Whole Genome Sequencing Confirmed SARS-CoV-2 Reinfections Among Healthcare Workers in India with Increased Severity in the Second Episode

Jayanthi Shastri[1+*] Swapneil Parikh[1*], Sachee Agrawal[1], Nirjhar Chatterjee[1], Manish Pathak[1], Chetan Sharma[2], Akshay Kanakan[3], Vivekanand A[3], Janani Srinivasa Vasudevan[3], Ranjeet Maurya[3], Saman Fatihi[3], Lipi Thukral[3], Anurag Agrawal[3], Lancelot Pinto[4], Rajesh Pandey[3], Sujatha Sunil[2+*]

* Contributed equally
+ Corresponding authors

Affiliations
1 Kasturba Hospital for Infectious Disease, Mumbai, India
2 International Centre for Genetic Engineering and Biotechnology, New Delhi, India
3 CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India

4 P. D. Hinduja Hospital, Mumbai, India

Corresponding authors:
Dr. Jayanthi Shastri
jsshastri@gmail.com
+91-9820549156

Dr. Sujatha Sunil
sujatha@icgeb.res.in
+91-9711491928

Keywords: SARS-CoV-2, COVID-19, Reinfections, Whole Genome Sequencing, Healthcare workers

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Summary:

Background:

Frontline healthcare workers (HCWs) are exceedingly exposed to SARS-CoV-2 and reinfections are a possibility. A RT-PCR positive test does not confirm reinfection. Whole genome sequencing (WGS) of the viral isolates from the different episodes can confirm a reinfection.

Methods

RNA was extracted from nasopharyngeal plus oropharyngeal samples from four HCWs who were SARS-CoV-2 RT-PCR positive in May/June and then again in July. Anti-NC antibody testing was performed after the second infection in three HCWs. The RNA was subjected to whole genome sequencing and comparative genome and protein-based functional annotation analyses were performed on the nucleotide and amino acid sequences.

Findings

Whole genome sequencing of the eight SARS-CoV-2 viral samples generated a genome coverage ranging from 82.55 to 98.23%. Phylogenetic analysis revealed that sequences belonged to the L clade and within this major clade; they clustered into India-specific A2a and A4 clades. A total of 39 mutations were identified within the eight genomes, including 22 non-synonymous, 16 synonymous, and 1 stop-coding substitutions. Comparative genomic and protein-based annotation analyses revealed differences in the presence and absence of specific mutations in the virus sequences from the first and second episode in all four paired samples. Three HCWs were negative for anti-NC antibodies after the second infection.

Interpretation

Genomic variations observed through whole genome sequencing coupled with clinical presentation confirm reinfections of SARS-CoV-2 in healthcare workers.

Funding

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Municipal Corporation of Greater Mumbai, ICGEB core funding, MLP-2005 and Fondation Botnar grants.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Research in context

Evidence before this study

September 4, 2020 we searched Google Scholar, for articles published since 2020, with the keywords "SARS-CoV-2" AND "whole genome sequencing" AND "reinfection". Our search retrieved 44 results. However only one report presented whole genome sequenced proven reinfection. An identical PubMed search returned 1 result of the same report from Hong Kong. A similar Google scholar search for "genomic evidence" AND "SARS-CoV-2" AND "reinfection" returned one result of a The Lancet preprint reporting whole genome sequencing (WGS) confirmed in Nevada with increased severity during the second episode. Adding "healthcare worker" to the above search returned no results. Based on available evidence, reinfections seem rare. However it is unclear if this is because of post-infection immunity or because we are still early in the pandemic and whole genome sequencing is not performed frequently on paired positive samples and correlated to clinical presentation.

Added value of this study

This study looks at reinfection in healthcare workers, a subset who are most exposed to infection risk. Previously reported WGS confirmed reinfections were single cases, this study reports four WGS confirmed reinfections.

Implications of all the available evidence

Our findings demonstrate that HCWs can get reinfected with SARS-CoV-2 with increased clinical severity in the second episode. SARS-CoV-2 reinfection risk in convalescent health-care workers should be addressed in health-care policy making.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Introduction

In December 2019, a novel coronavirus (n-CoV-19) sparked an outbreak in Wuhan, China. This virus was subsequently named SARS-CoV-2 and the disease COVID-19. On 11th March 2020, there were 1,18,000 cases in 114 countries with 4,291 deaths and the World Health Organisation (WHO) declared that COVID-19 was a pandemic.[1]

Health-care workers (HCWs) who have been on the frontlines of managing COVID-19 patients are highly exposed to SARS-CoV-2 infection, and have a much higher risk of infection than the general public.[2] For instance, in USA as of 25th Aug, there have been 1,43,743 confirmed cases and 660 casualties due to COVID-19 amongst HCWs.[3] In India as of 29th August, approximately 87,176 infections have been documented among HCWs and around 573 have succumbed to COVID-19.[4] HCWs are exposed to patients with varied clinical severity and possibly higher viral loads at different times points. The circulating viruses may also have a plethora of genomic variations within a patient or among patients.

In August, the first report of reinfection by a phylogenetically distinct strain of SARS-CoV-2 was confirmed in Hong Kong[5] and subsequently a preprint reported another reinfection in USA.[6] While there have been many reports of putative reinfections based on RT-PCR positivity, this has been confounded by prolonged shedding of viral RNA in the absence of replication competent virus[7] which can continue to cause RT-PCR positivity for up to at least 83 days.[8] A positive RT-PCR test can occur for a variable period of time following recovery from COVID-19, and may even occur after negative tests in the absence of true infection.[9] However, samples collected for RT-PCR can be sequenced and genomic analysis may demonstrate genetic variation that can't be explained by short term in vivo evolution, which when combined with epidemiological and clinical evidence, may confirm reinfection.[5,7]

The present study was undertaken using samples collected from HCWs tested for SARS-COV-2 as standard of care either for contact tracing or diagnostic purposes in symptomatic individuals. We

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

report a case series of four HCWs who had initial asymptomatic or mild RT-PCR proven COVID-19 followed by a second symptomatic RT-PCR positive episode with varying degrees of increased clinical severity. Whole genome sequencing of the viruses isolated during the different episodes revealed distinct genomic variations between the pair-wise isolates suggesting reinfection by different SARS-CoV-2 viruses. Our work highlights the presence of reinfections in India and the imminent threat to HCWs on the frontlines of the pandemic.

Methodology

Study Design and Participants

We identified four HCWs who had tested positive for SARS-CoV-2 RT-PCR in May or June 2020 and again tested positive by RT-PCR when they developed symptoms suggestive of COVID-19 in July. Kasturba Hospital was the diagnostic laboratory that tested three of the HCWs where as the fourth was tested at P D Hinduja Hospital and came to our attention when her attending physician requested us to verify her samples. Based on the RT-PCR results and clinical presentation of the HCWs we suspected reinfection with SARS-CoV-2. Upon confirmation of the RT-PCR findings, whole genome sequencing was performed on the stored paired samples. The study was approved by the Institutional Review Board of Kasturba Hospital of Infectious Diseases; IRB number 015/2020. The patients provided written informed consent.

Procedures

Sample Collection, Storage and RT-PCR

Nasopharyngeal (NP) plus oropharyngeal (OP) samples were collected from four HCWs in May or June as detailed in the supplement (S Table 1). Samples from the first positive RT-PCR in all HCWs were aliquoted and stored for future use as detailed in the supplementary Table 1. One of the aliquots was used for automated RNA extraction on three HCWs and manual extraction on the last HCW. All samples were tested by multiplex real time RT-PCR TaqPath™ COVID19 RTPCR kit for the qualitative

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

detection of nucleic acid of SARS-CoV-2 from Applied Biosystems. Additional details of RT-PCR testing are described in the supplement.

Serology

Antibody testing for COVID-19 was performed using chemiluminescence (CLIA) with Abbott Architect SARS-CoV-2 anti-NC IgG for patient A and B and with Roche SARS-CoV-2 anti-NC total antibody (IgM+IgG) for patient D.

SARS-CoV-2 Whole Genome Sequencing

Extracted RNA from all four paired stored samples was transported at -80 ° C to the sequencing team for whole genome sequencing. Sample preparation, sequencing and data analysis was performed by previously published protocols.[10] In brief, double-stranded cDNA was synthesized from 50ng of total RNA for all the SARS-CoV-2 positive samples. The first strand of cDNA was synthesized using Superscript IV followed by RNA digestion with RNase H for second strand synthesis using DNA Polymerase I Large fragment (Klenow fragment). 100ng of purified double-stranded cDNA was taken for forward using ARTIC tiling PCR (V3 primer pools) protocol. 200ng of each purified sample of multiplexed PCR amplicons obtained was taken for library preparation using Oxford Nanopore Technology (ONT). This included End Repair/dA tailing, Native Barcode Ligation, and Adapter Ligation of the PCR amplicons. 100ng of the pooled and purified library was sequenced using ONT's MinION Mk1B platform

Phylogenetic and Comparative genomic analysis

Samples were basecalled and demultiplexed using Guppybasecaller (https://community.nanoporetech.com). Reads having phead quality score <7 were discarded to filter the low-quality reads. The resulting fastq files were normalized by read length (300-500) and reads were aligned using Minimap2 (v2.17)[11] to the reference (MN908947.3). Variants were called using Nanopolish[12] from the aligned reads and further creating consensus fasta using bcftools (v1.8). Assembled fasta files from the SARS CoV-2 were aligned using CLC workbench and a UPGMA tree

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

was constructed using default parameters. A secondary tree was generated after downloading whole genome sequences from VIPR database from India submitted during the period from March 2020 to June 2020. Phylogentic Analysis was done on all the compiled datasets using Vipr. The tree generation algorithm used was PhyMl along with the HKY Model of Evolution.

Lineage analysis

Further, the assembled SARS-CoV-2 genomes were assigned lineages using the package Phylogenetic Assignment of Named Global Outbreak LINeages (PANGOLIN).[13]

Protein-based annotation

In order to categorize the specific amino acid variants present, the genomes were annotated by SnpEff version 4.5.[14] NC_045512 was taken as the reference genome of SARS-CoV-2.[15] The synonymous variants were filtered out from the analysis. The global frequency data for these 12 unique missense variations present across the four pairs was taken from cov-GLUE database which lists amino acid changes observed in GISAID SARS-CoV-2 sequences.[16,17] Total number of GISAID sequences retrieved at the time of analysis was 82,927, out of which 75,734 passed the exclusion criteria of CoV-GLUE.

Statistical Analysis

Due to the nature of our case series, no statistical analysis was performed on the patients' information.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

The molecular laboratory at Kasturba hospital is a reference molecular laboratory for the city of Mumbai and is one of the tertiary care hospitals assigned for SARS-CoV-2 testing. Clinical data was obtained from specimen referral forms submitted with the samples. No diagnostic intervention that required collecting a fresh sample was performed on any of the patients. For the present study, four patients were taken on the basis of their COVID-19 testing and were assigned the IDs - Patient A, Patient B, Patient C and Patient D and their follow up samples as Patient A f/u, Patient B f/u, Patient C f/u and Patient D f/u. Details of their clinical presentation during the two episodes, RT-PCR testing and serology are provided in Table 1. Details of all the qPCR analyses are included in Supplementary information 1.

Genome sequencing and assembly generated genome coverage of 82.55 to 98.23 percent of genome and an average depth of 233 was obtained ranging from a minimum of 90x to a maximum of 465x. The assembled genome was curated and was taken for further analysis. Phylogenetic analysis of the eight complete sequences with 52 other samples from India collected between the month of April-July revealed that these samples clustered together showing they were part of the same larger clade and aligned close to the Wuhan reference strain. Pairwise analysis of the first episode and second episode samples of the same patient showed that Patient A, B and C sub-clustered together forming small sub-clades with each other. However, Patient D and Patient D f/u clustered in different sub-clades (Fig 1). Additionally, analysis of lineage by PANGOLIN revealed distribution of lineages of the eight samples with B.1.1 lineage. In case of Patient A, a shift in the lineage was observed; from the lineage B.1 to the lineage B. Supplementary Table S2 summarizes the lineage distribution with its assignment probability for the samples.

Mutation analysis of the samples revealed distinct mutations in the samples (Table 2). In total, we found 39 mutations within eight genomes, including 22 non-synonymous, 16 synonymous, and one stop-coding substitutions. Variant analysis of the samples revealed that the samples could be classified into a major clade, A2a, based on the variants observed in the analysis. A2a clade defining variants include C14408T (ORF1b), C3037T (ORF1a), and A23403G (S-Protein). In addition to the

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

clade defining variants, HCW samples had many unique variations. This distribution of the variants across each sample type has been elucidated in the Table 2.

To evaluate amino-acid alterations, we performed protein-based annotation of the 22 non-synonymous mutations found from our genome analysis of the four pair of samples (Supplementary Figure 3). It was observed that Pair 1, ie, Patient A shows minor variations, with one mutation out of two unique mutations occurring within Nsp12. With respect to the other patients, interestingly, we found heterogeneity within mutations in both episodes. For instance, in Patient B, the mutations within Spike protein (D614G, Q677H) in the first episode were missing in the followup sample. Similarly, in Patients C and D, we found presence of additional mutations in samples of followup. Further, we also performed correlations of these mutations with viral genomes from world-wide populations (~82,000) to understand their relative frequency (Fig 2). While P323L mutation within nsp12 was found in all samples without exception, other frequent mutations showed abrupt patterns. In particular, D614G mutation within the Spike protein was consistently present in both infections in Patient C but was present only in one of the episodes in Patients B and D.

Discussion

The present study reports reinfection with SARS-CoV-2 virus among four HCWs that was confirmed using WGS. While the clinical presentation varied between the HCWs and between episodes in the same HCW, it was noteworthy that in all four HCWs the first episode was asymptomatic or mildly symptomatic and the second episode was marginally more clinically severe than the first. Furthermore, the WGS analysis revealed that the genomes from the samples in the two episodes had distinct mutations. The combination of the clinical findings, RT-PCR and WGS analysis confirmed reinfection.

The four HCWs, referred to as Patients A, B, C and D, were involved in the care of COVID-19 inpatients between May to July 2020. In May or June all four were SARS-CoV-2 RT-PCR positive for the first time. Patient A, B and C were asymptomatic at the time of testing. Two days after testing

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

positive 'A' developed mild upper respiratory symptoms that lasted two days. Patients 'B' and 'C' were asymptomatic during the first episode. All three tested negative subsequently. Patient D complained of upper respiratory symptoms and myalgia one day before testing positive. Her symptoms resolved in five days and she was not retested following the first episode. After their isolation period ended, they all rejoined clinical duties 65, 63 and 53 days after their first positive tests, A, B and D respectively developed symptoms suggestive of COVID-19. RT-PCR confirmed they had COVID-19. Tropical infection fever panel ruled out other common infections. Patient C developed symptoms just 16 days after his first positive test; RT-PCR done three days after symptom onset confirmed COVID-19. For all four HCWs, the second episode had more symptoms, with constitutional manifestations and illness that lasted longer than the first episode. All were hospitalized for observation and treatment; Patient C received convalescent plasma therapy. Patient D was unable to return to routine activities and work for three weeks. While none of the HCWs developed lower respiratory tract manifestations or breathlessness, this maybe explained by their young age. Older HCWs may experience more severe respiratory involvement.[18]

Whole genome sequence analysis of the viruses from both episodes from the four samples revealed that the genomes had distinct mutations in overlapping segments that could not be explained by short term in vivo viral evolution. SARS-CoV-2 has been reported to exhibit a low mutation rate[19] and thereby show low variability between different SARS-CoV-2 genomes.[20,21] In such a scenario, low-frequency variants also play important roles in deciphering whether a repeated infection in an individual is a case of fresh infection or a case of viral shedding of the previous infection. In the present study, WGS analysis of the virus from the first and second episode reveal that there are distinct mutations amongst the viruses collected at different time points. Patient A and B were part of the same clinical team and it is expected that the virus strains may be similar; however, genome analysis clearly reveal that the viruses belonged to different subclades with a distinct set of variations detected in the first episode, thereby hinting that they might have been infected through different sources. Also, their follow-up samples revealed different sets of mutations indicating that the second

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

episode were fresh acute infections and not a carryover of the previous infections. What can also be hypothesized is the source of infection pertaining to these two cases specifically is that, most probably their infections occurred from the existing circulating virus pool within the same region/time point as opposed to the case of Patient D where the viruses clustered in different subclades and exhibited a larger set of different mutations, clearly emphasizing reinfection of new SARS-CoV-2 strains. In case of Patient C, interestingly, in spite of the short period within the two episodes, the genetic variations were quite different implying possible reinfection and not virus shedding from the first infection.

In prior reported SARS-CoV-2 reinfections,[5,6] there was no evidence of seroconversion following the first episode. While it unclear if there is a threshold titre of neutralising antibodies that provides protection from reinfection, we hypothesize that those HCW's who do not seroconvert may not have the same degree of protection from reinfection as those with high titres of antibodies. This may be of utmost importance to HCW's who get infected and do not seroconvert as they are constantly exposed to the virus and may be at increased risk of reinfection. If reinfection can be more severe than the primary episode, convalescent HCWs need to be alerted to this possibility. Presently, there is no evidence to suggest widespread WGS confirmed re-infections of SARS-CoV-2, and it appears reinfections are rare. Of the few reported WGS confirmed reinfections, one reinfection has been less severe,[5] while another has been more severe.[6]

Studies indicate that SARS-CoV-2 infection induces both a neutralizing antibody response,[22] a cellular response with virus-specific T cells[23] and individuals who recover from COVID-19 appear to have memory B and T cells.[24]. However, not all individuals seroconvert,[22] milder infections may have less robust immune response and antibody tiers may decline with time.[25] Multiple annual reinfection with endemic human coronaviruses (HCoVs) are not uncommon.[26] It is unclear if post-COVID-19 immunity will be long lasting. If those who recover from mild COVID-19 have short-term immunity, reinfections may become more common in the future.

Immune enhancement of reinfection is well known in Dengue[27] and some have speculated that SARS-CoV-2 reinfections may result in antibody dependent enhancement (ADE).[28,29] Another hypothesis

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

is that some strains have mutations that have some bearing on clinical severity,[30] and therefore the primary infection may be mild or the reinfection severe depending on the mutations in the infecting strains. Our study neither suggests ADE nor mutations affecting clinical outcomes, but we suggest that should either ADE or clinically significant mutations occur, it is frontline HCWs in low incomes communities who are most vulnerable to reinfections. We must carefully look for reinfection in frontline HCWs and study their clinical outcomes.

Frontline HCWs have more than threefold higher risk of SARS-CoV-2 infection than the general community.[2] Globally, HCWs are facing PPE shortages, and it is often those HCWs dealing with the highest caseloads who are most likely to face critical PPE shortages.[2] There is an ever pressing need to protect HCWs from exposure to infection with appropriate PPE, and convalescence may not necessarily imply immunity.

This study was limited by the stored samples that were available and serological testing that could have been done after the first episode. Serological testing was performed just five and six days after the positive RT-PCR during the second episode in Patients A and B respectively which may have been too short a time frame for seroconversion. Patient D tested antibody negative 19 days after the second episode but this time period should have been sufficient for seroconversion. We did not test for neutralizing antibodies. Longitudinal serological assessment and reassessment would be useful. The short gap between the two episodes in Patient C could have represented an uncommonly long incubation period but WGS demonstrated mutations. It is possible Patient C's first infection occurred days to weeks before the positive test as he was asymptomatic Despite these limitations, the WGS and clinical presentations prove true reinfections.

While this study raises important questions, we are mindful that in the context of millions of infections, a few rare or uncommon presentations are not unexpected. With that caveat, we suggest that reinfection with SARS-CoV-2 is possible, that the second episode may be more clinically severe and that this is worthy of worldwide attention and surveillance for its implications on the danger to HCWs on the frontlines of the pandemic.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Contributors

Jayanthi Shastri conceptualised and designed the study. Jayanthi Shastri and Lancelot Pinto identified the study participants. Swapneil Parikh and Sachee Agrawal collected and compiled data from different sources. Nirjhar Chatterjee and Manish Pathak performed RNA extraction, aliquoting and RT-PCR. Rajesh Pandey, Vivekanand A, Janani Srinivasa Vasudevan, Akshay Kanakan, Ranjeet Maurya, Saman Fatihi, performed genome sequencing. Rajesh Pandey, Vivekanand A, Janani Srinivasa Vasudevan, Akshay Kanakan, Ranjeet Maurya, Saman Fatihi, Lipi Thukral, Sujatha Sunil and Chetan Sharma performed genomic and linage analyses. Anurag Agrawal, Jayanthi Shastri and Sujatha Sunil provided resources and participated in overall supervision. Swapneil Parikh, Sujatha Sunil, Jayanthi Shastri wrote the initial drafts of manuscript. All authors contributed to data interpretation, critically reviewed the manuscript, provided contributions to tables, figures and text in the manuscript and approved the final manuscript for submission.

Declaration of Interest

We declare no competing interests.

Acknowledgements

We thank Dean, BYL Nair Hospital and Municipal Corporation of Greater Mumbai (MCGM) for permission to publish. We thank MCGM for funding Kasturba Hospital's COVID-19 diagnostic laboratory. We acknowledge ICGEB core funding. CSIR-IGIB is grateful for support for the sequencing from CSIR project (MLP-2005) and Foundation Botnar project (CLP-0031). The authors acknowledge Dr. C Rodrigues for sample transfer and Dr. G Kang through THSTI for facilitating sequencing. for their involvement in sample transfer for sequencing. We thank the frontline HCWs who consented to participate and HCWs across the world who continue to serve humanity through the pandemic at great personal risk.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Figures and Tables

Table 1: Clinical course, RT-PCR and serology

Figure 1 : Circular Phylogram generated using UPMGA on MEGAx. A total of 60 sequences were used in the analysis. Each patient sample pairs are colored. Patient A and f/u light blue, Patient B and f/u dark blue, Patient C and f/u red and Patient A and f/u brown. Sequences downloaded from the public database are colored in Black.

Table 2: Mutation analysis of Healthcare workers with reinfections (n=4)

Figure 2: Mapping of amino-acid substitutions within n-SARS-CoV-2 genome of four pairs of samples. The upper plot demonstrates the seven proteins in different colors that harbour 12 mutations shown in dots. The Y-axis shows the four pair of patient samples. The blue and red dot indicates the presence of the mutation in the first and second episode of infections respectively. The lower plot shows the frequency of that particular mutation in 82,927 genomes deposited in GISAID.

Supplementary Information Appendix

Supplementary Table 1: Sample collection, RNA extraction, RT-qPCR, aliquots, Ct value

Supplementary table 2: PANGOLIN Analysis of lineage distribution with its assignment probability for the eight sequences.

Supplementary Figure 3: Mapping of amino-acid substitutions within n-SARS-CoV-2 genome of four pairs of samples. The upper plot demonstrates the seven proteins in different colors that harbour 22 mutations shown in dots. The Y-axis shows the four pair of patient samples. The blue and red dot indicates the presence of the mutation in the main, and re-infected cases.

References

1. WHO

Director-General's opening remarks at the media briefing on COVID-19.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020
Date: 11 March 2020
Date Accessed: 4 September 2020

2. Nguyen L, Drew D, Graham M., et al
Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study.
Lancet Public Health 2020;**5**:e475–e483

3. US CDC
Coronavirus Disease 2019 (COVID-19) in the U.S. Centers For Disease Control And Prevention.
https://covid.cdc.gov/covid-data-tracker/?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fcases-updates%2Fcases-in-us.html#health-care-personnel
Date: 4 September 2020
Date Accessed: 4 September 2020

4. Day S.
87k health staff infected with Covid, 573 dead; 74% cases from six states.
Times of India, Mumbai. The Times Of India 2020.
https://timesofindia.indiatimes.com/india/over-87k-health-workers-infected-with-covid-19-573-dead/articleshow/77814189.cms
Date: 29 August 2020
Date Accessed: 4 September 2020

5. To K, Hung I, Ip J., et al.
COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing.
Clin Infect Dis 2020. doi:10.1093/cid/ciaa1275

6. Tillett R, Sevinsky J, Hartley P., et al.
Genomic Evidence for a Case of Reinfection with SARS-CoV-2.
SSRN. 2020; (published on 31 August 2020) (preprint)
http://dx.doi.org/10.2139/ssrn.3680955

7. Wölfel, R., Corman, V.M., Guggemos, W. et al.
Virological assessment of hospitalized patients with COVID-2019.
Nature. 2020;**581:**465–469

8. Li N, Wang X, Lv T.
Prolonged SARS-CoV-2 RNA shedding: Not a rare phenomenon.
J. Med. Virol. 2020; doi:10.1002/jmv.25952

9. Alvarez-Moreno CA, Rodríguez-Morales AJ.
Testing Dilemmas: Post negative, positive SARS-CoV-2 RT-PCR - is it a reinfection?
Travel Med Infect Dis. 2020;**35**:101743.

10. Kumar P, Pandey R, Sharma P et al.
Integrated genomic view of SARS-CoV-2 in India.
Wellcome Open Res 2020;**5**:184

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

11. Heng Li
Minimap2: pairwise alignment for nucleotide sequences
Bioinformatics. 2018;**34(18)**:3094–3100

12. Loman, N., Quick, J., Simpson, J.
A complete bacterial genome assembled de novo using only nanopore sequencing data.
Nat Methods. 2015;**12**:733–735

13. Rambaut, A., Holmes, E.C., O'Toole, Á., et al.
A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology.
Nat Microbiol. 2020
https://doi.org/10.1038/s41564-020-0770-5

14. Cingolani P, Platts A, Wang le L, 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). 2012;**6(2)**:80–92.

15. Wang C, Liu Z, Chen Z, et al.
The establishment of reference sequence for SARS-CoV-2 and variation analysis.
J Med Virol. 2020;**92(6)**:667–674.

16. Singer J, Gifford R, Cotten M, Robertson D.
CoV-GLUE: A Web Application for Tracking SARS-CoV-2 Genomic Variation.
Preprints.org. 2020. (published on 18 June 2020) (preprint)
DOI: 10.20944/preprints202006.0225.v1.

17. Shu Y, McCauley J.
GISA ID: Global initia tive on sharing all influenza data – from vision to reality.
Euro Surveill. 2017;**22(13):**30494.

18. Sinclair A, Abdelhafiz A.
Age, frailty and diabetes – triple jeopardy for vulnerability to COVID-19 infection.
Eclinicalmedicine. 2020;**22**:100343.

19. Mercatelli D, Giorgi F.
Geographic and Genomic Distribution of SARS-CoV-2 Mutations.
Front. Microbiol. 2020;**11:**1800

20. Ceraolo C, Giorgi FM.
Genomic variance of the 2019-nCoV coronavirus.
J Med Virol. 2020;**92(5)**:522–528.

21. Lu R, Zhao X, Li J, et al.
Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding.
Lancet. 2020;**395(10224)**:565–574

22. Wajnberg A, Amanat F, Firpo A et al.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

SARS-CoV-2 infection induces robust, neutralizing antibody responses that are stable for at least three months
medRxiv. 2020 (published on 17 July 2020) (Preprint)
https://doi.org/10.1101/2020.07.14.20151126

23. Le Bert N, Tan AT, Kunasegaran K, et al.
SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls.
Nature. 2020;**584(7821)**:457–462.

24. Juno, J.A., Tan, H., Lee, W.S. et al.
Humoral and circulating follicular helper T cell responses in recovered patients with COVID-19.
Nat Med. 2020. https://doi.org/10.1038/s41591-020-0995-0

25. Ibarrondo F, Fulcher J, Goodman-Meza D, et al.
Rapid Decay of Anti–SARS-CoV-2 Antibodies in Persons with Mild Covid-19.
N Engl J Med. 2020. DOI: 10.1056/NEJMc2025179

26. Galanti M, Shaman J.
Direct Observation of Repeated Infections With Endemic Coronaviruses.
J Infect Dis. 2020. doi:10.1093/infdis/jiaa392

27. Levoir L, Goo L.
A class-switch in the antibody response may distinguish primary from secondary dengue virus infection.
Ebiomedicine. 2020;**56**:102790.

28. Arvin AM, Fink K, Schmid MA, et al.
A perspective on potential antibody-dependent enhancement of SARS-CoV-2.
Nature. 2020;**584(7821)**:353–363.

29. Parikh S, Desai M, Parikh R.
The Coronavirus : What You Need To Know About The Global Pandemic. 1st ed.
Mumbai, India. Penguin Random House India. 2020: p110–114

30. Young B, Fong S, Chan Y, et al.
Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study.
Lancet. 2020;**396**:603–611.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Table 1: Clinical Course, RT-PCR, Serology

	Age, Gender, Occupation, Pre-existing conditions	**Symptoms**	**Date of symptom onset**	**Duration of symptoms**	**Reason for testing**	**Date of RT-PCR positivity**	**Date of RT-PCR negativity**	**Duration Between First and Second Positive RT-PCR tests**	**Anti-SARS-CoV2 Serology by CLIA**
Patient A	27 years, Male, Resident Doctor treating COVID-19 patients, No pre-existing illnesses	Sore throat, nasal congestion and rhinitis	17/5/20	Two days	Contact tracing	15/5/20	19/05/2020	66	Negative after second episode on 23/07/20 (Abbott Anti-NC IgG)
Patient A f/u		Myalgia, fever, non-productive cough, fatigue	18/7/20	One week	Symptomatic	19/7/20	29/7/20		
Patient B	31 years, Male, Resident Doctor treating COVID-19 patients, No pre-existing illnesses	None	N/A	N/A	Contact tracing	15/5/20	18/5/20	65	Negative after second episode on 23/07/20 (Abbott Anti-NC IgG)
Patient B f/u		Myalgia, malaise	16/7/20	Two days	Symptomatic	18/7/20	25/7/20		

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Patient C	27 years, Male, Resident Doctor treating COVID-19 patients, No pre-existing illnesses	None	N/A	N/A	Screening prior to going home to visit parents	25/6/20	27/6/20	19	Not done
Patient C f/u		Fever, headache, myalgia and a non-productive cough	10/7/20	Six days	Symptomatic	13/7/20	27/7/20		
Patient D	24 years, Female, Staff Nurse treating COVID-19 patients, No pre-existing illnesses	Sore throat, rhinitis and myalgia	13/5/20	Five days	Symptomatic	14/5/20	N/A	55	Negative after second episode on 25/07/20 (Roche Anti-NC Total Antibody)
Patient D f/u		Fever, myalgia, rhinitis, sore throat, non productive cough and fatigue	5/7/20	Three weeks	Symptomatic	7/7/20	N/A		

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Table 2: Mutation Analysis of Healthcare workers with reinfections (n=4)

Name	Length	**Mutations**	% Genome Coverage
Patient A	29866	C241T,C3037T,G3231T,C3397T,C3634T,C14408T,C15324T,A22459T,C23185T,A23403G,G23593T,T23671C,C25710T,C28045T,C29095T	97.75
Patient A f/u	29866	C3037T,C3634T,C14408T,C15324T,A22459T,G23593T,T23671C	90.73
Patient B	29728	C3037T,G5857T,A9274G,C14408T,A23403G,G28881A,G28882A	90.05
Patient B f/u	29866	C1884T,C3037T,C3634T,C7604T,C14408T,A18262G,A23403G,C23613T	89.22
Patient C	29866	C3634T,C14408T,A15435G,C28866T	80.07
Patient C f/u	29532	C3037T,T8022G,C8175A,T22137C,A22374G	82.55
Patient D	29384	C3037T,C14408T,C20926T	84.74
Patient D f/u	29866	C241T,C313T,C3037T,T3442C,C5700A,C14120T,C14408T,G18213T,C18705T,A23403G,G28881A,G28882A,G28883C	98.23

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Supplementary Table 1 : Sample collection, RNA extraction, RT-qPCR, aliquots, Ct value

Patient names	**Date of RT-PCR positivity**	**HCW's Institute**	**Sample**	**Collection details**	**RT-qPCR done at**	**Aliquoting and storage**	**RNA extraction and RT-PCR**	**Ct Values**		
			Nasopharyngeal (NP) Oropharyngeal (OP)					**N gene**	**ORF1ab**	**S gene**
Patient A	15/5/20	BYL Nair Hospital	NP+OP	Collected at BYL Nair hospital in HiViral™ Transport Kit and transported in cold chain to Kasturba Hospital for testing.	Kasturba Hospital is the designated COVID-19 testing centre for BYL Nair Hospital and initial RT-PCR testing and subsequent verification were done	Samples were divided into four aliquots (1-4) and RT PCR was performed on aliquot 1. Aliquots 2-4 were stored at -80C degrees for future use.	Automated RNA extraction was performed on three HCWs' samples using Mylab's Maverick Magnetic Bead-based Extraction kit on KingFisher Flex Extraction System followed by multiplex real-time RT-PCR using	32	32	Nil
Patient A f/u	19/7/20	BYL Nair Hospital	NP+OP					25	23	23
Patient B	15/5/20	BYL Nair Hospital	NP+OP					33	Nil	32
Patient B f/u	18/7/20	BYL Nair Hospital	NP+OP					36	38	Nil
Patient C	25/6/20	BYL Nair Hospital	NP+OP					36	Nil	35

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Patient C f/u	13/7/20	BYL Nair Hospital	NP+OP		here.		TaqPath™ COVID19 CEIVD RTPCR kit for the qualitative detection of nucleic acid of SARS-CoV-2 from Applied Biosystems.	21	20	20
Patient D	14/5/20	PD Hinduja	NP+OP	Collected at PD Hinduja Hospital	PD Hinduja Hospital was its own COVID-19 diagnostic centre. Initial RT-PCR was performe d here and remains sample to transferre d Kasturba Hospital (referenc	Aliquot of extracted RNA transferred to our lab was utilised for whole genome sequencing by the sequencing group	Manual extraction was done and PCR was conducted on both times using the above kit & Xpert® Xpress SARS-CoV-2	32	34	35
Patient D f/u	7/7/20	PD Hinduja	NP+OP					17	18	21

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

					e lab for Mumbai)					

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Supplementary table 2

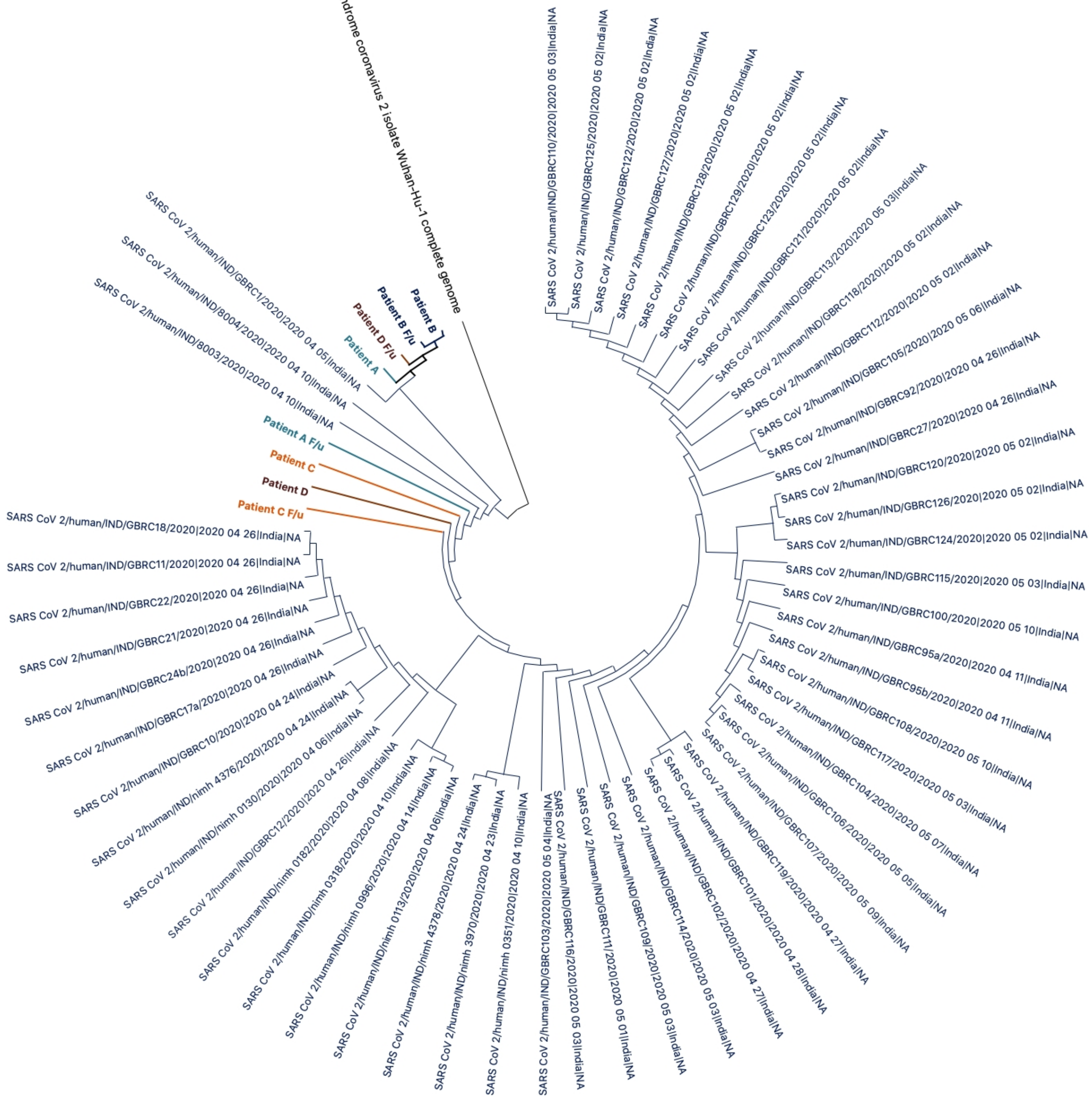
PANGOLIN Analysis of lineage distribution

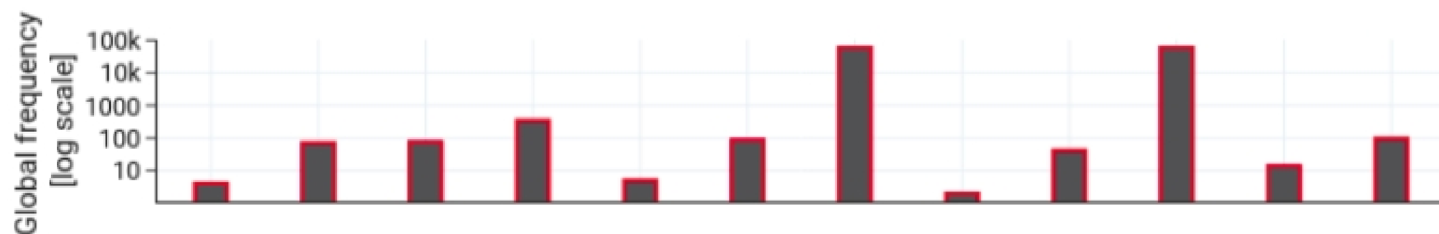
Sample name	Lineage Assigned	Lineage description	Assignment Probability	Number of taxa	Most common countries	Date range	Days since last sampling
Patient A	B.1	A large European lineage that corresponds to the Italian outbreak.	0.51	12883	USA, UK, Australia	February-03, June-21	74
Patient A f/u	B	Base of this lineage also lies in China, with many global exports, two distinct SNPs 8782TC and 28144CT define this lineage	0.42	1955	UK, China, USA	December-24, June-02	93
Patient B	B.1.1	New European lineage that's been assigned due to high support and 3 clear SNPs 28881GA,28882GA,28883GC. Note: Sub-lineages that previously existed within this lineage have been reassigned a new lineage name.	0.96	9643	UK, USA, Portugal	February-23, June-14	81
Patient B f/u	B.1.1	New European lineage that's been assigned due to high support and 3 clear SNPs 28881GA,28882GA,28883GC. Note: Sub-lineages that previously existed within this lineage have been reassigned a new lineage name.	0.77	9643	UK, USA, Portugal	February-23, June-14	81
Patient C	B.1.1	New European lineage that's been assigned due to high support and 3 clear SNPs 28881GA,28882GA,28883GC. Note: Sub-lineages that previously existed within this lineage have been reassigned a new lineage name.	0.77	9643	UK, USA, Portugal	February-23, June-14	81

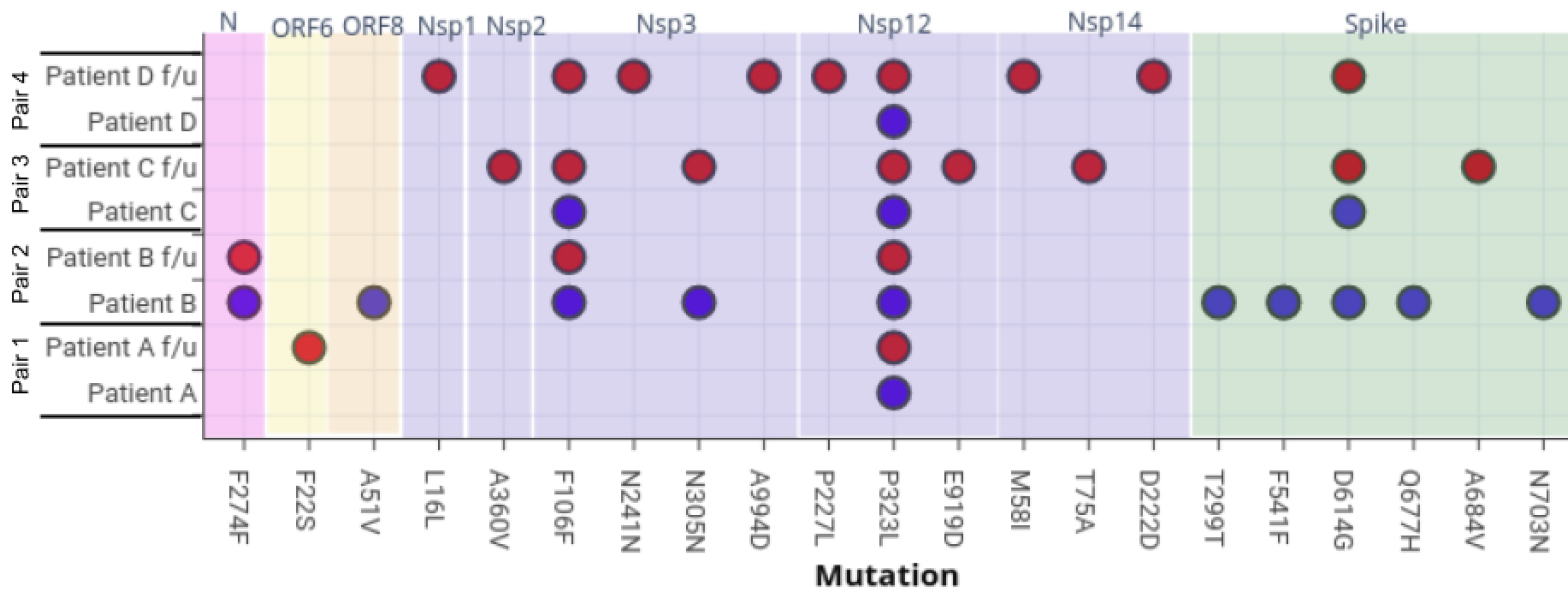
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220

Patient C f/u	B.1.1	New European lineage that's been assigned due to high support and 3 clear SNPs 28881GA,28882GA,28883GC. Note: Sub-lineages that previously existed within this lineage have been reassigned a new lineage name.	0.85	9643	UK, USA, Portugal	February-23, June-14	81
Patient D	B.1.1	New European lineage that's been assigned due to high support and 3 clear SNPs 28881GA,28882GA,28883GC. Note: Sub-lineages that previously existed within this lineage have been reassigned a new lineage name.	0.88	9643	UK, USA, Portugal	February-23, June-14	81
Patient D f/u	B.1.1	New European lineage that's been assigned due to high support and 3 clear SNPs 28881GA,28882GA,28883GC. Note: Sub-lineages that previously existed within this lineage have been reassigned a new lineage name.	0.91	9643	UK, USA, Portugal	February-23, June-14	81

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3688220


MN908947.3 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 complete genome
Patient B
Patient B F/u
Patient D F/u
Patient A
SARS CoV 2/human/IND/GBRC1/2020|2020 04 05|India|NA
SARS CoV 2/human/IND/8004/2020|2020 04 10|India|NA
SARS CoV 2/human/IND/8003/2020|2020 04 10|India|NA
Patient A F/u
Patient C
Patient D
Patient C F/u
SARS CoV 2/human/IND/GBRC18/2020|2020 04 26|India|NA
SARS CoV 2/human/IND/GBRC11/2020|2020 04 26|India|NA
SARS CoV 2/human/IND/GBRC22/2020|2020 04 26|India|NA
SARS CoV 2/human/IND/GBRC21/2020|2020 04 26|India|NA
SARS CoV 2/human/IND/GBRC24b/2020|2020 04 26|India|NA
SARS CoV 2/human/IND/GBRC17a/2020|2020 04 26|India|NA
SARS CoV 2/human/IND/GBRC10/2020|2020 04 24|India|NA
SARS CoV 2/human/IND/nimh 4376/2020|2020 04 24|India|NA
SARS CoV 2/human/IND/nimh 0130/2020|2020 04 06|India|NA
SARS CoV 2/human/IND/GBRC12/2020|2020 04 26|India|NA
SARS CoV 2/human/IND/nimh 0182/2020|2020 04 08|India|NA
SARS CoV 2/human/IND/nimh 0318/2020|2020 04 10|India|NA
SARS CoV 2/human/IND/nimh 0996/2020|2020 04 14|India|NA
SARS CoV 2/human/IND/nimh 0113/2020|2020 04 06|India|NA
SARS CoV 2/human/IND/nimh 4378/2020|2020 04 24|India|NA
SARS CoV 2/human/IND/nimh 3970/2020|2020 04 23|India|NA
SARS CoV 2/human/IND/nimh 0351/2020|2020 04 10|India|NA
SARS CoV 2/human/IND/GBRC103/2020|2020 05 04|India|NA
SARS CoV 2/human/IND/GBRC116/2020|2020 05 03|India|NA
SARS CoV 2/human/IND/GBRC111/2020|2020 05 01|India|NA
SARS CoV 2/human/IND/GBRC109/2020|2020 05 03|India|NA
SARS CoV 2/human/IND/GBRC114/2020|2020 05 03|India|NA
SARS CoV 2/human/IND/GBRC102/2020|2020 04 27|India|NA
SARS CoV 2/human/IND/GBRC101/2020|2020 04 28|India|NA
SARS CoV 2/human/IND/GBRC119/2020|2020 04 27|India|NA
SARS CoV 2/human/IND/GBRC107/2020|2020 05 09|India|NA
SARS CoV 2/human/IND/GBRC106/2020|2020 05 05|India|NA
SARS CoV 2/human/IND/GBRC104/2020|2020 05 07|India|NA
SARS CoV 2/human/IND/GBRC117/2020|2020 05 03|India|NA
SARS CoV 2/human/IND/GBRC108/2020|2020 05 10|India|NA
SARS CoV 2/human/IND/GBRC95b/2020|2020 04 11|India|NA
SARS CoV 2/human/IND/GBRC95a/2020|2020 04 11|India|NA
SARS CoV 2/human/IND/GBRC100/2020|2020 05 10|India|NA
SARS CoV 2/human/IND/GBRC115/2020|2020 05 03|India|NA
SARS CoV 2/human/IND/GBRC124/2020|2020 05 02|India|NA
SARS CoV 2/human/IND/GBRC126/2020|2020 05 02|India|NA
SARS CoV 2/human/IND/GBRC120/2020|2020 05 02|India|NA
SARS CoV 2/human/IND/GBRC27/2020|2020 04 26|India|NA
SARS CoV 2/human/IND/GBRC92/2020|2020 04 26|India|NA
SARS CoV 2/human/IND/GBRC105/2020|2020 05 06|India|NA
SARS CoV 2/human/IND/GBRC112/2020|2020 05 02|India|NA
SARS CoV 2/human/IND/GBRC118/2020|2020 05 02|India|NA
SARS CoV 2/human/IND/GBRC113/2020|2020 05 03|India|NA
SARS CoV 2/human/IND/GBRC121/2020|2020 05 02|India|NA
SARS CoV 2/human/IND/GBRC123/2020|2020 05 02|India|NA
SARS CoV 2/human/IND/GBRC129/2020|2020 05 02|India|NA
SARS CoV 2/human/IND/GBRC128/2020|2020 05 02|India|NA
SARS CoV 2/human/IND/GBRC127/2020|2020 05 02|India|NA
SARS CoV 2/human/IND/GBRC122/2020|2020 05 02|India|NA
SARS CoV 2/human/IND/GBRC125/2020|2020 05 02|India|NA
SARS CoV 2/human/IND/GBRC110/2020|2020 05 03|India|NA


ORF6
ORF8
Nsp2
Nsp3
Nsp12
Nsp14
Spike
Pair 4
Patient D f/u
Patient D
Pair 3
Patient C f/u
Patient C
Pair 2
Patient B f/u
Patient B
Pair 1
Patient A f/u
Patient A
F22S
A51V
A360V
A994D
E919D
P227L
P323L
T75A
M58I
D614G
A684V
Q677H
Mutation
Global frequency
[log scale]
100k
10k
1000
100
10


N
ORF6
ORF8
Nsp1
Nsp2
Nsp3
Nsp12
Nsp14
Spike
Pair 4
Patient D f/u
Patient D
Pair 3
Patient C f/u
Patient C
Pair 2
Patient B f/u
Patient B
Pair 1
Patient A f/u
Patient A
F274F
F22S
A51V
L16L
A360V
F106F
N241N
N305N
A994D
P227L
P323L
E919D
M58I
T75A
D222D
T299T
F541F
D614G
Q677H
A684V
N703N
Mutation