Mutational landscape and dominant lineages in the SARS-CoV-2 infections in the state of Telangana, India

Asmita Gupta[1], Radhakrishnan Sabarinathan[2*], Pratyusha Bala[1], Vinay Donipadi[1], Divya Vashisht[1], Madhumohan Rao Katika[3], Manohar Kandakatla[3], Debashis Mitra[1,¶], Ashwin Dalal[1], Murali Dharan Bashyam[1*†]

[1]Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India; [2]National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India; [3]Nizam's Institute of Medical Sciences, Hyderabad, India

[*] These authors are joint senior authors on this work

[†] Corresponding author

[¶] Present address: National Centre for Cell Science, Pune, India.

Abstract

The novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) causing COVID-19 has rapidly turned into a pandemic, infecting millions and causing ~7 million deaths across the globe. In addition to studying the mode of transmission and evasion of host immune system, analysing the viral mutational landscape constitutes an area under active research. The latter is expected to impart knowledge on the emergence of different clades, subclades, viral protein functions and protein-protein and protein–RNA interactions during replication/transcription cycle of virus and response to host immune checkpoints. In this study we have attempted to bring forth the viral genomic variants defining the major clade(s) as identified from samples collected from the state of Telangana, India.

Introduction

The outbreak of COVID-19 caused by the Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) in the Hubei province of China during late December 2019, has since taken the shape of a global pandemic, spreading across more than 200 countries, and resulting in approximately 7,75,581 deaths worldwide, as per recent statistics by World Health Organization (WHO) (https://covid19.who.int/, August 19, 2020). SARS-CoV-2 belongs to the subfamily *Coronavirinae* of the *Coronaviridae* family, classified under the order *Nidovirales*[1]. After host entry, the 29.9 kb positive sense, single-stranded, unsegmented RNA genome of this virus gives rise to 15 non-structural proteins (nsps, 1-15). The subsequent replication and transcription cycles produce a genomic RNA template along with nine subgenomic RNAs, which are further translated to form major structural proteins viz.Spike (S), Envelope (E), Nucleocapsid (N) and Membrane (M) proteins [2]. They share a high degree of sequence similarity with Severe Acute Respiratory Syndrome coronavirus (SARS-CoV, >80%) and a moderate sequence similarity with Middle East Respiratory Syndrome coronavirus (MERS-COV, >50%)[3]. The severity of the pandemic has demanded a concerted effort to comprehensively study the actively changing mutational landscape of the virus across multiple demographic locations and identify potential variant clusters or individual variants which might play a critical role in its spread and transmission dynamics.

The state of Telangana, located in south-central India, has seen an unusually high rate of infection and there appears to have been a sharp spike in the number of cases beginning from the second half of April 2020. In this study, we aimed at identifying the dominant lineages present in the samples collected in Telangana, among all identified lineages of SARS-CoV-2. We further endeavoured to draft a comprehensive mutational landscape of the viral genome from among patient samples in Telangana state. Towards this end, we applied next generation sequencing to determine the complete sequence of 210 SARS-CoV-2 RNA samples.

Data and Methods

Sample collection and processing

The Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, initiated reverse transcription-PCR (RT-PCR) based diagnostics for Covid-19 infection after approvals from Secretary, Department of Biotechnology (DBT), Government of India, Indian Council of Medical Research (ICMR) nominated nodal officer in Hyderabad, Telangana as well as from the Telangana state government. The work was initiated following approvals from the Institutional Bioethics committee and Biosafety committee. The samples were obtained as nasopharyngeal swabs collected in Viral Transport Medium from different parts of Telangana from patients with symptoms suggestive of Covid-19 as well as asymptomatic primary contacts of affected cases. The samples were transported to CDFD within 24 hours while maintaining a cold chain. Total RNA was isolated using the RNA isolation kit as per manufacturer's instructions (QIAmp Viral RNA Mini Kit; Cat #52906; Qiagen, Hilden, Germany). Each RNA sample was subjected to RT-PCR for multiple viral genes (including E-gene and RDRP gene) using the LabGun COVID-19 assay (Cat #CV9017B; LabGenomics, Republic of Korea) or the Allplex 2019-nCoV Assay (Cat #RP10250X, Seegene, Republic of Korea). The samples which tested positive in RT-PCR analysis were included for viral genome sequencing. Since RDRP consistently provided more robust amplification than E-gene and is a SARS-COVID-2 specific gene (unlike E-gene which is specific for all respiratory coronaviruses), we considered Ct values of RDRP gene alone for analysis. Samples exhibiting an RDRP Ct value greater than 10 and less than 35 were chosen for sequencing.

Sequencing protocol

Sequencing of SARS-CoV-2 RNA samples was performed using protocol described earlier (Quick et al., 2020) with slight modifications. Briefly, RNA isolated from nasopharyngeal swabs was reverse transcribed using random primer mix (New England Biolabs, Massachusetts, United States), and Superscript-IV (Thermofisher Scientific, Massachusetts, United States). The resulting cDNA was subjected to a 3-step multiplex PCR using nCoV-2019/V3 primer pools (Eurofins, India) 1, 2 and 3. The ~400 bp amplicons thus obtained in the pools were combined, purified using Agencourt AMPure XP beads (Beckman Coulter, California, United States) and eluted in 45µl elution buffer

(Qiagen, Hilden, Germany). DNA libraries for Illumina sequencing were prepared using the NEB Next Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Massachusetts, United States), according to the manufacturer's protocol. Paired-End Sequencing (2×250 bp) was performed on the Miseq FGx (Illumina Inc, California, United States) with a targeted depth of 0.5 million reads per sample (~4,000X coverage). Libraries for Nanopore sequencing were prepared using the Ligation sequencing kit (LSK-109; Oxford Nanopore Technologies, London, United Kingdom). Barcoded libraries were pooled (12-24 samples each) and sequenced on a MinION flow cell in GridION (Oxford Nanopore Technologies, London, United Kingdom). Sequencing was performed with a targeted depth of 0.1 million reads per sample (up to 24 hours).

Mutation analysis

All raw fastq files from Illumina were checked for overall sequencing quality, presence of adapters and bad quality reads using FastQC and Fastp[4]. The adapter sequences were trimmed using a wrapper script for Cutadapt[5] tool, called Trim Galore. The filtered reads were aligned to the reference strain NC_045512.1, Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, using bwa-mem[6] algorithm with default parameters. Mapping quality was assessed using samtools[7] and BAMStats. Post alignment, the reads were filtered, sorted and indexed using samtools, and any primer sequences were masked using iVar[8]. Subsequent mutation calling and generation of consensus sequence was performed using samtools mpileup and iVar (https://github.com/connor-lab/ncov2019-artic-nf/). The resulting VCF files were annotated using snpEff[9]. For processing the nanopore data, we followed the protocol suggested by ARTIC pipeline (https://github.com/artic-network/fieldbioinformatics, https://github.com/connor-lab/ncov2019-artic-nf/) for mutation calling as well as for assembling the reads for generating the consensus sequence. A schematic describing the entire workflow is shown in **Supplementary Fig S1.** Before analysing the obtained calls, we filtered all the problematic sites prone to errors by multiple sources as recommended by De Maio et al (https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473).

Phylogenetic analysis

The consensus fasta files generated for both Illumina and Nanopore data were subjected for phylogenetic analysis using the Nextstrain pipeline with recommended

default criteria for filtering, multiple sequence alignment (MSA) and nucleotide substitution calculations. To briefly summarize the workflow of the pipeline, all the consensus sequences having length < 27000 and Ns > 5% were filtered out. Three samples were removed from further analysis as their sequence data included Ns > 5%; thus all analyses were conducted on 207 individual patient viral genome sequences. A compendium of problematic sites as used earlier (https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473), was also provided to mask those sites prior to MSA by MAFFT[10]. Following MSA, the workflow constructed a time-resolved phylogenetic tree using the maximum likelihood based method IQ-TREE[11]. The resultant tree was pruned and internal nodes and ancestral traits were inferred from the dates of the sample collection using TreeTime[12]. The final tree in Newick format was then customized for visualization using iTol[13].

Results

General sample features

Our dataset consists of samples collected during late March to July, 2020. Interestingly, samples collected from late May till July represented a higher proportion of asymptomatic cases when compared to samples collected earlier (**Figure 1a**). A majority of our samples belonged to age group between15-62 years, with males (61%) dominating the profile distribution over females (39%) (**Figure 1b**). We also compared the distribution of cases with respect to Ct values, the latter being a proxy for viral load. Symptomatic cases appeared to be associated with higher Ct values (thus lower viral load), compared to asymptomatic ones, which was unexpected (**Figure 1c**). there was a reduction in the Ct values in samples, as we neared the end of June, 2020, implying that more recent samples seemed to carry a higher viral load than earlier samples (**Supplementary figure S2a)**. The distribution of Ct values with respect to age showed no clear correlation (**Supplementary figure S2b**). However, there was a consistent rise in the fraction of samples showing asymptomatic behaviour, as calculated from a cumulative increment, as opposed to symptomatic, starting from the end of May (**Figure 1d**). The locale distribution of the samples in and around the district of Hyderabad is indicated in **Supplementary figure S2c.**

Phylodynamic epidemiological clustering of SARS-Cov2 strains

The phylodynamic clustering of samples suggested presence of 2 minor (19A, 20A) and one major clade (20B). The phylogenetic clustering was rooted with respect to the sequence from the Wuhan Wu-1 strain (NC_045512.1), which has been annotated as the base clade 19A. One sample (collected in April, 2020) seemed to directly emerge from this base clade. Further, five samples harbored the C13730T mutation linking them to the base clade (establishing sub-clade 19A/C13730T). This split was marked by missense mutations G1820A (ORF1a, G519S), C6310A (ORF1a, S2015R), C6312A (ORF1a, T2016K), C28311T (N, P13L), and two synonymous mutations C19524T (ORF1a, L6425L) and C23929T (S, Y789Y). The first major split in our dataset was observed at around the starting of April, marking the onset of 20A clade, and was characterized by the appearance of co-occurring mutations C241T (5'-UTR), synonymous C3037T (ORF1ab, F924F), and missense C14408T (ORF1ab, P4720L) and A23403G (S, D614G). Seven samples belonged to this 20A clade, directly linked to

the dominant viral lineage established in Europe (Belgium, Italy, of which one exhibited a distinct 20A/18877T profile. This division retained none of the mutations found in the initial samples belonging to 19A and 19A/C13730T cluster (**Supplementary figure S3**). From the middle of April and onwards, the profile was strongly dominated by the 20B clade, which also formed the second cluster within this division, with a characteristic mutation of GGG28881/28882/28883>AAC (**Figure 2 and Supplementary figure S4.** Thus, of all the clades identified, our dataset was strongly populated by the presence of a single major clade 20B. From a time resolved mutational map calculated for all samples (**Supplementary figure S4**), we observed that the more recent samples, collected from the end of June onwards, did not show many of the mutations found in ORF1a, C5700A (nsp3, A1812D), C6573T (nsp3, S2103F) and C25528T (ORF3a, L46F). These samples also coincided with the most diverged samples along the phylogenetic time-tree, indicating towards a newer divergence path of the virus among the later infections. We checked whether these samples belonged to any single cluster within the phylogenetic tree, but they were found to be interspersed among different branches (**Figure 2, Supplementary figure S4**). Within the dominant 20B clade, a major proportion of samples between the age group of 15 and 50 were found to be asymptomatic, while the symptomatic patients mainly belonged to samples collected from age group >50. As also discussed before, the viral load, calculated in terms of RdRp gene, was found to be primarily associated with asymptomatic patients. This behaviour can explain the higher transmission rates of the virus among the Indian population, in general, as opposed to per capita mortality rates, which will be further elaborated upon in the next section. A location wise distribution indicating the neighbourhood origin of the samples in the phylogeny tree has been shown in **Supplementary figure S5**.

We performed Nextstrain analysis separately on our Illumina and nanopore datasets, in order to rule out the possibility of incorporation of any bias in cluster formation due to the sequencing platform. The clusters obtained did not show any segregation due to the platform used and were distributed across the clades (data not shown).

Mutational landscape of SARS-CoV-2

From the mutation analysis on the filtered, combined pool of 207 sequences from Illumina and Nanopore data, we obtained a total 302 mutations across the SARS-CoV-2 genome (**Supplementary data D1**). From this set, 17 mutations were consistently found to be present in >10% of samples (**Figure 3a**). The proportion of asymptomatic cases for each of these 17 high frequency mutations, as a fraction of the total number of cases, was found to be higher than symptomatic cases (**Figure 3b**). Further, for each of the 17 high frequency mutation identified in this set, we calculated the cumulative increment in the frequency of case symptoms as a function of time and noted a consistent increase in the frequency of asymptomatic cases specially for missense mutations (**Supplementary figure S6**). Of these 17 high frequency mutations, 5 viz. C241T (5'-UTR), C3037T (ORF1a), C14408T (nsp12, ORF1ab), A23403G (S), GGG28881-3AAC (N) were highly recurrent (present in >80% of the samples; **Figure 3a**). The A23403G (D614G) mutation in Spike protein was identified in samples as early as beginning of April, 2020. Although a highly recurrent mutation in multiple demographics, no clear correlation has been established between D614G mutation and severity of disease [14]. The D614G mutation has almost invariably been found to be associated with C241>T, C3037>T (a silent mutation) and a mutation in RNA dependent RNA polymerase gene, nsp12 C14408>T as has also been reported earlier [15]. The haplotype defined by the co-occurrence of these 4 mutations is the current dominant form circulating across the world. Apart from these four mutations, the nsp3 protein region within ORF1a also displayed a higher frequency of mutations, G4354A, A4372G, C5700A, C6027T, C6573T, the latter three being missense mutations. Of these, C5700 along with a silent C313T mutation has been reported to co-occur in samples collected from Western state of Maharashtra, India [16]. The nsp4 and nsp5 proteins each harbored one high frequency missense mutation each namely C9693T and C10815T, respectively. Similarly, the ORF3a region possessed one high frequency missense mutation C25528T.

Discussion

In the current study, we have presented a comprehensive map of the mutations identified from the confirmed COVID-19 cases collected from the southern state of India, Telangana. After a slow progress of the outbreak during the months of February-April, the state has been witnessing a constant upsurge in the number of infections and has been listed as one of the worst affected states in the country. Identifying the mutations from samples collected over a period of time, provides a way to assess the genomic diversity which the virus might have experienced during infection and transmission. With these aspects in our purview, we have attempted to characterize the genomic epidemiology of novel coronavirus and a comprehensive mutational landscape, using a dataset of 210 samples sequenced using both Illumina and Nanopore sequencing technologies. Most of our samples were collected from mid-March with the months of June-July reporting highest collections. A significantly higher association of samples with asymptomatic behaviour was noted, which were also associated with lower Ct values. We also observed an upsurge in the asymptomatic cases compared to symptomatic. A majority of our samples belonged to the 20B clade, with the clade seeming to appear in the beginning of April. Although we did not see a clear correlation between age and viral load, it was observed that samples collected from higher age groups frequently displayed symptomatic behaviour.

More importantly, mutational analysis revealed the presence of unique mutations in the samples from Telangana, especially in the nsp3, nsp4, nsp5 and ORF3a. The nsp3 is a papain-like protease (PLP2) and nsp5 is a 3C-like protease (3CLpro), both required for cleavage of polyproteins pp1a and pp1ab to generate 16 non-structural proteins (nsp1-16) [17]. Nsp3 is the largest multidomain protein encoded by SARS-CoV-2, and binds to viral RNA and nucleocapsid protein [18]. The mutations identified in nsp3 viz. C6027T (P1921L), C5700A (Ala1812Asp) and C6573T (S2103F) lie in the protease like (PLP2) and protease like non-canonical (PLnc) domains of the protein. The PLnc domain has been shown to interact with a large number of nsp partners during viral replication and transcription cycle and acts as a scaffold for membrane associated replication/transcription complex (RTC) formation. The formation of RTC also requires nsp3 association with nsp4 and nsp5 [19]. Hence, it could be speculated that these mutations can potentially alter the balance or stoichiometry of the complex formation,

perhaps causing viral RTC reprogramming. The RTC complexes in coronaviruses have been shown to perform multiple tasks during replication and transcription by being able to form compositionally diverse complexes, each specific to carry out a particular process. It is possible that these mutations could bring about a switch in this compositional programming [19].

SARS-CoV-2 ORF3a encodes an accessory, transmembrane, ion-channel protein, belonging to the class of proteins called Viroporins, which induce innate immune signalling receptor NLRP3, leading to production of cytokines IL-1β, IL-6 and tumor necrosis factor (TNF) [20]. This often results in tissue inflammation in a SARS-CoV-2 infection. The most common mutations identified in ORF3a among the prevalent lineages across the world is Q57H and G251V [21]. However, the C25528T (L46F) mutation observed in this study localizes to the same domain I as Q57H, in the N-terminal signal-peptide region of the protein. This region is critical for viral localization to the Golgi apparatus. Functional impact of these mutations on the location, and subsequent translation and assembly, might require additional experimental evidences.

In addition to these high frequency mutations, we also observed a few low frequency mutations in the Spike, N , nsp3 and nsp2 proteins. A detailed study of these mutations and their functional impact on the viral life cycle is currently underway.

In conclusion, we report the comprehensive mutation landscape of more than 200 individual SARS-CoV-2 viral genome sequences isolated from COVID-19 patients or primary contacts. We report the 20B clade as the major clade in Telangana which is similar to reports from West India. The data forms a starting point for the state government machinery to conduct further studies on virus transmission helping in taking informed public health decisions. The genomic mutations also inform on potential mechanisms being employed during evasion of the host's immune response and traces of higher or lower pathogenicity, if any, being developed over the span of time, thus significantly impacting efforts in vaccine development.

Acknowledgements

We are thankful to Drs Rashna Bhandari and R Harinarayanan, CDFD, Hyderabad, for co-ordinating the establishment of the COVID- 19 testing laboratory at CDFD. All volunteers and 'COVID warriors' from CDFD, Hyderabad, are gratefully acknowledge for their significant contribution in screening of samples. We are grateful to all Government district hospitals in the Telangana State for facilitating screening of suspected COVID-19 patients. We also thank the Telangana State Government, the Indian Council of Medical Research, Government of India, the University of Hyderabad, Hyderabad and CDFD, Hyderabad, for procurement of consumables and equipment to perform screening of patient samples and the Department of Biotechnology, Government of India for financial support through the National Genomics Core project (BT/INF/22/SP28169/2019,07/03/2019). R.S. acknowledges funding support from NCBS-TIFR and Ramanujan Fellowship (SB/S2/RJN-071/2018). A.G. acknowledges funding support from Science and Engineering Research Board, Department of Science and Technology (DST-SERB) in the form of National-Postdoctoral Fellowship (N-PDF).

Figures

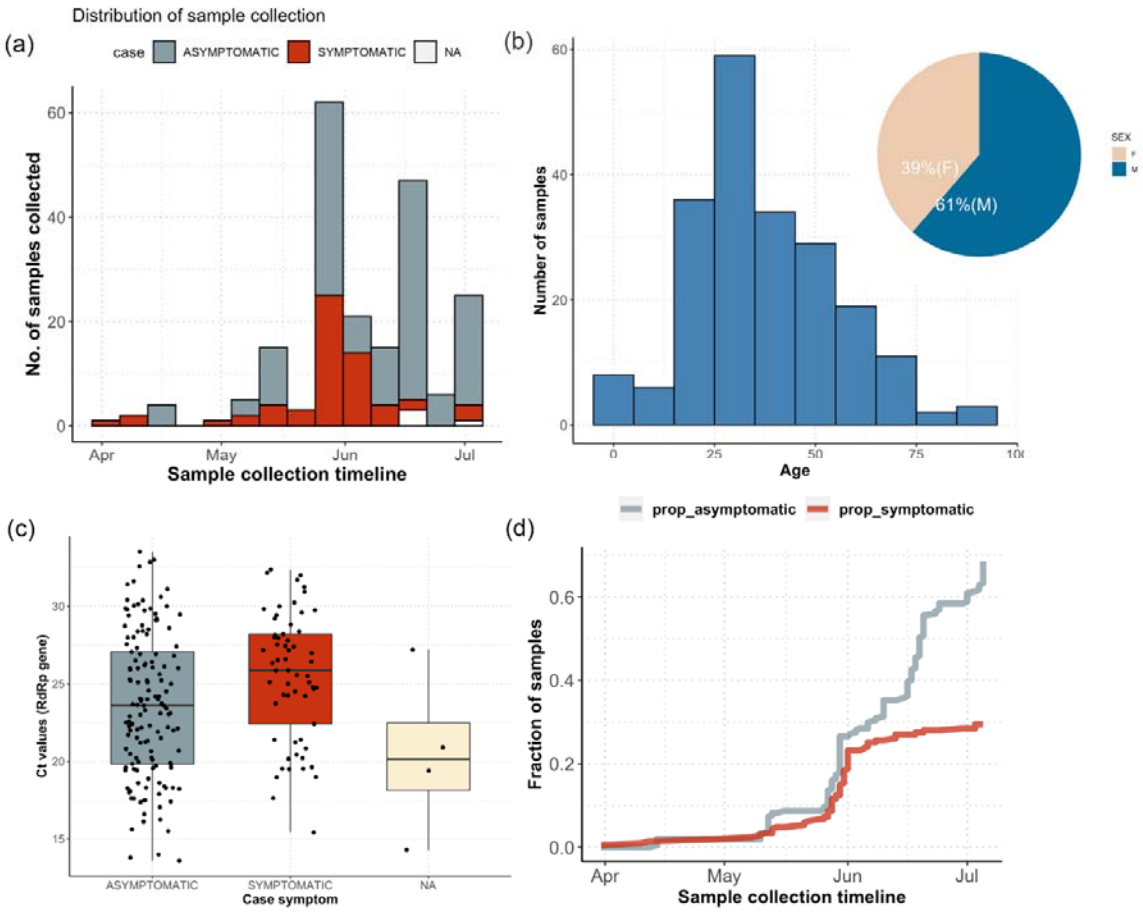


Figure 1: Characteristics of the dataset used in this study. (a) Distribution of sample collection timeline (b) age and gender distribution across samples (c) distribution of symptoms with respect to Ct values of RdRp gene (nsp12). (d) cumulative increment of symptomatic and asymptomatic fraction of samples along the sample collection timeline.

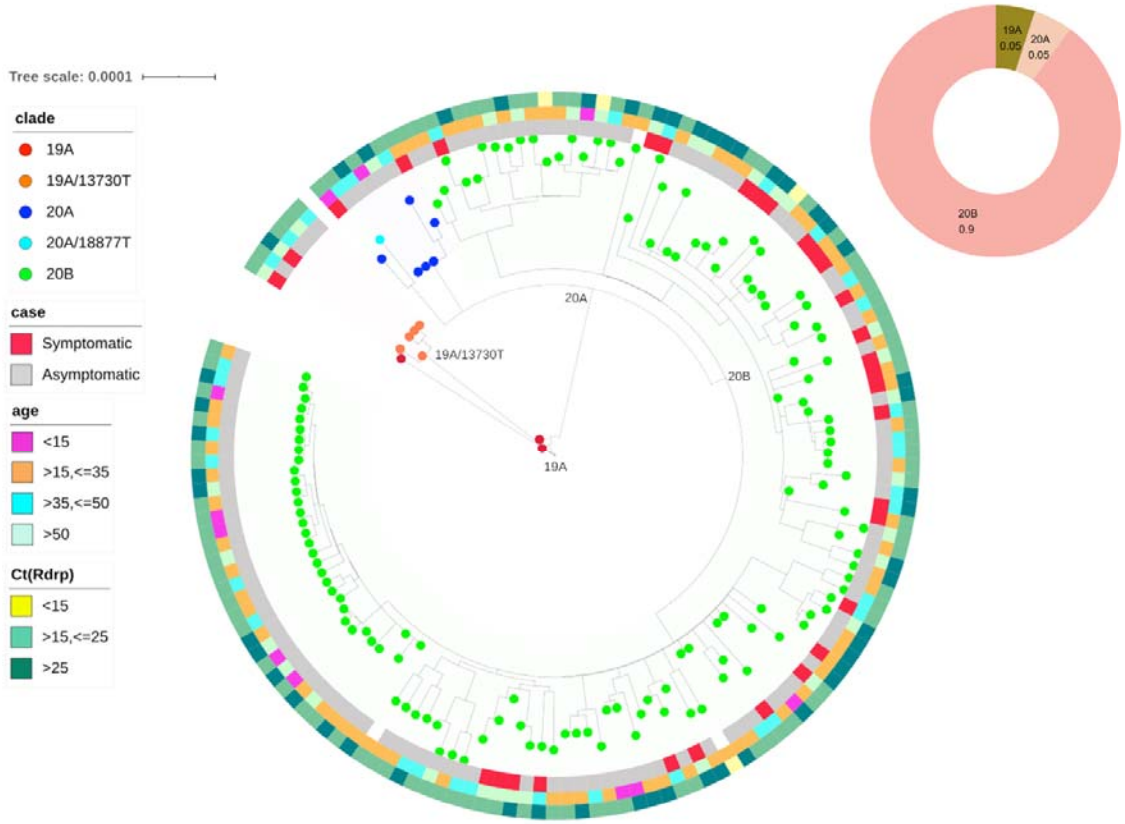
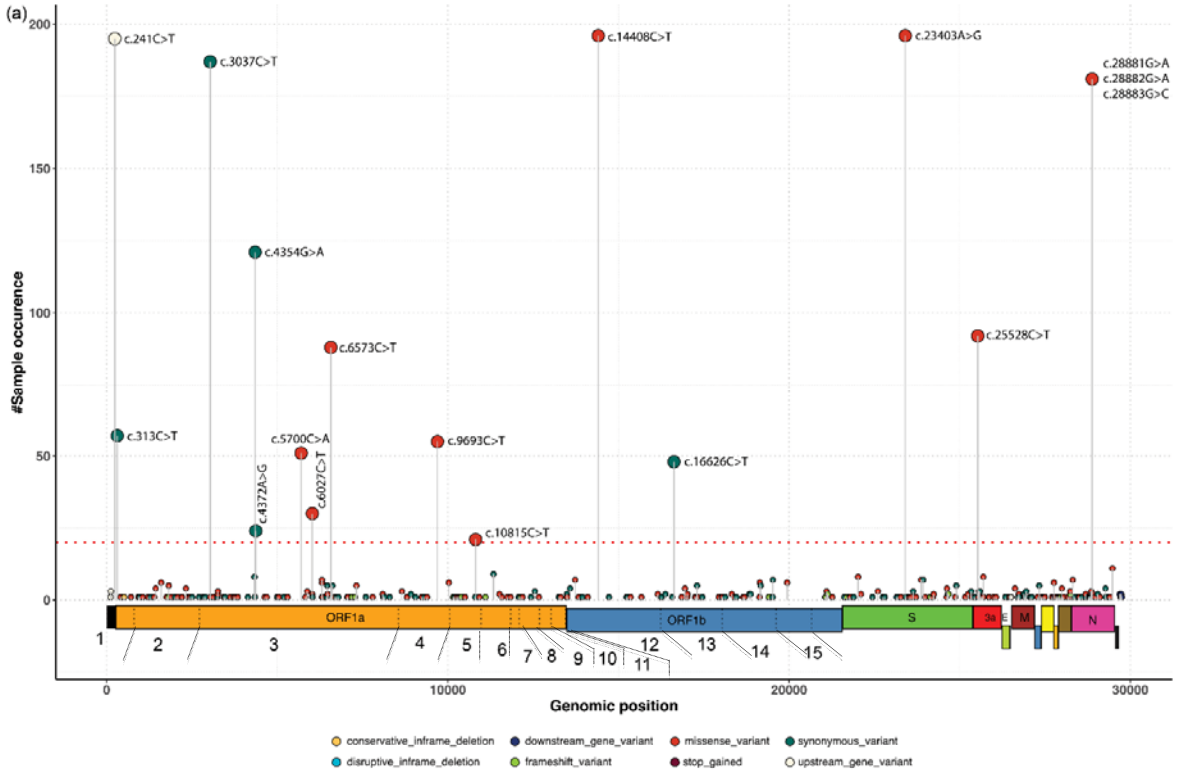


Figure 2: Phylodynamic tree of the samples as analyzed using Nextstrain (a) Time resolved phylogenetic tree representing all 205 samples which passed filtering criteria during Nextstrain analysis. (b) Distribution of clades across the dataset.



(b)

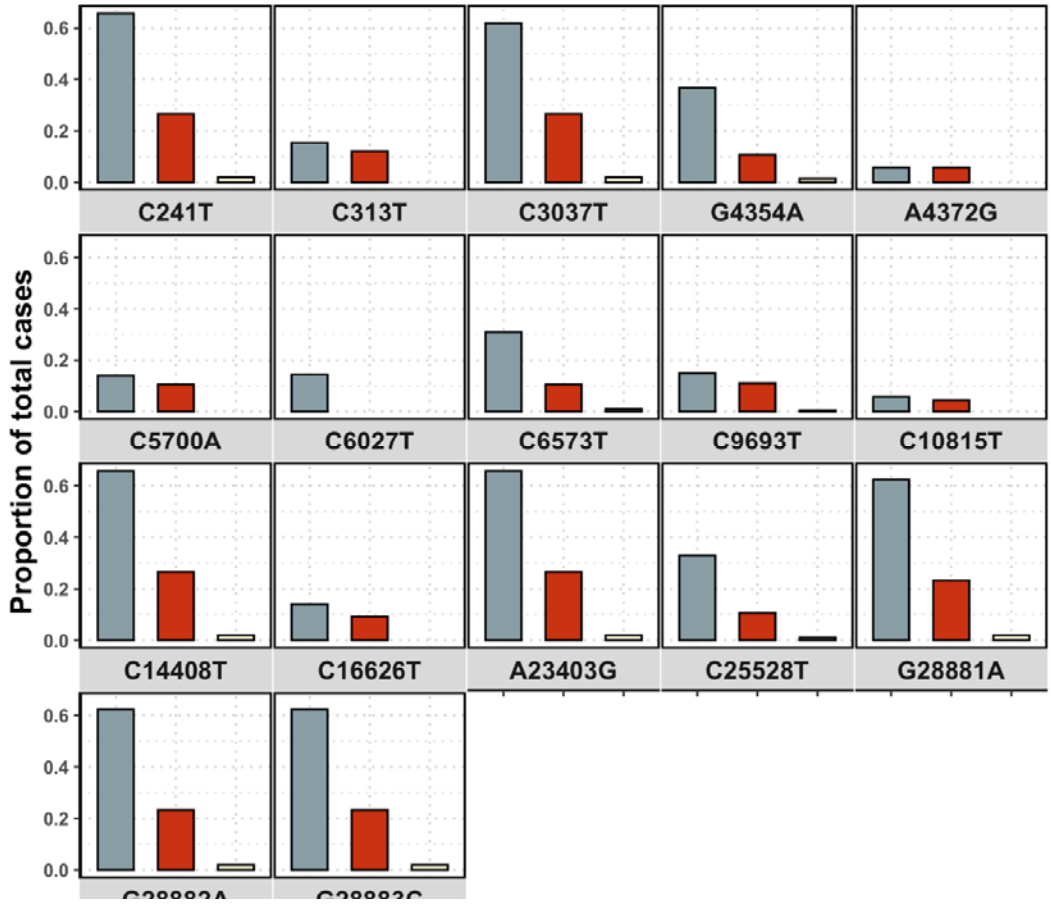


Figure 3: Distribution of all high frequency mutations as called on Illumina and nanopore sequencing data (a) genomic locations of all high frequency mutations (samples > 10%, indicated by red, dashed horizontal line) (b) case distribution of sample with respect to all indicated mutations. An associated cumulative frequency of cases with respect to mutations is provided in supplementary figure S6.

References

1. Wang, H. *et al.* The genetic sequence, origin, and diagnosis of SARS-CoV-2. *Eur. J. Clin. Microbiol. Infect. Dis.* (2020) doi:10.1007/s10096-020-03899-4.
2. Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**, 914-921.e10 (2020).
3. Kim, J.-M. *et al.* Identification of Coronavirus Isolated from a Patient in Korea with COVID-19. *Osong Public Health Res Perspect* **11**, 3–7 (2020).
4. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
5. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **17**, 10 (2011).
6. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
7. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
8. Grubaugh, N. D. *et al.* An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol* **20**, 8 (2019).
9. Cingolani, P. *et al.* 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 (Austin)* **6**, 80–92 (2012).
10. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059–3066 (2002).
11. Minh, B. Q. *et al.* Corrigendum to: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 2461–2461 (2020).

12. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol* **4**, vex042 (2018).

13. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* **47**, W256–W259 (2019).

14. Zhang, L. *et al.* The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv* (2020) doi:10.1101/2020.06.12.148726.

15. Korber, B. *et al. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2*. http://biorxiv.org/lookup/doi/10.1101/2020.04.29.069054 (2020) doi:10.1101/2020.04.29.069054.

16. Paul, D. *et al. Phylogenomic analysis of SARS-CoV-2 genomes from western India reveals unique linked mutations*. http://biorxiv.org/lookup/doi/10.1101/2020.07.30.228460 (2020) doi:10.1101/2020.07.30.228460.

17. Angeletti, S. *et al.* COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *J. Med. Virol.* **92**, 584–588 (2020).

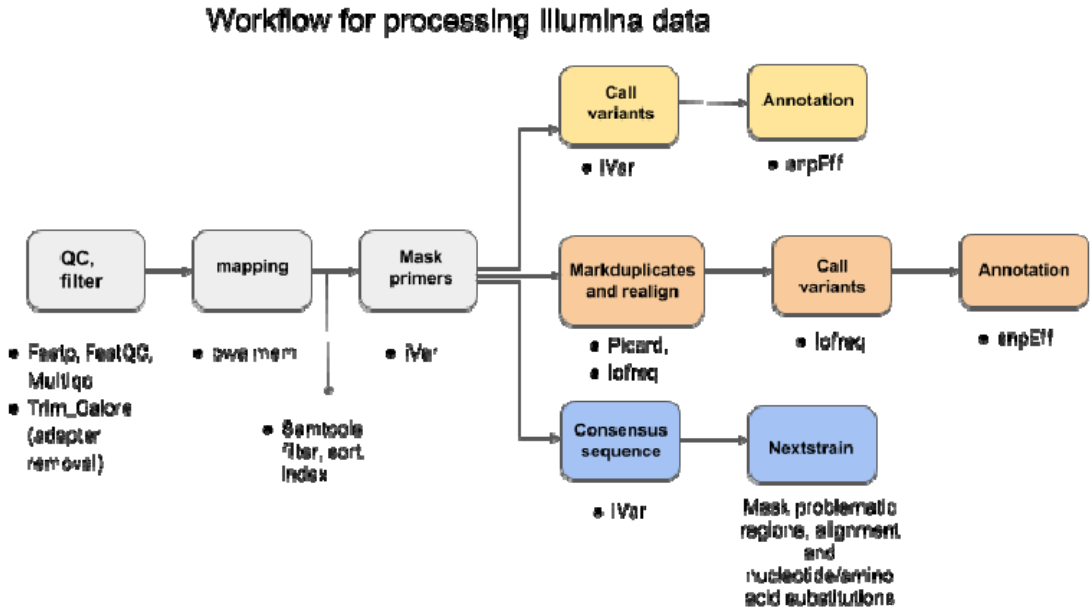
18. Lei, J., Kusov, Y. & Hilgenfeld, R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Research* **149**, 58–74 (2018).

19. Imbert, I. *et al.* The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. *Virus Res.* **133**, 136–148 (2008).

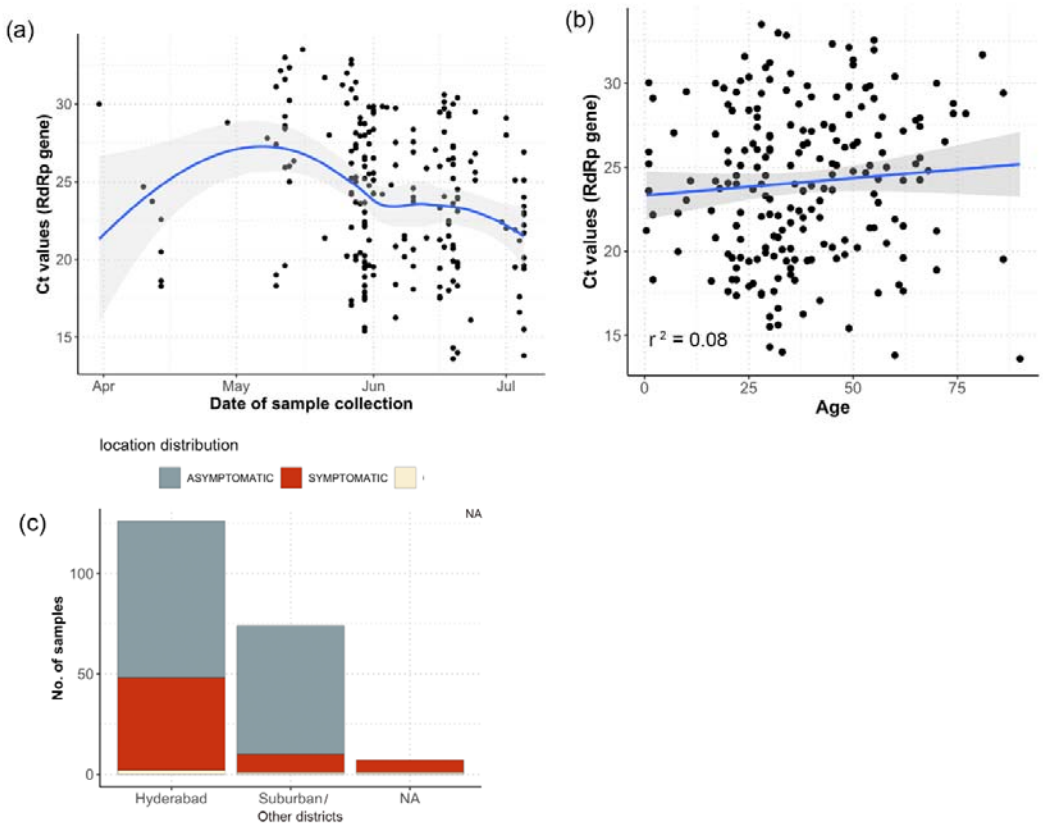
20. Siu, K. *et al.* Severe acute respiratory syndrome Coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. *FASEB j.* **33**, 8865–8877 (2019).

21. Issa, E., Merhi, G., Panossian, B., Salloum, T. & Tokajian, S. SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis. *mSystems* **5**, e00266-20, /msystems/5/3/msys.00266-20.atom (2020).

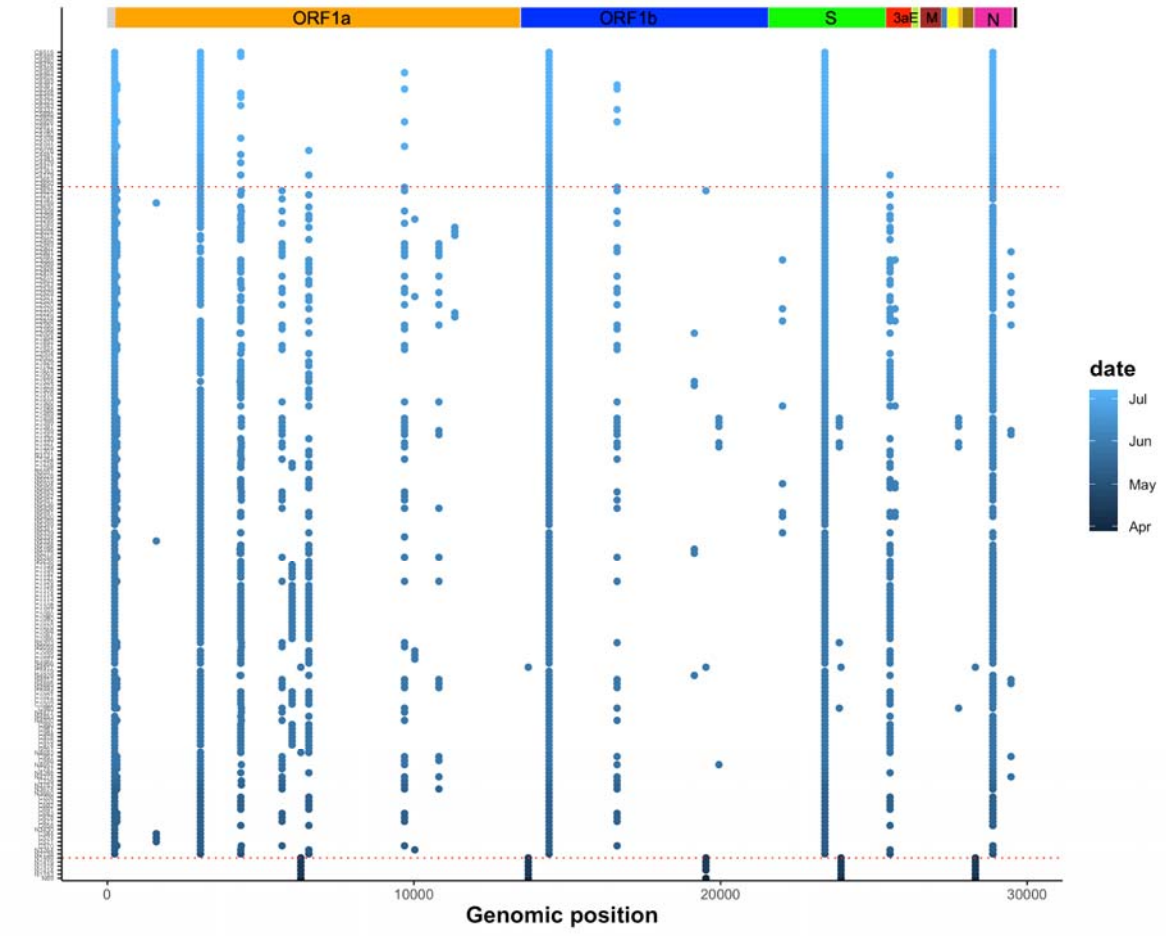
Supplementary data



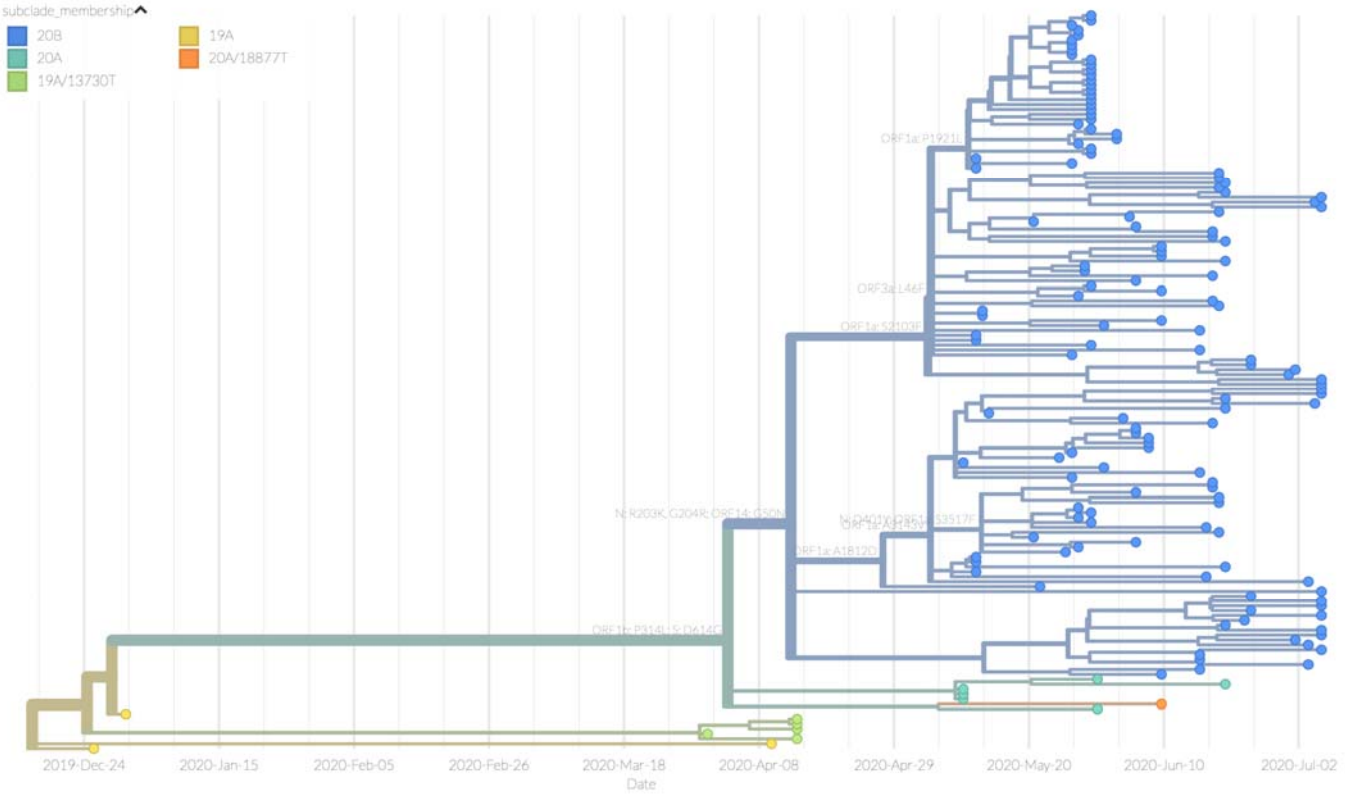
Supplementary figure S1: Detailed workflow of the methodology and various stages.



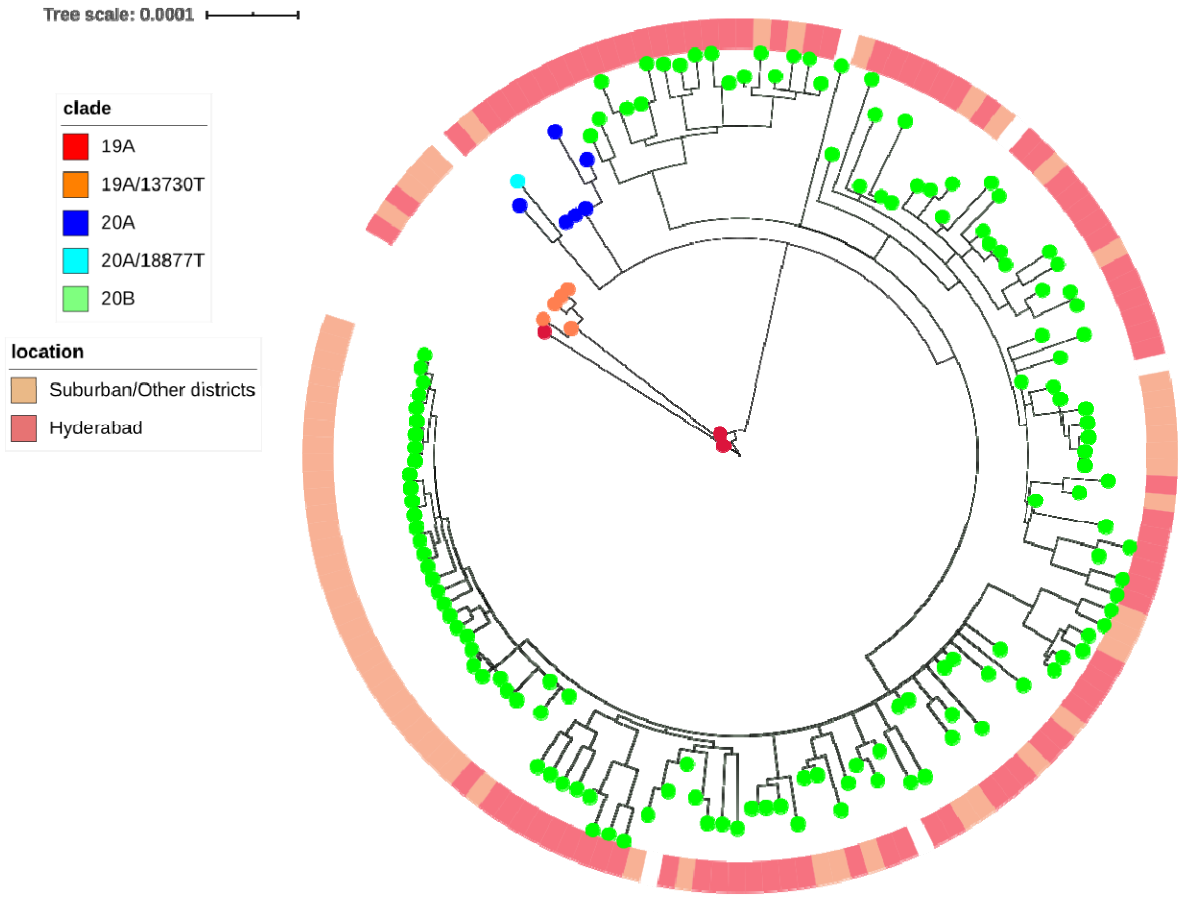
Supplementary figure S2: Sample features: (a) Ct value distribution of samples with respect to date of sample collection (b) Correlation between age and Ct values of the samples (c) Distribution of samples with respect to locations within the district of Hyderabad.



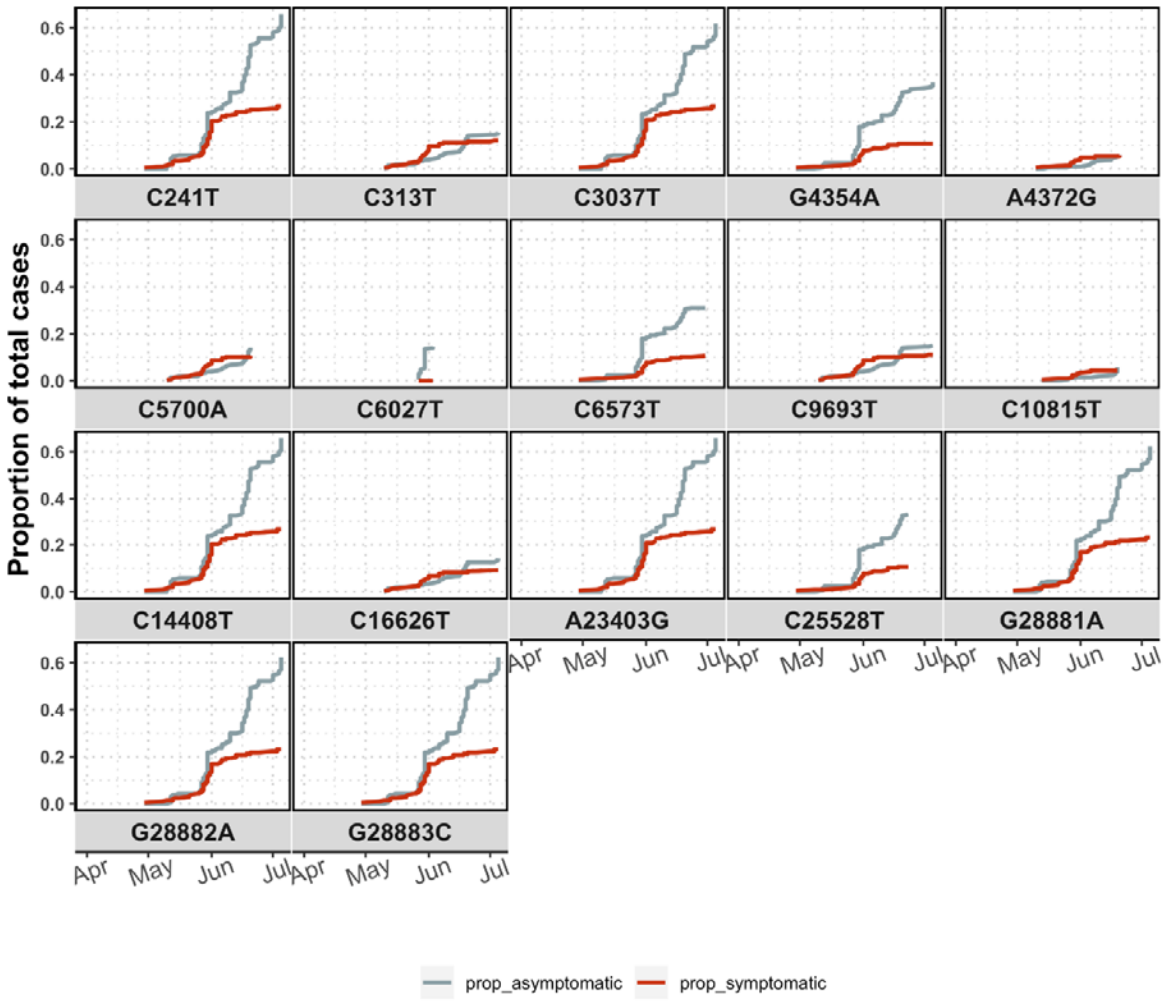
Supplementary figure S3: Mutational map across all the samples arranged with respect to date of collection. Horizontal red lines indicate the time at which major split were observed in the phylodynamic tree.



Supplementary figure S4: Time resolved phylogenetic tree created using Nextstrain.



Supplementary figure S5: Phylogenetic tree with local distribution of samples.



Supplementary figure S6: Cumulative frequency of samples showing asymptomatic and symptomatic behaviour for each of the 17 high frequency mutations.

Supplementary data D1: All the mutations identified in this study

POS	REF	ALT	cols2	Impact	Transcript	nucleotide change	AA change
117	C	T	upstream_gene_variant	MODIFIER	ORF1ab	c.-149C>T	
124	G	A	upstream_gene_variant	MODIFIER	ORF1ab	c.-142G>A	
241	C	T	upstream_gene_variant	MODIFIER	ORF1ab	c.-25C>T	
313	C	T	synonymous_variant	LOW	ORF1ab	c.48C>T	p.Leu16Leu
411	G	A	missense_variant	MODERATE	ORF1ab	c.146G>A	p.Gly49Asp
507	ATGGTCATGTTATGGT	A	disruptive_inframe_deletion	MODERATE	ORF1ab	c.245_259delGTCATGTTATGGTTG	p.Gly82_Val86del
517	TATG	T	conservative_inframe_deletion	MODERATE	ORF1ab	c.253_255delATG	p.Met85del
713	G	T	missense_variant	MODERATE	ORF1ab	c.448G>T	p.Gly150Cys
936	C	T	missense_variant	MODERATE	ORF1ab	c.671C>T	p.Thr224Ile
970	G	A	synonymous_variant	LOW	ORF1ab	c.705G>A	p.Glu235Glu
1059	C	T	missense_variant	MODERATE	ORF1ab	c.794C>T	p.Thr265Ile
1218	C	T	missense_variant	MODERATE	ORF1ab	c.953C>T	p.Ser318Leu
1288	C	T	synonymous_variant	LOW	ORF1ab	c.1023C>T	p.Cys341Cys
1397	G	A	missense_variant	MODERATE	ORF1ab	c.1132G>A	p.Val378Ile
1437	C	T	missense_variant	MODERATE	ORF1ab	c.1172C>T	p.Ser391Phe
1599	G	T	missense_variant	MODERATE	ORF1ab	c.1334G>T	p.Gly445Val
1758	C	T	missense_variant	MODERATE	ORF1ab	c.1493C>T	p.Ala498Val
1813	TA	T	frameshift_variant	HIGH	ORF1ab	c.1554delA	p.Gly519fs
1820	G	A	missense_variant	MODERATE	ORF1ab	c.1555G>A	p.Gly519Ser
1890	G	T	missense_variant	MODERATE	ORF1ab	c.1625G>T	p.Arg542Leu
2008	T	A	synonymous_variant	LOW	ORF1ab	c.1743T>A	p.Ile581Ile
2078	G	A	missense_variant	MODERATE	ORF1ab	c.1813G>A	p.Val605Ile
2207	A	G	missense_variant	MODERATE	ORF1ab	c.1942A>G	p.Ile648Val
2325	C	T	missense_variant	MODERATE	ORF1ab	c.2060C>T	p.Ala687Val
2397	C	T	missense_variant	MODERATE	ORF1ab	c.2132C>T	p.Thr711Met
2447	G	T	missense_variant	MODERATE	ORF1ab	c.2182G>T	p.Gly728Cys
2509	C	T	synonymous_variant	LOW	ORF1ab	c.2244C>T	p.Pro748Pro

2704	C	T	synonymous_variant	LOW	ORF1ab	c.2439C>T	p.Phe813Phe
2836	C	T	synonymous_variant	LOW	ORF1ab	c.2571C>T	p.Cys857Cys
2890	G	A	synonymous_variant	LOW	ORF1ab	c.2625G>A	p.Val875Val
3037	C	T	synonymous_variant	LOW	ORF1ab	c.2772C>T	p.Phe924Phe
3039	A	G	missense_variant	MODERATE	ORF1ab	c.2774A>G	p.Tyr925Cys
3122	G	A	missense_variant	MODERATE	ORF1ab	c.2857G>A	p.Asp953Asn
3241	C	T	synonymous_variant	LOW	ORF1ab	c.2976C>T	p.Asp992Asp
3259	G	T	missense_variant	MODERATE	ORF1ab	c.2994G>T	p.Gln998His
3261	C	T	missense_variant	MODERATE	ORF1ab	c.2996C>T	p.Thr999Ile
3431	G	T	missense_variant	MODERATE	ORF1ab	c.3166G>T	p.Val1056Leu
3472	A	T	synonymous_variant	LOW	ORF1ab	c.3207A>T	p.Gly1069Gly
3621	G	A	missense_variant	MODERATE	ORF1ab	c.3356G>A	p.Gly1119Asp
3729	G	A	missense_variant	MODERATE	ORF1ab	c.3464G>A	p.Gly1155Asp
3774	G	A	missense_variant	MODERATE	ORF1ab	c.3509G>A	p.Arg1170His
3846	A	G	missense_variant	MODERATE	ORF1ab	c.3581A>G	p.Lys1194Arg
4144	G	A	synonymous_variant	LOW	ORF1ab	c.3879G>A	p.Glu1293Glu
4331	C	T	synonymous_variant	LOW	ORF1ab	c.4066C>T	p.Leu1356Leu
4345	C	A	synonymous_variant	LOW	ORF1ab	c.4080C>A	p.Ile1360Ile
4354	G	A	synonymous_variant	LOW	ORF1ab	c.4089G>A	p.Glu1363Glu
4372	A	G	synonymous_variant	LOW	ORF1ab	c.4107A>G	p.Gly1369Gly
4631	G	T	missense_variant	MODERATE	ORF1ab	c.4366G>T	p.Gly1456Cys
4694	G	C	missense_variant	MODERATE	ORF1ab	c.4429G>C	p.Val1477Leu
4999	C	T	synonymous_variant	LOW	ORF1ab	c.4734C>T	p.Asn1578Asn
5045	T	C	missense_variant	MODERATE	ORF1ab	c.4780T>C	p.Phe1594Leu
5194	G	T	synonymous_variant	LOW	ORF1ab	c.4929G>T	p.Leu1643Leu
5305	G	C	missense_variant	MODERATE	ORF1ab	c.5040G>C	p.Leu1680Phe
5700	C	A	missense_variant	MODERATE	ORF1ab	c.5435C>A	p.Ala1812Asp
5826	C	T	missense_variant	MODERATE	ORF1ab	c.5561C>T	p.Thr1854Ile
5880	G	C	missense_variant	MODERATE	ORF1ab	c.5615G>C	p.Ser1872Thr
6026	C	T	missense_variant	MODERATE	ORF1ab	c.5761C>T	p.Pro1921Ser
6027	C	T	missense_variant	MODERATE	ORF1ab	c.5762C>T	p.Pro1921Leu

6034	A	T	synonymous_variant	LOW	ORF1ab	c.5769A>T	p.Ala1923Ala
6310	C	A	missense_variant	MODERATE	ORF1ab	c.6045C>A	p.Ser2015Arg
6312	C	A	missense_variant	MODERATE	ORF1ab	c.6047C>A	p.Thr2016Lys
6312	C	T	missense_variant	MODERATE	ORF1ab	c.6047C>T	p.Thr2016Ile
6330	C	T	missense_variant	MODERATE	ORF1ab	c.6065C>T	p.Ser2022Leu
6445	C	T	synonymous_variant	LOW	ORF1ab	c.6180C>T	p.Asp2060Asp
6466	A	G	synonymous_variant	LOW	ORF1ab	c.6201A>G	p.Lys2067Lys
6476	G	T	missense_variant	MODERATE	ORF1ab	c.6211G>T	p.Val2071Phe
6573	C	T	missense_variant	MODERATE	ORF1ab	c.6308C>T	p.Ser2103Phe
6616	A	G	synonymous_variant	LOW	ORF1ab	c.6351A>G	p.Leu2117Leu
6693	A	G	missense_variant	MODERATE	ORF1ab	c.6428A>G	p.Lys2143Arg
6736	T	C	synonymous_variant	LOW	ORF1ab	c.6471T>C	p.Val2157Val
6794	C	T	synonymous_variant	LOW	ORF1ab	c.6529C>T	p.Leu2177Leu
7050	A	G	missense_variant	MODERATE	ORF1ab	c.6785A>G	p.Tyr2262Cys
7070	G	T	missense_variant	MODERATE	ORF1ab	c.6805G>T	p.Gly2269Cys
7187	C	CA	frameshift_variant	HIGH	ORF1ab	c.6925dupA	p.Ile2309fs
7278	TC	T	frameshift_variant	HIGH	ORF1ab	c.7014delC	p.Tyr2339fs
7319	A	G	missense_variant	MODERATE	ORF1ab	c.7054A>G	p.Ser2352Gly
7594	C	T	synonymous_variant	LOW	ORF1ab	c.7329C>T	p.Gly2443Gly
7798	G	T	missense_variant	MODERATE	ORF1ab	c.7533G>T	p.Lys2511Asn
8090	C	T	missense_variant	MODERATE	ORF1ab	c.7825C>T	p.Leu2609Phe
8110	T	A	synonymous_variant	LOW	ORF1ab	c.7845T>A	p.Thr2615Thr
8131	G	T	missense_variant	MODERATE	ORF1ab	c.7866G>T	p.Lys2622Asn
8140	C	T	synonymous_variant	LOW	ORF1ab	c.7875C>T	p.Ser2625Ser
8296	T	C	synonymous_variant	LOW	ORF1ab	c.8031T>C	p.Tyr2677Tyr
8653	G	C	missense_variant	MODERATE	ORF1ab	c.8388G>C	p.Met2796Ile
8802	C	T	missense_variant	MODERATE	ORF1ab	c.8537C>T	p.Thr2846Ile
8917	C	T	synonymous_variant	LOW	ORF1ab	c.8652C>T	p.Phe2884Phe
9049	G	A	synonymous_variant	LOW	ORF1ab	c.8784G>A	p.Lys2928Lys
9234	A	G	missense_variant	MODERATE	ORF1ab	c.8969A>G	p.Glu2990Gly
9246	C	T	missense_variant	MODERATE	ORF1ab	c.8981C>T	p.Ala2994Val

9310	A	T	synonymous_variant	LOW	ORF1ab	c.9045A>T	p.Pro3015Pro
9389	G	A	missense_variant	MODERATE	ORF1ab	c.9124G>A	p.Asp3042Asn
9693	C	T	missense_variant	MODERATE	ORF1ab	c.9428C>T	p.Ala3143Val
10029	C	T	missense_variant	MODERATE	ORF1ab	c.9764C>T	p.Thr3255Ile
10156	C	T	synonymous_variant	LOW	ORF1ab	c.9891C>T	p.Asp3297Asp
10160	G	GT	frameshift_variant	HIGH	ORF1ab	c.9898dupT	p.Tyr3300fs
10198	C	T	synonymous_variant	LOW	ORF1ab	c.9933C>T	p.Asp3311Asp
10255	G	T	missense_variant	MODERATE	ORF1ab	c.9990G>T	p.Leu3330Phe
10296	C	T	missense_variant	MODERATE	ORF1ab	c.10031C>T	p.Ser3344Phe
10340	C	T	missense_variant	MODERATE	ORF1ab	c.10075C>T	p.Pro3359Ser
10386	C	CT	frameshift_variant	HIGH	ORF1ab	c.10126dupT	p.Ser3376fs
10500	G	T	missense_variant	MODERATE	ORF1ab	c.10235G>T	p.Gly3412Val
10802	C	T	synonymous_variant	LOW	ORF1ab	c.10537C>T	p.Leu3513Leu
10815	C	T	missense_variant	MODERATE	ORF1ab	c.10550C>T	p.Ser3517Phe
10924	A	T	missense_variant	MODERATE	ORF1ab	c.10659A>T	p.Glu3553Asp
11103	CT	C	frameshift_variant	HIGH	ORF1ab	c.10842delT	p.Phe3614fs
11335	G	T	synonymous_variant	LOW	ORF1ab	c.11070G>T	p.Val3690Val
11514	C	T	missense_variant	MODERATE	ORF1ab	c.11249C>T	p.Thr3750Ile
11579	A	G	missense_variant	MODERATE	ORF1ab	c.11314A>G	p.Thr3772Ala
11745	G	T	missense_variant	MODERATE	ORF1ab	c.11480G>T	p.Gly3827Val
11837	G	T	missense_variant	MODERATE	ORF1ab	c.11572G>T	p.Val3858Leu
12076	C	T	synonymous_variant	LOW	ORF1ab	c.11811C>T	p.Asn3937Asn
12134	T	C	missense_variant	MODERATE	ORF1ab	c.11869T>C	p.Phe3957Leu
12139	TAC	T	frameshift_variant	HIGH	ORF1ab	c.11875_11876delAC	p.Thr3959fs
12189	T	C	missense_variant	MODERATE	ORF1ab	c.11924T>C	p.Val3975Ala
12439	C	T	synonymous_variant	LOW	ORF1ab	c.12174C>T	p.Pro4058Pro
12547	A	T	synonymous_variant	LOW	ORF1ab	c.12282A>T	p.Ala4094Ala
12668	T	C	missense_variant	MODERATE	ORF1ab	c.12403T>C	p.Ser4135Pro
13255	C	T	synonymous_variant	LOW	ORF1ab	c.12990C>T	p.Cys4330Cys
13305	G	T	missense_variant	MODERATE	ORF1ab	c.13040G>T	p.Gly4347Val
13487	C	T	missense_variant	MODERATE	ORF1ab	c.13238C>T	p.Ala4413Val

13488	A	G	synonymous_variant	LOW	ORF1ab	c.13239A>G	p.Ala4413Ala
13501	C	T	missense_variant	MODERATE	ORF1ab	c.13252C>T	p.Pro4418Ser
13585	C	T	synonymous_variant	LOW	ORF1ab	c.13336C>T	p.Leu4446Leu
13730	C	T	missense_variant	MODERATE	ORF1ab	c.13481C>T	p.Ala4494Val
13837	C	T	missense_variant	MODERATE	ORF1ab	c.13588C>T	p.His4530Tyr
13944	C	T	synonymous_variant	LOW	ORF1ab	c.13695C>T	p.Asn4565Asn
14099	G	T	missense_variant	MODERATE	ORF1ab	c.13850G>T	p.Gly4617Val
14408	C	T	missense_variant	MODERATE	ORF1ab	c.14159C>T	p.Pro4720Leu
14724	C	T	synonymous_variant	LOW	ORF1ab	c.14475C>T	p.Phe4825Phe
15237	C	T	synonymous_variant	LOW	ORF1ab	c.14988C>T	p.His4996His
15273	C	T	synonymous_variant	LOW	ORF1ab	c.15024C>T	p.Asn5008Asn
15352	C	T	missense_variant	MODERATE	ORF1ab	c.15103C>T	p.Leu5035Phe
15672	G	C	missense_variant	MODERATE	ORF1ab	c.15423G>C	p.Glu5141Asp
16153	G	A	missense_variant	MODERATE	ORF1ab	c.15904G>A	p.Val5302Ile
16260	C	T	synonymous_variant	LOW	ORF1ab	c.16011C>T	p.Cys5337Cys
16374	T	C	synonymous_variant	LOW	ORF1ab	c.16125T>C	p.Asn5375Asn
16575	C	T	synonymous_variant	LOW	ORF1ab	c.16326C>T	p.Asp5442Asp
16626	C	T	synonymous_variant	LOW	ORF1ab	c.16377C>T	p.Leu5459Leu
16852	G	T	missense_variant	MODERATE	ORF1ab	c.16603G>T	p.Gly5535Cys
16946	C	T	missense_variant	MODERATE	ORF1ab	c.16697C>T	p.Ala5566Val
17010	C	T	synonymous_variant	LOW	ORF1ab	c.16761C>T	p.Ile5587Ile
17014	G	T	missense_variant	MODERATE	ORF1ab	c.16765G>T	p.Asp5589Tyr
17019	G	T	missense_variant	MODERATE	ORF1ab	c.16770G>T	p.Glu5590Asp
17246	G	T	missense_variant	MODERATE	ORF1ab	c.16997G>T	p.Arg5666Leu
17259	G	T	missense_variant	MODERATE	ORF1ab	c.17010G>T	p.Glu5670Asp
17304	C	T	synonymous_variant	LOW	ORF1ab	c.17055C>T	p.Val5685Val
17321	C	T	missense_variant	MODERATE	ORF1ab	c.17072C>T	p.Ala5691Val
17454	T	C	synonymous_variant	LOW	ORF1ab	c.17205T>C	p.Pro5735Pro
17944	G	T	missense_variant	MODERATE	ORF1ab	c.17695G>T	p.Val5899Leu
18029	C	T	missense_variant	MODERATE	ORF1ab	c.17780C>T	p.Ala5927Val
18052	A	G	missense_variant	MODERATE	ORF1ab	c.17803A>G	p.Thr5935Ala

18129	C	T	synonymous_variant	LOW	ORF1ab	c.17880C>T	p.Asp5960Asp
18322	C	T	missense_variant	MODERATE	ORF1ab	c.18073C>T	p.His6025Tyr
18395	C	T	missense_variant	MODERATE	ORF1ab	c.18146C>T	p.Ala6049Val
18452	C	T	missense_variant	MODERATE	ORF1ab	c.18203C>T	p.Ala6068Val
18453	T	C	synonymous_variant	LOW	ORF1ab	c.18204T>C	p.Ala6068Ala
18651	G	C	missense_variant	MODERATE	ORF1ab	c.18402G>C	p.Glu6134Asp
18695	C	T	missense_variant	MODERATE	ORF1ab	c.18446C>T	p.Thr6149Ile
18834	T	C	synonymous_variant	LOW	ORF1ab	c.18585T>C	p.Gly6195Gly
18877	C	T	synonymous_variant	LOW	ORF1ab	c.18628C>T	p.Leu6210Leu
19137	A	T	missense_variant	MODERATE	ORF1ab	c.18888A>T	p.Leu6296Phe
19145	C	T	missense_variant	MODERATE	ORF1ab	c.18896C>T	p.Ser6299Phe
19170	C	T	synonymous_variant	LOW	ORF1ab	c.18921C>T	p.Phe6307Phe
19406	GA	G	frameshift_variant	HIGH	ORF1ab	c.19161delA	p.Lys6387fs
19518	G	A	synonymous_variant	LOW	ORF1ab	c.19269G>A	p.Leu6423Leu
19524	C	T	synonymous_variant	LOW	ORF1ab	c.19275C>T	p.Leu6425Leu
19947	G	T	missense_variant	MODERATE	ORF1ab	c.19698G>T	p.Lys6566Asn
21057	C	T	synonymous_variant	LOW	ORF1ab	c.20808C>T	p.Asp6936Asp
21059	C	T	missense_variant	MODERATE	ORF1ab	c.20810C>T	p.Pro6937Leu
21066	TA	T	frameshift_variant	HIGH	ORF1ab	c.20822delA	p.Asn6941fs
21099	G	T	missense_variant	MODERATE	ORF1ab	c.20850G>T	p.Glu6950Asp
21133	C	CA	frameshift_variant	HIGH	ORF1ab	c.20888dupA	p.Leu6964fs
21268	G	T	missense_variant	MODERATE	ORF1ab	c.21019G>T	p.Ala7007Ser
21618	C	A	missense_variant	MODERATE	S	c.56C>A	p.Thr19Lys
21621	C	A	missense_variant	MODERATE	S	c.59C>A	p.Thr20Asn
21621	C	T	missense_variant	MODERATE	S	c.59C>T	p.Thr20Ile
21622	C	T	synonymous_variant	LOW	S	c.60C>T	p.Thr20Thr
21701	GT	G	frameshift_variant	HIGH	S	c.143delT	p.Leu48fs
21770	G	T	missense_variant	MODERATE	S	c.208G>T	p.Val70Phe
21792	A	T	missense_variant	MODERATE	S	c.230A>T	p.Lys77Met
21892	A	G	synonymous_variant	LOW	S	c.330A>G	p.Leu110Leu
22021	G	T	missense_variant	MODERATE	S	c.459G>T	p.Met153Ile

22104	G	T	missense_variant	MODERATE	S	c.542G>T	p.Gly181Val
22444	C	T	synonymous_variant	LOW	S	c.882C>T	p.Asp294Asp
22604	G	T	missense_variant	MODERATE	S	c.1042G>T	p.Ala348Ser
22747	C	T	synonymous_variant	LOW	S	c.1185C>T	p.Val395Val
22801	G	T	synonymous_variant	LOW	S	c.1239G>T	p.Gly413Gly
23150	TCTA	T	disruptive_inframe_deletion	MODERATE	S	c.1592_1594delCTA	p.Thr531del
23191	C	T	synonymous_variant	LOW	S	c.1629C>T	p.Phe543Phe
23277	C	T	missense_variant	MODERATE	S	c.1715C>T	p.Thr572Ile
23403	A	G	missense_variant	MODERATE	S	c.1841A>G	p.Asp614Gly
23422	C	T	synonymous_variant	LOW	S	c.1860C>T	p.Val620Val
23517	G	C	missense_variant	MODERATE	S	c.1955G>C	p.Gly652Ala
23523	A	G	missense_variant	MODERATE	S	c.1961A>G	p.Glu654Gly
23572	A	G	missense_variant	MODERATE	S	c.2010A>G	p.Ile670Met
23635	C	A	synonymous_variant	LOW	S	c.2073C>A	p.Ser691Ser
23718	C	T	missense_variant	MODERATE	S	c.2156C>T	p.Thr719Ile
23873	G	T	missense_variant	MODERATE	S	c.2311G>T	p.Ala771Ser
23929	C	T	synonymous_variant	LOW	S	c.2367C>T	p.Tyr789Tyr
23987	C	T	missense_variant	MODERATE	S	c.2425C>T	p.Pro809Ser
24130	C	T	synonymous_variant	LOW	S	c.2568C>T	p.Asn856Asn
24140	G	GT	frameshift_variant	HIGH	S	c.2582dupT	p.Leu861fs
24213	C	T	missense_variant	MODERATE	S	c.2651C>T	p.Ser884Phe
24236	G	T	missense_variant	MODERATE	S	c.2674G>T	p.Ala892Ser
24621	C	T	missense_variant	MODERATE	S	c.3059C>T	p.Ala1020Val
24626	G	T	missense_variant	MODERATE	S	c.3064G>T	p.Ala1022Ser
24652	A	T	synonymous_variant	LOW	S	c.3090A>T	p.Ser1030Ser
24672	CA	C	frameshift_variant	HIGH	S	c.3115delA	p.Arg1039fs
24867	G	T	missense_variant	MODERATE	S	c.3305G>T	p.Trp1102Leu
24904	C	T	synonymous_variant	LOW	S	c.3342C>T	p.Ile1114Ile
25164	A	T	missense_variant	MODERATE	S	c.3602A>T	p.Gln1201Leu
25244	G	T	missense_variant	MODERATE	S	c.3682G>T	p.Val1228Leu
25339	C	T	synonymous_variant	LOW	S	c.3777C>T	p.Asp1259Asp

25355	C	T	missense_variant	MODERATE	S	c.3793C>T	p.Leu1265Phe
25429	G	T	missense_variant	MODERATE	ORF3a	c.37G>T	p.Val13Leu
25437	G	T	missense_variant	MODERATE	ORF3a	c.45G>T	p.Leu15Phe
25449	A	G	synonymous_variant	LOW	ORF3a	c.57A>G	p.Glu19Glu
25528	C	T	missense_variant	MODERATE	ORF3a	c.136C>T	p.Leu46Phe
25549	C	T	missense_variant	MODERATE	ORF3a	c.157C>T	p.Leu53Phe
25563	G	T	missense_variant	MODERATE	ORF3a	c.171G>T	p.Gln57His
25587	C	G	synonymous_variant	LOW	ORF3a	c.195C>G	p.Leu65Leu
25596	A	G	synonymous_variant	LOW	ORF3a	c.204A>G	p.Arg68Arg
25599	G	T	missense_variant	MODERATE	ORF3a	c.207G>T	p.Trp69Cys
25609	C	T	missense_variant	MODERATE	ORF3a	c.217C>T	p.Leu73Phe
25635	C	T	synonymous_variant	LOW	ORF3a	c.243C>T	p.Cys81Cys
25654	G	T	missense_variant	MODERATE	ORF3a	c.262G>T	p.Val88Leu
25691	G	T	missense_variant	MODERATE	ORF3a	c.299G>T	p.Gly100Val
25693	C	T	missense_variant	MODERATE	ORF3a	c.301C>T	p.Leu101Phe
25855	G	T	missense_variant	MODERATE	ORF3a	c.463G>T	p.Asp155Tyr
25996	G	T	missense_variant	MODERATE	ORF3a	c.604G>T	p.Val202Leu
26062	G	T	missense_variant	MODERATE	ORF3a	c.670G>T	p.Gly224Cys
26335	C	T	synonymous_variant	LOW	E	c.91C>T	p.Leu31Leu
26456	C	T	missense_variant	MODERATE	E	c.212C>T	p.Pro71Leu
26501	G	T	upstream_gene_variant	MODIFIER	M	c.-22G>T	
26622	C	T	missense_variant	MODERATE	M	c.100C>T	p.Leu34Phe
26690	G	T	synonymous_variant	LOW	M	c.168G>T	p.Leu56Leu
26730	G	C	missense_variant	MODERATE	M	c.208G>C	p.Val70Leu
26735	C	T	synonymous_variant	LOW	M	c.213C>T	p.Tyr71Tyr
26882	C	T	synonymous_variant	LOW	M	c.360C>T	p.Leu120Leu
26894	C	T	synonymous_variant	LOW	M	c.372C>T	p.Leu124Leu
26895	C	T	missense_variant	MODERATE	M	c.373C>T	p.His125Tyr
26918	G	T	synonymous_variant	LOW	M	c.396G>T	p.Pro132Pro
26951	G	C	synonymous_variant	LOW	M	c.429G>C	p.Val143Val
27065	G	A	synonymous_variant	LOW	M	c.543G>A	p.Leu181Leu

27110	C	T	synonymous_variant	LOW	M	c.588C>T	p.Tyr196Tyr
27286	C	T	missense_variant	MODERATE	ORF6	c.85C>T	p.Leu29Phe
27331	C	T	synonymous_variant	LOW	ORF6	c.130C>T	p.Leu44Leu
27384	T	C	synonymous_variant	LOW	ORF6	c.183T>C	p.Asp61Asp
27416	C	T	missense_variant	MODERATE	ORF7a	c.23C>T	p.Ala8Val
27463	G	T	missense_variant	MODERATE	ORF7a	c.70G>T	p.Val24Phe
27494	C	T	missense_variant	MODERATE	ORF7a	c.101C>T	p.Pro34Leu
27703	G	A	missense_variant	MODERATE	ORF7a	c.310G>A	p.Val104Ile
27754	G	T	stop_gained	HIGH	ORF7a	c.361G>T	p.Glu121*
27762	G	C	missense_variant	MODERATE	ORF7b	c.7G>C	p.Glu3Gln
27816	G	A	missense_variant	MODERATE	ORF7b	c.61G>A	p.Val21Ile
27841	G	T	missense_variant	MODERATE	ORF7b	c.86G>T	p.Trp29Leu
27890	G	A	upstream_gene_variant	MODIFIER	ORF8	c.-4G>A	
27916	G	T	missense_variant	MODERATE	ORF8	c.23G>T	p.Gly8Val
27925	C	T	missense_variant	MODERATE	ORF8	c.32C>T	p.Thr11Ile
27966	TG	T	frameshift_variant	HIGH	ORF8	c.74delG	p.Cys25fs
27996	G	C	missense_variant	MODERATE	ORF8	c.103G>C	p.Asp35His
28028	G	C	missense_variant	MODERATE	ORF8	c.135G>C	p.Trp45Cys
28045	C	T	missense_variant	MODERATE	ORF8	c.152C>T	p.Ala51Val
28048	G	T	missense_variant	MODERATE	ORF8	c.155G>T	p.Arg52Ile
28178	G	T	missense_variant	MODERATE	ORF8	c.285G>T	p.Leu95Phe
28236	C	T	missense_variant	MODERATE	ORF8	c.343C>T	p.Arg115Cys
28239	G	GT	frameshift_variant	HIGH	ORF8	c.348dupT	p.Val117fs
28242	G	T	missense_variant	MODERATE	ORF8	c.349G>T	p.Val117Phe
28311	C	T	missense_variant	MODERATE	N	c.38C>T	p.Pro13Leu
28371	G	T	missense_variant	MODERATE	N	c.98G>T	p.Ser33Ile
28378	G	T	synonymous_variant	LOW	N	c.105G>T	p.Ala35Ala
28486	C	T	synonymous_variant	LOW	N	c.213C>T	p.Gly71Gly
28514	G	T	missense_variant	MODERATE	N	c.241G>T	p.Asp81Tyr
28582	T	C	synonymous_variant	LOW	N	c.309T>C	p.Asp103Asp
28688	T	C	synonymous_variant	LOW	N	c.415T>C	p.Leu139Leu

28690	G	T	missense_variant	MODERATE	N	c.417G>T	p.Leu139Phe
28765	A	G	synonymous_variant	LOW	N	c.492A>G	p.Gly164Gly
28795	A	G	synonymous_variant	LOW	N	c.522A>G	p.Glu174Glu
28821	C	A	missense_variant	MODERATE	N	c.548C>A	p.Ser183Tyr
28822	T	C	synonymous_variant	LOW	N	c.549T>C	p.Ser183Ser
28854	C	T	missense_variant	MODERATE	N	c.581C>T	p.Ser194Leu
28856	A	C	synonymous_variant	LOW	N	c.583A>C	p.Arg195Arg
28881	G	A	missense_variant	MODERATE	N	c.608G>A	p.Arg203Lys
28882	G	A	synonymous_variant	LOW	N	c.609G>A	p.Arg203Arg
28883	G	C	missense_variant	MODERATE	N	c.610G>C	p.Gly204Arg
28887	C	T	missense_variant	MODERATE	N	c.614C>T	p.Thr205Ile
28975	G	T	missense_variant	MODERATE	N	c.702G>T	p.Met234Ile
29044	G	A	synonymous_variant	LOW	N	c.771G>A	p.Lys257Lys
29109	C	T	missense_variant	MODERATE	N	c.836C>T	p.Pro279Leu
29272	C	T	synonymous_variant	LOW	N	c.999C>T	p.Tyr333Tyr
29284	C	T	synonymous_variant	LOW	N	c.1011C>T	p.Ile337Ile
29383	G	T	missense_variant	MODERATE	N	c.1110G>T	p.Lys370Asn
29402	G	T	missense_variant	MODERATE	N	c.1129G>T	p.Asp377Tyr
29474	G	T	missense_variant	MODERATE	N	c.1201G>T	p.Asp401Tyr
29692	G	T	downstream_gene_variant	MODIFIER	S	c.*4308G>T	
29711	G	T	downstream_gene_variant	MODIFIER	S	c.*4327G>T	
29723	A	T	downstream_gene_variant	MODIFIER	S	c.*4339A>T	
29734	G	T	downstream_gene_variant	MODIFIER	S	c.*4350G>T	
29736	G	T	downstream_gene_variant	MODIFIER	S	c.*4352G>T	
29742	G	T	downstream_gene_variant	MODIFIER	S	c.*4358G>T	
29747	G	T	downstream_gene_variant	MODIFIER	S	c.*4363G>T	