

https://dspace.org/

Molecular epidemiology of M and E protein coding genes from South African SARS-CoV-2 strains, 2020 to 2021

Marsden, Fabian

2024

https://hdl.handle.net/10539/37648

Downloaded from a DSpace repository. https://dspace.org/

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the current pathogen causing the COVID-19 pandemic across the world. While vaccines that elicit anti-SARS-CoV-2 antibodies have been developed and licenced, there is a reduced protection against variants of concern (VOCs) such as Beta, Delta and Omicron. This is due to the mutations within the spike (S) protein which is the antigen targeted by most vaccines. Other potential vaccine targets include the structural proteins namely the membrane (M) and envelope (E) proteins of SARS-CoV-2 which are more conserved. In this study we aimed to determine the extent of genetic diversity in the M and E protein genes from South African SARS-CoV-2 strains and its impact on predicted B and T cell epitopes. M and E gene sequences were extracted from South African SARS-CoV-2 genomes obtained from the Global Initiative on Sharing All Influenza Data (GISAID) database for the period 01March 2020 to 31 December 2021. Maximum-likelihood phylogenetic tree analysis shows that among South African E gene sequences only the Omicron VOC sequences form a distinct cluster. Similarly, the Omicron M gene sequences also form a distinct cluster compared to the Wuhan reference strain, Beta and Delta sequences. The predicted T cell and B cell epitopes of M and E proteins were identified with specific regions that have shown to have identical regions in both the variants and the reference strain, this shows the conserved nature of the M and E genes. SARS-CoV-2 are shown to have varying antigenic probabilities for M and E proteins from each of the variants considered as probable antigens. The allergenicity and toxicity of the M and E proteins was assessed in the context of potential vaccine development with certain peptides of each shown to have toxic properties. The predicted B and T cell epitopes show that despite the presence of mutations in the VOCs' derived protein sequences, there is a common epitopic region that is shared between the reference and the variants. There is a strong 9-mer coverage by the natural sequences despite some non-coverage due to non-silent mutations. The results from the epitope predication and HLA typing shows the conserved nature of the M and E proteins which highlights the potential use for the development of vaccines.