlated with the spread of the virus.

Figure 1. The genome structure of SARS-CoV-2 with known mutations in the Omicron variant highlighted. Mutations unique to the BA.1 lineage are highlighted in green, and those unique to the BA.2 lineage are highlighted in orange. Mutations common to both lineages of Omicron are in plain black font. The list of mutations was obtained from Tzou et al [12]. ORF: open reading frame; UTR: untranslated region.



Methods

Data Matrix and Cleanup

We downloaded 20,067 SARS-CoV-2 genome sequences belonging to the Omicron lineage, which were available up to January 31, 2022, from the GISAID (Global Initiative on Sharing All Influenza Data) repository (Multimedia Appendices 1 and 2), retaining only high-coverage genomes (<1% undetermined nucleotide bases; <0.05% unique amino acid mutations) and genomes with a collection date for this study. Only sequences obtained from humans were used for all analyses. We retained 20,067 genomes, which were further filtered for quality by using Nextclade CLI (Nextstrain) [14]. Nextclade examines each query sequence for flaws that could suggest sequencing or assembly errors and assigns a score for each sequence. The quality score of a sequence is determined by the number of undetermined bases, ambiguous sites, private mutations, and stop codons. All sequences classified as good-quality sequences by Nextclade were selected for further analysis. We retained 14,002 good-quality sequences, of which most were from Denmark (n=11,272, 80.5%); the rest of the sequences were from 43 countries. We aligned the filtered SARS-CoV-2 genomes to the Wuhan reference genome (accession ID: MN908947.1) in Nextalign CLI [14], using default parameters. Further, we assigned the lineage for each sequence by using the pangolin web server (versions 3.1.20 and 4.0.6; accessed on March 3 and May 6, 2022, respectively) [6].

Genetic Subdivision Analysis

We used two different approaches to understand the population subdivision within our SARS-CoV-2 data set. For the first approach, we reconstructed an evolutionary network by using the program VENAS (Viral Genome Evolution Network Analysis System) [15]. VENAS can analyze thousands of genomes in a short span of time (a few minutes) and is a useful tool for tracking changes across a transmission chain. It identifies mutations across alignments and constructs a network based on hamming distances. In VENAS, we first estimated the effective parsimony-informative sites and minor allele frequency, using default settings, and retained 5253 genomes. These were then used to construct an evolutionary network, which was viewed in Cytoscape 3.9.1 (Cytoscape Consortium)

XSL•FO
RenderX

by using the prefuse force-directed method [16]. Finally, we analyzed the BA.1 (n=260) and BA.2 (n=4993) lineages separately to understand the fine-scale subdivision within each lineage.

For the second approach, we used the discriminant analysis of principal components (DAPC) method to understand the fine-scale subdivision patterns observed in each lineage (based on the Pango lineage definitions mentioned in the *Data Matrix and Cleanup* section). The DAPC is a useful method for detecting subdivision, as it maximizes between-group differences while minimizing within-group variability [17]. It is a relatively fast method for detecting complex subdivision patterns from genomic data. We used the filtered genomes obtained from VENAS and performed a DAPC for both lineages by using the R *adegenet* package (R Foundation for Statistical Computing) [17,18]. We first identified the optimal number of clusters within each data set, using the K-means algorithm, and then performed the DAPC. We further identified the unique mutations for each of the DAPC clusters and only considered mutations that were present in at least 70% of the sequences belonging to the cluster.

Genomic Diversity and Selection Analysis

To estimate the level of genomic diversity within our data set, we characterized all substitutions in reference to the Wuhan genome by using VENAS. We considered all 14,003 good-quality sequences and identified the mutations for the 12 functional open reading frames (ORFs). We further estimated Tajima's D values for the spike protein sequences for all of the clusters identified in the DAPC, using MEGA (Molecular Evolutionary Genetics Analysis; version 10.2.6 [Pennsylvania State University]) [19]. The Tajima's D test is widely used to identify signatures of microevolutionary forces, such as population fluctuations and selections acting upon populations. We estimated the Tajima's D value for each genetic cluster identified by the DAPC to avoid confounding effects of population subdivision.

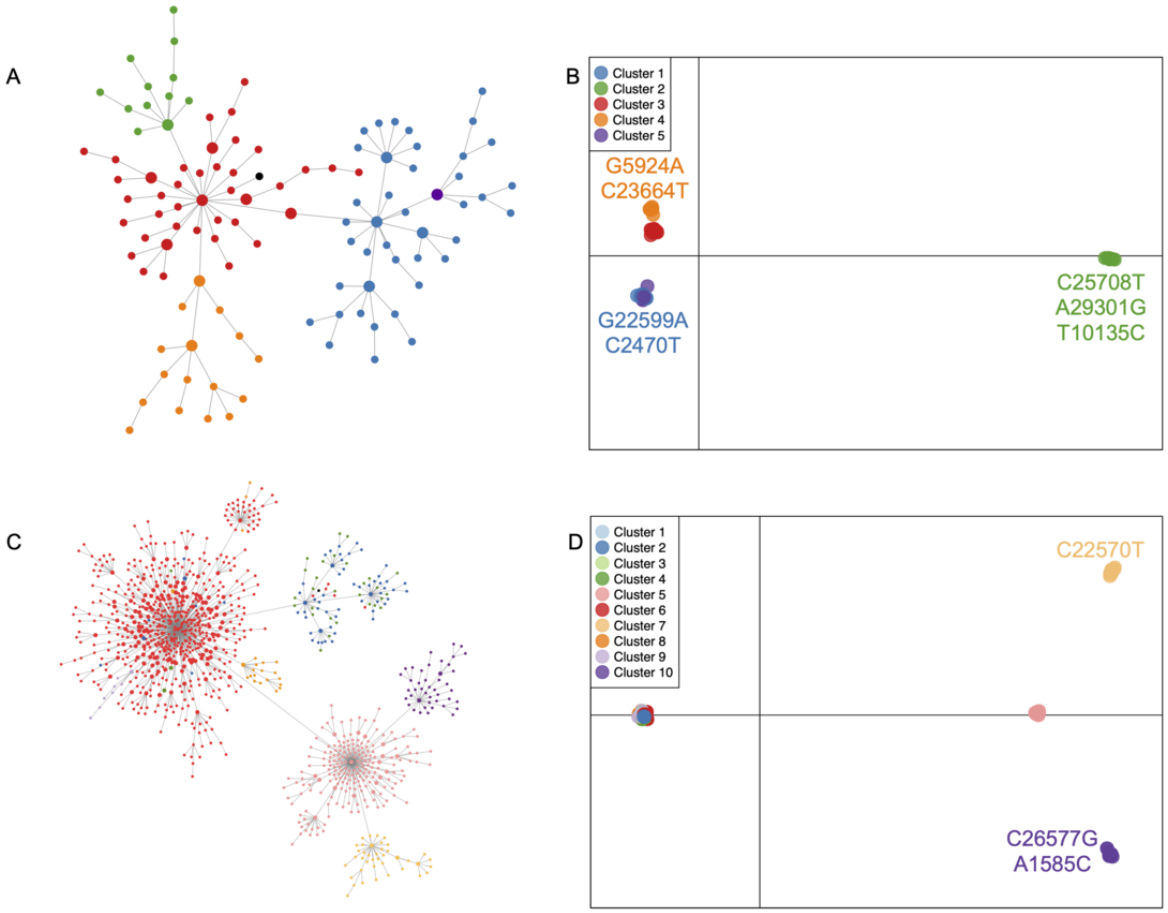
Results

Genetic Subdivision

We observed signatures of genetic subdivision in the BA.1 and BA.2 lineages of the Omicron variant. The patterns were broadly concordant between both approaches—the evolutionary network approach conducted in VENAS and the statistical approach using the DAPC. Although VENAS produced numerous nodes and groups (BA.1: n=111; BA.2: n=1046), these were nested within fewer, broader subdivisions retrieved by the DAPC (Figure 2). We identified 5 clusters for the BA.1 lineage and 10 clusters for the BA.2 lineage, using the DAPC. However, when we visualized our results, we observed that 2 clusters for the BA.1 lineage were clubbed together, and 4 clusters for the BA.2 lineage were clubbed together (Figure 2B and Figure 2D). Further, no sequences were assigned to clusters 1 and 3 for the BA.2 lineage. Thus, effectively, only 4 clusters for the BA.1 and BA.2 lineages were identified by the DAPC method. Private mutations were identified for 5 clusters (Figure 2B and Figure 2D).

XSL•FO
RenderX

Figure 2. The genetic subdivision observed in the Omicron lineage of SARS-CoV-2 based on haplotype networks (panel A: BA.1 lineage; panel C: BA.2 lineage). Panels B and D depict observed genetic subdivision based on the discriminant analysis of principal components (DAPC) for the BA.1 and BA.2 lineages, respectively. The Wuhan sequence is denoted by the black-colored node in panels A and C. Private mutations, if any, are depicted in the DAPC plot. A private mutation must be present in at least 70% of the sequences within the cluster.



Genomic Diversity

A detailed inspection of the pattern of substitution among the study genomes revealed that, as expected, the spike protein, ORF1a, and ORF1b harbored the maximum number of variations (Figure S1 in Multimedia Appendix 1). Gene coding for the envelope protein had the lowest rate of change. The most frequent mutations observed within our panel of genomes were C to T transition and G to T transversion (Figure S2 in Multimedia Appendix 1). Tajima's D values for the spike protein sequences were negative for all of the DAPC clusters, with significant values observed only for 4 clusters (Tajima's D<−2; Table 1).

XSL•FO
RenderX

Table 1. Tajima's D estimates for the various clusters that were identified by using the discriminant analysis of principal components.

Cluster ID	Number of sequences	Sampling location	Number of segregating sites	θ	Nucleotide diversity	Tajima's D
BA.1 lineage						
Cluster 1	41	Belgium, Germany, India, Japan, Mexico, Romania, South Africa, Switzerland, Taiwan, and United States of America	16	0.00293	0.00071	−2.425328[a]
Cluster 2	12	Denmark, Germany, India, Mexico, and United States of America	7	0.00182	0.00095	−1.866946
Cluster 3	40	Denmark, England, Germany, India, Japan, Slovenia, South Africa, Switzerland, Taiwan, Thailand, and United States of America	11	0.00203	0.00073	−1.948315
Cluster 4	16	Denmark, England, Germany, India, Romania, South Africa, Spain, and Switzerland	8	0.00189	0.00105	−1.598913
Cluster 5	1	South Africa	N/A[b]	N/A	N/A	N/A
BA.2 lineage						
Cluster 1	0	N/A	N/A	N/A	N/A	N/A
Cluster 2	61	Australia, Denmark, India, Singapore, and South Africa	8	0.001343	0.00028	−2.08083[a]
Cluster 3	0	N/A	N/A	N/A	N/A	N/A
Cluster 4	31	Denmark, India, Norway, and Singapore	2	0.000393	0.00010	−1.50558
Cluster 5	207	Denmark and Singapore	19	0.002527	0.00032	−2.31247[a]
Cluster 6	632	Denmark, Singapore, and South Africa	50	0.005591	0.00027	−2.59423[a]
Cluster 7	40	Denmark	3	0.000554	0.00019	−1.4309
Cluster 8	22	Denmark	1	0.000215	0.00007	−1.16240
Cluster 9	8	Denmark	1	0.000303	0.00020	−1.05482
Cluster 10	44	Denmark	3	0.000542	0.00027	−1.07839

[a]Tajima's D values of <−2 indicate significant demographic expansion or selection.

[b]N/A: not applicable.

Discussion

Study Overview

In this study, we investigated the effectiveness of population genomics methods to identify the fine-scale structure and demographic history of the Omicron lineage during the initial spread of the virus. Our study also highlights the utility of population genomics methods in handling large data sets and provides an analytical framework for future studies, which will help with understanding the genetic substructuring of the virus and identifying mutations that are potentially advantageous to the spread of the virus.

Fine-Scale Subdivision Within the Omicron Variant

Using a combination of population genetics methods, this study revealed cryptic, fine-scale substructuring within our data set. We observed a similar pattern of subdivision for each Omicron lineage, using both VENAS and the DAPC (Figure 2). Although both methods use different approaches, together they provide a robust understanding of the finer subdivision patterns within fast-evolving lineages. At the start of this study, Pango lineage definition 3.1.20 was available, which had divided the Omicron sequences into the BA.1, BA.1.1, and BA.2 lineages, and our population genomics–based clustering identified finer-level subdivision within these lineages. With the updated Pango lineage version 4.0.6, there is now a broad agreement in the lineage definitions between our methodology (DAPC and VENAS) and the Pango lineage.

We identified cluster-defining mutations that were later selected for Pango lineage definitions, such as the G22599A and G5924A mutations for BA.1.1 (clusters 1 and 5 in the DAPC) and BA.1.17 (cluster 4 in the DAPC), respectively (Figure 2B). The subdivision observed in our study, as well as some of the cluster-defining mutations (G5924A, G22599A, C2470T, and A29301G), also agrees with recent phylogenetic reconstructions (Figure 2B) [20,21].

We also recovered signatures of fine-scale subdivision within the updated Pango (version 4.0.6) definitions. For example, DAPC clusters 5, 7, and 10, which are all part of the BA.2.9 lineage from Denmark (Figure 2D), were segregated based on 3 unique mutations (Figure 2D). The mutation C22570T is unique to cluster 7, and the mutations C26577G and A1585C are unique to cluster 10 within our data set (Figure 2D). Thus,

XSL•FO
RenderX

our analytical regime could not only retrieve cluster-defining mutations in agreement with other methods but also identify finer subdivisions within existing Pango definitions and associated unique mutations.

Selection and Demographic History of the Omicron Variant

Tests for selection revealed that the evolution of the Omicron lineage could be attributed to microevolutionary processes, such as selection and demographic expansion. We used Tajima's D values to test for deviation of the identified clusters from neutrality. A negative Tajima's D value reflects a deficit of haplotypes in comparison to the number of segregating sites and is indicative of a recent selective sweep or a population expansion [22]. Significant negative Tajima's D values were observed for a subset of the DAPC clusters (BA.1 lineage: cluster 1; BA.2 lineage: clusters 2, 5, and 6; Table 1), suggesting that these clusters have undergone rapid expansion, experienced recent selective sweeps, or both. These attributes are indicators of greater transmissibility and thus make these clusters potential targets for surveillance and monitoring programs. For example, the spike protein mutation G22599A (S:R346K) is implicated in providing a transmission advantage and the antibody escape ability to the BA.1.1 variant. Although the population genomics framework adopted in our study identified this diagnostic mutation, which defines cluster 1 of the BA.1 lineage, the test for deviations from neutrality returned a significant negative value only for this cluster (Tajima's D=−2.425328), indicating the selective advantage, as well as signals of population expansion, of this cluster across the globe (Figure 2B) [23-25]. In addition to G22599A (S:R346K), we also identified the mutation C23664T (S:A701V), which, in conjunction with S:N501Y, provides a mild advantage to the virus by increasing the rate of infection [20]. However, the C23664T mutation is not unique to the Omicron lineage and is also observed in other SARS-CoV-2 variants of concern [20].

Interestingly, cluster 2 of the BA.1 lineage, which did not exhibit a signature of expansion or selection, also harbors 3 unique mutations (C25708T, A29301G, and T10135C), which have been independently identified as suppressor mutations associated with a reduction in the spread of the virus [26] (Figure 2A and Figure 2B; Table 1).

Future research efforts can use a similar analytical framework to swiftly identify mutations that are important for virus evolution, of which some might play an important role in facilitating the spread of a virus, while others may be detrimental to its transmission. We demonstrated that a combination of population genomics methods can be used to recover subtle variations within established lineage definitions and potentially aid in finding variants of concern. The identification of such target mutations is necessary from an epidemiological standpoint, as well as for vaccine development. This study provides an easy analytical framework that can be used by policy makers to identify variants of potential concern and understand the local demographic history and spread of a virus, thereby facilitating disease mitigation.

Conclusion

We provide an easy, computationally tractable framework for understanding the genetic subdivision and demographic history of SARS-CoV-2. Our framework can be quickly implemented to identify potentially important mutations that may be driving the spread of the virus. Such information can be very useful for deciphering the pattern of movement of variants and determining correlations with the local history of an outbreak.

Acknowledgments

BC acknowledges the startup funding from Trivedi School of Biosciences (TSB), Ashoka University, India. KMG acknowledges the support from the Department of Biotechnology-Ramalingaswami Fellowship (grant BT/HRD/35/02/2006). VL was supported by a TSB fellowship.

We gratefully acknowledge the authors from the originating laboratories, which were responsible for obtaining the specimens, and the submitting laboratories, where genetic sequence data were generated and shared via the GISAID (Global Initiative on Sharing All Influenza Data) Initiative, on which this research is based. A full acknowledgment table can be found in Multimedia Appendix 2.

Data Availability

The SARS-CoV-2 sequences analyzed during this study are available in the GISAID (Global Initiative on Sharing All Influenza Data) repository. Details of the sequences are available in Multimedia Appendix 2.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supporting information.
[DOCX File , 59 KB-Multimedia Appendix 1]

XSL•FO
RenderX

Multimedia Appendix 2

Acknowledgment table.
[PDF File , 3198 KB-Multimedia Appendix 2]

References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
2. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020 Mar;579(7798):270-273 [FREE Full text] [doi: 10.1038/s41586-020-2012-7] [Medline: 32015507]
3. World Health Organization. 2nd Global consultation on assessing the impact of SARS-CoV-2 variants of concern on public health interventions. World Health Organization. 2021 Jun 10. URL: https://www.who.int/publications/m/item/2nd-global-consultation-on-assessing-the-impact-of-sars-cov-2-variants-of-concern-on-public-health-interventions [accessed 2023-05-09]
4. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020 Feb 22;395(10224):565-574 [FREE Full text] [doi: 10.1016/S0140-6736(20)30251-8] [Medline: 32007145]
5. Brant AC, Tian W, Majerciak V, Yang W, Zheng ZM. SARS-CoV-2: from its discovery to genome structure, transcription, and replication. Cell Biosci 2021 Jul 19;11(1):136 [FREE Full text] [doi: 10.1186/s13578-021-00643-z] [Medline: 34281608]
6. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol 2021 Jul 30;7(2):veab064 [FREE Full text] [doi: 10.1093/ve/veab064] [Medline: 34527285]
7. McBroome J, Thornlow B, Hinrichs AS, Kramer A, De Maio N, Goldman N, et al. A daily-updated database and tools for comprehensive SARS-CoV-2 mutation-annotated trees. Mol Biol Evol 2021 Dec 09;38(12):5819-5824 [FREE Full text] [doi: 10.1093/molbev/msab264] [Medline: 34469548]
8. Chen C, Nadeau S, Yared M, Voinov P, Xie N, Roemer C, et al. CoV-Spectrum: analysis of globally shared SARS-CoV-2 data to identify and characterize new variants. Bioinformatics 2022 Mar 04;38(6):1735-1737 [FREE Full text] [doi: 10.1093/bioinformatics/btab856] [Medline: 34954792]
9. Sokhansanj BA, Rosen GL. Mapping data to deep understanding: Making the most of the deluge of SARS-CoV-2 genome sequences. mSystems 2022 Apr 26;7(2):e0003522 [FREE Full text] [doi: 10.1128/msystems.00035-22] [Medline: 35311562]
10. Paton RS, Overton CE, Ward T. The rapid replacement of the SARS-CoV-2 Delta variant by Omicron (B.1.1.529) in England. Sci Transl Med 2022 Jul 06;14(652):eabo5395 [FREE Full text] [doi: 10.1126/scitranslmed.abo5395] [Medline: 35503007]
11. Tian D, Sun Y, Xu H, Ye Q. The emergence and epidemic characteristics of the highly mutated SARS-CoV-2 Omicron variant. J Med Virol 2022 Jun;94(6):2376-2383 [FREE Full text] [doi: 10.1002/jmv.27643] [Medline: 35118687]
12. Tzou PL, Tao K, Pond SLK, Shafer RW. Coronavirus Resistance Database (CoV-RDB): SARS-CoV-2 susceptibility to monoclonal antibodies, convalescent plasma, and plasma from vaccinated persons. PLoS One 2022 Mar 09;17(3):e0261045 [FREE Full text] [doi: 10.1371/journal.pone.0261045] [Medline: 35263335]
13. Fan Y, Li X, Zhang L, Wan S, Zhang L, Zhou F. SARS-CoV-2 Omicron variant: recent progress and future perspectives. Signal Transduct Target Ther 2022 Apr 28;7(1):141 [FREE Full text] [doi: 10.1038/s41392-022-00997-x] [Medline: 35484110]
14. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation calling and quality control for viral genomes. J Open Source Softw 2021 Nov 30;6(67):3773 [FREE Full text] [doi: 10.21105/joss.03773]
15. Ling Y, Cao R, Qian J, Li J, Zhou H, Yuan L, et al. An interactive viral genome evolution network analysis system enabling rapid large-scale molecular tracing of SARS-CoV-2. Sci Bull (Beijing) 2022 Apr 15;67(7):665-669 [FREE Full text] [doi: 10.1016/j.scib.2022.01.001] [Medline: 35036033]
16. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003 Nov;13(11):2498-2504 [FREE Full text] [doi: 10.1101/gr.1239303] [Medline: 14597658]
17. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 2008 Jun 01;24(11):1403-1405 [doi: 10.1093/bioinformatics/btn129] [Medline: 18397895]
18. R: The R Project for Statistical Computing. R Foundation for Statistical Computing. URL: https://www.R-project.org/ [accessed 2023-05-09]
19. Kumar S, Tamura K, Nei M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Comput Appl Biosci 1994 Apr;10(2):189-191 [doi: 10.1093/bioinformatics/10.2.189] [Medline: 8019868]

XSL•FO
RenderX

20. Montaño RZ, Culasso ACA, Fernández F, Marquez N, Debat H, Salmerón M, et al. Evolution of SARS-CoV-2 during the first year of the COVID-19 pandemic in Northwestern Argentina. Virus Res 2022 Sep 28;323:198936 [FREE Full text] [doi: 10.1016/j.virusres.2022.198936] [Medline: 36181975]
21. Liu D, Cheng Y, Zhou H, Wang L, Fiel RH, Gruenstein Y, et al. Early introduction and community transmission of SARS-CoV-2 Omicron variant, New York, New York, USA. Emerg Infect Dis 2023 Feb;29(2):371-380 [FREE Full text] [doi: 10.3201/eid2902.220817] [Medline: 36692451]
22. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 1989 Nov;123(3):585-595 [FREE Full text] [doi: 10.1093/genetics/123.3.585] [Medline: 2513255]
23. Neher RA. Contributions of adaptation and purifying selection to SARS-CoV-2 evolution. Virus Evol 2022 Dec 10;8(2):veac113 [FREE Full text] [doi: 10.1093/ve/veac113]
24. Nutalai R, Zhou D, Tuekprakhon A, Ginn HM, Supasa P, Liu C, OPTIC consortium, ISARIC4C consortium, et al. Potent cross-reactive antibodies following Omicron breakthrough in vaccinees. Cell 2022 Jun 09;185(12):2116-2131.e18 [FREE Full text] [doi: 10.1016/j.cell.2022.05.014] [Medline: 35662412]
25. Zaman K, Shete AM, Mishra SK, Kumar A, Reddy MM, Sahay RR, et al. Omicron BA.2 lineage predominance in severe acute respiratory syndrome coronavirus 2 positive cases during the third wave in North India. Front Med (Lausanne) 2022 Nov 02;9:955930 [FREE Full text] [doi: 10.3389/fmed.2022.955930] [Medline: 36405589]
26. Yang HC, Wang JH, Yang CT, Lin YC, Hsieh HN, Chen PW, et al. Subtyping of major SARS-CoV-2 variants reveals different transmission dynamics based on 10 million genomes. PNAS Nexus 2022 Sep 01;1(4):pgac181 [FREE Full text] [doi: 10.1093/pnasnexus/pgac181] [Medline: 36714842]

Abbreviations

DAPC: discriminant analysis of principal components
GISAID: Global Initiative on Sharing All Influenza Data
MEGA: Molecular Evolutionary Genetics Analysis
ORF: open reading frame
VENAS: Viral Genome Evolution Network Analysis System

Edited by A Uzun; submitted 03.07.22; peer-reviewed by A Krishnan, A Alwin Prem Anand, S Sankar; comments to author 06.02.23; revised version received 07.03.23; accepted 05.05.23; published 12.06.23

Please cite as:
Garg KM, Lamba V, Chattopadhyay B
Genomic Insights Into the Evolution and Demographic History of the SARS-CoV-2 Omicron Variant: Population Genomics Approach
JMIR Bioinform Biotech 2023;4:e40673
URL: https://bioinform.jmir.org/2023/1/e40673
doi: 10.2196/40673
PMID: 37456139

©Kritika M Garg, Vinita Lamba, Balaji Chattopadhyay. Originally published in JMIR Bioinformatics and Biotechnology (https://bioinform.jmir.org), 12.06.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Bioinformatics and Biotechnology, is properly cited. The complete bibliographic information, a link to the original publication on https://bioinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX