**Table S4:** Summary of the number of categories and features used for the XGBoost analysis for each protein. Total samples per class indicate the number of samples within each branch class (hu-hu: human-to-human; sw-sw: swine-to-swine; hu-sw: human-to-swine; sw-hu: swine-to-human) before subsampling into training, validation and test datasets, Total columns before filtering indicate the total number of columns after one-hot-encoding of each protein position and amino acid combination, and Final number of features indicate the number of features used in the training and validation, after filtering out uninformative columns.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Segment (Protein) | Total samples per class | | | | Total columns before filtering | Final number of features |
| **hu-hu** | **sw-sw** | **hu-sw** | **sw-hu** |
| PB2 | 400 | 1230 | 74 | 38 | 15960 | 2331 |
| PB1 | 379 | 983 | 92 | 32 | 15918 | 2012 |
| PB1 (PB1-F2) | 379 | 983 | 92 | 32 | 2121 | 405 |
| PA | 443 | 1132 | 88 | 27 | 15057 | 2272 |
| PA (PA-X) | 443 | 1132 | 88 | 27 | 5481 | 934 |
| HA (H1) | 387 | 1873 | 116 | 39 | 11886 | 3285 |
| HA (H3) | 188 | 555 | 35 | 14 | 11907 | 2189 |
| NP | 362 | 862 | 66 | 19 | 10479 | 1368 |
| NA (N1) | 238 | 1044 | 60 | 37 | 9870 | 2181 |
| NA (N2) | 295 | 1281 | 57 | 22 | 9870 | 2128 |
| MP (M1) | 275 | 523 | 39 | 18 | 5292 | 535 |
| MP (M2) | 275 | 523 | 39 | 18 | 2058 | 368 |
| NS (NS1) | 321 | 944 | 59 | 35 | 4977 | 1144 |
| NS (NEP) | 321 | 944 | 59 | 35 | 2562 | 456 |