medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Genomic characterization and Epidemiology of an emerging SARS-CoV-2 variant in Delhi, India

Mahesh S Dhar[1]*, Robin Marwal[1]*, Radhakrishnan VS[1]*, Kalaiarasan Ponnusamy[1]*, Bani Jolly[2,3]*, Rahul C. Bhoyar[2,3]*, Saman Fatihi[2,3], Meena Datta[1], Priyanka Singh[1], Uma Sharma[1], Rajat Ujjainia[2], Salwa Naushin[2,3], Nitin Bhateja[2], Mohit Kumar Divakar[2,3], Viren Sardana[2,3], Manoj K Singh[1], Mohamed Imran[2,3], Vigneshwar Senthivel[2,3], Ranjeet Maurya[2,3], Neha Jha[2], Priyanka Mehta[2], Mercy Rophina[2,3], Arvinden VR[2,3], Urmila Chaudhary[1], Lipi Thukral[2,3], Rajesh Pandey[2,3], Debasis Dash[2,3], Mohammed Faruq[2,3], Hemlata Lall[1], Hema Gogia[1], Preeti Madan[1], Sanket Kulkarni[1], Himanshu Chauhan[1], Shantanu Sengupta[2,3], Sandhya Kabra[1], The Indian SARS-CoV-2 Genomics Consortium (INSACOG), Sujeet K Singh[1], Anurag Agrawal[2,3#], Partha Rakshit[1#]

[1]National Centre for Disease Control, Delhi, India
[2]CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India
[3]Academy for Scientific and Innovative Research, Ghaziabad, Uttar Pradesh, India
*Equal contribution; [#] Corresponding Author

Abstract: In April 2021, after successfully enduring three waves of the SARS-CoV2 pandemic in 2020, and having reached population seropositivity of about 50%, Delhi, the national capital of India was overwhelmed by the fourth wave. Here, we trace viral, host, and social factors contributing to the scale and exponent of the fourth wave, when compared to preceding waves, in an epidemiological context. Genomic surveillance data from Delhi and surrounding states shows an early phase of the upsurge driven by the entry of the more transmissible B.1.1.7 variant of concern (VOC) into the region in January, with at least one B.1.1.7 super spreader event in February 2021, relatable to known mass gatherings over this period. This was followed by seeding of the B.1.617 VOC, which too is highly transmissible, with rapid expansion of B.1.617.2 sub-lineage outpacing all other lineages. This unprecedented growth of cases occurred in the background of high seropositivity, but with low median neutralizing antibody levels, in a serially sampled cohort. Vaccination breakthrough cases over this period were noted, disproportionately related to VOC in sequenced cases, but usually mild. We find that this surge of SARS-CoV2 infections in Delhi is best explained by the introduction of a new highly transmissible VOC, B.1.617.2, with likely immune-evasion properties; insufficient neutralizing immunity, despite high seropositivity; and social behavior that promoted transmission.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

After escaping relatively unscathed during the first wave of the COVID-19 pandemic, and after a brief respite from November 2020 until Feb 2021, India witnessed a ferocious second Covid-19 wave, accounting for about half the cases worldwide in the first week of May 2021. SARS-CoV-2 had spread widely throughout India in the first wave, with initial results from the Indian Council of Medical Research third national serosurvey finding 21.4% of adults and 25.3% of 10 to 17 year children to be seropositive. Serosurveys carried out by State / Municipal authorities in Indian megacities had showed even higher seropositivity with the national capital, Delhi, reporting 56% seropositivity in February, 2021. Yet this seems to have offered little protection, with Delhi amongst the most affected cities in April 2021. Relatively well collected multi-modal data for Delhi, including continued genomic surveillance since April 2020 makes this an intriguing case study. Here, we examine the possible reasons why; dissecting the viral, host, and social factors contributing to the scale and exponent of the pandemic upsurge that totally overwhelmed the healthcare infrastructure of the national capital.

Results

Epidemiological characteristics of sequential SARS-CoV-2 outbreaks in Delhi

Records of all tests performed, positive cases, positivity rate and deaths, were accessed through the Integrated Disease Surveillance Programme (IDSP, NCDC) database. Since the first case of Covid-19 was detected in Delhi in March 2020, the city has experienced ups-and-downs in the number of cases and the positivity rates, with four distinct waves, in June, September, and November 2020, and now in April 2021. The first and the third waves were associated with spikes in mortality and stressed healthcare systems but were successfully managed with social restrictions and use of micro-containment strategies focusing on test, trace, treat. After touching the high of almost 9000 cases daily and a positivity rate of about 15% during the third wave in

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

November 2020, new cases steadily declined, with only 1% positivity between December 2020 and March 2021. This reversed and started increasing from third week of March, 2021, shooting to 30% by end of April, with almost 30,000 cases being reported per day. Fig 1 shows weekly cases, positives and positivity rate for 2021 leading up to and including the April 2021 surge. Further analysis focuses on the contributing factors.

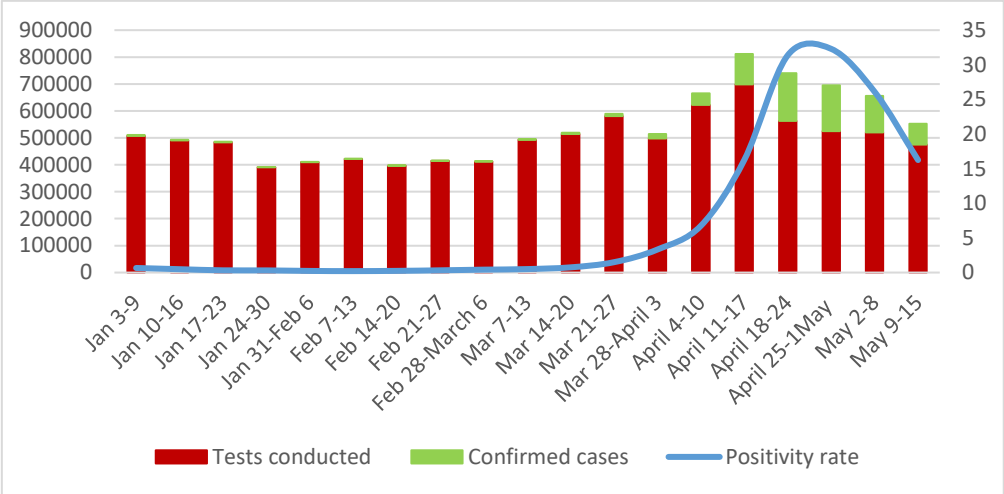


Figure 1. Weekly test numbers and positive fraction (thousands), plotted alongside positive rate since January 2021.

Seropositivity in Delhi

We have previously shown in a cohort-based serosurvey that high seropositivity was associated with declining test positivity rate across India in September, but there is a 20-30% decline in neutralizing activity over a 6 month period (Naushin et al., 2021). A follow-up study of the same cohort was conducted in January and February 2021, with semi-quantitative anti-nucleocapsid, quantitative anti-spike, and a neutralization surrogate assay being measured. 473 of 1115 employees were seropositive across ten Delhi sites where we have serial data. When compared to the prior serosurvey in August/September 2020, that seropositivity had increased across these sites in Delhi (Figure 2). The seropositivity was unevenly distributed, with samples collected from laboratories and public-office employees having much higher seropositivity

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

than samples collected from family members living at residential complexes. The median level of anti-spike antibodies in this sample from Delhi was well below 132 U/ml, the cutoff for convalescent plasma titres, suggesting that the protection against infection may be limited to a subset of the sero-positive fraction (Figure 2B) (USFDA, 2021). Further, in 35 sero-positive cases from 2020 under follow-up, retested in April 2021 and with three or more time-points since 2020, three showed a greater than 100 U and more than 50% increase in antibody levels after a decline, suggestive of possible reinfections and limited immunity.

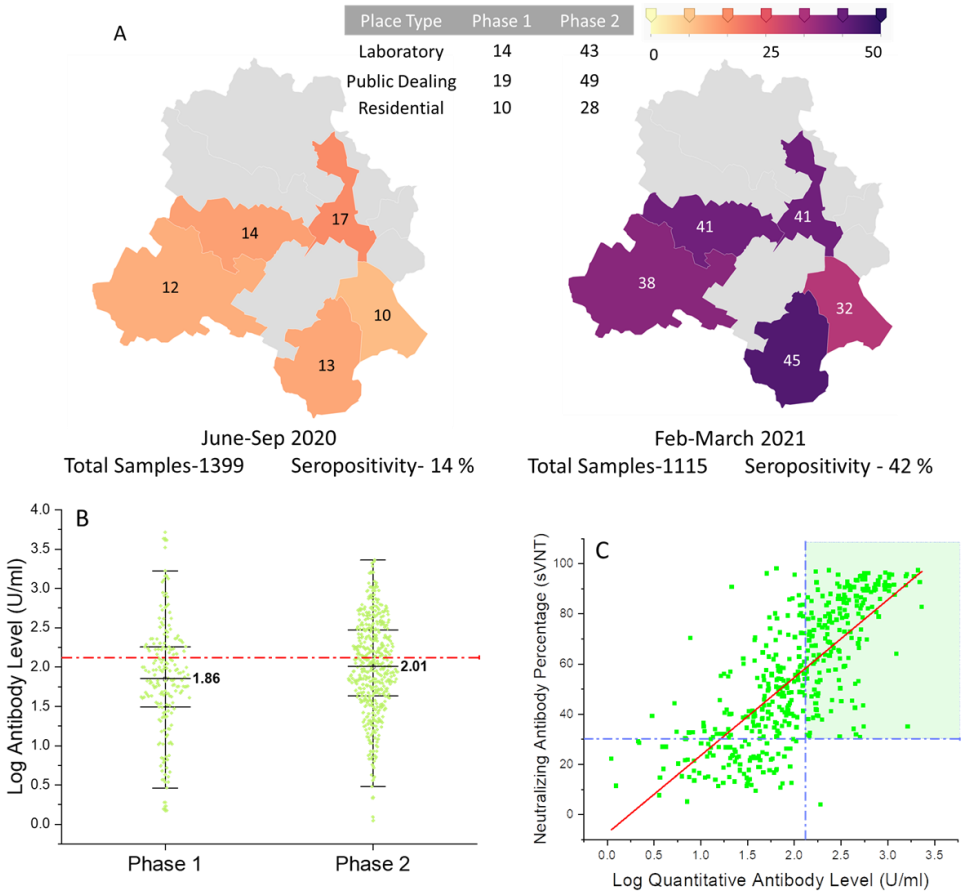


Figure 2. A) Seropositivity rates in Delhi in May-September 2020 and Feb-March 2021 of CSIR-Cohort (Map Created using Datawrapper). B) Quantitative Antibody Levels (U/ml) in Delhi during Phase 1 and Phase 2 of CSIR-Cohort. Red dash dot line denotes convalescent plasma titres considered to be sufficient for immunity: 132 U/ml C) Log Antibody Levels (x-axis) against surrogate Neutralizing Antibody Percentage (y-axis) in Delhi from Phase 2 data of CSIR-Cohort (Pearson's r - 0.74). Blue dash dot line along x-axis denotes 30% neutralization (cut-off) and along y-axis denotes plasma titres considered to be sufficient for immunity: 132 U/ml

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Genomic character of the SARS-CoV-2 outbreak in Delhi

To determine whether SARS-CoV2 variants may be responsible for the April 2021 outbreak in Delhi, we sequenced and analyzed the community samples from Delhi from the previous outbreak in November 2020 until May 2021 and related it to effective reproductive number (R_t). It can be seen in Fig 3A that detection of B.1.1.7 VOC was minimal until January 2021, increasing to about 20% in February and 40% in March 2021. This was associated with slight increase in R_t from about 1.2 to 1.6. B.1.617 lineages increased from below 5% in Feb 2021 to about 10% in March before overtaking B.1.1.7 in April and rising to about 60% of all samples. The sub-lineage B.1.617.2, in which the escape mutation E484Q is lost and a new mutation T478K is gained, showed the maximum rise going from less than 10% of B.1.617 to about 80%. The rise of B.1.617, more specifically B.1.617.2, was paralleled by a large increase in positivity rate and increase in R_t to about 2 (Fig. 3B). We analyzed the Ct values of random positive samples tested on a single high-throughput automated machine (Roche 6800) at a single center (NCDC) over this period to understand whether there was any evidence of the VOC being associated with higher viral load and thereby infectivity. As seen in Fig. 3B, when compared to the stable baseline from Dec 2020 to Feb 2021, Ct values dropped (99% CI) in March when B.1.1.7 was dominant, falling further in April (99% CI), with the rise of B.1.617 (Fig. 3C). The total change of about -6 Ct corresponds to about 50-fold increase in viral load. The B.1.1.7 variant has been previously reported to be 43 to 90% more transmissible than prior variants as the carriers are more infectious due to the increased viral load (Davies et al., 2021; Volz et al., 2021). Interestingly, the case fatality rate (CFR) was significantly reduced during the surge period in Delhi (95% CI). The R_t was hovering near 1 from December to February but increased significantly from late March to mid-April (95% CI). The CFR values, which were constant from December till February, 2021, witnessed a significant decrease with a simultaneous increase in R_t from March (Fig. 2A). This may be due to the sudden decrease of

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

B.1.1.7 in the Delhi region which reportedly causes high CFR (Davies et al., 2021). The CFR was rising again towards the end of the period. Since CFR may be due to an amalgamation of multiple factors, including the short-term collapse of the healthcare system, there is currently no clear evidence linking B.1.617.2 to change in CFR. This will be reexamined once final information is available on all deaths and COVID-related deaths over this period.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

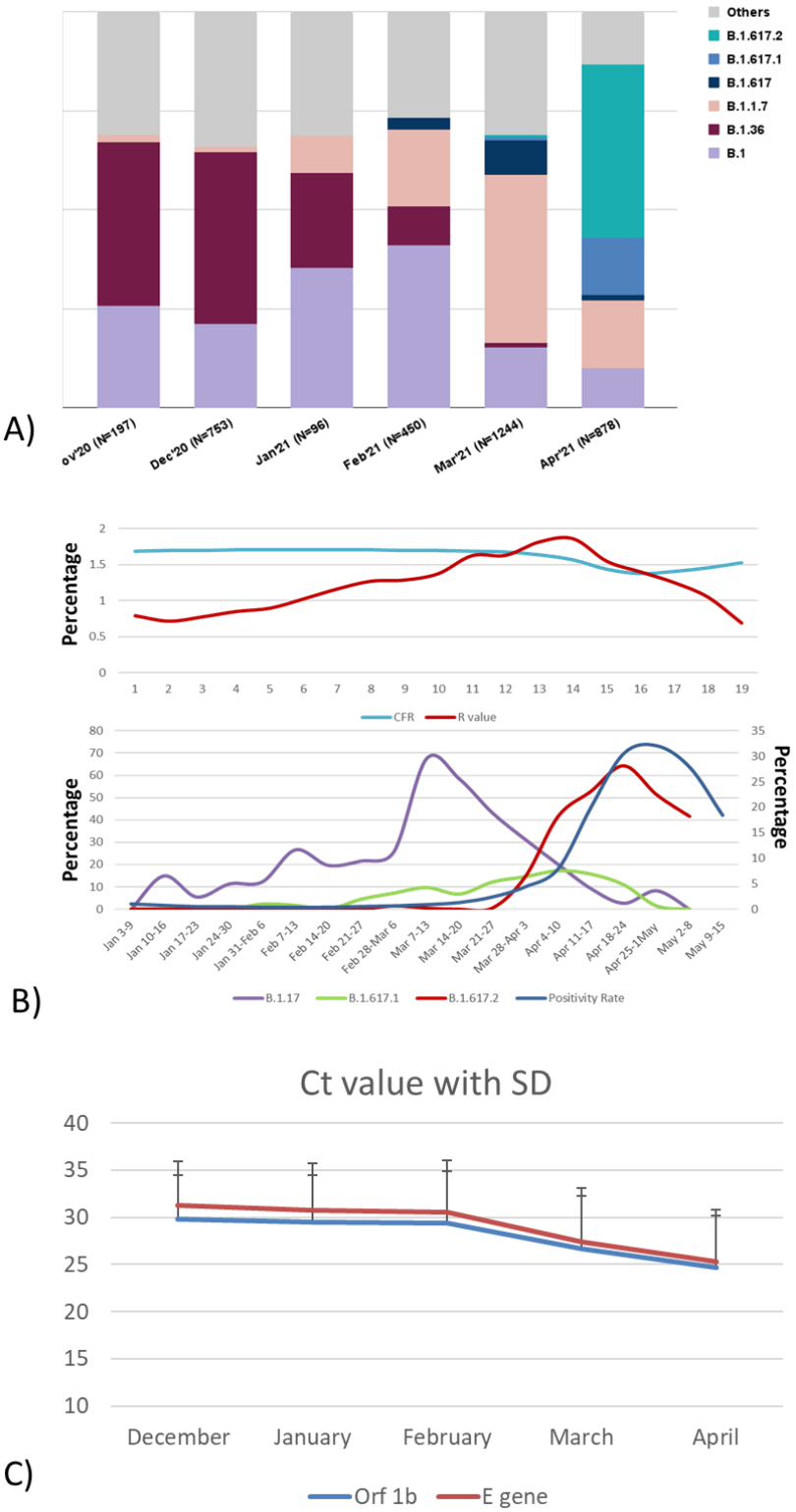


Fig 3. A) Normalized stacked bar graphs of main lineages in Delhi since the third wave. B) Comparison of the percent change of cases of B.1.1.7 and B.1.617.2 variants in Delhi in 2021 with positivity rate, CFR and R_t. C) Ct value of representative samples for each month since Dec 2020. The months of March and April show significant reduction in the Ct values, when compared to the last three months (P<0.01).

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

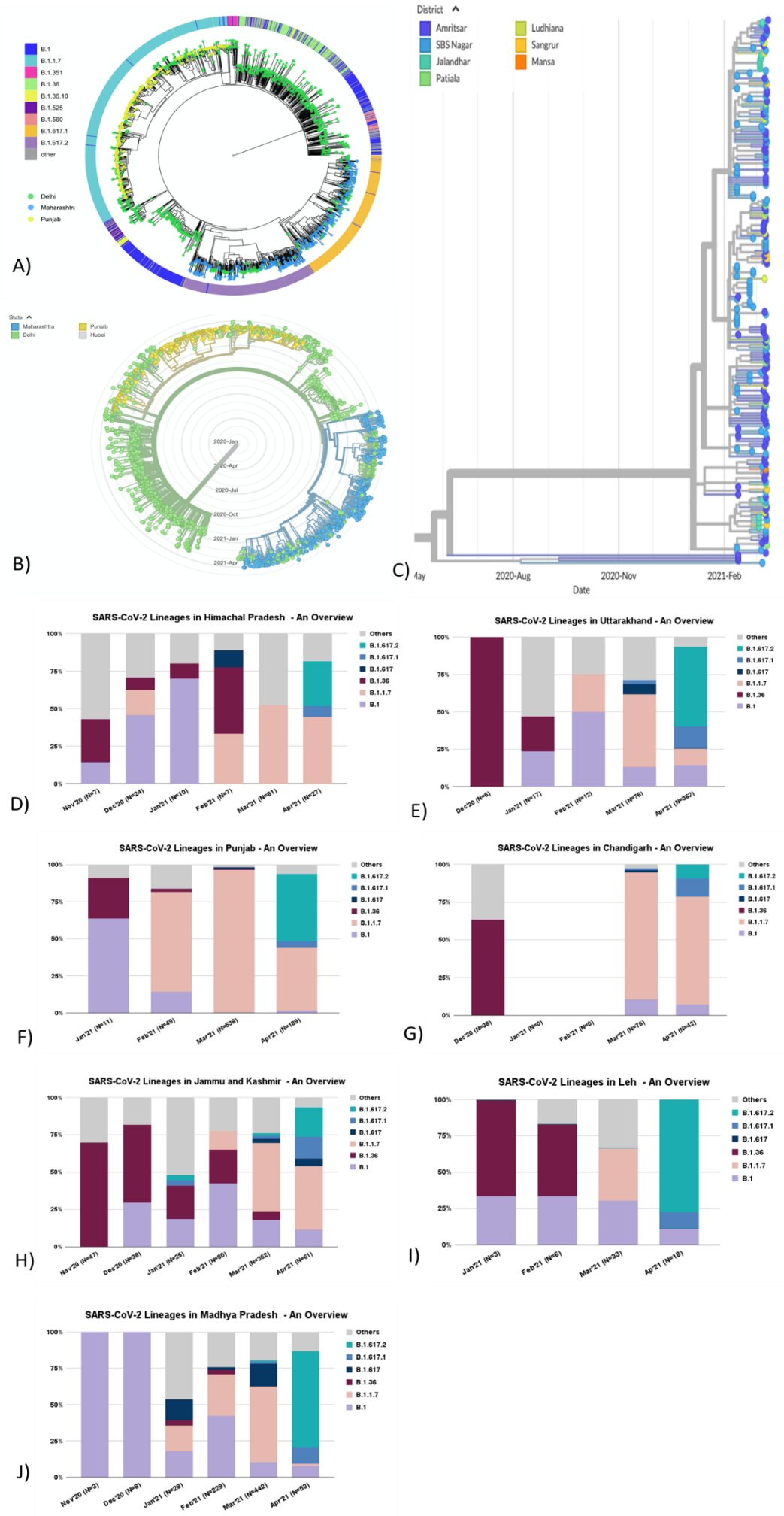
The origin of SARS-CoV-2 outbreaks in North India in 2021

The April 2021 outbreak in Delhi was preceded by outbreaks in the states of Kerala, Maharashtra and Punjab. While no VOC was identified in Kerala in Jan 2021, the outbreak in Maharashtra has been related to B.1.617.1 and in Punjab to the introduction of B.1.1.7. These were found to be phylogenetically related, with a strong phylogenetic connection between Delhi and Punjab for B.1.1.7, and between Delhi and Maharashtra for B.1.617 lineages, as shown in Figure 4A and B. The B.1.1.7 outbreak in Punjab was notable for multiple polytomies in the phylogenetic tree and clusters spanning multiple districts, suggesting a mass super-spreader event (Fig 4C). This is relatable to mass public gatherings and rallies held in different parts of North India since January 2021 and highlights the important role of social factors in SARS-CoV-2 outbreaks.

Expansion of B.1.1.7 and B.1.617 lineages in North India

To determine whether the outbreak in Punjab and Delhi may have led to seeding of B.1.1.7 across North India, for which these are the main travel hubs, we analyzed the temporal trend of the SARS-CoV-2 lineages (Figure 4 D-H). Despite limitations of lower sequencing than for Delhi, we see the same recurring pattern with initial seeding by B.1.1.7 in February to March and replacement by B.1.617.2 in April 2021.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 4. A) Phylogenetic relationships between samples from Delhi, Maharashtra and Punjab. B) Time tree. C) District-wise phylogenetic tree for samples from Punjab to determine common origins. D) – J) State-wise normalized stacked bar graphs of main lineages since Jan 2021.

Prevalent evidence of SARS-CoV-2 variant B.1.617.2 in post-vaccinated individuals

We sequenced all samples of vaccination-breakthrough cases at NCDC, Delhi, over the period of the study. Two VOC lineages were seen amongst 27 cases: B.1.617.1 (n=2; 8%), B.1.617.2 (n=19; 76%) Other samples had the background B.1 lineages (16%) (**Figure 5a**). It is noted that when compared to population prevalence (Figure 3A), B.1.617.2 was over-represented and B.1.1.7 was not even detected in vaccination breakthroughs, suggesting higher breakthrough risk of B.1.617.2 compared to B.1.1.7. There were some sequence variations within the lineages, as expected in an evolving outbreak. The lineage-defining mutations, frequent mutations and phylogenetic analysis are shown (**Fig 5b**).

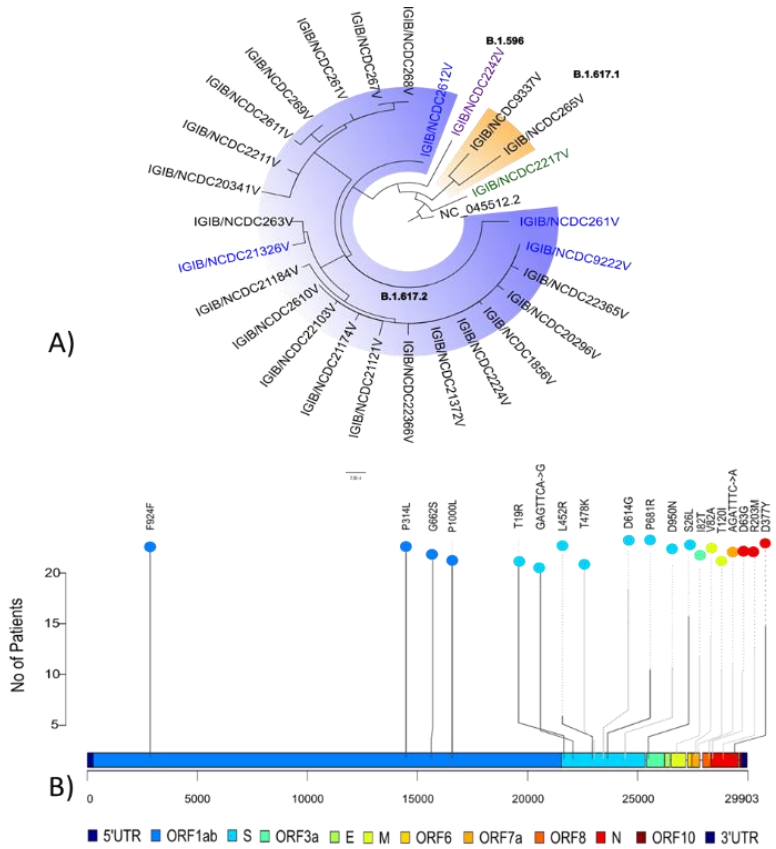


Figure 5. A) The Phylogenetic distribution of lineages in vaccinated samples. B) Top most frequent mutations detected in vaccine-breakthrough samples.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Structural insights into spike protein mutations showed unique features of B.1.617.2 that may contribute to increased transmissibility.

Our data shows that B.1.617.2 has higher transmissibility than B.1.1.7 in North India, with initial seeding by B.1.1.7 being consistently replaced by B.1.617.2, even in Punjab, where B.1.1.7 had reached near 100% prevalence. This is surprising since there is no known mutation in B.1.617 lineage or in B.1.617.2 that is associated with such high transmissibility. Stronger immune escape was initially considered a possible explanation for dominance of B.1.617 over B.1.1.7 (Hoffmann et al., 2021), given high seropositivity in Delhi. The lower representation of B.1.1.7 than B.1.617 lineages in vaccination breakthroughs, suggests that this is a plausible advantage to B.1.617 variants especially B.1.617.2. However, we see the same trend across various states including those with low seropositivity in the CSIR-cohort, such as Uttarakhand or Himachal Pradesh, suggesting that transmissibility is higher, independent of immune escape. Therefore, to understand the recent evolution of spike protein that may contribute to higher transmissibility, positional mapping of mutations on the structural model representing four lineages is shown (Figure 6). The B.1.1.7 lineage, first detected in September 2020, has been reported to carry several spike mutations, including two deletions in the N-terminal domain (NTD) and N501Y mutation at the critical receptor binding domain (RBD) that interacts with ACE2 human receptor (Rambaut et al., 2020). In addition, A570D, P681H, T716I, S982A and D1118H were also reported. In comparison, B.1.617 lineage carrying two unique RBD mutation combo (E484Q and L452R), however, lacked the NTD deletions (Ferreira et. al., 2021). The B.1.617.1 harbored similar set of two RBD mutations with slight changes. On the other hand, the rapidly spreading B.1.617.2 in India showed unique combination of seven spike mutations.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

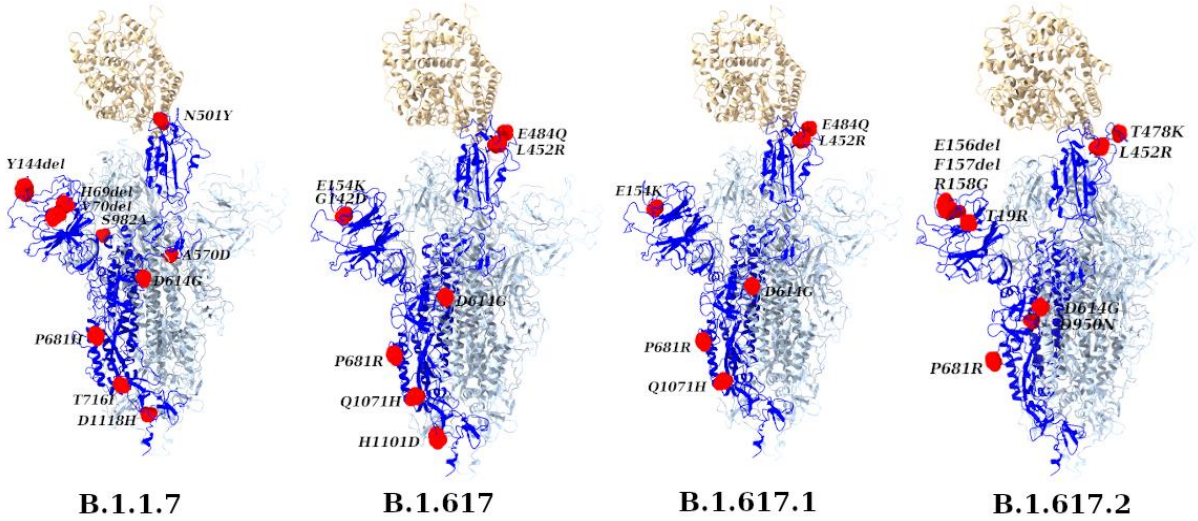


Figure 6: Lineage-wise mapping of SARS-CoV-2 Spike protein mutations. The structural models depict the head region of spike protein (14-1147 residues) and ACE2 human receptor in grey and golden, respectively. The residue location of mutations is marked in red color across four lineages. One of the three chains of spike protein is marked in blue color for clarity.

To understand the molecular interactions, we generated a structural model representing B.1.617.2 spike protein, as shown in Figure 7. The resulting map provides insights into the plausible mechanisms of regulation of virus entry and binding. It contains seven mutations in the spike protein, excluding the predominant D614G substitution. Three of these mutations, two substitutions (T19R and R158G) and one deletion (ΔE156-F157), were found in NTD region. The six nucleotides spanning the entire stretch of deletion (ΔE156-F157) was juxtaposed with R158G mutation, with one nucleotide of glycine contributed from E156 and the rest two nucleotides contributed by R158, resulting in GGA codon for glycine. The mutations within NTD occur on N1 and N3 loops that composes the prominent mABs recognition sites Chi et al, 2020; Naveenchandra, et al.,2021). In our mutated model, R19 and

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

G158 residues are surface accessible while in wild-type the T19, E156-F157 were relatively buried.

We also found a unique RBD mutation T478K, in addition to previously reported L452R in B.1.617.2. Previous reports have associated L452R amino-acid change with antibody binding and has been classified as an escape variant (McCallum et al.,202; Liu et al., 2021). T478, on the other hand, is previously unidentified and is present directly on the receptor binding motif (RBM). The inherent long side-chain of mutated lysine reduces the gap with the ACE2 receptor as compared to the wild type. The distance between spike: K478 and ACE2:L85 is 8.3 Å, while the threonine maintains a distance of 10 Å. However, these distances may vary with other conformational states of spike RBD. We also observed significant increase in number of ACE2 residues around these mutated sites (L452R, T478K) in B.1.617.2. In contrast with 10 residues, 24 residues of ACE2 were in close proximity to the mutated side chains of B.1.617.2 RBD mutations. In addition, we observed two mutations (P681R and D950N) in proximity to the S1/S2 cleavage site. Previous reports have also shown that the dominant D614G, although distal to the furin cleavage site has an allosteric effect on conformational changes leading to RBD opening (Gobeil et al. 2021). We speculate that a combination of these mutations at three critical sites leads to enhanced transmissibility of B.1.617.2 lineage.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

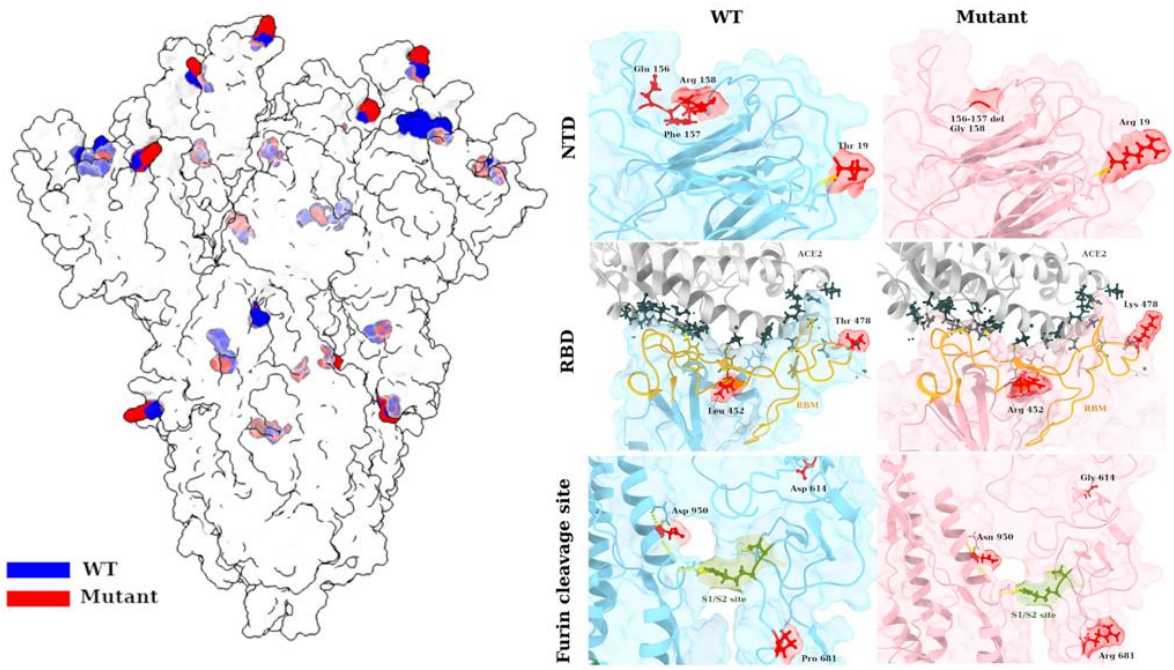


Figure 7: Mutant spike protein of B.1.617.2 lineage. The structural model of spike protein with seven mutations was generated and side-chains are highlighted in blue and red color to illustrate amino-acid substitutions. On the right panel, we show a zoomed snapshot of three critical regions namely, NTD, RBD, and furin cleavage sites. The location of mutated residues is marked in red and, and critical regions such as RBM and S1/S2 site is highlighted for clarity.

Conclusions

Northern India underwent seeding with B.1.1.7 in February 2021 followed by regional outbreaks that stretched the healthcare system but remained within its capacity. From March 2021 onwards, after introduction of B.1.617 and its sub-lineage B.1.617.2, there was a further rapid rise in new cases with displacement of B.1.1.7 and other B.1.617 lineages by B.1.617.2. Our data indicates B.1.617.2 shows high transmissibility and surges without any increase in CFR. We estimate the transmissibility to be as much as 50% greater than B.1.1.7. Viral load of B.1.617.2 appears to be higher than B.1.1.7 and based on data from India and UK, so does vaccination break-through rate. While immune escape seems less for B.1.617.2 compared to B.1.351 or P.1 (Li et al., 2021), overall, we note that B.1.617.2 is capable of creating very fast-

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

rising outbreaks with vaccination breakthroughs. We would re-emphasize that prior infections, high seropositivity and partial vaccination are insufficient impediments to its spread, as seen in Delhi, and strong public health response will be needed globally for its containment.

Methodology

Nasopharyngeal and throat swab samples from Covid-19 confirmed cases with CT value < 25 was collected and transported to Biotechnology Division, NCDC from the various testing sites across the States as per the sampling strategy of Central Surveillance Unit (CSU) of Integrated Disease Surveillance Programme (IDSP), NCDC. Twenty seven post-vaccinated Covid-19 positive samples were also included in this study. All the patient details were filled on patient identification form and was accompanied with the samples. The study was approved by the NCDC Ethics Review Committee vide approval 2020/NERC/14. More than 9000 samples were received at NCDC from more than 9 states and union territories from November, 2020 till May, 2021. The RNA was isolated using MagNA Pure RNA extraction system (Roche). The samples were processed as per the CovidSeq (Illumina) protocol for sequencing on the NextSeq 550 (Illumina) instrument according to the manufacturer's instructions. A total of 376 samples per lot were processed in batches of 94 with indexes A-D and 1.4pmol of the library was loaded on the 75cycle High Output flow cell. The run was over in 10 hours with approximately 20Gb data generated that was processed on the Dragen COVID Lineage Application v3.5.1 (Illumina). The raw data generated in binary base call (BCL) format from NextSeq 550 was demultiplexed to FASTQ file using bcl2fastq v2.20. Further the samples files in FASTQ format are provided as input for the COVID Lineage application. The reads are aligned to a SARS-CoV-2 reference genome (NC_045512). The minimum acceptable alignment score was set to 12, so any alignment results scoring lower are discarded. The coverage threshold and virus detection threshold were set to 20 and 5 respectively. The variant calling target coverage is set

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

to 50 which specifies the maximum number of reads with a start position overlapping any given position. A total of 8,477 sequences out of initial 11,001 were used for further analysis. The pipeline performs Kmer-based detection and then performs Map/Align, Variant calling and Consensus Sequence generation. The lineage/clade analysis performed using Pangolin and NextClade.

The lineage data was segregated according to date of collection (DOC), State, Age, Sex. State wise distribution of B.1.617.2 compared with B.1.617.1, B.1.1.7 and 'other variants' over a period of five months (January-May, 2021) was performed to ascertain the prevalence of each variant within these geographical areas. The number of tests, confirmed cases and positivity rate were accessed from the national databases. The case fatality ratio (CFR) was calculated as per the WHO recommended method, as follows:

$$\text{CFR (\%)} = \frac{\text{Number of deaths}}{\text{Number of deaths} + \text{Number of cases recovered}}$$

Serosurvey

The serosurvey was conducted through a voluntary participation wherein personnel working at CSIR labs/centers and their family members gave their blood samples in May -Sep 2020 (Phase 1) and Feb-March 2021 (Phase 2). The study was approved by the Institutional Human Ethics Committee of CSIR-IGIB vide approval CSIR-IGIB/IHEC/2019–20 and carried out in over 40 CSIR laboratories and centers spread across the country. In this study, samples from 10427 adult subjects were obtained of which; 1399 samples were from Delhi based laboratories and offices in Phase 1 and 9918 samples were obtained of which 1115 samples were from Delhi in Phase 2. Blood samples (6 ml) was collected in EDTA vials from each participant and analyzed on site or transported to CSIR-IGIB, New Delhi for Analysis. Elecsys Anti-SARS-CoV-2 kit from Roche Diagnostics was used to detect antibodies to SARS-CoV-2 NucleoCapsid antigen.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

It is a qualitative kit which was used for screening and a Cut-off index COI ≥1 was considered seropositive. Positive samples were further tested for quantitative antibody titers using the same manufacture's kit directed against the spike protein (S-antigen). An antibody levels >0.8 U/ml was considered sero-positive as per manufacturer's protocol. The detection range of this kit is from 0.4 U/ml to 250 U/ml. For samples, where values of >250 U/ml were obtained; appropriate dilutions were made. Neutralizing antibody (NAB) response directed against the spike protein (RBD site) was assessed using GENScript cPass kit which is a surrogate virus neutralization test (sVNT). A value of 30% or above was considered to have neutralizing ability.

Protein annotation and modelling

In total of 8,477 SARS-CoV-2 genomes were annotated for amino-acid substitutions by SnpEff version 4.5 (Cingolani et al., 2012). The annotation was done according to the known reference genome of SARS-CoV-2 i.e. NC_045512 in the NCBI database (Wang et al.,2020). The structural model of spike in 1 RBD-up state was generated using cryoEM structure of the spike 1 RBD-up state (PDB ID: 6VSB) as a template (Wrapp et al., 2020). To generate ACE2 bound structure, we took the X-ray structure of human ACE2 bound to the RBD domain with PDB ID: 6M0J (Lan Nature. 2020). Detailed modelling methodology is mentioned in our previous work (Fatihi et al, 2021). The structural mutant model of B.1.617.2 variant was generated using the structural model of ACE2-bound 1 RBD-up spike conformation as a reference. Each chain was mutated for missense mutations using ChimeraX (Goddard et al., 2018) whereas deletions in each chain were introduced by employing Coot (Emsley et al, 2010).

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Acknowledgments – Support from Ministry of Health and Family Welfare, Council of Scientific and Industrial Research, and Department of Biotechnology, India, is acknowledged.

References

1. Chi, X., Yan, R., Zhang, J., Zhang, G., Zhang, Y., Hao, M., . . . Yang, Y. (2020). A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science, 369*(6504), 650-655.
2. Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly, 6*(2), 80-92.
3. Davies, N. G., Jarvis, C. I., Edmunds, W. J., Jewell, N. P., Diaz-Ordaz, K., & Keogh, R. H. (2021). Increased mortality in community-tested cases of SARS-CoV-2 lineage B. 1.1. 7. *Nature*, 1-5.
4. Convalescent Plasma EUA Letter of Authorization USA March 9, 2021 [Available from: https://www.fda.gov/media/141477/download accessed April 2021.
5. Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography, 66*(4), 486-501.
6. Fatihi, S., Rathore, S., Pathak, A., Gahlot, D., Mukerji, M., Jatana, N., & Thukral, L. (2021). A rigorous framework for detecting SARS-CoV-2 spike protein mutational ensemble from genomic and structural features. *bioRxiv*.
7. Ferreira, I., Datir, R., Papa, G., Kemp, S., Meng, B., Rakshit, P., . . . Radhakrishnan, V. S. (2021). SARS-CoV-2 B. 1.617 emergence and sensitivity to vaccine-elicited antibodies. *bioRxiv*.
8. Gobeil, S. M.-C., Janowska, K., McDowell, S., Mansouri, K., Parks, R., Manne, K., . . . Edwards, R. J. (2021). D614G mutation alters SARS-CoV-2 spike conformation and enhances protease cleavage at the S1/S2 junction. *Cell reports, 34*(2), 108630.
9. Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., & Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science, 27*(1), 14-25.
10. Hoffmann, M., Arora, P., Groß, R., Seidel, A., Hörnich, B. F., Hahn, A. S., . . . Kempf, A. (2021). SARS-CoV-2 variants B. 1.351 and P. 1 escape from neutralizing antibodies. *Cell*.
11. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., . . . Zhang, L. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature, 581*(7807), 215-220.
12. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics, 34*(18), 3094-3100.
13. Li, Q., Nie, J., Wu, J., Zhang, L., Ding, R., Wang, H., . . . Zhang, M. (2021). SARS-CoV-2 501Y. V2 variants lack higher infectivity but do have immune escape. *Cell, 184*(9), 2362-2371. e2369.
14. Loman, N. J., Quick, J., & Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature methods, 12*(8), 733-735.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.02.21258076; this version posted June 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

15. McCallum, M., Bassi, J., De Marco, A., Chen, A., Walls, A. C., Di Iulio, J., . . . Saliba, C. (2021). SARS-CoV-2 immune evasion by variant B. 1.427/B. 1.429. *bioRxiv*.
16. Naushin, S., Sardana, V., Ujjainiya, R., Bhatheja, N., Kutum, R., Bhaskar, A. K., . . . Rawat, B. S. (2021). Insights from a Pan India Sero-Epidemiological survey (Phenome-India Cohort) for SARS-CoV2. *Elife, 10*, e66537.
17. Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., . . . Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature microbiology, 5*(11), 1403-1407.
18. Suryadevara, N., Shrihari, S., Gilchuk, P., VanBlargan, L. A., Binshtein, E., Zost, S. J., . . . Chen, E. C. (2021). Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell, 184*(9), 2316-2331. e2315.
19. Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., . . . Ferguson, N. M. (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*. doi:10.1038/s41586-021-03470-x
20. Wang, C., Liu, Z., Chen, Z., Huang, X., Xu, M., He, T., & Zhang, Z. (2020). The establishment of reference sequence for SARS-CoV-2 and variation analysis. *Journal of medical virology, 92*(6), 667-674.
21. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., . . . McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science, 367*(6483), 1260-1263.