Genome variation analysis and strategic clustering to sub-lineage of double mutant strain B.1.617 of SARS-CoV-2

Anjana Ghelani[1], Vishal Mevada[2], Rajesh Patel[3]

[1]Ramkrishna Institute of Computer Science and Applied Sciences,
Sarvajanik University, Surat 395007, India

[2]DNA Division, Directorate of Forensic Science,
Gandhinagar 382007, India

[3]Bioinformatics Laboratory, Super Computing Facility, Department of
Biosciences, Veer Narmad South Gujarat University, Surat 395007, India

ghelanianjana@gmail.com

SARS-CoV-2 is an RNA coronavirus responsible for Acute Respiratory Syndrome (COVID-19). In January 2021, the re-occurrence of COVID-19 infection was at its peak, considered the second wave of epidemics across the world. In the initial stage, it was considered a double mutant strain due to two significant mutations observed in their Spike protein (E484Q and L452R). Although it was first detected in India, later on it was spread to several countries of Asia, Europe and other continents, causing high fatality due to this evolved strain. In the present study, we investigated the spreading of B.1.617 strain worldwide through 822 genome sequences submitted in GISAID on 21 April 2021. Submitted sequence data were extracted and uploaded to AnCovid19 Database (http://covid19.vnsguhpc.co.in/) for analysis. AnCovid19 was developed during study in Drupal 7.78. All genome sequences were analyzed for variations in genome sequences based on their effects due to changes in nucleotides. At Allele frequency 0.05, there were a total of 47 variations in ORF1ab, 22 in Spike protein gene, 6 variations in N gene, 5 in ORF8 and M gene, four mutations in Orf7a, and one nucleotide substitution observed for ORF3a, ORF6 and ORF7b gene. The clustering for similar mutations mentioned B.1.617 sub-lineages.

The outcome of this study established relative occurrence and spread worldwide. The study's finding represented that "double mutant" strain is not only spread through traveling but it is also observed to evolve naturally with different mutations observed in B.1.617 lineage. The information extracted from the study helps to understand viral evolution and genome variations of B.1.617 lineage. The results support the need of separating B.1.617 into sub-lineages.

The results describe that B.1.617 was not spread through India to other countries but eventually observed as a sub-lineage of B.1.617.1, B.1.617.2 and B.1.617.3. The variations E154K, E484Q, L452R, P681R and Q1071H were observed in most samples with allele frequency beyond 0.85. These variations might be responsible for several cases during the wave of COVID-19 infections. The recent submissions to NCBI GenBank database and GISAID Epi-Flu Database will elucidate more variations belonging to B.1.617 and its sub-lineages. The resulting continuous tracking of such variations will generate a complete picture of epidemiology and transmission of SARS-CoV-2 during the second wave of COVID-19 worldwide.