# Analyzing Hemagglutinin Protein Sequences of H3N2 Influenza Viruses to Predict Antigenic Variants

Trepte, Matthew

Influenza viruses are responsible for tens of thousands of deaths each year, and vaccines are used as a primary tool for reducing their impact and severity. The rapid mutation of the virus's Hemagglutinin antigen, however, poses a challenge to developing effective vaccines. Each year, Hemagglutinin antigen protein sequences of circulating influenza virus strains are compared to vaccine strains to determine whether the strains are immunologically distinct and, therefore, if the circulating strain should be included in the updated vaccine. Conventional computational methods focus on the mutations of a few specific, 'significant' amino acid positions of the Hemagglutinin protein sequences when comparing strains. Here, we consider a novel, more sensitive computational algorithm (Trepte Pivot Decision Tree or TPDT algorithm, written in JAVA) that examines many more positional significance variables for comparing strains and identifying those that are antigenic. The algorithm is calibrated using a Training dataset and then tested on the Training dataset and two independent datasets. Testing shows that the TPDT algorithm is more precise and outperforms previously published algorithms, especially when predicting similar viruses. Future efforts will seek to expand the size of the dataset of virus strain sequences and antigenic distances in order to create a more statistically robust algorithm.