Research Article

Whole Genome Sequencing of Omicron Identified Multiple Outbreaks and Introduction Events in India during November 2021 and January 2022

Shreedhanya D. Marathe, Varun Shamanna, Geetha Nagaraj, Nischita S, Muthumeenakshi Bhaskaran, K.L Ravi Kumar[*]

Department of Biotechnology, Kempegowda Institute of Medical Sciences, Bengaluru, India

ABSTRACT

The Variant of Concern (VOC), Omicron is the predominant variant circulating throughout the world of the SARS CoV-2 pandemic during the third wave including India. The world health organisation has designated this highly mutated variant as a VOC due to its high transmissibility and risk of reinfection. Whole genome sequencing and analysis were performed for SARS-CoV-2 PCR positive samples between December 2021 to January 2022. From the 133 Omicron variants detected, genomic analysis was carried out by contextualizing them with 1586 complete genomes of Omicron from India obtained from GISAID. The Omicron variant prevalence in India has increased in a log phase within 3 months in most of the metropolitan cities. With the limited sequencing data available, it was identified that the sublineage BA.1 was predominant in India from November 2021 to January 2022. Further, the first sequence of BA.2 sublineage was submitted from Delhi only in the mid of December 2021. The two outbreaks observed were of BA.2 variant and were found to spread to multiple cities in a short time. The rapid spread and specific mutations in the outbreak samples of Omicron indicate that the variant is highly transmissible when compared to previous variants. The study shows the importance of genomic sequence to identify the emergence of clusters and take actions to prevent further spreading events.

Keywords: SARS-CoV-2; Omicron; India; Epidemiology; Whole genome sequencing

INTRODUCTION

Coronavirus is a virus that infects the nose, sinuses, and upper respiratory tract, most of them not causing severe effects. After an outbreak in Wuhan, China in December 2019, the World Health Organisation (WHO) recognized Severe Acute Respiratory Syndrome-Coronavirus 2 (SARS-CoV-2) as a novel coronavirus that was a serious threat to humans. The SARS-CoV-2 virus, like other coronaviruses, has been undergoing continuous mutations, resulting in multiple Variants of Concern (VOC) [1].

On November 25, 2021, the WHO identified the Omicron (B. 1.1.529) variation of SARS-CoV-2 to be a variant of concern after its initial identification in gauteng province, South Africa. The Omicron variant's rapid spread sparked the fourth wave of SARS-CoV-2 infections in South Africa, with daily diagnosed cases exceeding totals reported in the country during all preceding periods. Various reports from across the globe have shown that there is a significant increase in the spread of the omicron variant when compared to the delta. As of January 2022, the Omicron variant has disseminated worldwide including India, where 90% of all SARS-CoV-2 infections diagnosed in the metro cities in January 2022 were estimated to be caused by the Omicron variant. The rise in the positivity rate resulted in the declaration of a third wave crisis in the country. The spread of SARS-CoV2 widely and causing a high risk of infection has led to a greater opportunity for mutation and possible positive evolution. This variant is known to contain numerous mutations in the RBD of the spike protein which has led to its maximum transmissibility. It has been hypothesized that omicron would have emerged from Alpha that had

stemmed from a persistently infected patient who got infected with multiple variants. But as the number of cases is increasing and mutation analysis is becoming more robust, we are deviating from this hypothesis and concluding that omicron is emerging with multiple mutations from different variants. At least 50 mutations differentiate the Omicron variant from the SARS-CoV-2 reference strain. The spike protein alone possesses 30 of the mutations with some overlapping mutations with previously identified strains as well as mutations specific to omicron. Preliminary laboratory studies have shown that Omicron can escape extensively, but incompletely, from neutralizing antibodies in vaccinated and convalescent sera [2].

COVID-19 pandemic is the first where real time Whole Genome Sequencing (WGS) was performed to track transmission, identify vaccine targets, and design therapeutics for the virus. The accurate identification and sequencing of SARS-CoV-2 in patient samples provides extensive information, ranging from reducing false negatives and contact tracing to evaluating the suitability of diagnostic assays (particularly nucleic acid-based) and determining whether vaccines are likely to be or remain effective despite genetic recombinations, diverse mutations during virus replication [3-6].

We performed and analyzed WGS of SARS-CoV2 samples sequenced as a part of INSACOG and compared with existing genome sequences of omicron variant available in GISAID from India. The geospatial distribution and demographic characterization of the sequences submitted were analyzed. Sublineage proportions and distinct SNPs were identified [7-10].

MATERIALS AND METHODS

Sequencing of SARS-COV-2 PCR positive isolates

Clinical specimens: The study was conducted after obtaining ethical approval vides letter number KIMS/IEC/A012/M/2022. Central research laboratory is an ICMR recognized COVID-19 testing laboratory set up in Kempegowda institute of medical sciences, a tertiary care hospital and a postgraduate medical college situated in the heart of Bengaluru city. Nasopharyngeal swabs collected from 01 December 2021 up to the 10th of January 2022 as a part of routine diagnostics were tested for COVID-19 using the ICMR approved RT-PCR test. The positive samples with a CT Value between 20 to 30 were taken for sequencing as a part of the national SARS COV-2 surveillance program INSACOG. The epidemiological metadata was obtained from the SRF forms submitted during the collection of the sample [11].

Genome sequencing: The 192 samples positive for COVID-19 from RT-PCR were sequenced on the nanopore MK1C platform. Initially, 1200 bp amplicons were generated using midnight PCR primers by doing as per the manufacturer's specification. The concentrations of cDNA were determined using the dsDNA HS Assay Kit with Qubit 4 fluorometer. RBK-110-96 rapid barcoding kit was used to barcode each sample and the sequencing were performed on MIN-106 flow cells as per the SARS-CoV-2 genome sequencing protocol described in the protocols [12].

Sequence analysis: Guppy 3.6.0 was used for base calling and demultiplexing the samples with a 450 bp fast base calling configuration. The demultiplexed reads were assembled using nf-core viralrecon pipeline with medaka g303 model for read polishing. A minimum of 250 bp and a maximum of 1400 bp were used as read cutoff as per the standards used for the midnight protocol. The consensus assembly obtained was used further for variant calling and lineage analysis [13].

Retrieval of genomic data from the database

From the Global Initiative for Sharing All Influenza Data (GISAID), we downloaded 1586 high coverage, complete VOC Omicron GRA (B.1.1.529+BA.*) SARS-CoV-2 genome sequences submitted from different centers across India as of 19th January 2022. Also, the patient information and metadata for the downloaded sequences were retrieved. The sequences with large gaps and more than 50 ambiguous bases were removed. The sequence of the established SARS-CoV-2 reference genome was downloaded from NCBI for further variant analysis and gene annotations.

SNP identification and phylogenetic analysis

SNP identification and lineage prediction: The consensus sequence files from in house sequence analysis and those downloaded from GISAID were combined and submitted to get the type defining markers and clade assignment. The open source web based tool provided mutation detection, clade assignment, quality checks, and phylogenetic placement of the isolates. The genomic dataset was classified into sub lineages using PANGOLIN vv3.1.20 (Phylogenetic Assignment of Named Global Outbreak Lineages) [14].

Phylogenetic tree construction: The consensus sequences were aligned to the reference genome to create multiple sequence alignment using MAFFT and the alignment positions tagged as inappropriate for phylogenetic analysis were removed. Maximum likelihood phylogenetic reconstruction was performed with the curated alignment and a General Time Reversible (GTR) model using fast tree with 1000 bootstrap replicates [15].

Visualisation of the tree

The microreact was used to view the tree and analyze the SNPs and distribution of the sub lineages across various states in the span of 3 months *i.e,* November 2021, December 2021, and January 2022 [16].

RESULTS

Genomic data

A total of 3429 in-house samples were tested on RT-PCR for COVID-19 from December 2021 to January 2022 from Bengaluru, Karnataka. Among them, 235 samples were tested positive and 192 samples that had CT values between 20-30 were taken for sequencing. The sequence data provided even genome coverage with high quality viral genomes to be reconstructed at 500x average depth coverage with an average of

0.1 million reads. Out of 192, the Pangolin tool classified 133 isolates as omicron variants. In addition, 1586 isolates corresponding to the omicron variant are included from GISAID. Thus for further analysis, we included 1719 isolates. The metadata and variant details are provided in supplementary data [17].

Geospatial distribution

On the 10th of November 2021, the first Omicron isolate was collected and tested from Mumbai, Maharashtra. By the time WHO declared Omicron as the VOC, it had already spread to multiple states like Delhi, Kerala, Gujarat, and Karnataka. As shown in Figure 1A maximum sequences were uploaded in the month of December 2021 out of 29% were classified as Omicron. Then in the month of January up to the 19th of 2022, 84% of the total sequences were classified as Omicron. We can observe the takeover of the omicron variant with respect to other variants. By the beginning of January 2022, the virus had extended its reach to further states like Telangana, Punjab, Andhra Pradesh, etc, and community transmission was seen and the state wise distribution is shown in Figure 1B [18,19].
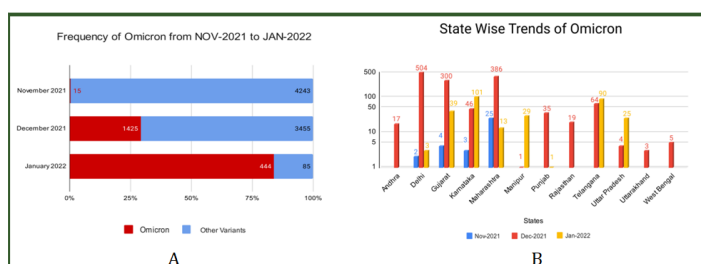

Figure 1: A) Number of samples in different months. The bar represents the total samples in that month and the red color shows the percentage of samples belonging to the omicron variant; B) Number of Omicron samples distributed across different states of India. The height of the bar is proportionate to the number of samples.

Lineage prevalence

Pango lineage B.1.1.529 was the first to be declared as the Omicron VOC. Due to acquisitions of mutations, as they spread globally, sub lineages emerged as BA.1, BA.2, and BA.3.

Until the mid of December 2021, we observed the prevalence of BA.1 sublineage along with B.1.1.529 in different states of India as shown in Figure 2 Karnataka and Delhi were the first states to report this sublineage in the mid of November 2021.

The first case of BA.2 sublineage was found in Delhi on the 16th of December 2021 with a cluster of 22 isolates from a single area. We analyzed the SNP difference within this cluster and found that they had 0-1 SNP, which confirmed it to be a potential outbreak. BA.2 was observed in other states like Gujarat, Telangana, Maharashtra, etc with an incidence of 18% (n/total=312/1719) by December end. N:S413R, S:A27S, S:V213G, ORF1a:L3027F nonsynonymous mutations were specific to the outbreak. However, it was found in less than 30% of the total isolates from India [20].
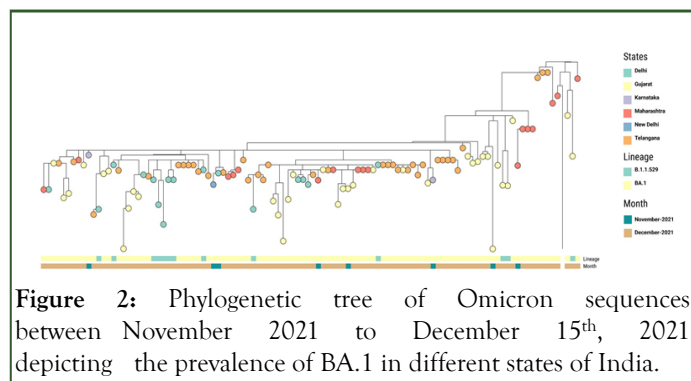

Figure 2: Phylogenetic tree of Omicron sequences between November 2021 to December 15th, 2021 depicting the prevalence of BA.1 in different states of India.

We observed a second cluster from multiple states with an SNP difference of 0-1. The first case in this cluster was from Maharashtra which originated on the 20th of Dec 2021 and spread to Karnataka and Punjab simultaneously. Out of 51 samples in the cluster, 32 of them were from Bengaluru which were out in house samples. The remaining 18 isolates belonged to Punjab where it was observed after 4 days of the first detection. This could potentially mean that there were multiple introductions to Bengaluru and Punjab from other metropolitan cities that drove the BA.2 spread. SNP analysis showed that the predominant deletions such as N:E31-, N:R32-, N:S33-, ORF9b:E27-, ORF9b:N28-, ORF9b:A29- and substitutions N:P13L, ORF9b:P10S were detected only in 2% (1/51) and 17.6% (9/51) respectively (Figure 3).
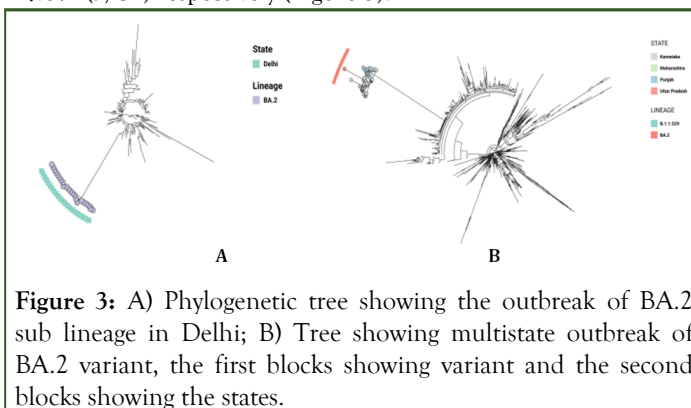

Figure 3: A) Phylogenetic tree showing the outbreak of BA.2 sub lineage in Delhi; B) Tree showing multistate outbreak of BA.2 variant, the first blocks showing variant and the second blocks showing the states.

DISCUSSION

Omicron has quickly spread to become the world's dominant variant of SARS-CoV-2 by the end of December 2021. The Omicron variant belongs to PANGO lineage B.1.1.529, in comparison to the original SARS-CoV-2 strain, this has 30 amino acid modifications, three modest deletions, and one short insertion in the spike protein. In addition to these mutations, a number of substitutions and deletions in other genomic areas are also present in Omicron. Our study showed the presence of specific mutations in the outbreak isolates and the microevolution of the sublineages. The B.1.1.529 lineage was given the name Omicron after it was designated as a variant of concern. Later, it was discovered that this lineage had its own small variations, and the three most common were designated as B.1.1.529.1, B.1.1.529.2, B.1.1.529.3, or simply known as BA.1, BA.2, and BA.3. In the month of January, BA.1 is the prevalent

Omicron sub variant in circulation while there is a significant rise in the proportion of BA.2 simultaneously.

According to the WHO, >98 percent of all Omicron genetic sequences submitted in global databases till January 15 were that of the BA.1 variant. BA.2 holds most of the characteristics of BA.1 but has some additional mutations that can give it a distinguishing feature. Some of the important mutations in BA. 2 that are different from other sub-lineages include the spike protein mutations such as T19I, 24-26 deletion, V213G, T376A, R408S. BA.2 has been discovered in 49 countries, according to the outbreak. Info website, which tracks the prevalence of different lineages of this virus around the world. Since the first sequence was available to the public on the 10th of November 2021, we have seen varied proportions of sublineages circulating in India. As seen in the global trend, India also follows the same route.

Omicron was introduced to India in early November 2021. Reports say that it was first seen in Karnataka but the first sequence collected and submitted to GISAID was on the 10th of November 2021 from Maharashtra. At the beginning of December 2021, sequences of Omicron were 1% of the total sequences from India. Within a month the Omicron replaced all other variants to become 98% of total sequences. At present 100% of the cases are due to Omicron in India. Even though the reports in early January 2022 say that Omicron is in the community transmission stage the data in GISAID shows that the community transmission was seen in mid of December 2021 and was a dominant variant by end of December 2021. Even though in the month of December 2021 the number of omicron cases remained less than other variants it had already extended to most of the cities in India signifying community transmission. While it took a month to increase its frequency to 50% of the total cases in the African continent, Europe, Asia, and the American continent saw 75% of the total cases reported as Omicron in the month of December 2021.

We have BA.1 prevailing in different states with 61% (n=1046/1719) of the total cases analyzed. Following this, BA.2 (n=545/1719) was seen in 37% of the cases. At different time frames, we have noticed the introduction of sublineages in different parts of the country. The first outbreak cluster involved BA.2 introduction in the mid of December 2021 which then spread across different states within a few days. We found that both the outbreak clusters were caused by the BA.2 sub lineage. It indicates that BA.2 is highly variable when compared to BA.1. This is causing higher transmission leading to the community spread of Omicron. During the analysis period, BA.3 sublineage was not found in India.

Even though the omicron sequences were obtained at different time points, the sequencing data pertaining to the route of transmission was scarce. And therefore the results discussed in the present study relate to the number of cases and the sequence data reported during the time of study and not with respect to the epidemiological data perse.

CONCLUSION

The study provides an in depth view of the spread and transmission of the omicron variant in India and showed that community transmission had begun in the mid of December 2021 itself. WGS significantly enhanced understanding of SARS-CoV-2 omicron variant transmission and dynamics of outbreaks, identifying spatial and mutational variation of the same variant and within sub lineages. Examination of the genomic epidemiology of Omicron variant across India showed the multiple introduction events at different time spans which led to the 3rd wave of the COVID-19 pandemic.

REFERENCES

1. Weiss SR, Leibowitz JL. Coronavirus pathogenesis. Adv Virus Res. 2011;81: 85-164.

2. Zhu N, Zhang D, Wang W, Li X, Yang B, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382:727-733.

3. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, et al. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol. 2021;19:409-424.

4. Thakur V, Ratho RK. Omicron (B.1.1.529): A new SARS-CoV-2 variant of concern mounting worldwide fear. J Med Virol. 2022;94:1821-1824.

5. Doucleff M. New evidence shows omicron likely spreads twice as fast as delta in South Africa. NPR. 2021;30:1-2.

6. Cele S, Jackson L, Khoury DS, Khan K, Moyo-Gwete T, Tegally H, et al. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. Nature. 2022;602: 654-656.

7. Meredith LW, Hamilton WL, Warne B, Houldcroft CJ, Hosmillo M, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health care associated COVID-19: A prospective genomic surveillance study. Lancet Infect Dis. 2020;20:1263-1272.

8. Freed NE, Vlkova M, Faisal MB, Silander OK. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. Biol Methods Protoc. 2020;5.

9. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 2019;20:129.

10. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J. The nf-core framework for community curated bioinformatics pipelines. Nat Biotechnol. 2020;38:276-278.

11. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, et al. Nextstrain: Real time tracking of pathogen evolution. J Bioinform. 2018;34:4121-4123.

12. Toole A, Scher E, Underwood A, Jackson B, Hill V. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. 2021;7:64.

13. Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059-3066.

14. Price MN, Dehal PS, Arkin AP. Fast Tree 2 approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5:9490.

15. Argimon S, Abudahab K, Goater RJE, Fedosejev A, Bhai J. Microreact: Visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom. 2016;2:000093.

16. Desingu PA, Nagarajan K. Omicron BA.2 lineage spreads in clusters and is concentrated in Denmark. J Med Virol. 2022;94:23600-2364.

17. Haseltine WA. Birth of the omicron family: BA.1, BA.2, BA.3. Each As Different As Alpha Is From Delta. Forb Mag. 2022;26:42-46.

18. Saxena SK, Kumar S, Ansari S, Paweska JT, Maurya VK, et al. Characterization of the novel SARS-CoV-2 Omicron (B.1.1.529) variant of concern and its global perspective. J Med Virol. 2022;94:1738-1744.

19. Chen C, Nadeau S, Yared M, Voinov P, Xie N. CoV-Spectrum: Analysis of Globally Shared SARS-CoV-2 Data to Identify and Characterize New Variants. Bioinform. 2021.

20. Garg R, Gautam P, Suroliya V, Agarwal R, Bhugra A, et al. Evidence of early community transmission of Omicron (B1.1.529) in Delhi-A city with very high seropositivity and past exposure. Travel Med Infect Dis. 2022;46:102276.