Evaluation of sars Cov-2 variants among RT-PCR tested positive COVID-19 patients at our tertiary care hospital in Telangana State, India

Dr. Ramavath Usha Rani
Associate Professor, Dept. of Microbiology, SVS Medical College, Yenugonda, Mahabubnagar, AP
Corresponding author email: ramavathusha@gmail.com

Dr. Velagapalli Ramadevi
Professor and HOD, Dept. of Microbiology, Government Medical College, Mahabubnagar, AP

Dr. Qader Ahmed Jalily
Associate Professor, Dept. of Microbiology, MIMS, Shivareddypet, Vikarabad, Telangana

Dr. Dinesh Eshwar Mummareddi
MIMS, Shivareddypet, Vikarabad, Telangana

Abstract---Background: The initial reports of coronavirus disease 2019 (COVID-19) came out with an outbreak in Wuhan, China on 30 December 2019 [1,2], SARS-CoV-2 has been spreading worldwide and, as of 15 July 2021, there have been 188.13 million confirmed infections and 4.06 million deaths [3]. New variants are thought to be responsible for re-infections, either after natural infection or after vaccination, as observed in Brazil and the United States, respectively. There is evidence that re-infections are already happening in India; a recent survey identified a re-infection proportion of 4.5% from a pool of 1,300 participants infected between January 2020 and October 2020. Another concern about the emergence of new variants is the potential failure of RT-PCR tests for diagnostics. Failure to target the gene encoding the spike protein was observed during detection of the 501Y.V1 variant in the UK2. Aim and Objectives of the study: The aim and objective of the current study are to identify the lineages and calculate the prevalence among our population. Methodology: The work was initiated following approvals from the Institutional Ethical Committee. The samples were obtained as nasopharyngeal swabs collected in Viral Transport Medium from patients with symptoms suggestive of COVID-19 as well as asymptomatic primary contacts of
affected cases from different parts of Telangana. RT-PCR tested COVID
19 samples were aliquoted and stored at -80 and these samples were
transported in cold chain to The Centre for DNA Fingerprinting and
Diagnostics CDFD, Hyderabad once in a week for gene sequencing.
Results: Our data sheet consisted of total 345 samples collected and
tested for RT-PCR found positive with the Ct values ranging from >15
to <35 during July 2021 to February 2022. Out of the 345 samples
200 (57.9%) were males and 145 (42%) females in the age group of 0-
80 years. The data is from the month of June 2021 to February 2022.
The distribution of cases as per month wise shows that in July 31
cases (16 males, 15 females), August 34 cases (20 males, 14 females),
in September 59 cases (40 males, 19 females), in October 44 cases (27
males, 17 females), in November 38 cases (22 males, 16 females), in
December 17 cases (12 males, 5 females), in January 21 cases (16
males, 5 females) and February we had 101 cases (53 males and 48
females). The total vaccinated cases were 141 out of 345. In the
current study, 14 lineages were observed by using Phylogenetic
Assignment of Named Global Outbreak Lineage (Pangolin 3.1.4,
http://www.pangolin.cog-uk.io). Out of 345 samples, 213 showed the
gene sequencing rest were failed. The overall lineage distribution
highlighted the occurrence of B.1.617.1 (1.4%), B.1.167.2 (n=147),
B.1.1.306 (1.4%), B.1.36.29 (8.45%), B.1.1 (1.87%), B.1.36.8 (1.87%),
AY.4 (n =1.87%), AY.20 (2.81%), AY.39 (3.75%), AY.125 (2.34%),
AY.122 (1.4%), AY.102 (1.4%), AY.120 (1.4%) as mentioned in Table 2
and Figure 3. Discussion and Conclusion: We found that the most
common variant lineage was B.1.167.2 accounting for 69% percent
followed by B1.1 in our District of Telangana State. B.1.617.2, with
mutations L452R, T478K and P681R, is highly prevalent in the state
of Gujrat in India. There was enrichment for the mutation T478K or
T478R when SARS-CoV-2 was subjected to weak neutralizing
antibodies, which indicates this mutation may lead to antibody
escape. This variant needs strict monitoring and hence the prevention
of transmission.

Keywords---corona virus disease, reverse transcriptase polymerase
chain reaction, gene sequencing.

Introduction

The initial reports of coronavirus disease 2019 (COVID-19) came out with an
outbreak in Wuhan, China on 30 December 2019 [1,2], SARS-CoV-2 has been
spreading worldwide and, as of 15 July 2021, there have been 188.13 million
confirmed infections and 4.06 million deaths [3]. This was the reason that SARS-
CoV-2 epidemic was declared a global pandemic by World Health Organization
(WHO) on 11 March 2020 [4]. SARS-CoV-2 is a single stranded positive-sense
RNA virus belonging to the genus Betacoronavirus, and subgenus Sarbecovirus.
The genome size of SARS-CoV-2 is approximately 30kb and its genomic structure
has followed the characteristics of known genes of the coronavirus [5]. Albeit, RNA
viruses are highly prone to mutations but Coronaviruses (CoVs) have an
outstanding and distinctive feature of intrinsic proofreading mechanism [6]. However, CoVs encode a protein called non-structural protein 14 (nsp14) that possesses a 30 to 50 exonuclease (ExoN) activity which is necessary for replication fidelity and proofreading activity [7]. Owing to the presence of large and complex genome, this proofreading mechanism is considered critical for maintaining normal functioning and fitness of CoVs. Perhaps, this is the reason that mutation rate in SARS-CoV-2 is 10-fold lesser compared with other RNA viruses.

However, even then geneticists had reported a rate of 33 mutations/year in SARS-CoV-2 genome and now scientists are using these mutations to categorize different variants of this virus into different clades, lineages, and sublineages [4]. To date the SARS-CoV-2 has been divided in more than 81 lineages, based on variations in its genome, which vary significantly in their transmissibility and virulence [8]. All the existing lineages including variants of concerns are the descendants of two ancestral lineages, A and B, discovered from China in the start of the pandemic. However, the European Center for Disease Prevention and Control (ECDC) has demonstrated five variants of concerns (VOCs) due to their high transmissibility, pathogenicity, and effects on the vaccine efficacy, which is why they are known as VOCs. The alpha variant (B.1.1.7) also known as Alpha VOC, was first detected in the UK in September 2020 [9]. This variant is highly variable and contains more than a dozen mutations compared with wild-type lineages.

It has also been reported to have 50–70% high transmissibility [10], greater severity of the disease [11,12], and effects on the efficacy of the vaccine compared with wild-type lineage. The beta variant of concern, B.1.351, also known as 20H/S01Y.V2 was identified from South Africa. It shares several mutations with B.1.1.7 and has the following major amino acid mutations, K417N, E484K, N501Y, D614G, and A701V in its spike protein. This lineage is also reported to have increased transmissibility and even have the advantage of escaping from immunity due to the presence of E484K mutation in its spike protein [13].

Likewise, B.1.617.2 lineage (Delta variant), also known as 20A/S:478K, was identified in India at the end of 2020 and was reported to have the following major spike mutations, L452R, T478K, D614G, and P681R, which are involved in enhancing its transmissibility and risk of hospitalization compared with B.1.1.7 [14–16]. Likewise, P.1 lineage (Gamma), also known as 20J/501Y.V3, was identified from Brazil [17]. It has K417T, E484K, N501Y, D614G, and H655Y as the most noteworthy amino acid mutations which impact on its transmissibility and immunity [18]. B.1.427 and B.1.429 lineages (Epsilon), also collectively known as 20C/S452R [19] are two more variants of concern which are based on several spike protein mutations, including L452R, which is associated with increased cell entry and reduced susceptibility to neutralization by convalescent and vaccine recipient plasma in vitro [20].

The current outbreak of COVID-19 in India, which started in early March 2021, has created a new world record even beyond the outbreaks in the UK, the United States and Brazil. Prior to March 2021, less than 0.7% of the Indian population was infected with COVID-19. This current second wave took only 2 months to
infect an additional ~0.36% of the population, and India is now recording over 0.4 million new cases per day (as of 23 April 2021). The true number is probably even higher, with some estimates putting the number of daily new cases at over 1 million, more than five times the officially recorded number [21]. The sudden surge in COVID-19 cases in India coincides with high prevalence of more-transmissible variants, associated with diagnostic test failures and antibody escape [22]. These coronavirus SARS-CoV-2 variants of concern—B.1.1.7 (501Y.V1), B.1.351 (501Y.V2) and B.1.1.28.1 (501Y.V3; also known as P.1)—were observed during the sudden surge in COVID-19 cases in the UK, South Africa and Brazil, respectively, with subsequent local transmission across the world [22, 23]. In India, the frequency of 501Y.V1 is higher than that of 501Y.V2 and 501Y.V.3. The recently designated variant of concern B.1.617 and variant of interest B.1.618 have also been gaining attention in India. Variant B.1.617.1 shows co-occurrence of three key mutations in sequence encoding the viral spike protein: L452R, E484Q and P681R. L452R raised concerns in the United States as part of the California variants B.1.427 and B.1.429 and conferred resistance to the neutralizing monoclonal antibodies X593 and P2B-2F64. E484Q shares antibody-escape features similar to those of mutation E484K, seen in variants 501Y.V2, 501Y.V3 and B.1.6182. P681R may enhance processing by host proteases by extending the polybasic ‘RRAR’ motif, which results in a greater viral load and the potential for increased transmission. Another similar variant of concern, B.1.617.2, with mutations L452R, T478K and P681R, is highly prevalent in the state of Gujrat in India. There was enrichment for the mutation T478K or T478R when SARS-CoV-2 was subjected to weak neutralizing antibodies, which indicates this mutation may lead to antibody escape. Variant B.1.618, which has the E484K mutation, is prevalent in the state of West Bengal, India [23].

New variants are thought to be responsible for re-infections, either after natural infection or after vaccination, as observed in Brazil and the United States, respectively. There is evidence that re-infections are already happening in India; a recent survey identified a re-infection proportion of 4.5% from a pool of 1,300 participants infected between January 2020 and October 2020. Another concern about the emergence of new variants is the potential failure of RT-PCR tests for diagnostics. Failure to target the gene encoding the spike protein was observed during detection of the 501Y.V1 variant in the UK2. According to the latest guidelines of the Indian Council of Medical Research, approved test kits employ multiplex RT-PCR assays, as tests assessing only the spike protein may fail and thereby underestimate the true number of cases [24,25].

During mid-January 2021, vaccination programs commenced in India with the BBV152 (Covaxin) and ChAdOx1 (Covishield) vaccines. The potential for new variants to increase transmission and reduce the effectiveness of vaccines means that it is crucial to scale up Indian sequencing efforts and characterize the distribution of variants across all states [26]. Hence, we have taken up this study to assess the variants of SARS-CoV-2 at our tertiary care centre in collaboration with Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad.

**Aim and objective of the study**

The aim and objective of the current study are to identify the lineages and
calculate the prevalence among our population.

**Methodology**

Source of data: This study was conducted in RT-PCR tested positive COVID-19 patients of our hospital during the period from July 2021 to February 2022 at Government Medical College, Mahabubnagar, Telangana in collaboration with The Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad

Duration of study: The study was conducted during the period from July 2021 to February 2022.

Type of study: It is a prospective observational study.

Inclusion criteria:

RT-PCR tested positive COVID-19 samples having Ct values >10 and <35.

Exclusion Criteria:

Ct values <10 and >35.

**Sample Collection and Analysis**

The work was initiated following approvals from the Institutional Ethical Committee. The samples were obtained as nasopharyngeal swabs collected in Viral Transport Medium from patients with symptoms suggestive of COVID-19 as well as asymptomatic primary contacts of affected cases from different parts of Telangana. Total RNA was isolated using the RNA isolation kit as per manufacturer's instructions (QIAmp Viral RNA Mini Kit; Cat. No. 52906; Qiagen, Hilden, Germany). Each RNA sample was subjected to RT-PCR for multiple viral genes [including E-gene, RNA-dependent RNA polymerase (RDRP) gene, and N gene] using the LabGun COVID-19 assay (Cat. No. CV9017B; LabGenomics, Republic of Korea) or the Allplex 2019-nCoV Assay (Cat. No. RP10250X, Seegene, Republic of Korea). RT-PCR tested COVID 19 samples were aliquoted and stored at -80 and these samples were transported in cold chain to CDFD, Hyderabad once in a week for gene sequencing.

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 samples that 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-CoV2 specific gene (unlike E-gene, which is specific for all respiratory coronaviruses), we considered Ct (threshold cycle) values of RdRp alone for analysis. Samples exhibiting a Ct value greater than 10 and less than 35 were sent for sequencing.

**Sequencing protocol**

Sequencing of SARS-CoV-2 RNA samples was performed using the protocol described earlier (nCoV-2019 sequencing protocol; https:// dx. doi. org/ 10. 17504/ protocols. io. bdp7i5rm) with slight modifications. Briefly, RNA isolated from nasopharyngeal swabs was reverse transcribed using random primer mix
DNA libraries for Illumina sequencing were prepared using the NEB Next Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, MA, USA), according to the manufacturer’s protocol. Paired-end sequencing (2Å~250 bp) was performed on the Miseq FGx (Illumina, CA, USA) with a targeted depth of 0.5 million reads per sample (~4000Å~ coverage). Libraries for Nanopore sequencing were prepared using the Ligation sequencing kit (LSK-109; Oxford Nanopore Technologies, London, UK). Barcoded libraries were pooled (12–24 samples each) and sequenced on a MinION flow cell in GridION (Oxford Nanopore Technologies, London, UK). Sequencing was performed with a targeted depth of 0.1 million reads per sample (up to 24 h).

Phylogenetic analysis

The consensus fasta files generated for both Illumina and Nanopore data were subjected to 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 <27 000 and N’s >5 % were filtered out. 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 [17]. Following MSA, the workflow constructed a time-resolved phylogenetic tree using the maximum-likelihood-based method IQ-TREE [18], using the default nucleotide substitution model (‘GTR’ or general time reversible [19] implemented in IQ-TREE with 1000 bootstrap cycles). The resultant tree was pruned and internal nodes and ancestral traits were inferred from the dates of the sample collection using TreeTime [20]. The final tree in Newick format was then customized for visualization using iTol [21]. The nomenclature of Nextstrain-assigned phylogenetic clades is based on the naming scheme proposed by Rambaut et al. This scheme uses a year-letter nomenclature with numbers indicating the year of emergence of viral strain and letters A and B corresponding to the two root or reference sequences used for creating phylogenetic trees. Letter ‘A’ refers to the Wuhan/WH04/2020 (GISAID database accession – EPI_ISL_406801; GISAID-Global Initiative on Sharing All Influenza Data), while ‘B’ represents Wuhan-Hu-1 (GISAID accession – EPI_ISL_402125) sequence. These two sequences represent the earliest sampled strains of SARS-CoV-2, which were submitted to the GISAID database.

Results

Our data sheet consisted of total 345 samples collected and tested for RT-PCR found positive with the Ct values ranging from >15 to <35 during July 2021 to
February 2022. Out of the 345 samples 200 (57.9%) were males and 145 (42%) females in the age group of 0-80 years.

Table 1: Shows Age-wise and Gender-wise distribution of the Cases

<table>
<thead>
<tr>
<th></th>
<th>0-20</th>
<th>21-40</th>
<th>41-60</th>
<th>&gt;60</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>35</td>
<td>76</td>
<td>66</td>
<td>23</td>
<td>200</td>
</tr>
<tr>
<td>Females</td>
<td>37</td>
<td>60</td>
<td>35</td>
<td>13</td>
<td>145</td>
</tr>
<tr>
<td>Total</td>
<td>72</td>
<td>136</td>
<td>101</td>
<td>36</td>
<td>345</td>
</tr>
</tbody>
</table>

It is evident from the above table that, out of 345 cases, 200 were males and 145 were females.

Figure 1: Shows month wise, age wise and gender wise distribution of vaccinated and non-vaccinated males

Figure 2: Shows month wise, age wise and gender wise distribution of vaccinated and non-vaccinated females

It is evident form figure 1 and 2 that in each age group males were in predominance compared to females. The data is from the month of June 2021 to February 2022. The distribution of cases as per month wise shows that in July 31 cases (16 males, 15 females), August 34 cases (20 males, 14 females), in
September 59 cases (40 males, 19 females), in October 44 cases (27 males, 17 females), in November 38 cases (22 males, 16 females), in December 17 cases (12 males, 5 females), in January 21 cases (16 males, 5 females) and February we had 101 cases (53 males and 48 females). The total vaccinated cases were 141 out of 345.

Table 2: Shows SARS-COV2 Variants by Gene Sequencing

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>B.1.617.1</th>
<th>B.1.617.2</th>
<th>B.1.1.306</th>
<th>B.1.36.29</th>
<th>B.1.1.326</th>
<th>B.1.36.8</th>
<th>AY.4</th>
<th>AY.29</th>
<th>AY.39</th>
<th>AY.12</th>
<th>AY.122</th>
<th>AY.102</th>
<th>AY.120</th>
</tr>
</thead>
<tbody>
<tr>
<td>July</td>
<td>31</td>
<td>0</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>August</td>
<td>34</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>16</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>September</td>
<td>59</td>
<td>3</td>
<td>34</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>October</td>
<td>44</td>
<td>0</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>November</td>
<td>38</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>December</td>
<td>17</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>January</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>February</td>
<td>101</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>345</td>
<td>3</td>
<td>147</td>
<td>3</td>
<td>18</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 3: Shows SARS-COV2 Variants by Gene Sequencing

In the current study, 14 lineages were observed by using Phylogenetic Assignment of of Named Global Outbreak Lineage (Pangolin 3.1.4, http://www.pangolin.cog-uk.io). Out of 345 samples, 213 showed the gene sequencing rest were failed. The overall lineage distribution highlighted the occurrence of B.1.617.1 (1.4%), B.1.167.2 (n=147), B.1.1.306 (1.4%), B.1.36.29 (8.45%), B.1.1 (1.87%), B.1.36.8 (1.87%), AY.4 (n =1.87%), AY.20 (2.81%), AY.39 (3.75%), AY.125 (2.34%), AY.122 (1.4%), AY.102 (1.4%), AY.120 (1.4%) as mentioned in Table 2 and Figure 3.
Discussion

Our data sheet consisted of total 345 samples collected and tested for RT-PCR found positive with the Ct values ranging from >15 to <35 during July 2021 to February 2022. Out of the 345 samples 200 (57.9%) were males and 145 (42%) females in the age group of 0-80 years. The data is from the month of June 2021 to February 2022. The distribution of cases as per month wise shows that in July 31 cases (16 males, 15 females), August 34 cases (20 males, 14 females), in September 59 cases (40 males, 19 females), in October 44 cases (27 males, 17 females), in November 38 cases (22 males, 16 females), in December 17 cases (12 males, 5 females), in January 21 cases (16 males, 5 females) and February we had 101 cases (53 males and 48 females). The total vaccinated cases were 141 out of 345.

In the current study, 14 lineages were observed by using Phylogenetic Assignment of of Named Global Outbreak Lineage (Pangolin 3.1.4, http://www.pangolin.cog- uk.io). Out of 345 samples, 213 showed the gene sequencing rest were failed. The overall lineage distribution highlighted the occurrence of B.1.617.1 (1.4%), B.1.167.2 (n=147), B.1.1.306 (1.4%), B.1.36.29 (8.45%), B.1.1 (1.87%), B.1.36.8 (1.87%), AY.4 (n =1.87%), AY.20 (2.81%), AY.39 (3.75%), AY.125 (2.34%), AY.122 (1.4%), AY.102 (1.4%), AY.120 (1.4%) as mentioned in Table 2 and Figure 3. Across the world, there are multiple variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes coronavirus disease 2019 (Covid-19). SARS-CoV-2 variants have been classified by the Centers for Disease Control and Prevention (CDC) as variants of interest, variants of concern, and variants of high consequence. Three new variants that have rapidly become dominant within their countries have aroused concerns: B.1.1.7 (also known as VOC-202012/01), 501Y.V2 (B.1.351), and P.1 (B.1.1.28.1).

The B.1.617 variant of SARS-CoV-2 carries two mutations, E484Q and L452R. Both are separately found in many other coronavirus variants, but they have been reported together for the first time in India. The two mutations are found in the virus’s spike protein. The spike protein helps the virus to bind itself to the human cell’s receptors and gain entry into a host cell. The E484Q mutation is similar to E484K, a mutation found in the United Kingdom (lineage B.1.1.7) and South Africa (B.1.351) variants of the coronavirus. The L452R mutation has been found in fast spreading variants in California (B.1.427 and B.1.429). It can increase the binding power of spike proteins with ACE2 receptors on human cells, making it more transmissible. L452R can also potentially enhance viral replication [21-23].

Conclusion

We found that the most common variant lineage was B.1.167.2 accounting for 69% percent followed by B1.1 in our District of Telangana State. B.1.167.2, with mutations L452R, T478K and P681R, is highly prevalent in the state of Gujrat in India. There was enrichment for the mutation T478K or T478R when SARS-CoV-2 was subjected to weak neutralizing antibodies, which indicates this mutation may lead to antibody escape. This variant needs strict monitoring and hence the prevention of transmission.
References


