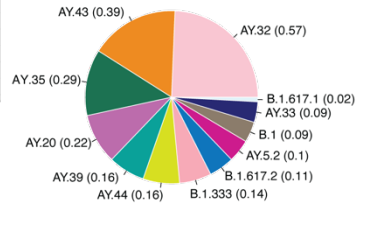
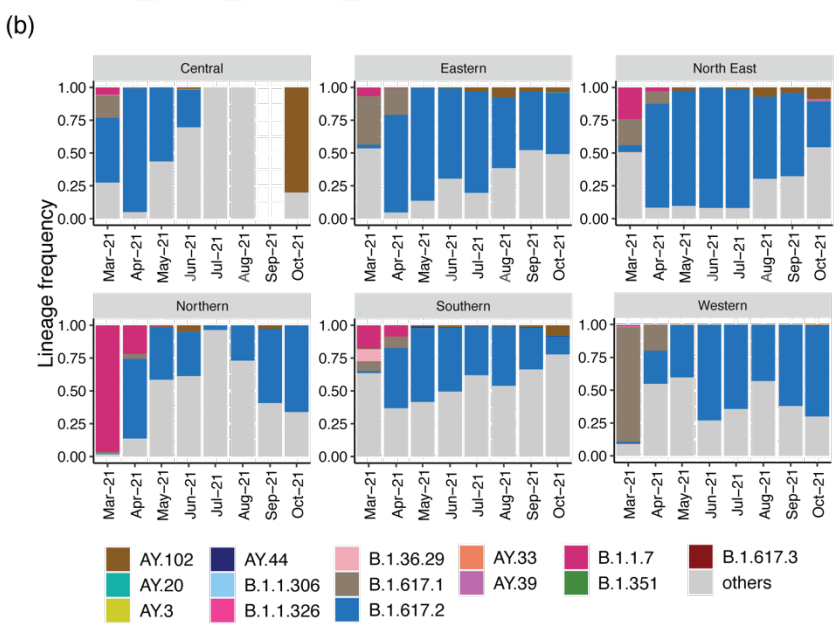
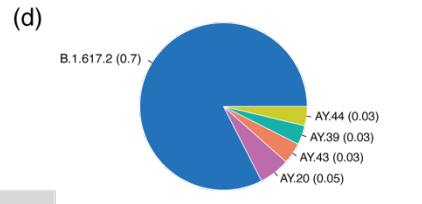
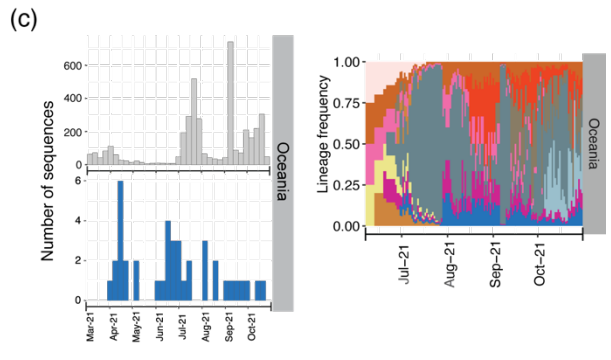
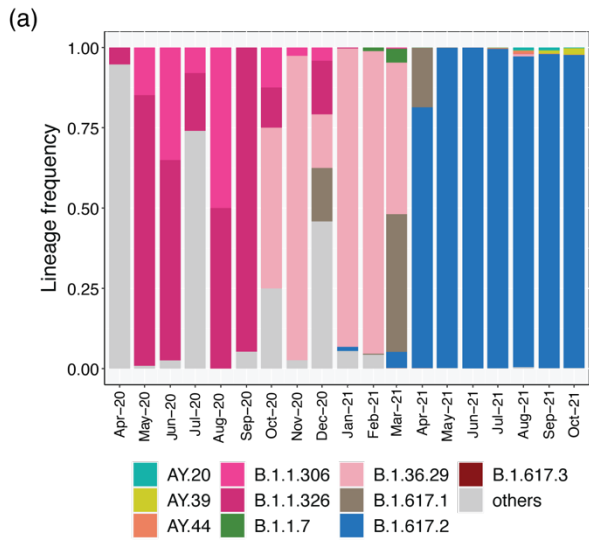


1 Supplementary figures

2 Figure S1 Timeline of major SARS-CoV-2 lineage distribution in Telangana, India
3 and the world. (a) SARS-CoV-2 lineage timeline for Telangana, India. (b) Dominant
4 lineages observed in other parts of India; states are pooled into distinct geographical
5 regions viz. Northern (comprising of Rajasthan, Delhi, Uttar Pradesh, Haryana,
6 Punjab, Uttarakhand, Himachal Pradesh, Ladakh, Jammu and Kashmir,
7 Chandigarh), Western (comprising Gujarat, Maharashtra, Goa, Dadar & Nagar
8 Haveli), Southern (comprising Andhra Pradesh, Telangana, Karnataka, Tamil Nadu,
9 Kerala), Central (Madhya Pradesh, Chhattisgarh), Eastern (West Bengal, Bihar,
10 Jharkhand, Odisha), North East (Mizoram, Tripura, Manipur, Nagaland, Assam,
11 Arunachal Pradesh, Assam, Meghalaya, Sikkim). (c) Timeline of changing frequency
12 of the Delta variant (blue) and total (grey) cases in Oceania, along with frequency
13 change in the Delta sublineages (sequence contribution coming mostly from
14 Australia; color key for lineages is same as in figure 1b). (d) (top) Distribution of
15 lineages assigned to samples designated as vaccination breakthroughs (bottom)
16 fraction of total lineages which were assigned to vaccination breakthrough cases
17 (lineages which were present in <5 samples were excluded).

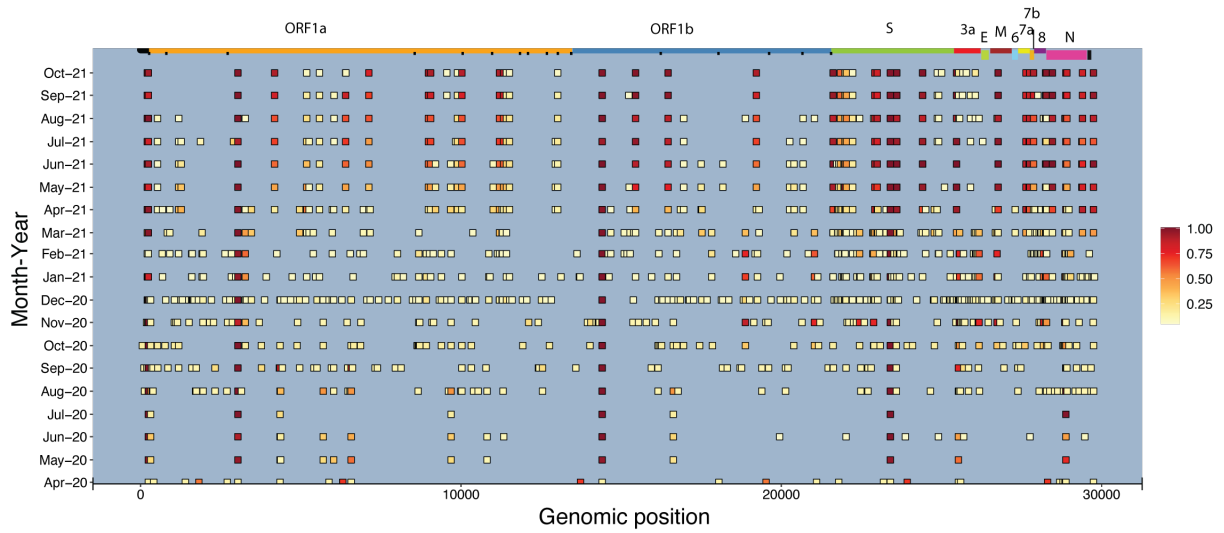


18

19

20

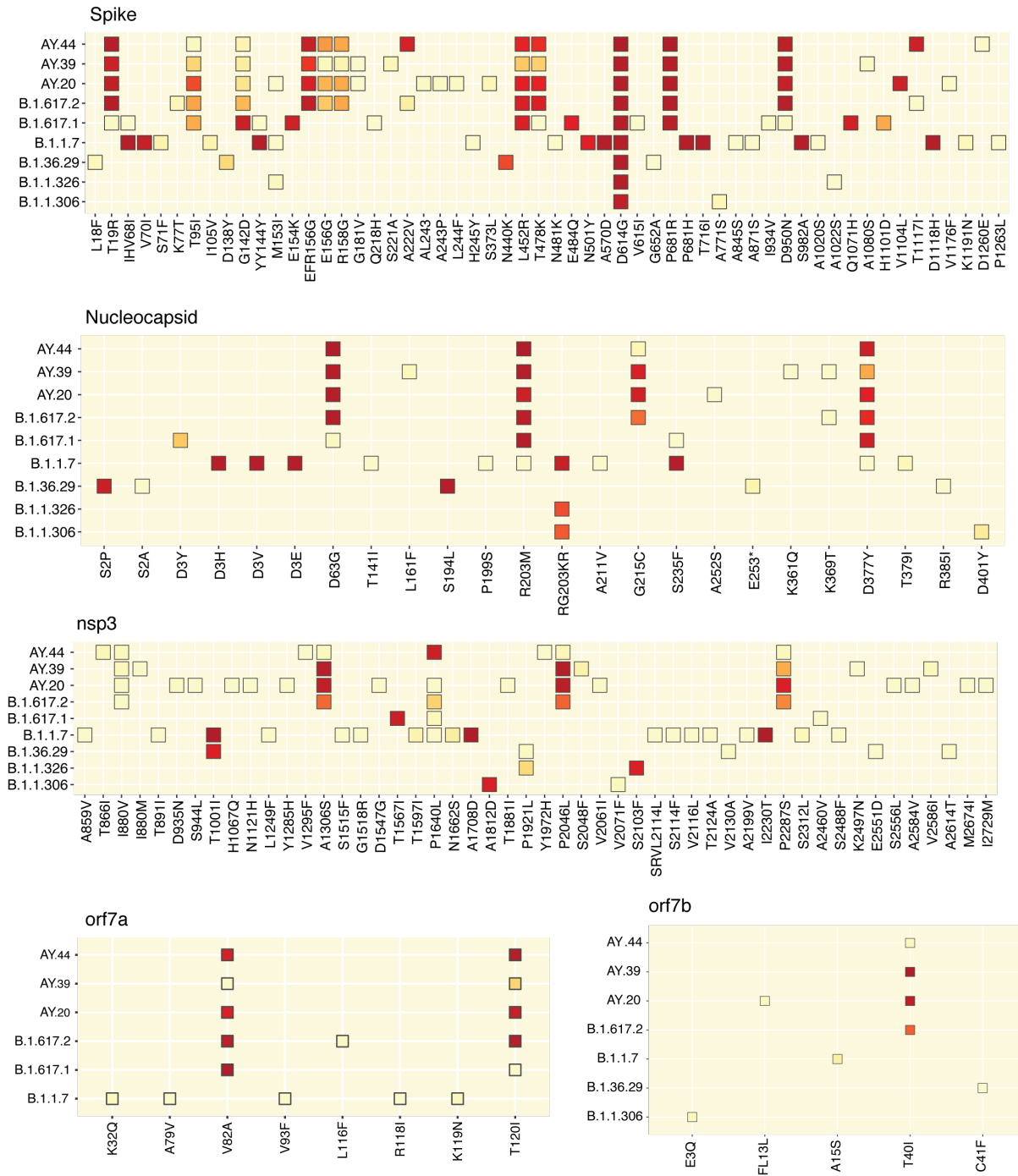
21 Figure S2: Profile of all SARS-CoV-2 NVs identified in >3% samples from
22 Telangana, India. The position of different proteins are shown on top (nsp
23 boundaries in ORF1a and ORF1b are indicated as small vertical black bars).



24

25

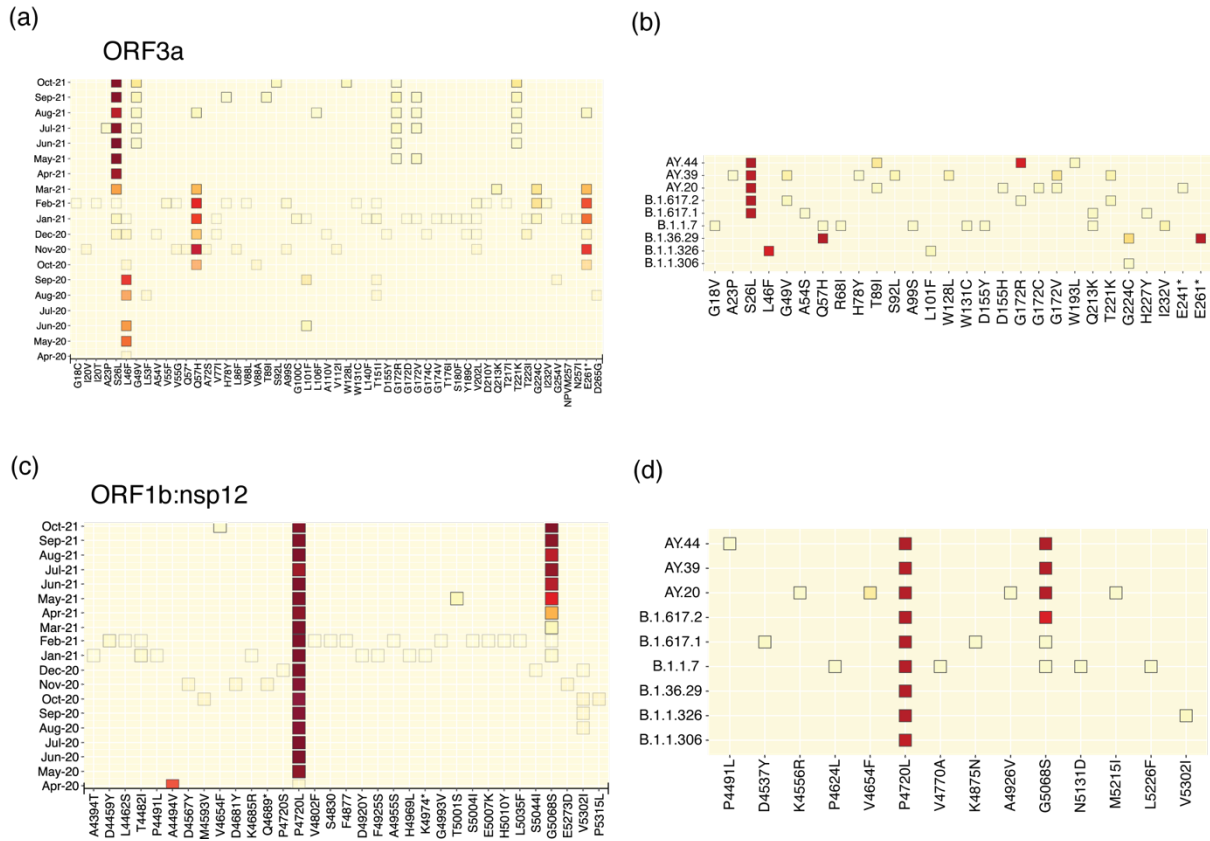
26 Figure S3: Frequency of all missense NVs (found in >3% of samples) in different
 27 virus lineages. The lineages on the y-axis are arranged in the order (bottom to top)
 28 they appeared in the population. Color key of frequency is same as in Figure 2.



29

30

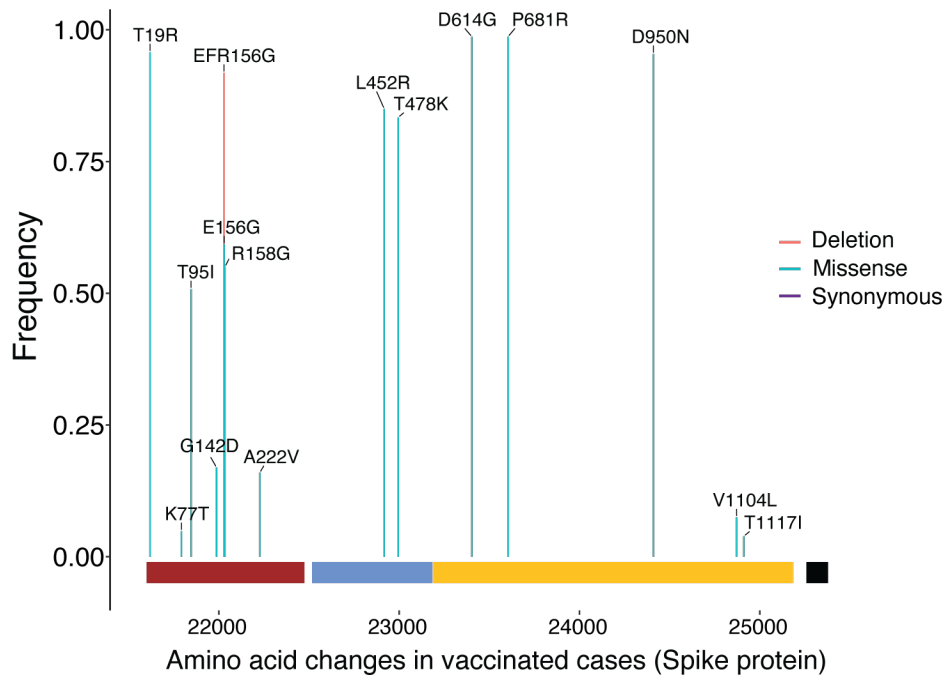
31 Figure S4: Timeline figure indicating month-wise (a, c) and lineage-wise (b, d)
 32 distribution of amino acid variants in ORF3a (a, b) and nsp12 (c, d). Colour scheme
 33 is same as shown in Figure 2.



34

35

36 Figure S5: Frequency of amino acid changes found in >3% of all vaccinated
37 samples.

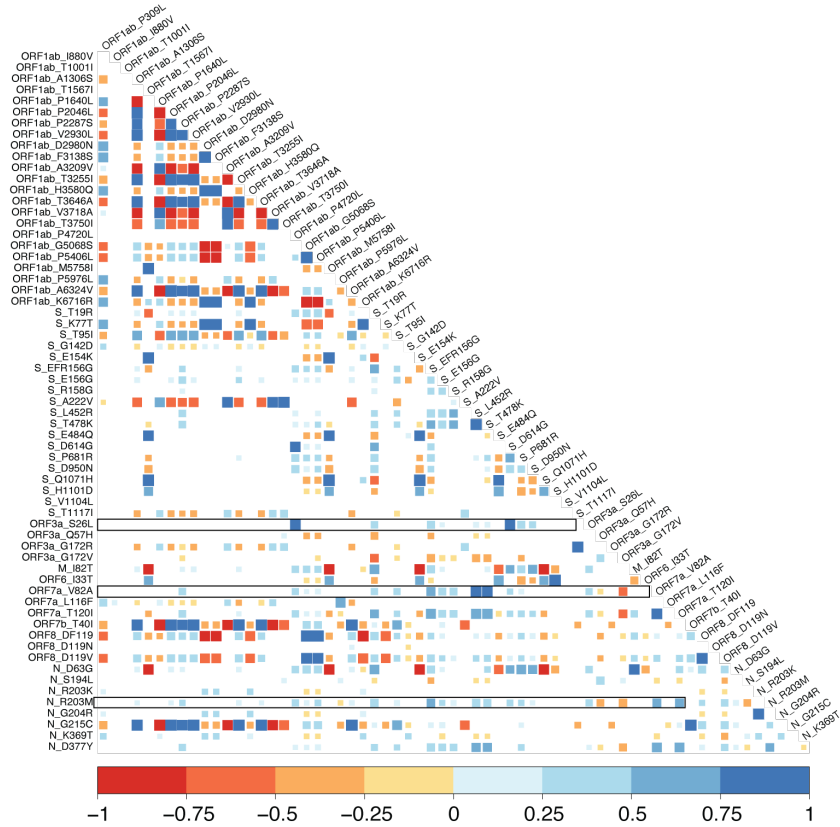


38

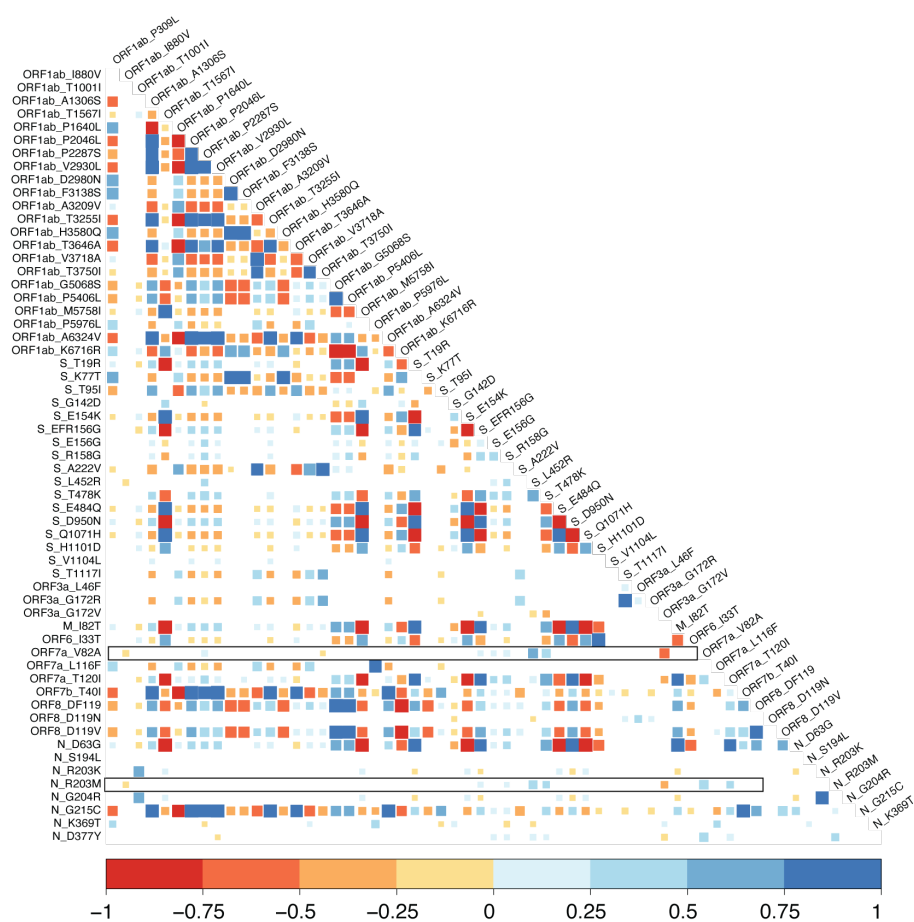
39

40 Figure S6: Comparative NV cross-correlation analysis in vaccinated and non-
41 vaccinated cases. (a) Pairwise NV cross-correlation between all S protein missense
42 NVs found in >3% of samples in vaccination breakthrough cases. (b) NV cross-
43 correlation calculated from a random subset of 350 samples from March 2021
44 onwards consistent with the timeline of vaccination breakthrough samples. Colour
45 scheme is same as in Figure 3c.

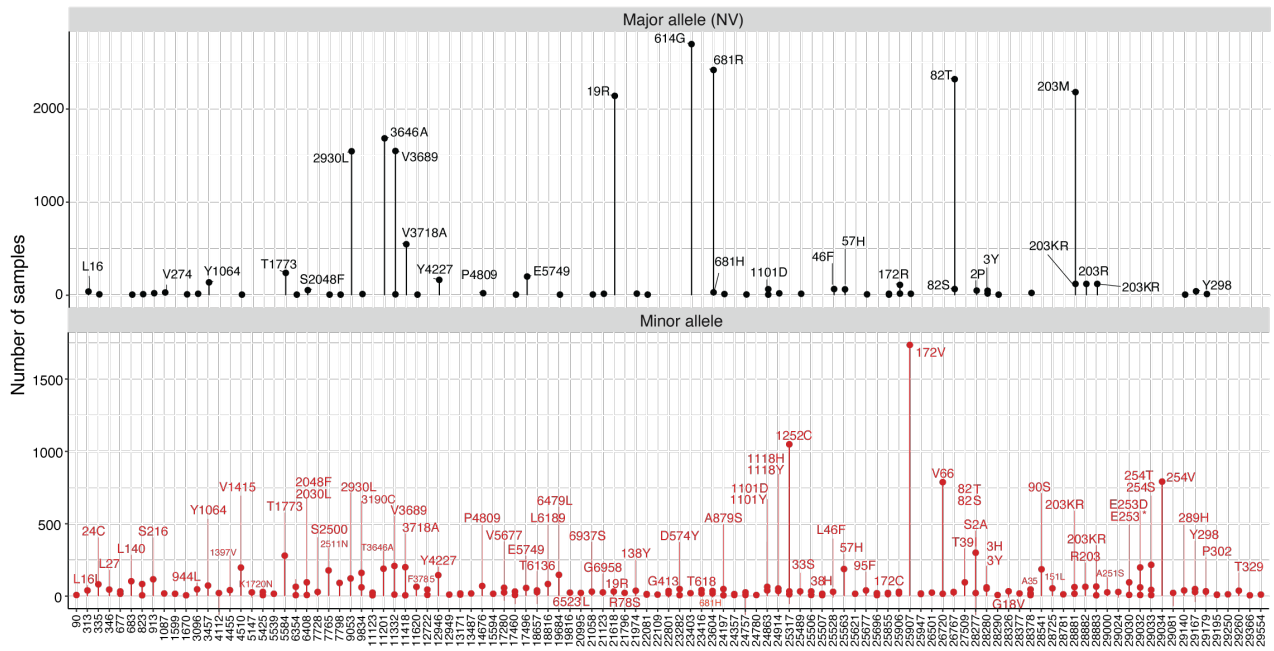
(a)



(b)

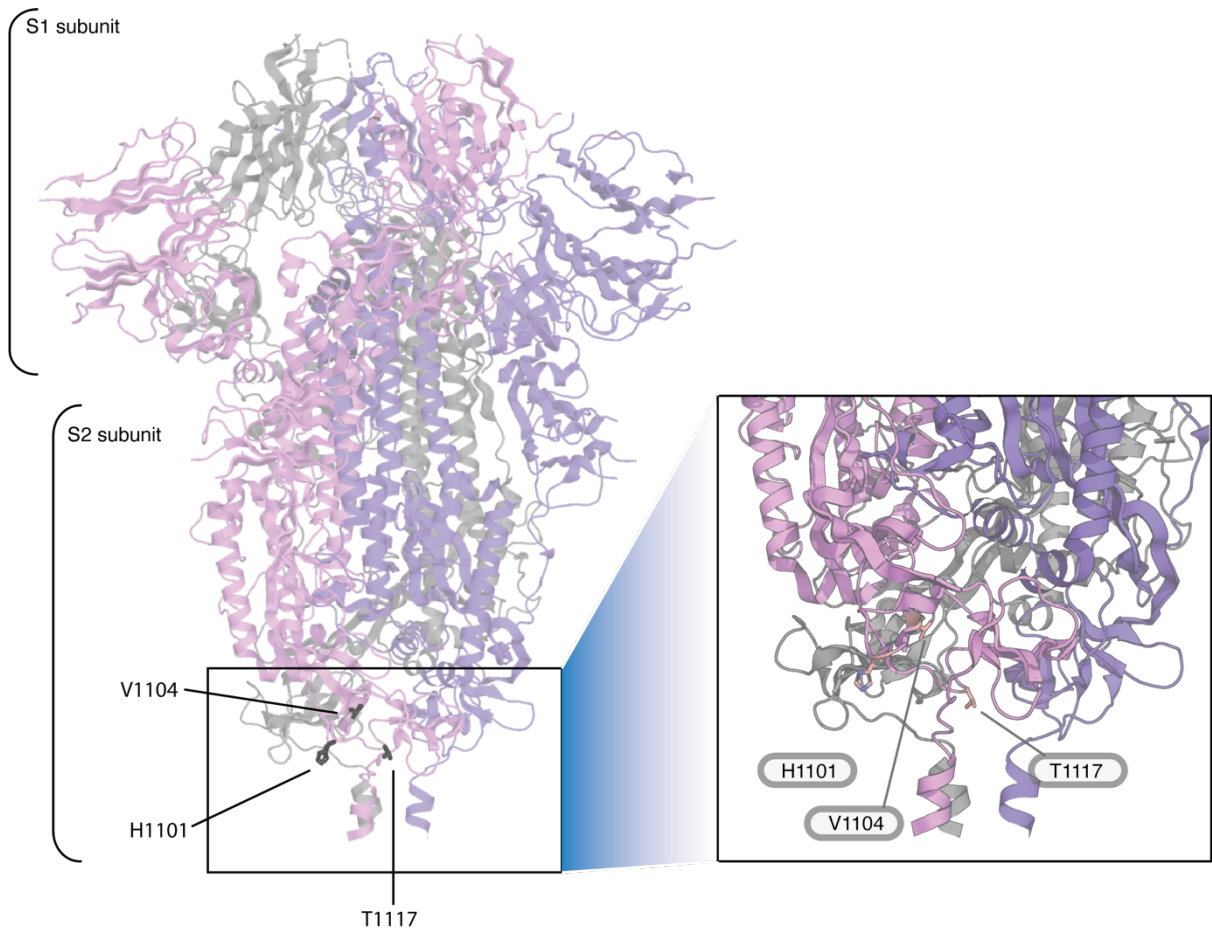


47 Figure S7: Sample wise distribution of all NVs and alternative/minor alleles
 48 generated as a results of iSNVs at genomic loci with shared iSNV and NV (only
 49 alterations, either major or minor, present in > 5 samples are shown; all loci with > 25
 50 samples are labelled)



51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62

63 Figure S8: Location of V1104, H1101, and T1117 mapped on the S protein structure.



64

65

66 **Supplementary data**

67 Table S1A: Metadata associated with all 3543 samples analyzed in this work.

68 Table S1B: Metadata of vaccination breakthrough samples from Tamil Nadu.

69 Table S1C: Metadata of vaccination breakthrough samples from Uttar Pradesh.

70 Table S2: List of accession IDs of all the sequences submitted to GISAID.

71 Table S3: Distribution of all lineages observed in the population till October 2021.

72 Table S4: Odds ratio point estimates and associated p-values for the NVs found in
73 >5% of samples.

74 Table S5: Ct values (for RdRp) of all samples showing presence of >9 iSNVs.

75 Table S6: Distribution of the number of iSNV sites per gene, normalized by the gene
76 size

77

78