Utilizing Homologous Simplicial Complexes to Model Genomic Data

Wang, Jessica

Liu, Ying

Our project involves analyzing sets of genomic sequences and the underlying similarities and interconnected relations between them. Our project makes use of homology groups and the Smith-Waterman similarity algorithm in order to represent strings of genomic data as points within an topological field. We can then use persistence homology groups to analyze evolutionary trends between genomic sequences. Using a new and innovative method of combining the Smith-Waterman algorithm and persistence homology, we were able to create a high specialized program for comparing genomic sequences. Our program has been tested and verified by analysis of various serotypes of the influenza A virus. Our program has produced significant results, providing insight on the evolutionary patterns of various horse breeds and chimpanzees. In our influenza A analysis, we were able to identify three different strains and the common mutations between each, allowing us to produce more efficient immunizations and prevention methods. Our discoveries and research have broad applications in the field of bioinformatics and can assist in identifying single-nucleotide polymorphisms that cause genetic diseases (i.e. Alzheimers, Sickle Cell Anemia) as well as aid in selective breeding within agriculture or aquaculture to produce more efficient and stable food sources.

Awards Won: