RESEARCH ARTICLE

Genomics of postvaccination SARS-CoV-2 infections during the Delta dominated second wave of COVID-19 pandemic, from Mumbai Metropolitan Region (MMR), India

Kayzad Nilgiriwala[1] 🔵 | Pratibha Kadam[1] | Grishma Patel[1] | Ambreen Shaikh[1] |
Tejal Mestry[1] | Smriti Vaswani[1] | Shalini Sakthivel[1] | Aruna Poojary[2] |
Bhavesh Gandhi[2] | Seema Rohra[2] | Zarir Udwadia[2] | Vikas Oswal[3] |
Daksha Shah[4] | Mangala Gomare[4] | Kalpana Sriraman[1] | Nerges Mistry[1]

[1]The Foundation for Medical Research, Mumbai, India

[2]Breach Candy Hospital (BCH) Trust, Mumbai, India

[3]Vikas Nursing Home, Mumbai, India

[4]Municipal Corporation of Greater Mumbai (MCGM), Mumbai, India

Correspondence
Kayzad Nilgiriwala, The Foundation for Medical Research (FMR), Dr. Kantilal J. Sheth Memorial Bldg, 84-A, R. G. Thadani Marg, Worli, Mumbai 400 018, India.
Email: fmr@fmrindia.org;
Website: www.fmrindia.org

Funding information
Members of the Harvard Business School Alumni Club of India and general donation from Zoroastrian Charity Funds of Hongkong, Canton and Macau

Abstract

The present study was initiated to understand the proportion of predominant variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in postvaccination infections during the Delta dominated second wave of coronavirus disease 2019 (COVID-19) in the Mumbai Metropolitan Region (MMR) in India and to understand any mutations selected in the postvaccination infections or showing association with any patient demographics. Samples were collected ($n = 166$) from severe/moderate/mild COVID-19 patients who were either vaccinated (COVISHIELD/COVAXIN—partial/fully vaccinated) or unvaccinated, from a city hospital and from home isolation patients in MMR. A total of 150 viral genomes were sequenced by Oxford Nanopore sequencing and the data of 136 viral genomes were analyzed for clade/lineage and for identifying mutations. The sequences belonged to three clades (21A, 21I, and 21J) and their lineage was identified as either Delta (B.1.617.2) or Delta+ (B.1.617.2 + K417N) or sub-lineages of Delta variant (AY.120/AY.38/AY.99). A total of 620 mutations were identified of which 10 mutations showed an increase in trend with time (May–October 2021). Associations of six mutations (two in spike, three in orf1a, and one in nucleocapsid) were shown with milder forms of the disease and one mutation (in orf1a) with partial vaccination status. The results indicate a trend toward reduction in disease severity as the wave progressed.

KEYWORDS
epidemiology, genetics, mutation, SARS coronavirus, virus classification

1 | INTRODUCTION

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2)—the causative agent of the novel coronavirus related disease (COVID-19), since its occurrence in late 2019, has been continuously evolving and

leading to the emergence of new variants. In light of the COVID-19 pandemic, the development of various types of vaccines was achieved at a remarkable speed, since vaccines are the most potent weapon for controlling the pandemic.[1] The national COVID-19 vaccination program in India was launched in January 2021, with

approval for the vaccines—COVISHIELD (ChAdOx1-S/nCoV-19) (Serum Institute of India) and COVAXIN (BBV152) (Bharat Biotech). Clinical trials have shown that both vaccines are reported to have >70% efficacy against symptomatic infections.[2] The immune response in postvaccination infections seems to favor vaccine-escape mutations. Several studies in India have reported break-through infections and the prevalence of variants of concern (VOC) in vaccinated and unvaccinated COVID-19 positive individuals.[3] The earliest evidence reported breakthrough infections in 19 out of 113 employees (16.8%) at a noncommunicable disease healthcare facility in Delhi.[4] Another study reported the viral genomic characterization of six viral variants isolated from healthcare workers with break-through infections,[5] whereas the study by Gupta et al.[6] provided evidence that the majority of the breakthrough COVID-19 cases in India were infected with the Delta variant, with only 9.8% cases requiring hospitalization, and 0.4% fatalities, indicating that vaccina-tion is helpful in reducing hospitalization and mortality.

Since the release of the first SARS-CoV-2 genome sequence in January 2020,[7] whole-genome sequencing (WGS) has proved to be a powerful tool to identify genomic characteristics and for the development of newer diagnostics, genomic surveillance, and contact tracing.[8] India experienced the second wave of COVID-19 during April–June 2021, with a peak in mid-May 2021[9] whilst the first case of the Delta (B.1.617.2) was reported in Maharashtra in March 2021,[10] the first case of the Delta variant was detected in the Mumbai Metropolitan Region (MMR) only in May 2021. Moreover, during the peak of the second wave in May 2021 only 2005 complete genomes of the Delta variant of SARS-CoV-2 were reported from India and only four genomes from MMR were reported in the Global Initiative on Sharing All Influenza Data (GISAID) database.[11]

Hence, considering the limited genome sequences from the MMR and to obtain a better representation of viral genome sequences from the MMR towards understanding the prevalence of the predominant VOC in postvaccination SARS-CoV-2 infections, This study was conducted between May and October 2021. Clinico-demographic information from postvaccination SARS-CoV-2 infec-tions in MMR was collected in an attempt to understand the association (if any) of the variants with disease severity, vaccination status (vaccinated/unvaccinated), vaccine type (COVISHIELD/COV-AXIN), and vaccination dose status (partially vaccinated/fully vacci-nated). Phylogenetic and mutation analysis of the variants was performed in the cohort to get insights into the genomic changes in the virus and to understand any longitudinal mutation-specific trends during the Delta dominant second wave.

2 | MATERIALS AND METHODS

2.1 | Patients and sample collection

This study was conducted between May and October 2021 by the Foundation for Medical Research (FMR), in collaboration with the

Municipal Corporation of Greater Mumbai (MCGM), Mumbai, and Breach Candy Hospital (BCH) Trust, Mumbai. Ethical clearance for the study was obtained from the Institutional Ethics Commit-tees at FMR (FMR/IREC/C19/01/2021 and FMR/IREC/C19/02/2021) and BCH (P6/2021). A total of 166 patients were recruited through referrals from study collaborators in the public and private systems, out of which 153 patients were postvaccination COVID-19 cases and 13 patients were unvaccinated. All the COVID-19 cases were confirmed based on reverse-transcription polymerase chain reaction (RT-PCR). Patients were recruited from the MMR, of whom 74 patients were hospitalized (fully vaccinated) and 92 patients were in home isolation (vaccinated with one dose, vaccinated with two doses, or unvaccinated). For hospitalized patients, samples were collected by the tertiary care hospital doctors from Breach Candy Hospital ($n = 71$)—a private hospital and Kasturba Hospital for Infectious Diseases ($n = 3$)—referred by MCGM, and in the case of home isolation partici-pants, the samples were collected by a trained field researcher from cases in MCGM ($n = 88$) and three other Municipal Corporations—Thane ($n = 2$), Kalyan–Dombivali ($n = 2$), and Navi Mumbai ($n = 2$) of MMR. Written informed consent was obtained from all the patients during recruitment regarding the collection of swab samples and patient metadata (Table 1). Naso/oro-pharyngeal swabs ($n = 74$) and nasopharyngeal swabs ($n = 92$) were collected from the patients, and samples were categorized based on disease severity defined as per ICMR guidelines dated May 17, 2021.[12] The vaccinated group had representation from both the adenoviral vaccine (COVISHIELD) and the inactivated whole virus vaccine (COVAXIN). The swabs were collected in a viral transport medium (VTM) and transported at 4°C to the FMR. A detailed case history was recorded at the time of recruitment.

2.2 | RNA isolation and RT-PCR

Viral RNA from the swab samples were isolated with QiaAmp viral RNA mini kit (Qiagen GmBH) as per the manufacturer's protocol. The RT-PCR was carried out in the Bio-Rad CFX96 (Bio-Rad Laboratories), real-time PCR Detection System, and SARS-CoV-2 specific genes (N and ORF1) were detected using the COVIpath™ COVID-19 RT-PCR kit (Applied Biosystems- Invitrogen Bioservices India Pvt. Ltd.) as per the manufacturer's protocol.

2.3 | cDNA synthesis and multiplex PCR

Subsequent to RT-PCR, RNA samples with Ct < 33 (150/166) were subjected to reverse transcriptase PCR to convert the SARS-CoV-2 RNA into cDNA for sequencing, using LunaScript RT SuperMix Kit (Cat no: E3010, NEB) as per manufacturer's protocol. The cDNA were amplified by ARTIC primers (version 3) for 27 initial samples and later by MIDNIGHT primers (Version: PCTR_9125_v110_revB_24-Mar2021) for 123 samples. Two separate PCR amplification reactions

TABLE 1 Patient demographics and clinical parameters

Characteristics	COVISHIELD Vaccinated with one dose*	Vaccinated with two doses*	COVAXIN Vaccinated with one dose*	Vaccinated with two doses*	Unvaccinated	Total cases
	n = 21	n = 80	n = 2	n = 26	n = 7	n = 136
Age (years)						
Median	35	64	24	41	34	54
Gender						
Male	11	48	1	14	4	78 (57%)
Female	10	32	1	12	3	58 (43%)
Disease severity						
Mild	19	62	2	23	7	113
Moderate	2	9	0	2	0	13
Severe	0	9	0	1	0	10
Symptoms						
Fever	12	61	1	21	1	96 (71%)
Cough	18	45	3	20	6	92 (68%)
Cold	5	5	0	3	1	14 (10%)
Chills	1	7	0	9	1	18 (13%)
Shortness of breath or difficulty in breathing	1	17	0	3	0	21 (15%)
Sore throat	7	11	2	7	6	33 (24%)
Loss of taste	6	9	0	9	3	27 (20%)
Loss of smell	6	10	0	12	2	30 (22%)
Headache	3	13	0	2	1	19 (14%)
Muscle aches	6	15	1	6	0	28 (21%)
Weakness	0	10	0	1	0	11 (8%)
Nausea	0	2	0	1	0	3 (2%)
Diarrhea	0	1	0	3	0	4 (3%)
Comorbidities						
Diabetes	0	34	0	5	1	40 (29%)
Hypertension	0	36	0	1	0	37 (27%)
Hypothyroid	0	8	0	0	0	8 (6%)
Cardiovascular disease	0	13	0	2	0	15 (11%)

(with pool-1 and pool-2 primers) for each sample were conducted in case of both protocols.

2.4 | Genome sequencing by Oxford nanopore

The amplified products (from pool-1 and pool-2) of respective samples were confirmed by agarose gel electrophoresis. Samples that showed amplification for both the primer pools were considered optimal for Nanopore library preparation and sequencing. Amplified DNA were subjected to barcoding and adapter ligation using either the rapid barcoding MIDNIGHT protocol or native barcoding ARTIC protocol.[13,14] Quality control of sequencing was done by including one positive control and a no template control during each sequencing run. DNA libraries were sequenced using the SpotON flow cell (FLO-MIN106, Oxford Nanopore Technologies) in a MinION MK1B sequencer using MinKNOW operating software for primary data acquisition (Oxford Nanopore Technologies).

TABLE 2 Distribution of samples based on clinical parameters (disease severity, vaccination status, vaccine type, and vaccine dose status)

Category	Sub-class 1	Sub-class 2	Total
Disease severity	Severe/moderate (n = 23)	Mild (n = 113)	136
Vaccination status	Vaccinated (129)	Unvaccinated (7)	136
Vaccine type	COVISHIELD (n = 101)	COVAXIN (n = 28)	129
Vaccine dose status	Partial vaccinated (23)	Fully vaccinated (106)	129

2.5 | Genomic data analysis

2.5.1 | Raw data processing

Base-calling and demultiplexing were conducted using Guppy (v5.0.17) in high accuracy mode.[15] The resulting.fastq files were normalized by read length. The processed reads were filtered with the *field bioinformatics pipeline* (v1.2.1).[16] Reads were aligned using Minimap2 (v2.17)[17] to the reference genome (MN908947.3). Variants were called using Medaka (v.1.5.0)[18] from the aligned reads and consensus FASTA were created using *samtools* (v1.14).[19] SnpEff (version latest core) was used to annotate the discovered variants with reference strain NC_045512.

2.5.2 | Lineage analysis

The assembled SARS-CoV-2 genomes were assigned lineages using Phylogenetic Assignment of Named Global Outbreak LINeages (PANGOLIN) (v3.1.17)[20] with Ultrafast Sample Placement on Existing Trees (UShER) model (v1.2.121).[21]

2.5.3 | Phylogenetic analysis

The consensus FASTA files from the SARS-CoV-2 were aligned using *MAFFT* (v7.489)[22] and clustered using *Augur* (v13.0.0).[23] Maximum likelihood trees were constructed with default parameters using *IQ-TREE* (v2.1.3)[24] and visualized with *Auspice* (v2.32.1).[25] Nextclade[26] was used to assign clades to the sequences. A secondary tree was generated to compare the sequences from this study with whole-genome sequences from the GISAID database from MMR between Decemebr 2020 and October 2021 the GISAID database from MMR between Decemebr 2020 and October 2021.

2.5.4 | Mutation association analysis

The association of each mutation (including lineage-defining mutations) with clinical parameters such as disease severity (severe/moderate/mild), vaccination status (vaccinated/unvaccinated), vaccine type (COV-ISHIELD/COVAXIN), and vaccination dose (partially/fully vaccinated) were analyzed using the Chi-square test in *GraphPad Prism 6*.

To analyze the association of the mutations with clinical parameters, four categories were considered: disease severity (severe/moderate/ mild), vaccination status (vaccinated/unvaccinated), vaccine type (COV-ISHIELD/COVAXIN), and vaccine dose (partial/fully vaccinated). Since the severe and moderate patients required oxygen supplementation, which indicated extensive lung pathology, The patient samples were grouped into two categories: severe/moderate and mild, to understand the association with disease severity (Table 2). In the data set mentioned above, 620 mutations (including non-lineage and lineage-defining mutations) were present from 136 patient samples. Each mutation was considered a feature in the analysis. Mutations found in less than 5% of patients that may not provide meaningful association were excluded from the analysis—any association of mutations with a clinical parameter having p value < 0.05 was considered statistically significant.

3 | RESULTS

3.1 | Demographics and clinical features of COVID-19 patients

Of the 166 patients recruited, 16 patient samples were excluded due to high Ct values (Ct > 33). Of the 150 samples sequenced, the sequence data of 136 samples were considered for further analysis. Fourteen samples were excluded due to low sequence coverage either because the Ct values were high (n = 13; Ct between 30 and 33) or there was a high number of ambiguous bases were observed (n = 1). All further analyses were carried out on the genomic data of the 136 samples. There were 57% (78/136) males and 43% (58/136) females among the analyzed patients. The patients' ages ranged between 18 and 89 years, with a median age of 54 years. The demographics of 10 severe, 13 moderate, and 113 mild cases are provided in Table 1.

3.2 | Genome sequencing statistics

The sequence reads generated for the samples ranged between 0.1 and 1.17 million reads with a median sequence coverage (depth) of 5327x. The GC content of the genomes ranged between 38% and 40%, and the sequenced genome length ranged from 28 918 to 29 890 bases. The genome coverage ranged from 96% to 100%. A range of 26–56 mutations per genome (SNP/Indels) were identified in the sequenced samples. The number of mutations ranged between 26 and 51 for mild cases, 31–56 for moderate between 31 cases, and 29–48 for severe between 29 and 48 cases.

3.3 | Phylogeny of SARS-CoV-2 genomes

The genomic data were analyzed for viral clades and lineages. The samples were found to be distributed in three clades based on the Nextclade classification, that is, 21A (Delta), 21I (Delta), 21J (Delta). Six samples belonged to Clade 21A defined by positions 21618 (S: T19R), 23403 (S: D614G), 26767 (M: I82T), and 28461 (N: D63G). Thirty-two samples belonged to Clade 21I defined by positions 5184 (nsp3:P1640L), 9891 (nsp4:3209V), 21618 (S: T19R), 22227 (A222V), 23403 (S:D614G), 26767 (M:I82T), and 28461 (N:D63G). Ninety-eight samples belonged to Clade 21J defined by positions 11332 (nsp6:V3689V), 19220 (nsp14:A6319V), 21618 (S: T19R), 23403 (S:D614G), 26767 (M:I82T), and 28461 (S: T19R).

Lineages were determined by PANGOLIN (Usher) interface, where 120 samples were found to be Delta (B.1.617.2), and 16 samples were Delta sub-lineages (AY-series and Delta+) (Table 3). A phylogenetic tree of the Clades 21A, 21I, and 21J is shown in

TABLE 3 Lineage distribution of Delta and its sublineages in the study

Lineage/sublineage	No. of samples
B.1.617.2	120
AY.120	13
AY.38	1
AY.99	1
B.1.617.2 + K417N (Delta+)	1

(Figure 1). An additional phylogenetic tree was generated to compare genomes sequenced in this study with other sequenced genomes from MMR available in the GISAID database (Delta variants from MMR between May and October 2021). Sequences generated in this study and other MMR sequences (from GISAID) showed no grouping or presence of clusters in the phylogenetic profile. The Delta variants for samples collected during May–August 2021 belonged to either Clade 21A, 21I, or 21J, whereas samples from September 2021 onwards belonged exclusively to 21J Clade. Additionally, the variants in Clade 21J showed an increase in the number of mutations compared to those in Clades 21A and 21I.

3.4 | Association of mutations with clinical parameters

A total of 620 mutations (SNP/Indels) were observed in 136 genomes in this study. An increase in the frequency of 10 mutations over time was observed (May–October 2021) (Figure 2) which are as follows: (1) Orf1a: P2287S, (2) Orf1a: T3255I, (3) Orf1a: A1306S, (4) Orf1a: P2046L, (5) Orf1a: T3646A, (6) Orf1b: A6319V and (7) Orf7b: T40I, (8) S: G142D, (9) S: T95I, and (10) N: G215C. These mutations showed a progressive increase in frequency (since Apr 2021), as also seen in the genome sequences available on the open-source database of COVID-19 resources and epidemiology data (https://outbreak.info/) (Figure 2). In this study, these mutations were prevalent in over 70% of Delta variants, including Delta sub-lineages.

To analyze the association of the mutations with clinical parameters, four categories were considered: disease severity (severe/moderate/
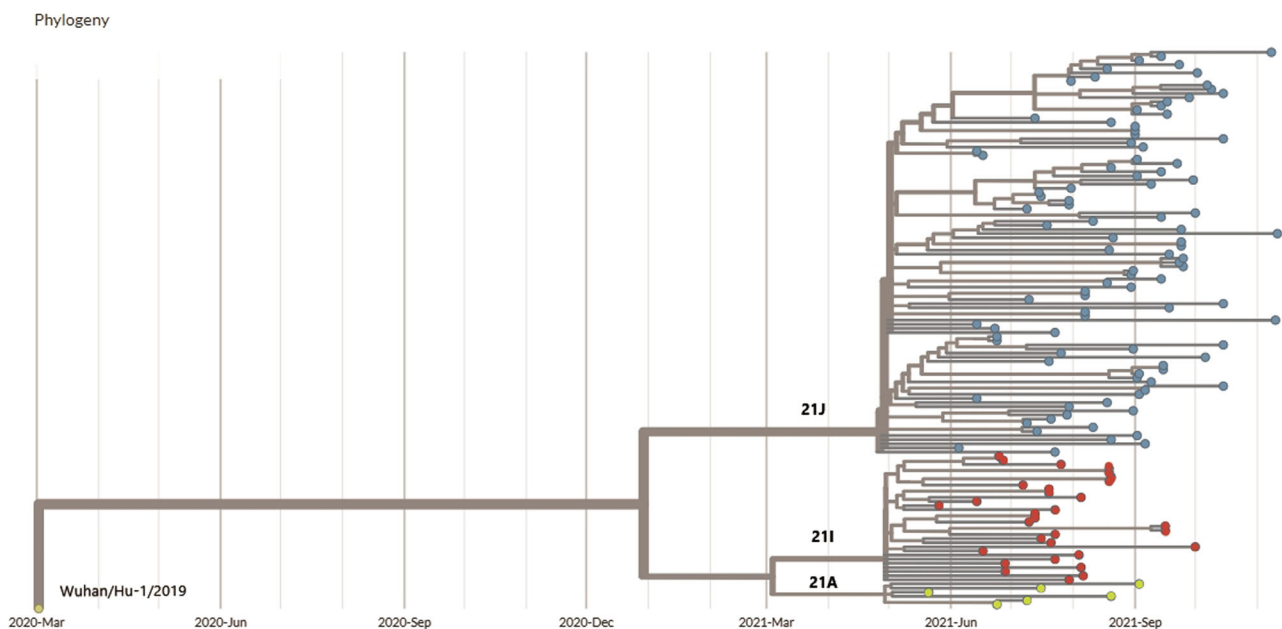

FIGURE 1 Phylogenetic tree representing 136 viral genome sequences from the study. The Clade 21A (parrot green), 21I (red), and 21J (blue) are shown in a phylogenetic tree. The sequences from May to August 2021 belonged to either Clade 21A, 21I, or 21J, and September 2021 onwards belonged to Clade 21J.
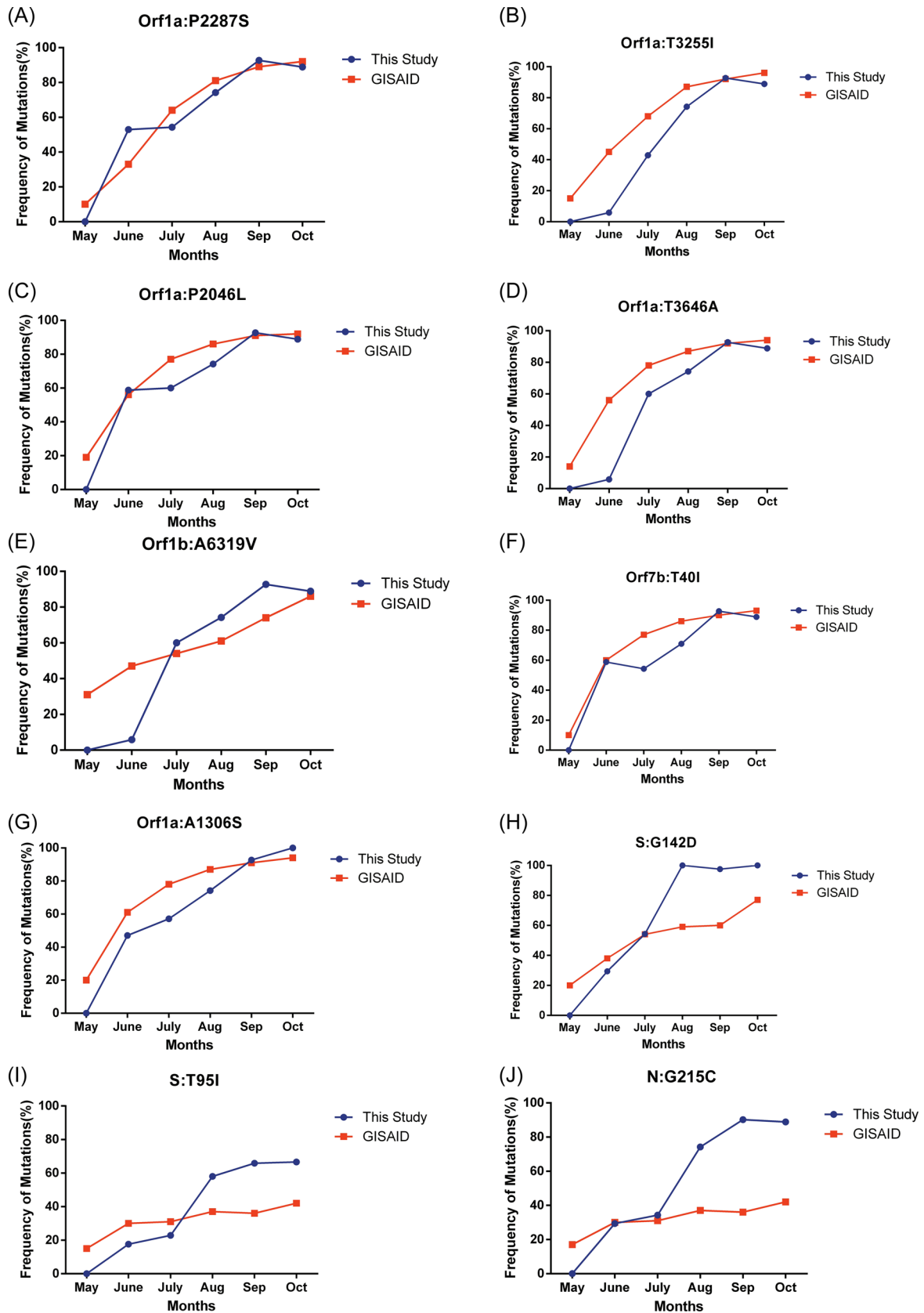
FIGURE 2 Mutations showing an increase in frequency from May to October 2021. (A) Orf1a: P2287S, (B) Orf1a: T3255I, (C) Orf1a: P2046L, (D) Orf1a: T3646A, (E) Orf1b: A6319V, (F) Orf7b: T40I, (G) Orf1a: A1306S, (H) S: G142D, (I) S: T95I, and (J) N: G215C. The graphs were generated using GraphPad Prism 6. The trendline in blue indicates the proportion of the respective mutations from this study, whereas the red trendline indicates the proportion of the respective mutations in the GISAID repository.

TABLE 4 Mutations showing association with disease severity and vaccine dose

Association of mutations with in severe/moderate versus mild cases					
Mutation	Locus	Type of mutation	Severe/moderate (n = 23)	Mild (n = 113)	p value
21618|C|G	S	T19R	15 (65.2%)	102 (90.2%)	0.0045
212987|G|A	S	G142D	13 (56.5%)	93 (82.3%)	0.0117
7124|C|T	ORF1a	P2287S	12 (52.2%)	85 (75.2%)	0.0409
10029|C|T	ORF1a	T3255I	9 (39.1%)	83 (73.5%)	0.0027
7646|G|A	ORF1a	G2461S	7 (30.4%)	68 (60.2%)	0.0114
28881|G|T	N	R203M	15 (65.2%)	99 (87.6%)	0.0135
Association of mutations with partial versus fully vaccinated patients					
Mutation	Locus	Type of mutation	Partial vaccinated (n = 23)	Fully vaccinated (n = 106)	p value
7646|G|A	ORF1a	G2461S	20 (87%)	50 (47.2%)	0.0004

Note: The shaded mutations showed an increase in frequency over time (refer Figure 3).

mild), vaccination status (vaccinated/unvaccinated), vaccine type (COVISHIELD/COVAXIN), and vaccine dose (partial/fully vaccinated). In the data set, as mentioned above, 620 mutations were present from 136 patient samples, and each mutation was considered a feature in the analysis.

A comparison between the presence of mutations with the disease severity categories (severe/moderate vs. mild) showed an association of six mutations (2 in S gene, 3 in orf1a, and 1 in N gene) (Table 4) with mild cases (p value < 0.05). Three mutations (1 in spike—G142D and 2 in orf1a—P2287S and T3255I) out of the six mutations showing association with the mild cases were also observed to have an increased frequency over time (between May and October 2021) (Figure 2), and with two mutations found to be lineage defining for Delta (Table 4).

A comparison between the presence of mutations in partially vaccinated and fully vaccinated patients showed an association with a single mutation in orf1a with partially vaccinated patients (p value < 0.001) (Table 4). A comparison between vaccinated and unvaccinated patients and between vaccine types (COVISHIELD vs. COVAXIN) did not show any statistically significant association.

4 | DISCUSSION

Overall it is evident that COVID-19 is more lethal in unvaccinated people than those fully vaccinated, with vaccines providing the necessary protection from the causative virus.[27] Although vaccines can help reduce the severity of the disease in patients, they do not stop infection and do not reduce transmission. With the ever-changing virus during a pandemic, it becomes necessary to understand the pattern of its transformation, about vaccination rates, and disease severity. The world has seen a repertoire of VOCs during this pandemic leading to the realization of the important role of different

variants in manifesting varying levels of disease severity and transmission, creating abrupt surges (or waves) in the number of cases in turn increasing transmissions in various countries. To get insights into the newly acquired mutations in Delta variants from MMR, The mutations were screened that showed an increase in trend from May to October 2021. It was found that 10 mutations showed increased frequency with time. The comparison of mutations with the Mumbai sequencing data (available on the GISAID) indicated that these mutations appear to stabilize over time (Figure 2) and may have potential implications in tuning transmission and infection levels of the virus, possibly leading to the generation of milder viral variants in the future.

It was found that the association of six mutations with mild cases (Table 4); three of those mutations were also observed to have an increased frequency with time (between May and October 2021) (Figure 2). Two mutations in the spike protein (T19R and G142D) (Figure 3) were associated with mild disease. The spike mutation T19R is a lineage-defining mutation found in most Delta variants. In contrast, the G142D mutation has been observed in 49% Delta variants and 69% Delta Plus variants in the global database.[28] It was found 117/136 (86%) samples with co-occurrence of the T19R and G142D mutations in the spike; these mutations are known to change the supersite epitope that binds the N-terminal domain directed antibodies leading to immune evasion.[28,29] Moreover, 41% of the samples had co-occurrence of T95I and G142D mutations in the spike. The co-occurrence of these mutations is reported to increase the viral load significantly.[30] The increase in viral load due to increased transmission fitness or immune escape may possibly lead to a tradeoff with disease severity leading to a milder disease. The presence of orf1a: T3255I, S: T95I, and S: G142D mutations in all the Omicron (B.1.1.529) variants—a variant supposed to lead to a milder disease reiterates the association of these mutations with reduced severity.[31,32]
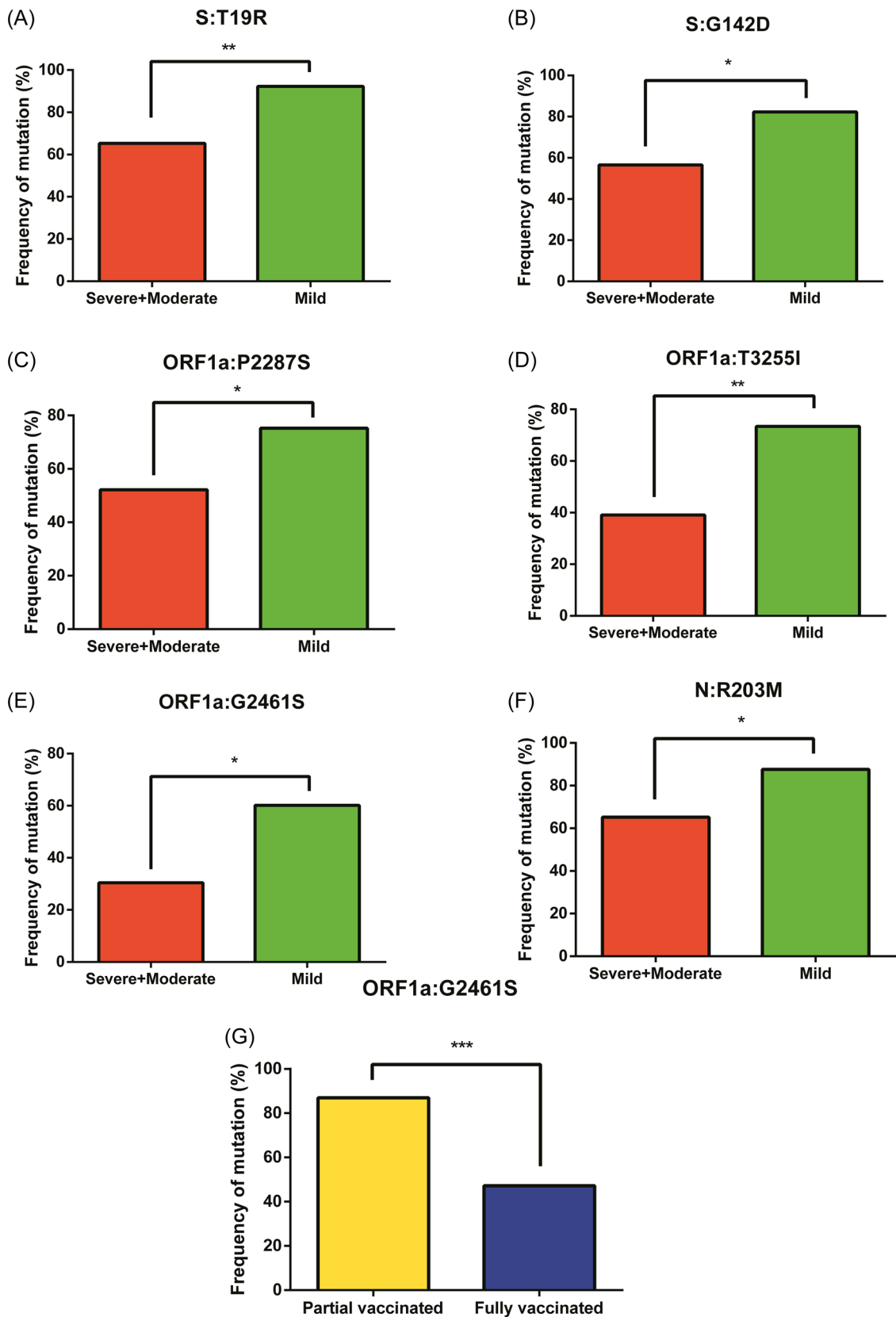
FIGURE 3 Association of mutations with disease severity and vaccination status (partially/fully vaccinated). Six mutations showed association with mild cases: (A) S: T19R, (B) G142D, (C) Orf1a: P2287S, (D) Orf1a: T3255I, (E) orf1a: G2461S, and (F) N: R203M. One mutation (G) orf1a: G2461S showed association with partially vaccinated cases.

It was also found that an association of three mutations in orf1a (P2287S, T3255I, and G2461S) with the mild disease, of which two mutations in orf1a (P2287S and T3255I) were also observed to increase with time (Figures 2 and 3). The orf1a: P2287S mutation is present in the nsp3 PLpro papain-like protease whose function is to block host innate immune response and promote cytokine expression.[33] The increasing frequency of orf1a: P2287S mutation and its association in mild cases could possibly indicate an enhanced host immune response and reduced cytokine expression (potentially mitigating cytokine storm), leading to less severe disease. Another mutation in orf1a associated with mild disease in nsp3 PLpro papain-like protease (orf1a: G2461S) could possibly play a similar role in causing less severe disease by enhancing host immune response and mitigation of cytokine storm. Interestingly, the orf1a: G2461S mutation also showed association in partially vaccinated cases indicating selection of the mutation after the first dose of vaccine. However, the limitation of the meager number of partially vaccinated patients (n = 23) in this study should be noted before drawing any conclusion. The third mutation in orf1a (T3255I) is present in the nsp4 protein, whose function is to produce double-membrane vesicles required to form a replication-transcription complex.[33] This mutation may play a role in reducing the formation of active replication-transcription complexes in the host cell, in turn reducing the viral load and explaining its association with the milder disease form.

In contrast, another mutation in the nucleocapsid protein (N: R203M) was observed to associate with the milder form of the disease (Figure 3). It has already been reported that the N: R203M mutation leads to increased packaging of the viral RNA genome producing a 50-fold higher viral load.[34] The association of the N: R203M mutation with the milder disease could possibly explain the increase in viral load due to an increase in variant transmission fitness, leading to a tradeoff with disease severity. The increase in the frequency of the mutations orf1a (T3255I) (leading to a potential decrease in viral load) and N: R203M (leading to a possible increase in viral load) in the recent Delta variants and their association with milder disease indicates their significant role in determining the transmission fitness of the newer variants and tradeoff with disease severity.

A continued global effort towards ever-expanding viral genome sequencing will help to further the knowledge of potential novel mutations of interest and to understand the evolution of the virus regarding its behavior in transmission and epidemiology.

AUTHOR CONTRIBUTIONS

Conceptualization: Kayzad Nilgiriwala, Aruna Poojary, Zarir Udwadia, Vikas Oswal, Kalpana Sriraman, and Nerges Mistry. *Data Curation*: Kayzad Nilgiriwala, Pratibha Kadam, Grishma Patel, Tejal Mestry, Ambreen Shaikh, Smriti Vaswani, Aruna Poojary, Bhavesh Gandhi, Seema Rohra, and Kalpana Sriraman. *Formal Analysis*: Kayzad Nilgiriwala, Pratibha Kadam, Ambreen Shaikh, Smriti Vaswani, and Kalpana Sriraman. *Funding Acquisition*: Kalpana Sriraman and Nerges Mistry. *Investigation*: Pratibha Kadam, Grishma Patel, Tejal Mestry, Ambreen Shaikh, Smriti Vaswani, Shalini Sakthivel, Bhavesh Gandhi, and Seema Rohra. *Methodology*: Kayzad Nilgiriwala, Aruna Poojary, Kalpana Sriraman, and Nerges Mistry. *Project Administration*: Kayzad Nilgiriwala, Pratibha Kadam, Grishma Patel, Tejal Mestry, Ambreen Shaikh, Aruna Poojary, Kalpana Sriraman, and Nerges Mistry. *Resources*: Kayzad Nilgiriwala, Pratibha Kadam, Aruna Poojary, Zarir Udwadia, Vikas Oswal, Daksha Shah, Mangala Gomare, and Nerges Mistry. *Software*: Kayzad Nilgiriwala and Pratibha Kadam. *Supervision*: Kayzad Nilgiriwala, Ambreen Shaikh, Aruna Poojary, Zarir Udwadia, Daksha Shah, Kalpana Sriraman, and Nerges Mistry. *Validation*: Kayzad Nilgiriwala, Pratibha Kadam, Grishma Patel, Tejal Mestry, Ambreen Shaikh, and Smriti Vaswani. *Visualization*: Kayzad Nilgiriwala, Pratibha Kadam, Grishma Patel, and Tejal Mestry. *Writing—Original Draft Preparation*: Kayzad Nilgiriwala, Pratibha Kadam, Grishma Patel, and Tejal Mestry. *Writing—Review and Editing*: Kayzad Nilgiriwala, Pratibha Kadam, Grishma Patel, Tejal Mestry, Ambreen Shaikh, Smriti Vaswani, Aruna Poojary, Zarir Udwadia, Daksha Shah, Kalpana Sriraman, and Nerges Mistry.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in SRA at https://www.ncbi.nlm.nih.gov/sra, reference number PRJNA813693. The genomic data generated in this study can be found and downloaded under GISAID accession numbers (EPI_ISL_8164107 - EPI_ISL_8164235 and EPI_ISL_8173172 - EPI_ISL_8173178) from the GISAID repository[11]; SRA accession numbers (SRR18426699 - SRR18426834); and BioProject accession number PRJNA813693.

ORCID

Kayzad Nilgiriwala http://orcid.org/0000-0002-5458-4916

REFERENCES

1. Pollard AJ, Bijker EM. A guide to vaccinology: from basic principles to new developments. *Nat Rev Immunol*. 2021;21(2):83-100.
2. Sharma K, Koirala A, Nicolopoulos K, Chiu C, Wood N, Britton PN. Vaccines for COVID-19: where do we stand in 2021? *Paediatr Respir Rev*. 2021;39:22-31.
3. Thangaraj JWV, Yadav P, Kumar CG, et al. Predominance of delta variant among the COVID-19 vaccinated and unvaccinated individuals, India, May 2021. *J Infect*. 2022;84(1):94-118.

4. Tyagi K, Ghosh A, Nair D, et al. Breakthrough COVID19 infections after vaccinations in healthcare and other workers in a chronic care medical facility in New Delhi, India. *Diab Metab Syndr*. 2021;15(3): 1007-1008.

5. Philomina J B, Jolly B, Jolly B, et al. Genomic survey of SARS-CoV-2 vaccine breakthrough infections in healthcare workers from Kerala, India. *J Infect*. 2021;83(2):237-279.

6. Gupta N, Kaur H, Yadav PD, et al. Clinical characterization and Genomic analysis of COVID-19 breakthrough infections during second wave in different states of India. *Viruses*. 2021;13(9):1782.

7. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269.

8. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nature Med*. 2020;26(4):450-452.

9. Yang W, Shaman J. COVID-19 pandemic dynamics in India, the SARS-CoV-2 Delta variant, and implications for vaccination. *medRxiv*. Published online November 21, 2021.

10. Cherian S, Potdar V, Jhadav S, et al. Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. Cold Spring Harbor Laboratory. Published online April 24, Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. Cold Spring Harbor Laboratory. Published online April 24, 2021.

11. GISAID Database. Available from: https://www.gisaid.org

12. Clinical Guidance for Management of Adult COVID-19 patients (ICMR). 2021.

13. Quick J nCoV-2019 sequencing protocol v3 (LoCost) V.3. 2020; Available from: https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye.

14. Nilgiriwala K, Mandal A, Patel G, et al. Genome Sequences of Five SARS-CoV-2 Variants from Mumbai, India, Obtained by Nanopore Sequencing. *Microbiol Resour Announc*. 2021;10(15):e00231-21.

15. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol*. 2019;20(1):129.

16. The ARTIC field bioinformatics pipeline. Available from: https://github.com/artic-network/fieldbioinformatics.

17. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)*. 2018;34(18):3094-3100.

18. Nanopore/medaka. Available from: https://github.com/nanoporetech/medaka.

19. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*. 2009;25(16): 2078-2079.

20. O'Toole Á, Scher E, Underwood A, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol*. 2021;7(2):veab064.

21. Turakhia Y, Thornlow B, Hinrichs AS, et al. Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. *Nature Genet*. 2021;53(6):809-816.

22. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772-780.

23. Huddleston J, Hadfield J, Sibley TR, et al. Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. *J Open Source Softw*. 2021;6(57):2906.

24. Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol*. 2016;65(6):997-1008.

25. McBroome J, Thornlow B, Hinrichs AS, et al. A daily-updated database and tools for comprehensive SARS-CoV-2 mutation-annotated trees. *Mol Biol Evol*. 2021;38(12):5819-5824.

26. Aksamentov I, Roemer C, Hodcroft E, Neher R. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *J Open Source Softw*. 2021;6(67):3773.

27. Dyer O. Covid-19: Unvaccinated face 11 times risk of death from delta variant, CDC data show. *BMJ*. 2021;374:n2282.

28. Kannan SR, Spratt AN, Cohen AR, et al. Evolutionary analysis of the Delta and Delta Plus variants of the SARS-CoV-2 viruses. *J Autoimmun*. 2021;124:102715.

29. Planas D, Veyer D, Baidaliuk A, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature*. 2021;596(7871):276-280.

30. Shen L, Triche TJ, Bard JD, Biegel JA, Judkins AR, Gai X. Spike protein NTD mutation G142D in SARS-CoV-2 Delta VOC lineages is associated with frequent back mutations, increased viral loads, and immune evasion. medRxiv. Published online September 15, 2021.

31. Micheli V, Bracchitta F, Rizzo A, et al. First identification of the new SARS-CoV-2 Omicron variant (B.1.1.529) in Italy. *Clin Infect Dis*. 2022

32. Madhi SA, Kwatra G, Myers JE, et al. Population Immunity and Covid-19 Severity with Omicron Variant in South Africa. *N Engl J Med*. 2022;386:1314-1326.

33. Perez-Gomez R. The development of SARS-CoV-2 Variants: the gene makes the disease. *J Dev Biol*. 2021;9(4):58.

34. Syed AM, Taha TY, Tabata T, et al. Rapid assessment of SARS-CoV-2–evolved variants using virus-like particles. *Science*. 2021;374(6575):1626-1632.