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

Authors: Mahesh S Dhar[1#], Robin Marwal[1#], Radhakrishnan VS[1#], Kalaiarasan Ponnusamy[1#], Bani Jolly[2,3#], Rahul C. Bhoyar[2#], Viren Sardana[2,3], Salwa Naushin[2,3], Mercy Rophina[2,3], Thomas A Mellan[4], Swapnil Mishra[4], Charlie Whittaker[4], Saman Fatihi[2,3], Meena Datta[1], Priyanka Singh[1], Uma Sharma[1], Rajat Ujjainiya[2,3], Nitin Bhateja[2], Mohit Kumar Divakar[2,3], Manoj K Singh[1], Mohamed Imran[2,3], Vigneshwar Senthivel[2,3], Ranjeet Maurya[2,3], Neha Jha[2], Priyanka Mehta[2], Vivekanand A[2,3], Pooja Sharma[2,3], Arvinden VR[2,3], Urmila Chaudhary[1], Lipi Thukral[2,3], Seth Flaxman[5], Samir Bhatt[4,6], 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), Ravindra K. Gupta[7,8], Sujeet K Singh[1], Anurag Agrawal[2,3]*, Partha Rakshit[1]*

Affiliations:

[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

[4]Medical Research Council (MRC) Centre for Global Infectious Disease Analysis, Jameel Institute, School of Public Health, Imperial College London, UK.

[5] Department of Mathematics, Imperial College London, London, UK.

[6] Section of Epidemiology, Department of Public Health, University of Copenhagen, Denmark.

[7] Department of Medicine, Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Cambridge, UK.

[8] Africa Health Research Institute, KwaZulu-Natal, South Africa

# These authors contributed equally to this work

*Corresponding authors. Email: partho_rakshit@yahoo.com (Partha Rakshit), a.agrawal@igib.in (Anurag Agrawal)

Abstract: After multiple SARS-CoV2 outbreaks in 2020 and a population seropositivity of about 50%, Delhi was overwhelmed by the fourth wave. Multi-modal data, including genomic surveillance, shows that a controlled phase characterized by Alpha variant (B.1.1.7) dominance, became uncontrollable during its displacement by the new Delta variant (B.1.617.2). Bayesian modeling supports a combination of 40-60% increased transmissibility of the Delta variant and 10-50% reduction in immunity elicited by prior infection. A cohort with 45% seropositivity in March, showed 85% positivity in June 2021, with 14% reinfection based on uptick in titers after decline. High transmissibility and evasion of prior immunity by the Delta variant, coupled to social behavior that promoted transmission, led to the overwhelming pandemic outbreak in Delhi. This may become a global pattern.

One-Sentence Summary: The Delta variant of SARS-CoV2, is extremely transmissible and can evade immunity elicited by infections of prior lineages, leading to severe outbreaks in regions with high seropositivity.

Main Text:

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 (1). Serosurveys carried out by State / Municipal authorities in Indian megacities had shown even higher seropositivity with the national capital, Delhi, reporting 56% seropositivity in February, 2021 (2). 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 has enabled an analysis of the dynamics of a fourth wave 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 four distinct waves, in June, September, and November 2020, and now in April 2021. The first and the third waves were large, associated with stressed healthcare systems and spikes in mortality, but were successfully managed with social restrictions and use of micro-containment strategies focusing on test, trace, treat. After touching a high of almost 9000 cases daily and a positivity rate of about 15% during

the third wave in November 2020, new cases steadily declined, with only 1% positivity between December 2020 and March 2021. This reversed and started increasing from the third week of March, 2021, reaching 30% by end of April, with 20,000-25,000 cases being reported per day at the peak (Fig 1A-C). Deaths rose proportionately (Fig 1B) and the case-fatality ratio (CFR), estimated as the scaling factor between time-advanced cases and deaths (Fig 1C), was stable (1.9 $\pm$ 0.3%), with a dip in February when both cases and death were at their lowest. Since cases declined faster than deaths, a modest upward revision of CFR estimates is plausible.

Serosurvey results from multiple national laboratories of the Council of Scientific and Industrial Research (CSIR), India, have previously approximated that of national cohorts, with the additional advantage of more detailed antibody stability studies (3, 4). A 20-30% decline in neutralization capacity by 6 months had been reported in previously seropositive subjects (3). A repeat sero-survey was conducted in Feb-Mar 2021, with semi-quantitative anti-nucleocapsid, quantitative anti-spike antibodies (Ab), and a validated surrogate viral neutralization test (sVNT) being measured. When compared to the prior serosurvey in August/September 2020, seropositivity had increased with 473 of 1115 employees being seropositive (Fig 1D). The seropositivity rate was highly uneven, across CSIR laboratories, public-offices, and residential complexes, possibly reflecting varying exposures. The median level of anti-spike antibodies in Feb-Mar 2021, was about 100 U/ml, which would normally be considered a high level close to that of high-titre plasma (132 U/ml) (5). However, there was wide variation in titres as well as neutralization efficiency, with 19% of seropositive samples in the negative range of the sVNT assay, including one case classified as high-titre based on anti-spike Ab (Fig 1 E-F). The sVNT assay is based on wildtype spike protein and neutralization efficiency may thus be lower for variants, or for Ab induced by variants. Thus, the fraction of subjects with robust neutralization activity was substantially lower than simple seropositivity would indicate.

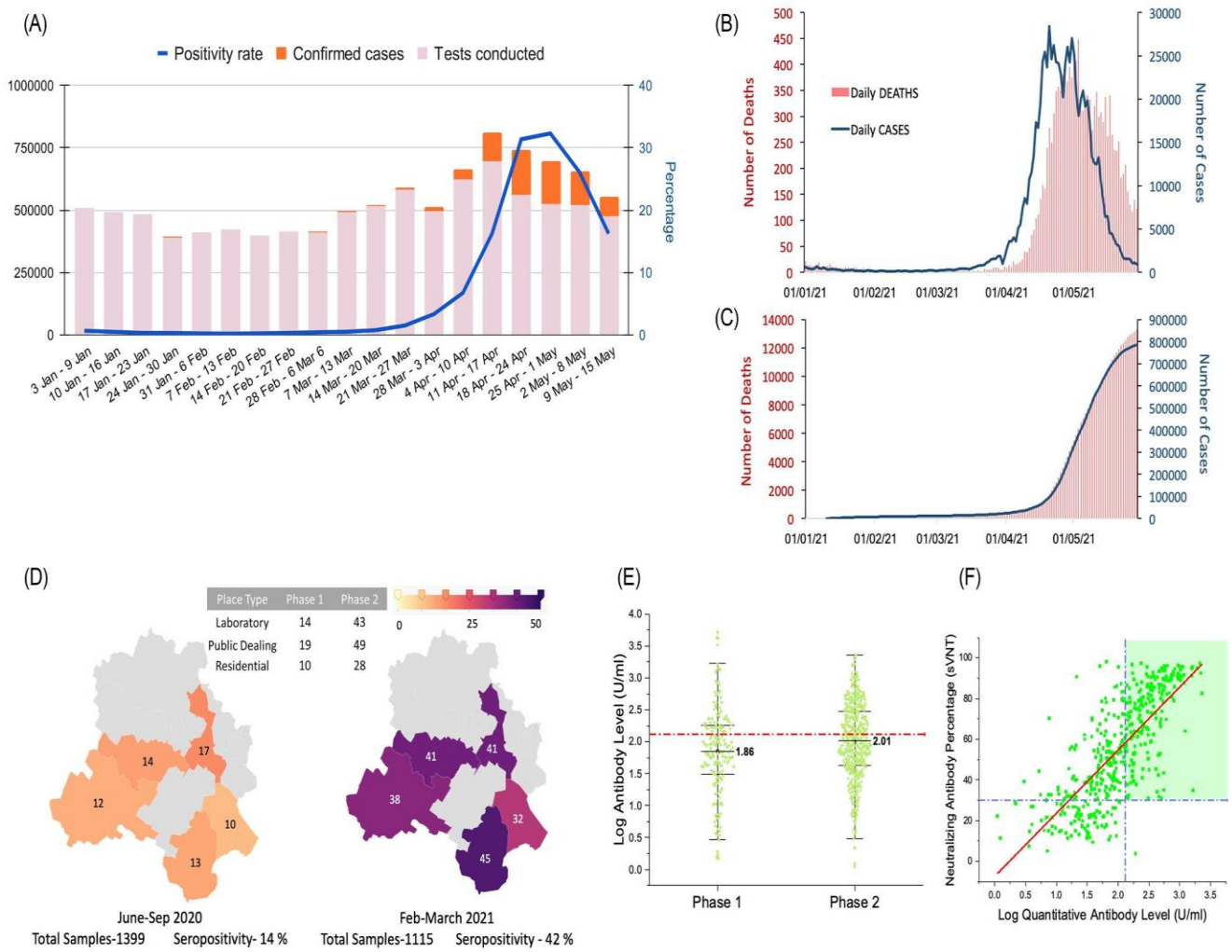


Fig. 1. Epidemiology of the 2021 outbreak in Delhi. A) Weekly tests, new cases and test positivity rates in 2021 up to and after the peak. B) Daily cases and deaths. C) Time advanced and scaled cumulative cases, fitted to cumulative deaths, with CFR estimated as the scaling factor. D) Seropositivity rates in Delhi in Phase I (June-September 2020) and Phase II (Feb-March 2021) of CSIR-Cohort (Map Created using Datawrapper). E) Quantitative Antibody Levels (U/ml) in Delhi during Phase 1 and Phase 2 of CSIR-Cohort. Red dashed dotted line denotes 132 U/ml, a cut-off for defining high-titre plasma. F) Results of sVNT assay, shown as a function of quantitative titres (Horizontal blue line denotes 30 percent cut off for sVNT).

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

To determine whether SARS-CoV-2 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. We also analyzed the mean monthly threshold cycle (Ct) values from a single high-throughput automated machine (Roche 6800) at a single center over this period (Fig 2A).

The outbreaks in 2020 in Delhi were unrelated to any VOC (Fig 2B). The Alpha variant (B.1.1.7) was identified only occasionally, primarily in foreign travellers (not shown), until January 2021. Alpha variant outbreaks related to a super-spreader event were noted in the nearby state of Punjab in February, where it quickly established near complete dominance (Fig. S1). The Alpha variant increased in Delhi to about 20% in February and 40% in March, before it was displaced by a rapid increase in the Delta variant (B.1.617.2) in April (Fig 2A). This was paralleled by falling Ct values, first in March with the rise of Alpha and further in April after displacement by Delta. The total change of about -6 Ct corresponds to about 25 to 50 fold increase in viral load. To exclude the possibility that this is an artefact related to an outbreak and to confirm that there is an increase in high viral load samples we looked at the frequency of samples with Ct $\leq$ 20 at one centre with a campus surveillance program. The frequency of such samples increased from 9% (27 of 297) in the November 2020 outbreak to 15% in the April 2021 outbreak (53 of 358) (χ^2- 4.94, p = 0.02). This may be a relevant high transmissibility factor, since higher transmissibility of the Alpha variant has been previously related to increased viral load (6, 7).

Phylogenetic analysis showed common origins between Alpha variants in Delhi and Punjab, and between Kappa (B.1.617.1) or Delta variants in Delhi and Maharashtra, where Kappa and Delta were first sequenced (Fig 2C). To determine whether the major genomic and epidemiologic

trends, showing the correlation between the rise in Delta variant with increased Rt and outbreaks, were isolated to Delhi, or part of a broader trend, we also analyzed genomic trends in surrounding North Indian states that had outbreaks in April-May (Fig. 2D-F, Fig S2). In each of these, the same pattern was seen – initial entry and increase of the Alpha variant, followed by the Delta variant displacing the Alpha variant, coinciding with a surge in cases.

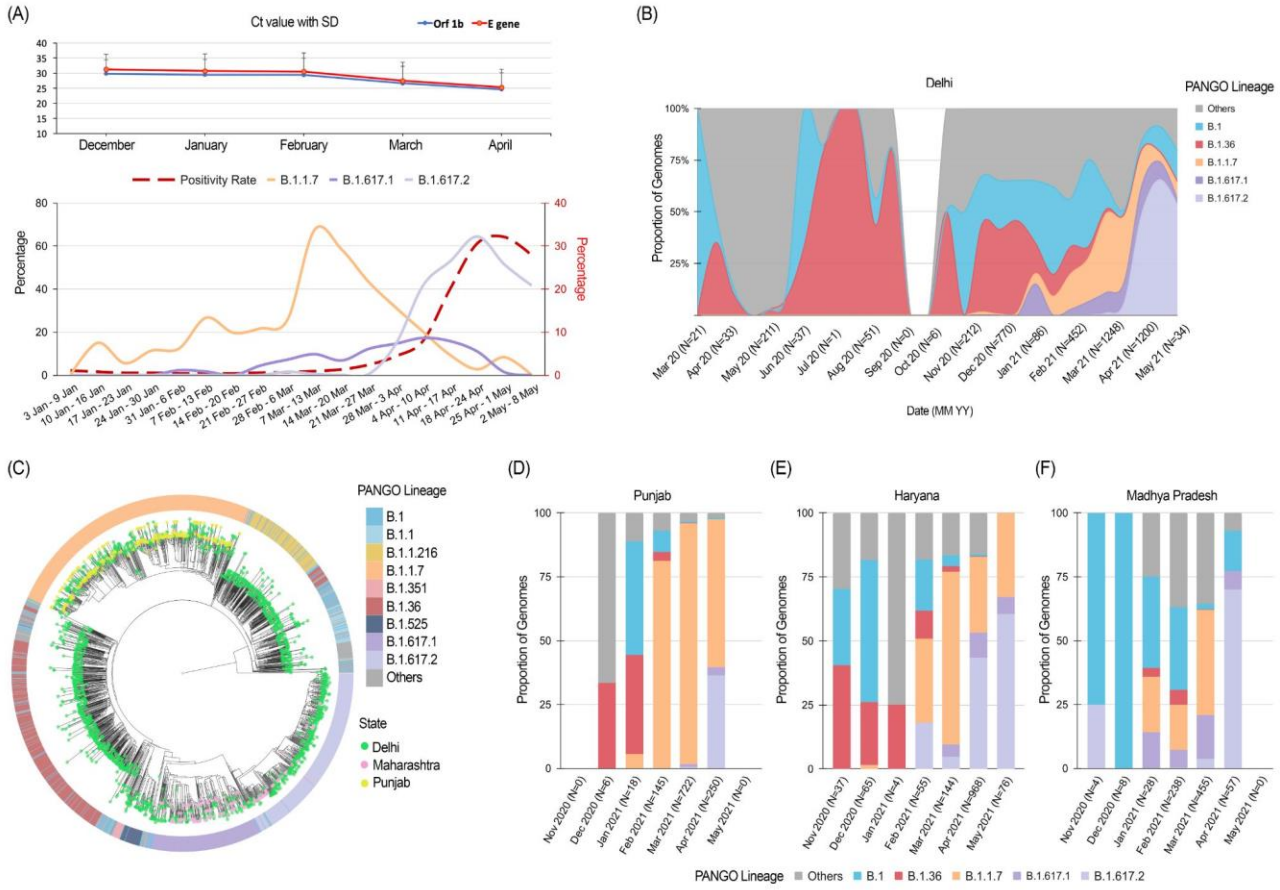


Fig. 2. Genomic-Epidemiologic correlations. A) Time-trends of Ct values, Positivity rate, and prevalence of Alpha, Delta, and Kappa variants in Delhi in 2021. B) Graph of main lineages in Delhi from March 2020 to May 2021. Data is very low or missing for some periods in 2020 C) Phylogenetic analysis for VOC strains between Delhi and states (Punjab and Maharashtra) with known VOC outbreaks before April 2021. D) – F) State-wise normalized stacked bar graphs of main lineages for neighbouring states.

Combined epidemiological analysis for Delhi based on genomic, seroprevalence, and death data

To better characterise the epidemiological properties of the Delta variant in Delhi and how these might differ to other SARS-CoV-2 lineages previously circulating in the city, we utilised a Bayesian model of SARS-CoV-2 transmission and mortality that simultaneously models the dynamics of two-categories of virus ("B.1.617.2" and "non-B.1.617.2") (8), whilst also explicitly incorporating natural waning of immunity derived from prior infection, with the duration of immunity based on the results of recent longitudinal cohort studies (9, 10). This model is fitted to COVID-19 mortality data(11), genomic sequence data presented here and from GISAID (with lineage classification carried out used Pangolin (12), (https://pangolin.cog-uk.io/), and serological data presented here alongside an additional longitudinal serosurvey carried out in the city in the period July-December 2020 (13) . Substantial uncertainties remain as to the date of B.1.617.2's introduction into Delhi and the degree of COVID-19 death under-ascertainment. We therefore explored a range of different scenarios varying under-ascertainment (10%, 33%, 50%, and 66%) and introduction dates (14 January 2021, 14 February 2021 and 28 February 2021).

Using this framework, and for an introduction date of 14 February 2021 and mortality under-ascertainment of 33%, our results suggest that the Delta variant is 1.4-1.6 (50% bCI) fold more transmissible than earlier/co-circulating SARS-CoV-2 lineages in Delhi, including the Alpha variant. Importantly, the model also suggests that the Delta variant is able to partially evade protective immunity elicited by prior infection, with prior infection providing only 50 to 90% (50% bCI) of the protection against infection with Delta variant that it provides against previous lineages. These results are robust to our assumptions surrounding under-ascertainment and the date of introduction - for the full set of results spanning the range of scenarios considered, see Supplementary Information (Table S1). The results are valid for a population where the majority

of immunity arose from prior infection (rather than vaccination), which is true for Delhi. Previous work has shown differences in the breadth and quality of immunity elicited by natural infection compared to immunity, and the degree to which these results generalise to vaccine-derived immunity is likely limited. Uncertainty in death reporting and start date may quantitatively alter the presented results, although not the general conclusions.

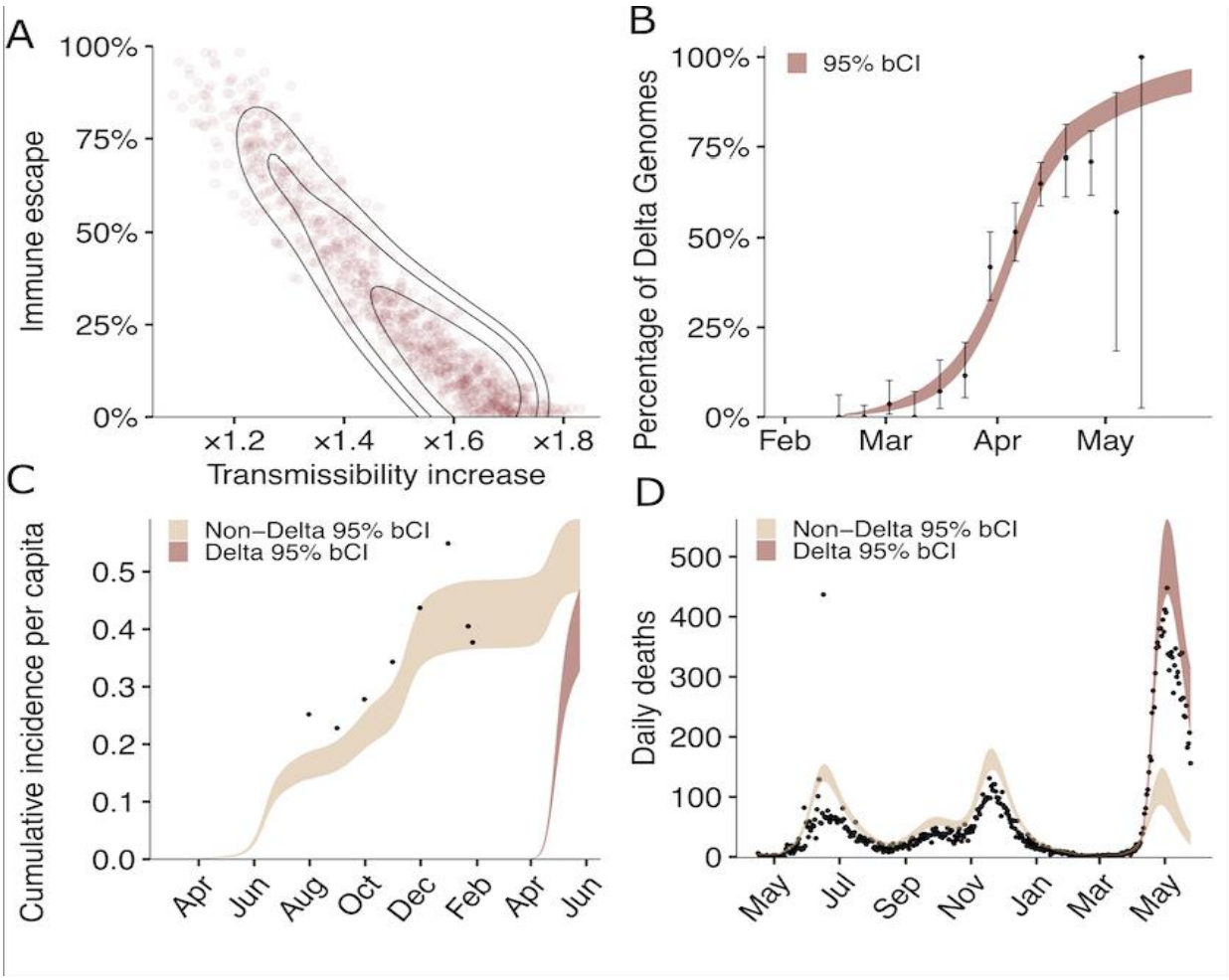


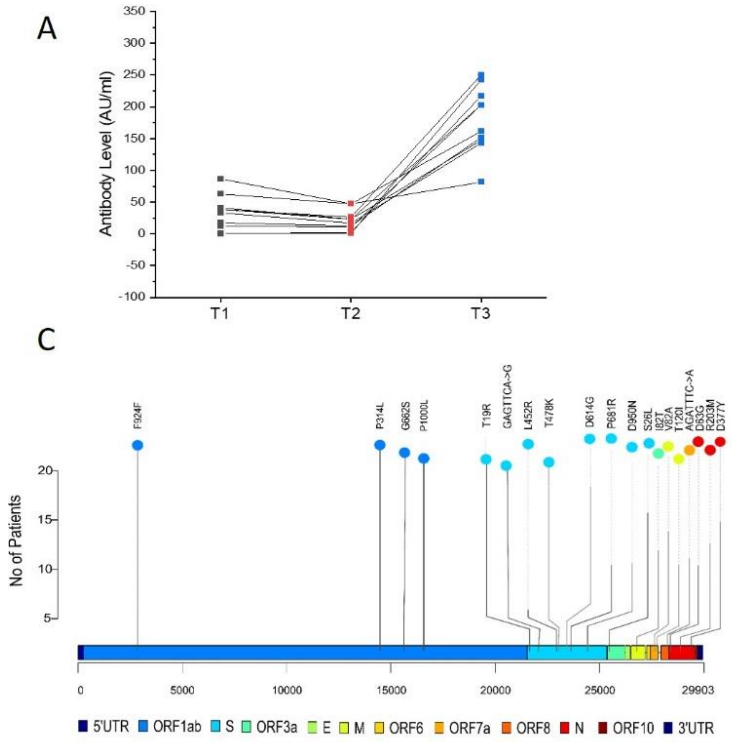
Fig. 3. Estimates of the epidemiological characteristic of the Delta variant. Values were inferred from a two category Bayesian transmission model fitted to mortality, serosurvey and genomic data from Delhi, India. A. Joint posterior distribution of the Delta variant immune escape and transmissibility increase relative to non-Delta categories. B. Delta fraction over time inferred

by the model and genome sampling data points. C. Serosurvey data and inferred cumulative incidence for Delta and non-Delta variant categories. D. Mortality data and inferred deaths assuming 33% under reporting. Other under-ascertainment scenarios are presented in the Supplementary Information

Ascertainment of reinfections and vaccination break-throughs

To determine whether there was actual evidence of reinfection during the outbreak, as predicted from the model, we started serology testing in June for the CSIR cohort (Phase 3) and completed it for one laboratory. In February, 172 (45%) of 379 subjects participating in Phase 2 had been seropositive (45%). In June, the corresponding number was 196 (85%) of 231, $p<0.0001$). Restricting the analysis further to only cohort participants with prior participation, we noted that 65 (86%) of 76 were seropositive, excluding participation bias for Phase 3. Interestingly amongst these 65, there was serologic evidence of reinfection in 9 cases by examining three serial seropositive samples (Fig. 4B).

We additionally sequenced all samples of vaccination-breakthrough cases at a single centre over the period of the study. Amongst 27 cases, 19 (76%) were Delta and 2 Kappa (8%) (Fig. 4). It is noted that when compared to population prevalence (Fig. 2), Delta was over-represented while Alpha was not detected in vaccination breakthroughs. This is consistent with a higher vaccination breakthrough risk of Delta compared to Alpha. The lineage-defining mutations, new mutations and phylogenetic analysis are shown (Fig 4C and D).



B

	Phase 2 and Phase 3 Data Summary			
	Naïve	Infected	Reinfected	Total
Phase 2	63	70	1	134
Phase 3	11	56	9	76

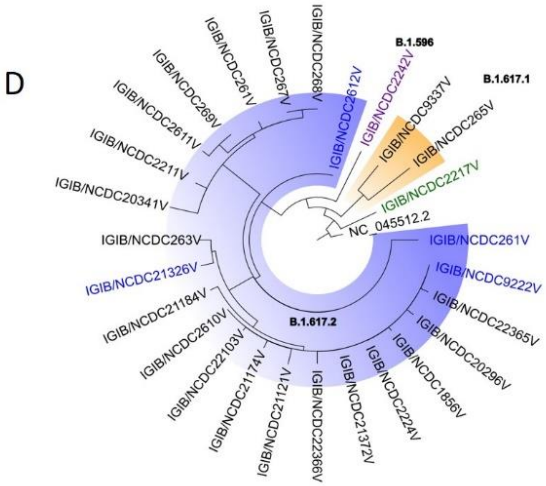


Fig. 4. Increased reinfections and vaccination breakthrough. A) Participants with 3 or more datapoints above the threshold for anti-spike Ab positivity, with higher titres after previous downtick (n=10). B) Phase 2 (P2) and Phase 3 (P3) based disaggregation of reinfections of such participants. C) Lollipop plot of mutations seen in sequenced samples from vaccination breakthrough cases D) Phylogenetic analysis of mutations and lineage matching showing mostly Delta variant in vaccination breakthroughs.

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

Our data shows that the Delta variant has both higher transmissibility and immune escape than other variants in Delhi, including the Alpha variant. This is somewhat surprising since it has lost the E484 mutation that is associated with immune escape and L452R was not expected to be associated with such high transmissibility. Therefore, to understand the recent evolution of spike protein that may contribute to higher transmissibility, we generated a structural model representing B.1.617.2 spike protein (Figure S3 and Supplementary information). 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 the 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 compose the prominent mABs recognition sites (14, 15). In our mutated model, R19 and 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 (16, 17). 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 a significant increase in the 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 (18). We speculate that a combination of these mutations at three critical sites leads to enhanced transmissibility of B.1.617.2 lineage.

Conclusions

Northern India underwent seeding with the Alpha variant in February 2021 followed by regional outbreaks that stretched the healthcare system but remained within its capacity. From March 2021 onwards, after introduction of Kappa and Delta variants, there was a further rapid rise in new cases with displacement of Alpha and Kappa lineages by Delta. A catastrophic 10-fold rise in new cases in April, from about 2500 to 25,000 daily cases, within a fortnight, severely stressed the healthcare system, with daily deaths spiking to levels three-fold higher than previous waves. The genomic and serologic evidence presented here indicates that the Delta variant was associated with high viral load, transmissibility, vaccination breakthrough, and reinfection. Models incorporating all data infer that Delta transmissibility may be 30-60% higher than the underlying genetic background in Delhi prior to introduction of Delta. Importantly, it is estimated that protection from prior infections by lineages excluding B.1.617, would be reduced by 12 to 60%. Overall, we note that the Delta variant is capable of creating very fast-rising outbreaks with vaccination breakthroughs. We would re-emphasize that prior infections, high seropositivity and low

vaccination coverage are insufficient impediments to its spread, as seen in Delhi, and a strong public health response will be needed globally for its containment.

References

1. ICMR sero survey: One in five Indians exposed to Covid-19 2021 [Available from: https://www.bbc.com/news/world-asia-india-55945382.

2. Sero survey shows 56% have antibodies 2021 [Available from: https://www.thehindu.com/news/cities/Delhi/sero-survey-shows-56-have-antibodies/article33722284.ece.

3. Naushin S, Sardana V, Ujjainiya R, Bhatheja N, Kutum R, Bhaskar AK, et al. Insights from a Pan India Sero-Epidemiological survey (Phenome-India Cohort) for SARS-CoV2. eLife. 2021;10:e66537.

4. Murhekar MV, Bhatnagar T, Selvaraju S, Saravanakumar V, Thangaraj JWV, Shah N, et al. SARS-CoV-2 antibody seroprevalence in India, August–September, 2020: findings from the second nationwide household serosurvey. The Lancet Global Health. 2021;9(3):e257-e66.

5. Convalescent Plasma EUA Letter of Authorization USA March 9, 2021

6. Davies NG, Jarvis CI, van Zandvoort K, Clifford S, Sun FY, Funk S, et al. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. Nature. 2021;593(7858):270-4.

7. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature. 2021;593(7858):266-9.

8. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DdS, Mishra S, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science. 2021;372(6544):815-21.

9. Hansen CH, Michlmayr D, Gubbels SM, Mølbak K, Ethelberg S. Assessment of protection against reinfection with SARS-CoV-2 among 4 million PCR-tested individuals in Denmark in 2020: a population-level observational study. The Lancet. 2021;397(10280):1204-12.

10. Hall VJ, Foulkes S, Charlett A, Atti A, Monk EJM, Simmons R, et al. SARS-CoV-2 infection rates of antibody-positive compared with antibody-negative health-care workers in England: a large, multicentre, prospective cohort study (SIREN). The Lancet. 2021;397(10283):1459-69.

11. "COVID19-India API" [Available from: http://api.covid19india.org.

12. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nature Microbiology. 2020;5(11):1403-7.

13. Velumani A, Nikam C, Suraweera W, Fu SH, Gelband H, Brown P, et al. SARS-CoV-2 Seroprevalence in 12 Cities of India from July-December 2020. medRxiv. 2021:2021.03.19.21253429.

14. Chi X, Yan R, Zhang J, Zhang G, Zhang Y, Hao M, et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. Science. 2020;369(6504):650-5.

15. Suryadevara N, Shrihari S, Gilchuk P, VanBlargan LA, Binshtein E, Zost SJ, et al. Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. bioRxiv. 2021:2021.01.19.427324.

16. McCallum M, Bassi J, Marco A, Chen A, Walls AC, Iulio JD, et al. SARS-CoV-2 immune evasion by variant B.1.427/B.1.429. bioRxiv. 2021.

17. Li Q, Nie J, Wu J, Zhang L, Ding R, Wang H, et al. SARS-CoV-2 501Y.V2 variants lack higher infectivity but do have immune escape. Cell. 2021;184(9):2362-71.e9.

18. Gobeil SMC, Janowska K, McDowell S, Mansouri K, Parks R, Manne K, et al. D614G Mutation Alters SARS-CoV-2 Spike Conformation and Enhances Protease Cleavage at the S1/S2 Junction. Cell Reports. 2021;34(2):108630.

Acknowledgments: Suggestions by Prof Narendra K Arora, Prof G. Padmanabhan and member of the Scientific Advisory Group of INSACOG are acknowledged

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

Competing interests: Authors declare that they have no competing interests

Data and materials availability: All data used in the analysis is either deposited into GISAID or available to any researcher upon request."

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S3

Tables S1

References (*1-24*)