SARS-CoV-2 Variant Analysis Using Discrete Fourier Transform

Wang, Aidan (School: University School of Milwaukee)

The spread and transmission of SARS-CoV-2 and other viruses can lead to various mutations that cause significant spikes in infections. It is essential for researchers to study virus genome sequences to analyze these mutations. However, prior to conducting genome analysis, researchers must use sequence alignment tools. Current sequence alignment tools are limited in their ability to handle large amounts of data, requiring substantial time and memory resources. To address these challenges, we propose a novel method that utilizes Discrete Fourier Transform based on dinucleotides to define a distance score that quantifies the similarity and dissimilarity between genome sequences. This method can be used to categorize new variants of viruses and identify potential mutations. Our alignment-free method for genome sequence comparison demonstrates its effectiveness through successful phylogenetic analysis. This approach makes it possible to analyze large sets of sequence data within a reasonable time frame. The proposed method was also applied to analyze other viruses, including influenza A virus, as well as bacteria which have much larger genome sequences composed of millions of nucleotide bases. While traditional sequence alignment tools are not feasible for such long genomes, this method successfully categorized 30 bacterial sequences in just 25 minutes. With the continuous advancements in sequencing technology, there is an ever-increasing amount of genome sequence data becoming available. Our alignment-free method provides significant benefits for large-scale genome analysis to enable researchers to analyze variations between sequences, gain insights into the efficacy of vaccines, and potentially create models to predict sequence variation in the future.

Awards Won:
Second Award of $2,000
Long Island University: Presidential Scholarships