

Diversity and Evolution of Coronaviruses: Whole Genome and Pan-genome Perspective

Vaishnavi Kolte, Shenu Hudson B., Azra Khan, Arohi Parlikar, Kishan Kalia, Shruti Sinha, Sucheta Patnaik,
Neeraj Sharma, Sai Gayatri Vemuri,
Gaurav Sharma*

Affiliation: Institute of Bioinformatics and Applied Biotechnology (IBAB), Bengaluru, Karnataka, India

E-mail: gaurav.amit30@gmail.com, gauravsharma@ibab.ac.in

Abstract:

The emergence of COReNaVirus Disease 2019 (COVID-19) has sent people from across the world into a state of high alert. Since December 2019, numerous SARS-CoV-2 genomes have been sequenced allowing researchers to address critical questions regarding its origin, transmission, epidemiology, vaccines design, and detection kits. In the first part of this study, we analyzed 167 SARS-CoV-2, 312 SARS-CoV, and 5 Pangolin CoV genomes to understand their genomic variability, evolution, and mutation hotspots. Whole-genome-based phylogeny of subgenus *Sarbecovirus* genomes reasserted the fact that SARS-CoV-2 strains evolved from common ancestors putatively residing in bat or pangolin hosts. We predicted a few country-specific patterns of relatedness and identified mutational hotspots with high, medium, and low probability based on genome alignment of 167 SARS-CoV-2 strains. 100-nucleotide segment-based homology studies revealed that the SARS-CoV-2 genome is composed of segments close to Bat CoV and Pangolin CoV; however, several segments are distinctive in nature, therefore, making this virus unique as compared to other close relatives. The pan-genome of genus *Betacoronavirus* was identified as open indicating the vast diversity contributed by the novel viruses emerging within this group. In the second part of this study, we identified 30 variants showing contrasting mutational probabilities in the span of four months amongst 864 Indian SAR-CoV-2 strains. Amongst these, 25 variants are showing increasing mutational probabilities suggesting their propagation with time due to their unexplored evolutionary advantages. Whereas, mutational probabilities of five variants have significantly decreased in May/June as compared to March/April, suggesting their termination with time. Overall, the exploration of the diversity of these isolates, mutational hotspots, and pan-genome shed light on the evolution and pathogenicity of SARS-CoV-2.