Original article

GENOMIC CHARACTERIZATION OF SEVERE ACUTE RESPIRATORY SYNDROME CORONA VIRUS-2 (SARS-COV-2) STRAINS CIRCULATING IN CHENNAI

Shanmugam Siva Kumar [1], Selvaraj Anbalagan [1], Palani Nandhini [1], Nesakumar Manohar [1], Haresh Adikesavalu [1], Hanna Luke E [1], Padmapriyadarsini Chandrasekaran [1], Jagadeesan M [2], Varghese Alby John [2], V P Brintha[2], Narayanan Manikandan [3], Potdhar Varsha [4]

[1] ICMR-National Institute for Research in Tuberculosis
[2] Greater Chennai Corporation
[3] BIRDS Group, IITM, Chennai
[4] ICMR-National Institute for Virology

DOI: https://doi.org/10.3126/saarctb.v22i1.72462
Received: 23rd May Accepted: 5th June Published: 31st Dec
This article is available at: https://www.saarctb.org/wp-content/uploads/2024/08/STAC_Journal_2023_VOL21_2024-AUGUST-2.pdf

ABSTRACT

Introduction: COVID-19 cases began from January 30, 2020, and continued transmission has ensued in multiple states including Tamil Nadu. During the epidemic, the SARS CoV-2 genome has accumulated several mutations. The evolution of viral variants has significant implications for epidemiology, diagnosis, and management, as well as the emergence of re-infection. In this study, we undertook Whole Genome Sequencing (WGS) of SARS CoV-2 genomes collected from SARS CoV-2 infected persons at two time points one year apart, May 2020 and April/May 2021, from one of the metropolitan cities of India. RNA was extracted from oropharyngeal and nasopharyngeal swab specimens collected from 85 COVID-19 infected individuals with a CT value of less than 30. WGS of SARS CoV-2 was performed using Oxford Nanopore Technology (ONT).

A striking change was noticed in the major circulating strains of SARS-CoV-2 from May 2020 (B.1 and B.1.1.1) to May 2021 (B.1.617.1 and B1.1.7). Our findings reveal a clear shift in the distribution of sub-lineages during the last year in the local population, with the newly identified double mutant/variants of concern overtaking the local epidemic. We noticed that infection with the newly emerged mutant strains was more common in the younger age group (18-44 years) as compared to older individuals (>45 Years).

We recommend continuous genomic surveillance to identify the emergence of newer strains in the city and detect early warning signs for an upsurge in COVID-19 cases due to new variants. Under the given circumstances, sustained efforts to strengthen enforcement of physical distancing, use of masks and sanitization to contain virus transmission, besides vaccination of younger individuals is advisable

Key words: SARS CoV-2, Tuberculosis, Oxford Nanopore Technology, COVID-19

INTRODUCTION

COVID-19 cases began from January 30, 2020 and since then more than twenty million people have died in the country[1]. When a virus replicates there is a possibility for it to change its characteristics by undergoing mutations. A virus with one or several new mutations is referred to as a "variant" of the original virus. Since it was first identified, the SARS CoV-2 genome has accumulated several mutations during the infection of human hosts resulting in a number of variant strains in different parts of the globe [2,3]. This has significant implications for disease surveillance and management of infection as well as re-infection, through vaccines and drugs.

Correspondence:
Dr. S Siva Kumar
Scientist E and HOD
ICMR-National Institute for Research in Tuberculosis
No. 1, Mayor Sathyamoorthy Road, Chetput, Chennai 600031
Ph: 044-2836-9682
Email: shanmugamsiva27@gmail.com

The genome sequence data of SARS-CoV-2 has classified it as a member of the genus Betacoronavirus, subgenus Sarbecovirus[3]. Epidemiological data indicate that SARS-CoV-2 virus spread widely from the Wuhan in China to other parts of the globe[4].

The sudden outbreak of a new SARS-CoV-2 variant in India has led to much concern. Establishing genomic surveillance protocols to better understand and curb the spread of the virus in a rapid and robust manner is the need of the hour. The government of India established the Indian SARS-CoV-2 Genomics Consortium to expand whole-genome sequencing of the SARS-CoV-2 viruses across the nation, in order to understand how the virus spreads and evolves. The Consortium detected 771 variants of concerns (VOCs) in a total of 10787 positive samples shared by States/UTs. These included 736 samples positive for the UK variant (B.1.1.7), 34 samples positive for the South African variant (B.1.351) and one sample positive for the Brazilian variant (P.1). The VOCs were identified in 18 states of the country[5]. In the present study, we sequenced the whole genome of SARS-CoV-2 from oropharyngeal and nasopharyngeal swab specimens collected at two different time points (one year apart) since the origin of the pandemic, with a view to understanding the molecular epidemiology of the local pandemic in Chennai, one of the large metropolitan cities in India.

METHODOLOGY

Upper respiratory tract samples including oropharyngeal and nasopharyngeal swab specimens were collected from suspected COVID-19 infected individuals residing in the Greater Chennai Corporation of Chennai city of Tamil Nadu, India. As part of the routine diagnostic tests samples were tested for the existence of SARS-CoV-2 using real-time RT-PCR. Left-over specimens were stored at -800C for repeat testing and as part of biorepository in India (ICMR). For the present study and analysis, 85 positive samples with CT value <25 were randomly selected among samples collected at 2 different time points, May 2020 and April/May 2021. Viral RNA was freshly extracted from 200ul of supernatant of the stored clinical samples using an automated RNA extractor (Purifier 96, Genetix) and subjected to whole-genome sequencing on the MinION platform Figure 1.

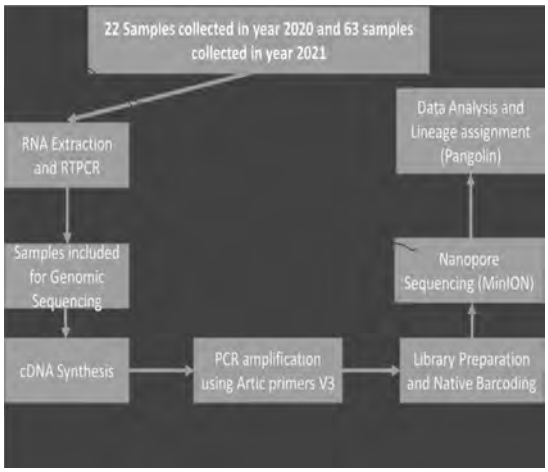

Figure 1. The study design is retrospective and stored samples were subjected to RNA extraction and Nanopore sequencing.

We used the tiling primer-based methodology for whole genome sequencing as described by the ARTIC Network using the Primal Scheme [https://protocols.io/view/ncov-2019-sequencing-protocol-v3-locost]. Briefly, we used the V3 primer set (96 pairs with amplicons of about 400 base pairs (bp) spanning the whole genome except 31bp at the 5' end and a part of the 3' UTR. PCR was performed by pooling adjacent/overlapping primers into different pools so as to prevent preferential amplification of short fragments between adjacent primer pairs. The resulting PCR amplicons were used for preparing libraries for Nanopore sequencing using the native barcoding (NBD196) approach combined with the ligation sequencing kit (SQK-LSK109). 12–60 samples were barcoded and included in a single run. The resulting DNA was cleaned up and added to the FLO-MIN-106 flow cell and sequenced on the MinION [https://protocols.io/view/ncov-2019-sequencing-protocol-v3-locost].

In addition, a no-template negative control from the PCR amplification step was prepared in parallel and sequenced on each flow cell. We also performed multiple re-runs of the same sample for quality control. The MiniKow software package was used to monitor sequencing performance in real-time; runs proceeding until a minimum of 100,000 reads were achieved across all barcodes. At this point, the run was terminated, and the flow-cell was washed using the ONT Flow Cell Wash kit, allowing re-use in subsequent runs. The fastq data obtained from

each barcode was subjected to Fast QC and an in-house pipeline for the generation of the FASTA file. The fasta file was used as an input for Pangolin COVID-19 Lineage Assigner. For lineage analysis, we performed genome analysis as described by Rambaut et al., 2020[6], (Pango lineage: pangolin v2.4.2 accessed on 24th May 2021). Pango Lineage is designed to capture the emerging edge of the pandemic and is at a fine-grain resolution suitable for genomic epidemiological surveillance and outbreak investigation[6] SARS-CoV-2, which is associated with severe pneumonia/disease (COVID-19). The quality check of the raw reads is done using FastQC (v 0.11.9) [https://www.bioinformatics.babraham.ac.uk/publications.html], followed by adapter trimming using Porechop (v 0.2.4) tool. The bwa mem (v 0.7.17)[7] tool is used to align raw reads to the SARS-CoV-2 reference genome (NC_045512), followed by deduplication of reads using the GATK (v 4.1.8.0) Mark Duplicates tool. Samtools mpileup, vcfutils.pl, and bcftools tools[8] are used for variant calling and to generate consensus sequences. The variants that do not satisfy the depth criteria (>5 or <300) are filtered. The online version of Pangolin (v 2.4.2) (https://github.com/cov-lineages/pangolin) tool is used to determine the consensus sequences' lineages. Pangolin uses a multinomial logistic regression model trained using the known lineages of sequences from GISAID to assign lineage to a new sequence. Phylogenetic analysis of the sequences is performed using the NGPhylogeny workflow[9], which includes a set of tools, MAFFT, BMGE and FastME, for multiple sequence alignment, alignment curation and phylogenetic tree inference respectively.

RESULTS

A total of 85 SARS CoV-2 positive samples collected at two-time points, May 2020 and April/May 2021 were selected for this analysis. All individuals were symptomatic and had either fever, breathlessness, or cough at the time of testing. Eleven of these individuals were vaccinated; 5 had Covisheild and 6 had Covaxin. Six of these individuals were above 45 years of age and the rest of them were below 45 years of age. One required hospitalization (age 50) while the others required only home quarantine. The hospitalized patient had taken the first dose of the vaccine in the last week of April 2021. Of the remaining 64 who were non-vaccinated, 5 required hospitalizations. There was no mortality in the study group.

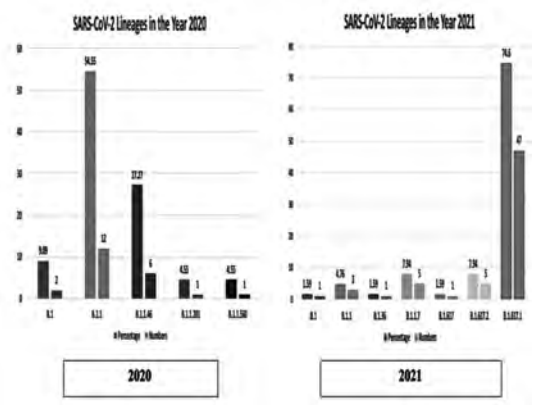

Figure 2: Distribution of SARS-CoV-2 Lineages between 2020 and 2021.

We were able to sequence 85 genomes and deposited them in the GISAID database (The ID for the deposit is mentioned at the end of the manuscript). Of the 85 genomes sequenced, all 85 sequences were complete (>92% at 10X and >85%at 50X) and have been deposited in the GISAID database. The distribution of lineages and the time points are provided in Figure 2. The major lineages identified included B.1.1 (2020: 54.5%, 2021: 4.76%), B.1 (2020: 9.1%, 2021: 1.6%), B.1.1.46 (2020: 27.27%, 2021: 0%), Alpha variant B.1.1.7 (2020: 0%, 2021: 7.94%), B.1.617 (2020: 0%, 2021: 1.59%), Kappa Variant B.1.617.1(2020: 0%, 2021: 74.6%) and Delta variant B.1.617.2 (2020: 0%, 2021: 7.94%). Out of the 63 samples collected in 2021, 47 samples showed the presence of the Kappa variant B.1.617.1 lineage, suggesting major transmission of this lineage in the community presently. 30 of the 47 samples were from individuals aged less than 45 years of age, and 17 were aged more than 45 years. The rest of the 38 samples had different lineages as mention above, 8 individuals were greater than 45 and 26 were less than 45 years old.

We also performed a phylogenetic analysis of the assembled sequences and found a clear difference in prevalent strains isolated between last year (May 2020) and this year (April/May 2021). Closer clustering was observed with the recent SARS CoV-2 viral genomes than the older isolates, indicating faster transmission within close contacts in the current scenario (Fig 3). The most common

variants B.1.617, Kappa variant B.1.617.1, Delta Variant B.1.617.2 identified in our study have been classified as variants of concern recently by WHO. These variants have been reported to be associated with a potential reduction in neutralization by some EUA monoclonal antibody treatment and slightly reduced neutralization by post-vaccination sera. The B.1.1.7 Alpha variant (UK variant) also classified as a variant of concern was noticed less frequently in the Chennai population.

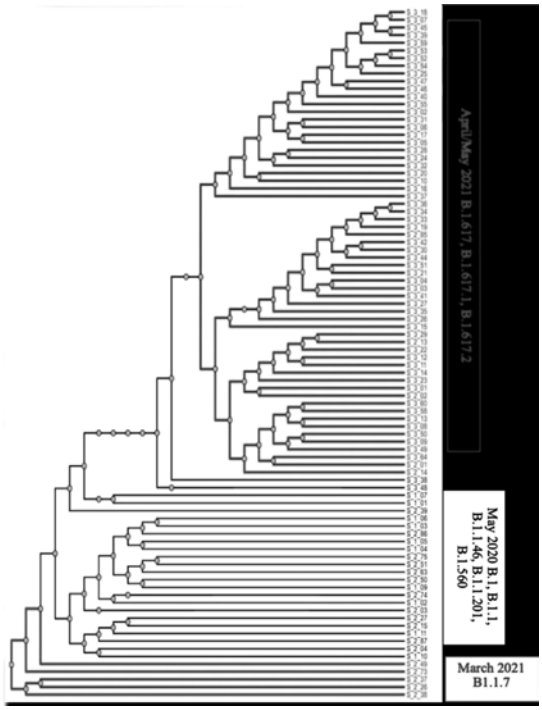

Figure 3: Linear Dendrogram of Fasta files of the genomic sequences of SARS CoV-2 with the lineages. The red color lines denote samples from the 2nd wave (2021). Other lines represent samples collected during the first wave (2020).

DISCUSSION

Systematic analysis of the SARS CoV-2 genome across time would help monitor changes in the genome and provide an opportunity to comprehend the genetic epidemiology and evolution of the virus and allow us to track its spread in a region, country, and globally. Incidentally, most of the mutations leading to the emergence of new variants have taken place only in spike protein, which enables the virus to attach to human cells as first step towards infection. This segment of the genome thus operates as a key region for monitoring mutations. We employed Nanopore sequencing for the present analysis as it offers several advantages. It is relatively inexpensive, highly portable as it requires minimal associated laboratory infrastructure, enables rapid generation of sequence data, and permits real-time data analysis[1].

A study from Karnataka, India, which analyzed 91 genomes of SARS-CoV-2 collected during the period March to May 2020 showed the clustering of the sequences into seven Pangolin lineages (A, B, B.1, B.1.80, B.1.1, B.4 and B.6). The present study also included samples from the same time but found B.1.1 (54.55%) as the most common lineage. The other lineages present were B.1 and B.1.46, suggesting some common introduction of strains, along with strains evolved in the community. A similar study all over the country has shown similar data[10].

Two variants as defined in the Pangolin nomenclature—B.1.1.7, and B.1.351 have recently been reported from India in the 2nd wave [23, 24]. These variants are of concern due to antigenic drift, increased transmissibility, and immune escape (particularly for B.1.351). Recently, a new PangoLIN lineage (B.1.617) was identified in the Indian SARS-CoV-2 sequences, with two significant mutations, E484Q and L452R in the spike protein (commonly referred to as the double mutant), and is believed to have a higher transmission rate[11]. The present study demonstrates a preponderance of strains categorized as VOC in the local population, strongly implying that this could be one of the major reasons for the rapid spread of the virus in the community since these variants harbor significant mutations in the spike protein that enable the virus to transmit rapidly.

In the present study, we had 11 individuals who had been infected in spite of vaccination; of these, 5 had taken Covishield and 6 had Covaxin. One of these individuals (aged 50 years) required hospitalization while the others required only home quarantine. These are termed as "breakthrough" infections, suggesting that the virus has been able to overcome the defence mechanism mounted by the vaccine. Our data suggests that vaccination reduces the requirement for hospitalization and/or severe disease. One limitation of the study is that the sequencing is limited on one zone of the city, which may present an incomplete picture of overall circulating viruses in the city and increase the proportion of the some that are more readily sequenced.

In summary, our data highlight a sharp and significant increase in the frequency of the B.1.617.1 Kappa lineage in Chennai. Whether this numbers is the result of epidemiological linkages such as more travel, sustained local transmission chains or super-spreader events remains to be probed. Our data highlights for a continued genomic surveillance, including sequencing of SARS-CoV-2 genomes from travellers, outbreaks and breakthrough infections in a systematic manner to rapidly detect newer emerging variants. Rapid identification of such variants could go a long way in preparing the healthcare system for an upsurge in cases, revisiting vaccine strategy, diagnostic tests, and guide public health measures.

ACKNOWLEDGEMENT

The authors would like to thank the laboratory personnel of department of HIV/AIDS and Bacteriology for the support provided for the study. We also thank the Commissioner Greater Chennai Corporation for the financial support

CONFLICT OF INTEREST

None

FUNDING

The research was funded by Greater Chennai Corporation and SARRC TB and HIV/AIDS Centre.

REFERENCES

1. Pattabiraman C, Habib F, Harsha PK, Rasheed R, Prasad P, Reddy V, et al. Genomic epidemiology reveals multiple introductions and spread of SARS-CoV-2 in the Indian state of Karnataka. PLoS One [Internet]. 2020;15(12 December):1–15. Available from: http://dx.doi.org/10.1371/journal.pone.0243412

2. Kaushal N, Gupta Y, Goyal M, Khaiboullina SF, Baranwal M, Verma SC. Mutational frequencies of SARS-CoV-2 genome during the beginning months of the outbreak in USA. Pathogens. 2020;9(7):1–16.

3. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet [Internet]. 2020;395(10224):565–74. Available from: http://dx.doi.org/10.1016/S0140-6736(20)30251-8

4. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science (80-). 2020;368(6489):395–400.

5. Genomic Surveillance for SARS-CoV-2 In India Indian SARS-CoV-2 Genomics Consortium (INSACOG). :1–18.

6. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol [Internet]. 2020;5(11):1403–7. Available from: http://dx.doi.org/10.1038/s41564-020-0770-5

7. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

9. Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S, et al. NGPhylogeny.fr: New generation phylogenetic services for non-specialists. Nucleic Acids Res. 2019;47(W1):W260–5.

10. Yadav PD, Nyayanit DA, Majumdar T, Patil S, Kaur H, Gupta N, et al. An Epidemiological Analysis of SARS-CoV-2 Genomic Sequences from Different Regions of India. 2021;

11. Pattabiraman C, Prasad P, George AK, Sreenivas D, Rasheed R, Reddy NVK, et al. Importation, circulation, and emergence of variants of SARS-CoV-2 in the South Indian State of Karnataka. medRxiv [Internet]. 2021;65:2021.03.17.21253810. Available from: http://medrxiv.org/content/early/2021/03/20/2021.03.17.21253810.abstract