Distinctive Mutation Profiles of SARS-CoV-2 Spike Protein in Different Geographic Regions of the United States

Guo, Ellen (School: Mounds View High School)

Mutations on the SARS-CoV-2 spike protein have gained increasing attention. However, its mutation pattern in different geographic locations within a country has not been investigated. In this study, over 16,000 sample spike protein sequences from eight states in the US were analyzed by comparing the sample sequences with the reference using the NCBI blastp. The results showed a varying number of mutations and overall mutation rates among different states, with MN leading both. Only two mutations (L5F and D614G) were shared by all these states. The L5F mutation frequency ranged from 0.5% to 1.95% for seven states but reached 7.66% for MN. Most of states had more than a 90% frequency for D614G, except the Western states of CA (82.82%) and WA (66.38%). Some mutations were shared by two to four states, but CA and WA had no additional mutual mutations. Surprisingly, around 50% of the mutations were unique to seven states with A845D exclusive to MA reaching a 7.22% high frequency, while MI had no unique mutation. The E780Q mutation was observed only in WI and MN with a much higher frequency of 9.64% for WI, although structure modeling did not indicate its significant impact on spike protein conformation and stability. Furthermore, E780Q mutants from WI and MN differed with an additional mutation. Thus, the results revealed distinctive SARS-CoV-2 spike protein mutation landscapes in different and even the same US geographic regions, highlighting the urgent needs of nationwide genome surveillance for detecting emerging SARS-CoV-2 variants.

Awards Won: