Project 3

Prediction of Pathogenicity based on Non-structural protein 1

Background
Highly pathogenic avian influenza viruses (HPAIV) pose a threat for yet another epidemic or pandemic, which can potentially result in severe consequences for both animal and human life. To infect, the surface glycoprotein hemagglutinin (HA) precursor, HA0, needs to be cleaved by cellular proteases into functional HA1 and HA2 subunits. Lowly pathogenic avian influenza viruses (LPAIV) carry a monobasic cleavage site that is recognized only by trypsin-like proteases thus limiting the infection to the respiratory and gastrointestinal tracts. High pathogenicity has been previously linked to insertions in the cleavage site of HA. These insertions allow the HA0 to be cleaved by ubiquitously expressed intracellular proteases such as furin leading to a systematic infection and lethal disease with mortality rates being as high as 100%. Recent studies have shown that the insertions in the cleavage site of HA may not be sufficient to render the virus highly pathogenic (HP). Evidence is accumulating that the HPAIV also carry virulence determinants other than the elongated cleavage sites of the HA protein i.e. on other expressed proteins.

Aim of the study
The goal of this study is to find these virulence markers in the Nonstructural protein 1 (NS1) proteins.

Description of the dataset
The data is a decision table where columns are coming from a multiple alignment of protein sequences and the rows are the viral sequences. These sequences are labeled as 1 for representing high pathogenicity or 0 for low pathogenicity.

Ref: Khaliq, Zeeshan, et al. "A complete map of potential pathogenicity markers of avian influenza virus subtype H5 predicted from 11 expressed proteins." BMC microbiology 15.1 (2015): 128.
https://bmcmicrobiol.biomedcentral.com/track/pdf/10.1186/s12866-015-0465-x