Evolutionary Dynamics of SARS-CoV-2 Lineages B.1.1.7, B.1.351, and P.1

Testing the possibility of the mutator phenotypes

Aysevil Pektas, 201903938

Master's Thesis in Bioinformatics (30 ECTS)

June 2021

Supervisors:

Assoc. Prof. Thomas Bataillon and Assoc. Prof. Palle Villesen

Bioinformatics Research Centre (BiRC)

Aarhus University

Denmark

ABSTRACT

Severe acute respiratory syndrome coronavirus (SARS-CoV-2) has spread all around the world and led to pandemics. During the second wave of infections, the emergence of new lineages in SARS-CoV-2 was identified as B.1.1.7, B.1.351, and P.1 referred to variants of concern. The variants of concern replaced previous circulating lineages in their geographic regions, spread to other countries worldwide, and were taken under surveillance. Some of the research claimed an increased mutation rate and a skew in the mutation spectrum could be due to the mutator phenotypes that could occur by mutations in the RdRp and nsp14. This study hypothesized that the variants of concern might have possessed mutator phenotypes that lead to alteration in the evolutionary dynamics of these lineages, which might explain the increased transmissibility of variants of concern. To test the hypothesis, mutation accumulation rates and skews in the distributions in the mutation spectrum of the variants of concern were compared with the B.1 lineage, which was the ancestor lineage of the variants of concern. The mutation spectrum and the rate of the mutation accumulation were investigated in different aspects that were nonsynonymous and synonymous mutations, transitions and transversions, different mutation types, and the relationship between strong and weak mutations. Among all the mutation accumulation rate analyses, the only significant differences were detected in mutation types. For the B.1 lineage, there was a decrease in the rate of the A>G, C>T, and G>T, whereas, for variants of concern, significant increases were identified in not only the mutation types found in B.1 lineage but also the other mutation types. Two common mutation types were G>A, and T>C for variants of concern, but P.1 lineage had an increased mutation accumulation rate for almost all mutation types. It was obtained skews in the mutation spectrum of the mutation types for the lineages of interest. The most explicit one was for the P.1 lineage. Also, a higher amount of nonsynonymous to synonymous ratio was detected in P.1 lineage than B.1 lineage, where the opposite results were obtained for the B.1.1.7 and B.1.351 lineages. However, it was not obtained a signal for the lineage-specific direction of selection. In conclusion, even skews in the mutational spectrum of the variants of concern were detected, it was hard to get a clear outcome to explain the presence of the mutator phenotypes in the variants of concern since I did not observe a high overall mutation accumulation rate in the variants of concern rather than B.1 lineage, and I could not also detect a mutation that could the function of RdRp or nsp14.

TABLE OF CONTENTS