Comparative analyses of SAR-CoV2 genomes from different geographical locations and other coronavirus family genomes reveals unique features potentially consequential to host-virus interaction and pathogenesis

doi: https://doi.org/10.1101/2020.03.21.001586

Abstract

The ongoing pandemic of the coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV2). We have performed an integrated sequence-based analysis of SARS-CoV2 genomes from different geographical locations in order to identify its unique features absent in SARS-CoV and other related coronavirus family genomes, conferring unique infection, facilitation of transmission, virulence and immunogenic features to the virus. The phylogeny of the genomes yields some interesting results. Systematic gene level mutational analysis of the genomes has enabled us to identify several unique features of the SARS-CoV2 genome, which includes a unique mutation in the spike surface glycoprotein (A930V (24351C>T)) in the Indian SARS-CoV2, absent in other strains studied here. We have also predicted the impact of the mutations in the spike glycoprotein function and stability, using computational approach. To gain further insights into host responses to viral infection, we predict that antiviral host-miRNAs may be controlling the viral pathogenesis. Our analysis reveals nine host miRNAs which can potentially target SARS-CoV2 genes. Interestingly, the nine miRNAs do not have targets in SARS and MERS genomes. Also, hsa-miR-27b is the only unique miRNA which has a target gene in the Indian SARS-CoV2 genome. We also predicted immune epitopes in the genomes.