



Article

Predicting Zoonotic Risk of Influenza A Viruses from Host Tropism Protein Signature Using Random Forest

Christine L. P. Eng^{1,*}, Joo Chuan Tong² and Tin Wee Tan³

¹ Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 117597 Singapore, Singapore

² Institute of High Performance Computing, A*Star, 138632 Singapore, Singapore; tongjc@ihpc.a-star.edu.sg

³ National Supercomputing Centre, 138632 Singapore, Singapore; tinwee@nsc.sg

* Correspondence: christine@bic.nus.edu.sg; Tel.: +65-6516-3357

Academic Editor: Susanna K. P. Lau

Received: 14 March 2017; Accepted: 19 May 2017; Published: 25 May 2017

Abstract: Influenza A viruses remain a significant health problem, especially when a novel subtype emerges from the avian population to cause severe outbreaks in humans. Zoonotic viruses arise from the animal population as a result of mutations and reassortments, giving rise to novel strains with the capability to evade the host species barrier and cause human infections. Despite progress in understanding interspecies transmission of influenza viruses, we are no closer to predicting zoonotic strains that can lead to an outbreak. We have previously discovered distinct host tropism protein signatures of avian, human and zoonotic influenza strains obtained from host tropism predictions on individual protein sequences. Here, we apply machine learning approaches on the signatures to build a computational model capable of predicting zoonotic strains. The zoonotic strain prediction model can classify avian, human or zoonotic strains with high accuracy, as well as providing an estimated zoonotic risk. This would therefore allow us to quickly determine if an influenza virus strain has the potential to be zoonotic using only protein sequences. The swift identification of potential zoonotic strains in the animal population using the zoonotic strain prediction model could provide us with an early indication of an imminent influenza outbreak.

Keywords: influenza; zoonosis; machine learning

1. Introduction

Influenza A viruses primarily reside in avian species, yet in recent years, there have been an increasing number of documented zoonotic infections in humans. After the first highly pathogenic H5N1 outbreak in 1997 in Hong Kong, there were subsequently many more local epidemic outbreaks from H5N1 viruses, especially in Asia and Africa [1–3]. There have also been a smaller number of human infections involving other avian influenza subtypes including H7N7 in United Kingdom and Netherlands [4–6], H9N2 in China [7,8], as well as the recent H7N9 outbreak in China [9,10]. Most of these zoonotic infections emerged in a similar manner, with patients having contracted the virus upon direct contact with poultry or other avian species [2,11,12]. While there was no direct evidence of human transmissibility or stable adaptation in humans, many of these zoonotic infections particularly, of H5N1 and H7N9 subtypes, cause severe illnesses, with the mortality rate for H5N1 estimated to be as high as 60% [13]. These zoonotic strains originated from avian species, having acquired sufficient mutations or new segments from reassortment to overcome host range restriction and successfully cause infections in humans.

Despite many years of intensive research, current surveillance technologies for influenza viruses remain limited as there are still no reliable measures in predicting zoonotic strains that can cause the next zoonotic outbreak or pandemic. Current surveillance efforts focus on detection, assessment and

response following an outbreak [14,15]. Antigenic and genetic characterization of the new strains by phylogenetic analyses with existing strains are performed to understand how the outbreak started as well as to formulate effective response and treatment [16,17]. There have been increasing efforts in surveillance recently, with disease surveillance in wild birds and poultry farms where influenza sequence data are collected and deposited online [18,19]. Yet, the computational methods to identify possible zoonotic strains remain rudimentary, with the reliance on host-associated genetic markers [20]. A number of avian- or human-specific residues at certain amino acid positions have been identified to differentiate between avian and human strains [21,22], most notably the polymerase basic protein 2 (PB2) E627K host range determinant which shows a strong selection for the amino acid lysine (K) in human strains and some zoonotic strains as opposed to glutamate (E) carried by avian strains [23,24]. More recent bioinformatics approaches have identified diversity motifs or combinations of interacting amino acid residues to distinguish between avian or human strains [25,26]. However, these approaches are context-specific and generally do not apply to novel influenza subtypes [20,27], because mutations identified as critical in a particular zoonotic event may or may not be detected in other events. The World Health Organization (WHO) and the United States Center for Disease, Control and Prevention (CDC) have in recent years introduced influenza risk assessment tools to evaluate potential pandemic risks of influenza A viruses circulating in animal species [28,29]. Both tools consist of several evaluation criteria in three categories of viral properties, population attributes, and ecology and epidemiology to characterize the risk of a virus. While the tools are comprehensive, several evaluation criteria such as antiviral treatment resistance, receptor binding properties, and lab animal transmission require time and extensive testing in the laboratory. As such, it is still a challenge to predict potential zoonotic strains based on sequence information alone.

There are also attempts in the development of machine learning approaches to predict zoonotic transmission. Qiang and Kou first developed a computational prediction model based on an artificial neural network (ANN) to predict interspecies transmission of influenza A viruses based on molecular patterns found in protein sequences [30]. The model utilized a wavelet packet decomposition method to extract energy feature vectors from protein sequences in the training process, distinguishing avian species with the capability to cross host species barrier from those that do not possess the zoonotic capability. Another paper by Wang et al. also described a prediction model developed from a support vector machine (SVM) to classify avian and human influenza A sequences [31]. The model employed position-specific entropy profiles of avian and human protein sequences [21], which were then transformed into feature vectors encoded with amino acid physicochemical properties. Both prediction models apply protein sequences from six influenza inner proteins: three viral polymerases polymerase acidic protein (PA), polymerase basic protein 1 (PB1), and PB2, nucleoprotein (NP), non-structural protein 1 (NS1), as well as matrix protein 1 (M1). While both models reported high accuracy in prediction, the accuracies in predicting past zoonotic strains from influenza outbreaks have not been verified.

To achieve this goal, we have constructed a zoonotic strain prediction model using the machine learning classifier random forest, capable of predicting avian, human or zoonotic influenza virus strains in this study. Our previous work on host tropism of individual influenza virus proteins has resulted in the construction of a host tropism prediction system [32]. The system consists of individual protein prediction models of 11 influenza A virus proteins: hemagglutinin (HA), M1, matrix protein 2 (M2), neuraminidase (NA), NP, NS1, non-structural protein 2 (NS2), PA, PB1, accessory protein F2 translated from PB1 segment (PB1-F2) and PB2, which independently predicts avian or human host tropism of each protein based on protein sequences translated into amino acid physicochemical properties feature vectors. We next combined the protein prediction results into a host tropism protein signature for each influenza virus strain, which is defined as an influenza viral proteome profile of 11 independent host tropism predictions of avian or human influenza virus proteins. The host tropism protein signature analysis of 12,624 strains has led to the discovery of distinct host tropism protein signatures between avian, human and zoonotic strains [33]. Furthering this finding, we utilized the host tropism protein

signatures to build a computational prediction model which is able to predict zoonotic strains capable of causing human infections. Instead of the conventional avian versus human strains approach generally adopted [21,22,25,26], we have defined zoonotic strains as a separate category distinct from typical avian and human strains, resulting in a three-class classification of avian, human and zoonotic strains. We then additionally validated the capability of the zoonotic strain prediction model whereby avian strains shown to be possible sources of zoonotic outbreaks by previous studies were accurately identified by the prediction model. This represents a significant validation to the capability of the zoonotic strain prediction model in using protein sequences to detect zoonotic strains that can lead to an influenza outbreak.

2. Results

2.1. Sufficient Distinction in Host Tropism Protein Signatures to Characterize Zoonotic Strains

Host tropism protein signatures obtained for the influenza virus strains in the dataset demonstrate the distinct signatures between avian, human and zoonotic strains. This is consistent with earlier findings where typical avian and human strains show almost unanimous host tropism predictions of avian or human proteins respectively, while suspected and confirmed zoonotic strains typically display a mixture of avian and human protein predictions [33]. As compared to the previous study however, the signatures generated in this study are of a higher resolution, owing to the avian and human probability distribution being used instead of binary predictions of either avian or human. Each host tropism protein prediction is associated with a probability estimate which represents the confidence of the prediction by each individual protein prediction model, loosely describing how “avian-like” or “human-like” the proteins are, as illustrated by the intensity of the color (Figure 1). This allows us to inspect with greater detail the host tropism protein signature of an influenza virus strain, which could provide a clue as to how much it has deviated from a typical strain.

The host tropism protein signatures indeed provided sufficient distinction for the classification of avian, human and zoonotic influenza virus strains. Based on the training samples in the dataset, the random forest zoonotic strain prediction model achieved very high prediction performance, 99.20% prediction accuracy and 1.000 weighted area under the receiver operating characteristic curve (AUC; Table 1). This represents the correct avian, human or zoonotic strain classification by the prediction model for 374 out of 377 strains in the training dataset. As identification of zoonotic strains are of greater emphasis in this study, the prediction accuracy for zoonotic strains, while slightly lower at a 98.40%, is still of satisfactory performance (Table 1). This could be attributed to the zoonotic strains having a much more diverse range of avian and human protein predictions in their signatures as compared to typical avian and human strains, hence amounting to the increase in difficulty to predict these strains. Nevertheless, the prediction performance by the random forest zoonotic strain prediction model is still significantly better than random three-class classification, highlighting that the host tropism protein signatures of zoonotic strains are sufficiently distinct from typical avian and human strains. This therefore enables the prediction model to identify zoonotic strains with a high standard of accuracy.

Table 1. Performance evaluation of random forest zoonotic strain prediction model. AUC: area under the receiver operating characteristic curve.

Stage	Overall Accuracy	Weighted AUC	Zoonotic Accuracy	Zoonotic Sensitivity	Zoonotic Specificity	Zoonotic AUC
Training	99.20	1.000	98.40	0.984	0.996	1.000
Testing	99.06	1.000	97.14	0.971	1.000	0.999

Independent validation of the prediction model with a separate testing dataset further affirms the high predictive performance of the model. The prediction model achieved a 99.06% prediction accuracy

with 1.000 weighted AUC even when tasked to predict strains which were not included in the training process (Table 1). All but one of the zoonotic strains in the testing dataset including those isolated from H5N1 outbreaks in Asia and H7N9 outbreaks in China were correctly identified by the prediction model (Figure 2), resulting in a zoonotic prediction accuracy of 97.14% (Table 1). Results from this demonstrate that the prediction model was able to predict novel avian, human and zoonotic strains with high accuracy from the host tropism protein signature, even when presented with a diverse range of signatures. Most of the zoonotic strains were predicted with high zoonotic probabilities exceeding 0.8, with the remaining predicted with low to moderate zoonotic probabilities of 0.517 to 0.682, as well as one incorrect avian prediction with 0.315 zoonotic probability. What came as a surprise were the zoonotic strains which carried signatures of all avian tropism yet predicted accurately as zoonotic strains by the prediction model. This seems to suggest that zoonotic strains need not acquire human proteins to cause human infections. Results from this independent validation of the prediction model thus substantiate the capability of the model to accurately identify zoonotic strains in the future.

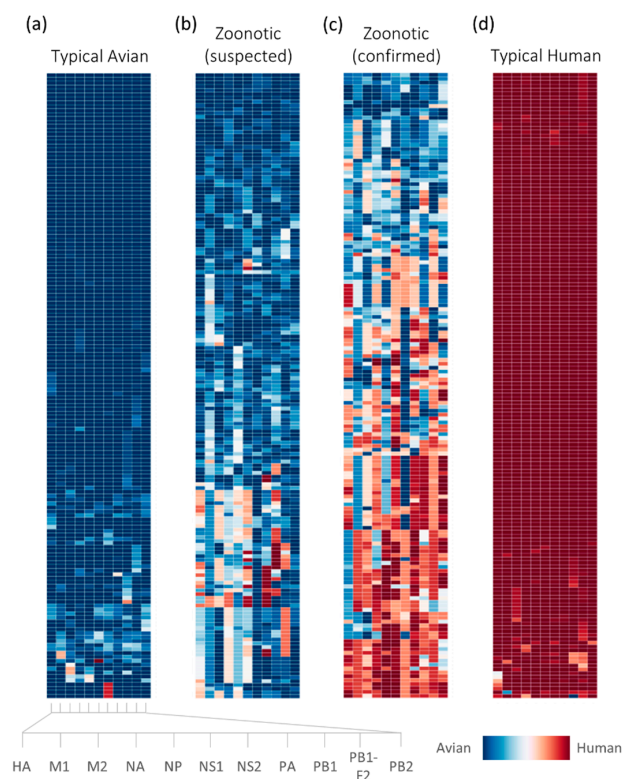


Figure 1. The dataset used in the study represented by host tropism protein signatures of influenza virus strains: 159 typical avian (a); 160 human-isolated confirmed zoonotic (c), and 164 typical human (d) strains were used in the construction of the zoonotic strain prediction model. Additionally, the signatures of a random 165 of 1047 avian-isolated suspected zoonotic strains (b) subsequently analyzed using the prediction model are also shown. Each individual virus strain in the bar is represented by the host tropism protein signature laid across the row, with the independent predictions of 11 proteins depicted in each column (HA, M1, M2, NA, NP, NS1, NS2, PA, PB1, PB1-F2, and PB2). Avian protein predictions are illustrated in blue, while human proteins are in red. The confidence of the avian or human host tropism prediction is expressed by the intensity of the color, based on the prediction probability estimates found in Supplementary S1 and S2 datasets. HA: hemagglutinin; M1: matrix protein 1; M2: matrix protein 2; NA: neuraminidase; NP: nucleoprotein; NS1: non-structural protein 1; NS2: non-structural protein 2; PA: polymerase acidic protein; PB1: polymerase basic protein 1; PB1-F2: accessory protein F2 translated from PB1 segment; PB2: polymerase basic protein 2.

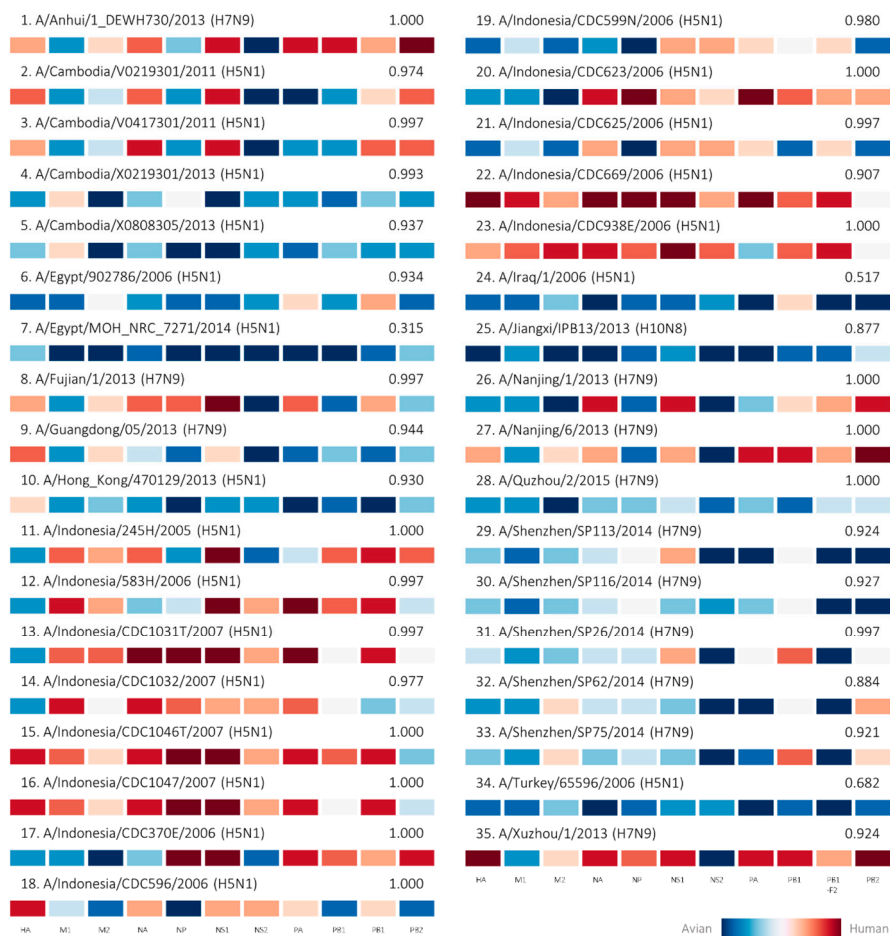


Figure 2. Human-isolated confirmed zoonotic strains prediction results by the prediction model during the independent validation process. A total of 35 zoonotic strains represented by their host tropism protein signatures were predicted by the zoonotic strain prediction model with the estimated zoonotic probability describing the confidence of the prediction. The strains are labelled with consecutive numbers for reference in the main text.

2.2. Retrospective Analysis of Avian Strains from Outbreaks Demonstrate Capability of Zoonotic Strain Prediction Model

By employing the zoonotic strain prediction model to perform an analysis of avian strains isolated from zoonotic outbreaks, we validated the capability of the prediction model to identify potential zoonotic strains circulating in avian species. Early studies on phylogenetic analyses from the H7N9 outbreak in China identified several avian-isolated strains sharing almost identical sequences with strains isolated from one of the first few human infections at the start of the outbreak [34,35], two of which in our dataset were successfully predicted as zoonotic by the prediction model with estimated zoonotic probabilities of 0.967 and 0.940 (Figure 3a). This corroborates earlier findings that the H7N9 outbreak originated from poultry and avian sources [34–36], as many strains isolated from avian species subsequently during the outbreak display classic zoonotic host tropism protein signatures with very high estimated zoonotic probabilities predicted by the prediction model. This would suggest that zoonotic H7N9 viruses had circulated among many avian species, covertly asymptomatic [35,36], to spread across China causing severe human infections in many states.

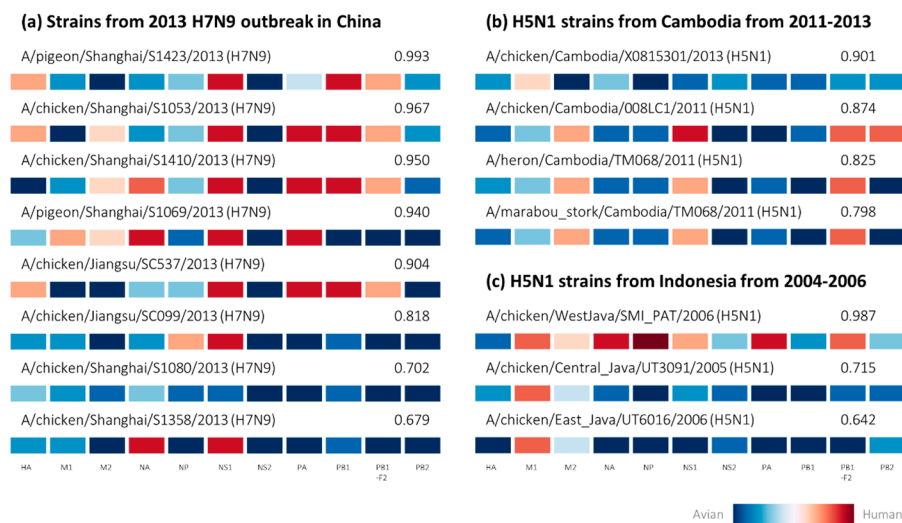


Figure 3. Avian-isolated strains from zoonotic influenza outbreaks represented by host tropism protein signatures with estimated zoonotic probability predicted by the zoonotic strain prediction model. The strains were cross-referenced with studies on phylogenetic analyses, demonstrating that some of these strains share high sequence similarities with zoonotic strains isolated from human patients during influenza outbreaks of (a) novel outbreak of H7N9 in China in 2013 and (b) H5N1 outbreaks in Cambodia from 2011 to 2013 as well as (c) Indonesia from 2004 to 2006.

We also cross-referenced an additional six avian-isolated H7N9 strains from a recent study investigating the import of H7N9 human infections into Taiwan [37], which were all predicted as zoonotic by the prediction model. Four of the six strains were predicted with high zoonotic probabilities exceeding 0.8, while the remaining two strains were predicted with moderate zoonotic probabilities of 0.679 to 0.702. Intriguingly, we again observe a strain having all avian proteins in its host tropism protein signature being predicted as zoonotic with 0.702 zoonotic probability (Figure 3a). While we cannot confirm if this strain did indeed cause human infections during the outbreak, it is possible that it may evolve to be more zoonotic based on previous observation of several confirmed zoonotic strains also carrying all avian signatures (Figure 1).

An additional four avian-isolated strains from Cambodia predicted as zoonotic by the prediction model (Figure 3b) were also observed from phylogenetic analyses of another study to share the same clades as human-isolated strains from H5N1 outbreaks in Cambodia from 2011 to 2013 [38]. Surprisingly, closer observation shows that the host tropism protein signature of the strain with the highest zoonotic risk of 0.901 actually contains the least number of human proteins among all four Cambodian strains (Figure 3b), with only the M1 protein having slight human tropism. Indeed, a cross examination showed that the zoonotic strains isolated from human patients during the H5N1 outbreak in Cambodia carried similar host tropism protein signatures (Appendix A Figure A1). This demonstrates that zoonotic strains from the same outbreaks carried similar host tropism protein signatures.

Results from our analysis also suggest that not all influenza viruses isolated from avian species during the outbreaks are zoonotic strains capable of causing human infections. Of the three H5N1 strains isolated from chicken in Indonesia, one was predicted with very high zoonotic probability of 0.987, with the remaining two strains predicted with lower zoonotic probability estimates of 0.642 and 0.715 (Figure 3c). Phylogenetic analysis of the strains by another study demonstrated the close evolutionary relationships of the HA and NA glycoproteins to confirmed zoonotic strain isolated from human patients in 2005 [39,40]. Nevertheless, this indicates that not all avian species or poultry sources were infected with zoonotic strains, as some avian strains of the same subtype circulating in the same region might in fact not have the capability to cause human infections. Taken together, this might

present an exciting prospect in the future where we can monitor avian influenza strains to determine their zoonotic risks in causing human infections.

3. Discussion

This study describes the successful use of machine learning on influenza sequence data to predict avian-to-human transmission of influenza viruses. Using host tropism protein signatures of influenza viruses which are predicted from protein sequences, the zoonotic strain prediction model can accurately distinguish between typical avian strains found in avian species, seasonal influenza circulating in humans, and zoonotic strains originating from avian species that have caused human infections. Almost all known zoonotic strains from past influenza outbreaks with complete proteome were accurately predicted as zoonotic, regardless of their HA and NA subtypes, which also includes the less common subtype of H10N8 in addition to H5N1 and H7N9 subtypes (Figure 2). As compared to the context-specific application of host-associated genetic markers, the zoonotic strain prediction model is not restricted by this limitation and can be applied for prediction across all influenza subtypes.

The design of this study employs a systems approach which includes two layers of machine learning on influenza protein sequences to predict zoonotic strains. This is a departure from most other studies looking into avian- and human- specific amino acid residues which are primarily based on sequence alignments of common influenza subtypes such as H1N1, H3N2 and H5N1 [21,22], as well as machine learning approaches to predict avian or human sequences using the host-specific residues [25,31]. Here, we defined zoonotic strains as a third, separate category in addition to avian and human strains. These zoonotic strains are recognized as an intermediate between avian and human strains, where they may have started evolving to overcome the host species barrier, but have not adequately adapted to humans yet. These changes can be reflected in the mixture of avian and human proteins in their host tropism protein signatures (Figure 1). By using the host tropism protein signatures which are in turn host tropism predictions based on global amino acid physicochemical properties descriptors, the zoonotic prediction model can recognize potential zoonotic strains, regardless of subtypes. This is in contrast with conventional approaches investigating amino acid positions which show strong selection for either avian or human strains. While the host-associated genomic markers are useful in providing clues behind the mechanism of avian-to-human transmission, the zoonotic prediction model constructed in this study aims to complement existing tools by providing a rapid prediction of zoonotic strains using a machine learning approach. This would allow a swift detection of possible zoonotic strains circulating in avian species, which can then be further analyzed for their host-associated genomic markers.

Analysis of the zoonotic strains using the zoonotic strain prediction model illustrates that zoonotic events are truly a complex process. Similar to results from our previous study [33], there are no universal host tropism protein signatures for zoonotic strains with strains from the same influenza outbreaks sharing similar signatures (Appendix A Figure A1). Additionally, we have encountered some puzzling observations from the analysis of the host tropism protein signatures, where some zoonotic strains were observed to be carrying all avian tropism in their signatures. As these confirmed zoonotic strains isolated from human patients carrying predominantly avian signatures can also be found in the training dataset (Appendix A Figure A1), this resulted in high zoonotic probability predictions by the zoonotic strain prediction model in the testing phase (Figure 2) as well as the subsequent analysis for similar strains (Figure 3). This is due to the supervised training process by the machine learning classifier where the prediction model learns from training examples provided in the training dataset. Nevertheless, it can be observed crudely that the avian proteins for these zoonotic strains seem to carry less avian tropism compared to typical avian strains (Figure 1). Despite that, our goal in this study is to use protein sequences to detect zoonotic strains that pose a risk to cause human infections circulating in avian species, and we have successfully shown with these findings the potential of the zoonotic strain prediction model in using the underlying host tropism protein signatures to predict zoonotic strains with a high degree of accuracy. We aim to provide a tool capable of predicting zoonotic strains

circulating in avian species rapidly using only protein sequences. Understanding how zoonotic strains are generated to cause outbreaks however, is a subject of further intensive investigation, requiring much more data on zoonotic strains with the aid of in-depth phylogenetic analysis.

While we are no closer to dissecting exactly which proteins are required for zoonotic influenza strains to make that zoonotic leap, we are slowly beginning to understand that we need to approach the problem from a systems perspective by looking at contributions of all influenza virus proteins. In using the host tropism protein signature, it is now possible to predict zoonotic strains as well as estimate the zoonotic risk. Nonetheless, the zoonotic prediction model is still in its infancy stages with limited data on confirmed zoonotic strains owing to the scarcity of complete genomic sequences from earlier zoonotic outbreaks. This is evident from the training dataset containing only zoonotic strains isolated since the beginning of the 21st century (Supplementary S1 dataset), as zoonotic strains prior to that do not have the complete the host tropism protein signature required for prediction. The strength of the zoonotic strain predictor lies in taking into consideration the contribution of all influenza virus proteins and by using them to distinguish between avian, human and zoonotic strains.

The zoonotic strain prediction model has been validated with confirmed zoonotic strains from past influenza outbreaks, where most zoonotic strains were predicted correctly as zoonotic with high zoonotic probability estimates. Due to the probabilistic nature of the random forest classifier with the classification output determined by majority voting of the random trees in the forest [41], it is possible to manually define a threshold for the probability estimates in each classification. Based on the predicted zoonotic probability estimates from the confirmed zoonotic strains in the independent testing dataset as well as the analysis from the avian-isolated suspected zoonotic strains, we propose a zoonotic risk table to aid in the interpretation of each strain prediction by the zoonotic strain prediction model (Table 2). This is again in preliminary stages based on the data in this study, with the sensitivity of zoonotic strain detection at 0.988 by defining the threshold for zoonotic probability estimate at 0.7. With the increase in influenza surveillance and sequencing of complete genome in the future, the zoonotic strain prediction model can only improve with the collection of more data for continuous training of the prediction model. This would also help us further understand which proteins are required for interspecies transmission. In the meantime, the zoonotic strain prediction model could prove to be a valuable addition to influenza virologic surveillance to complement traditional analytical methods through the monitoring of influenza strains in avian species and poultry, and by providing swift prediction on the zoonotic risks of influenza virus strains using sequence data.

Table 2. Proposed zoonotic risk interpretation table based on zoonotic probability estimates predicted by the zoonotic strain prediction model.

Zoonotic Probability Estimate	Zoonotic Risk
≥ 0.7	High
< 0.7	Low to none

4. Materials and Methods

4.1. Data Collection and Preparation

Influenza A virus protein sequence data was acquired from Influenza Research Database (<http://www.fludb.org> (accessed on 27 October 2015)) [42]. The data was next processed to retain only influenza A virus strains with complete proteome, comprising complete full-length sequences of 11 proteins (HA, M1, M2, NA, NP, NS1, NS2, PA, PB1, PB1-F2, PB2). This included the removal of invalid protein sequences with non-standard amino acids or of incomplete lengths, as well as the removal of strains with multiple contradictory sequences of the same protein. The complete dataset consisted of 13,998 strains with 7592 avian strains and 6406 human strains.

We next identified zoonotic strains which were distinct from typical avian and human strains. Based on published literature on avian or zoonotic influenza outbreaks, WHO reports and CDC reports,

a total of 160 confirmed zoonotic strains were identified from strains isolated from human cases during influenza outbreaks from 1997 to 2015 (Appendix A Table A1) [3–10,38,43–68]. An additional 1047 avian-isolated strains collected during the same period as the outbreaks around the geographic region were also identified and designated as avian-isolated suspected zoonotic strains. The dataset used in this study is thus categorized into three groups of avian, human and zoonotic strains, with the avian-isolated suspected zoonotic strains excluded from the following prediction model construction process for subsequent analysis.

4.2. Host Tropism Protein Signature Feature Transformation

Host tropism protein signatures for all influenza strains were next obtained using the host tropism protein prediction system (<http://fluleap.bic.nus.edu.sg> (accessed on 7 December 2015)) [32]. The system provides independent avian or human host tropism predictions of 11 influenza virus proteins. For each individual protein, the host tropism prediction model predicts avian or human host tropism based on protein sequence input. The protein sequences were represented by 146 feature vectors comprising 20 standard amino acid compositions and global descriptors of six amino acid physicochemical properties of hydrophobicity, normalized van der Waals volume, polarity, polarizability, charge and solvent accessibility [32]. Based on these, the avian or human host tropism prediction results for 11 proteins are integrated as a host tropism protein signature for each strain. Each prediction by the respective protein prediction model in the system is predicted with avian and human probability distribution to indicate the confidence of host tropism prediction based on the protein sequence. In summary, each influenza virus strain is thus represented by 22 avian and human probability distributions of 11 host tropism predictions of each protein (Figure 1). The 22 avian and human probability distributions therefore compose the training dataset for the subsequent machine learning process.

4.3. Construction of Zoonotic Strain Prediction Model

The influenza virus strains represented by the host tropism protein signatures were next used for machine learning to build a zoonotic strain prediction model in the classification of three groups of avian, human and zoonotic strains. The zoonotic strains in this training process consist only of the confirmed zoonotic strains isolated from human patients during influenza outbreaks. As the number of avian and human strains were disproportionately greater than the number of confirmed zoonotic strains, the method of down-sampling was introduced to prevent an imbalanced dataset. An imbalanced dataset may result in bias in the training process which may affect the performance evaluation. In the down-sampling process, avian and human strains were randomly removed to result in approximate equal number of strains in the three groups of avian, human and zoonotic strains. Following that, the final dataset was partitioned into separate training (80%) and testing (20%) datasets (Table 3).

Table 3. Final dataset of avian, human and zoonotic strains for construction of the zoonotic strain prediction model.

Category	Training Dataset	Testing Dataset	Total Samples in Group
Avian	125	34	159
Human	127	37	164
Zoonotic	125	35	160
Total samples in dataset	377	106	483

The machine learning algorithm employed in the construction of the prediction model is random forest. Random forest is an ensemble of decision trees, where the random trees are grown using the bagging technique, in which a randomly selected subset of features from the entire feature space is selected to split each leaf node in the tree [41]. Random forest has been shown to consistently

achieve high performance and is also the most suited for this as the dataset was obtained from host tropism predictions made on random forest protein prediction models as well [32]. This was performed on the WEKA machine learning platform [69], the Waikato Environment for Knowledge Analysis software containing a suite of machine learning algorithms for data mining and classification tasks. Ten-fold cross-validation training was applied to minimize the effect of overfitting. In this process, the training dataset is randomly partitioned into nine training subsets and one testing subset over ten iterations. The algorithm will train with nine training subsets and evaluate the prediction model with the remaining testing subset for every iteration, with each subset used exactly once as testing. Results for the performance evaluation are taken as an average of ten iterations, and the model with the best results is chosen.

In addition, a parameter optimization process was also performed in the training process. The optimized random forest parameters were the number of trees in the random forest and the number of features to use in random selection. These were fine-tuned using the grid search approach where each parameter in a manually defined subset of a maximum of 500 trees and 22 features, is exhaustively applied to select for the parameters producing the best results. This approach ensures that the best parameters were chosen to maximize the performance in constructing the prediction model. The final random forest prediction model was constructed with 302 trees in the random forest, with 1 random feature at each branch split.

The prediction model was next assessed with several performance measures. This includes overall prediction accuracy and AUC. The prediction accuracy measures the number of predictions correctly made from the total number of strains in the training dataset. AUC, on the other hand, describes the probability of a randomly chosen positive sample ranking higher than a randomly chosen negative sample by the model [70,71]. As this study involves a three-class classification problem with an approximately balanced dataset, the models were evaluated primarily with overall prediction accuracy and weighted AUC of the three groups of avian, human and zoonotic prediction. Furthermore, the prediction accuracy and AUC for zoonotic strains were also taken into account as the primary concern of this study is in the prediction of zoonotic strains. This was implemented through the generalization of the three-class classification into a binary classification of zoonotic versus non-zoonotic comprising both avian and human strains.

The completed zoonotic strain prediction model was finally independently validated with the testing dataset, consisting of strains which were excluded from the initial training process. Performance of the model in predicting strains from the separate testing dataset could help establish whether overfitting has occurred in the training process. This would hence determine if the model is robust for accurate prediction of novel strains in the future.

The zoonotic strain prediction model classifies a strain as avian, human, or zoonotic from the feature vectors represented by the host tropism protein signatures. The random forest algorithm is, by nature, a probabilistic classifier where the outputs are continuous decision values determined based on voting by the random trees in the random forest [41]. Therefore, each strain prediction by the random forest prediction model has an avian, human and zoonotic probability estimate as calculated from the number of votes by the random trees out of the total number of trees in the forest. This represents the confidence of the prediction by the random forest prediction model. The final predicted avian, human or zoonotic classification of a strain would thus be the class with the highest probability estimate.

4.4. Analysis of Avian-Isolated Suspected Zoonotic Strains

The zoonotic strain prediction model was then tasked to analyze the zoonotic risks of avian-isolated suspected zoonotic strains. This group of strains were excluded initially from both the training and testing process as not all the avian strains isolated from influenza outbreaks contributed to the onset of the outbreak [40]. Thus, the zoonotic capability of these strains cannot be established with certainty. The strains, represented by their host tropism protein signatures, were provided to the

zoonotic strain prediction model for prediction. The resulting avian, human or zoonotic classifications, along with the estimated probability distributions, were analyzed in conjunction with the host tropism protein signatures.

5. Conclusions

Our study demonstrated the successful use machine learning trained on host tropism protein signatures to predict zoonotic strains having the capability to cause human infections. The zoonotic strain prediction model is proposed as an influenza virologic surveillance tool to detect changes in protein sequences in avian strains that may indicate a zoonotic jump event. As influenza sequence data are already regularly sampled and collected [18,19], this tool could complement existing methods to rapidly screen for possible zoonotic strains. Future work to integrate geographical and ecological data [72] would bring more significant advancements in predicting future influenza outbreaks beyond current sequence prediction capabilities. The detection of possible zoonotic strains in avian species in the future could grant us precious time in formulating appropriate responses before they can reach the human population to start devastating outbreaks. This would ultimately not only benefit public health, but also reduce the economic impact to the agriculture industry in the event of an influenza outbreak. The zoonotic strain prediction model is available for prediction online at <http://fluleap.bic.nus.edu.sg> (accessed on 20 May 2017).

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/1422-0067/18/6/1135/s1. Dataset S1. Training dataset for the zoonotic strain prediction model. Dataset S2. Independent testing dataset for the zoonotic strain prediction model. Dataset S3. Dataset for analysis of avian-isolated suspected zoonotic strains using the zoonotic strain prediction model.

Acknowledgments: The work was supported by the National University of Singapore Research Scholarship awarded to Christine L. P. Eng from the National University of Singapore.

Author Contributions: Christine L. P. Eng designed the project, performed the analysis, and wrote the paper. Joo Chuan Tong and Tin Wee Tan provided supervision, contributed to the analysis, and the writing of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ANN	Artificial neural network
AUC	Area under the receiver operating characteristic curve
CDC	Center for Control, Disease and Prevention
HA	Hemagglutinin
M1	Matrix protein 1
M2	Matrix protein 2
NA	Neuraminidase
NP	Nucleoprotein
NS1	Non-structural protein 1
NS2	Non-structural protein 2
PA	Polymerase acidic protein
PB1	Polymerase basic protein 1
PB1-F2	Accessory protein F2 translated from PB1 segment
PB2	Polymerase basic protein 2
SVM	Support vector machine
WHO	World Health Organization

Appendix A

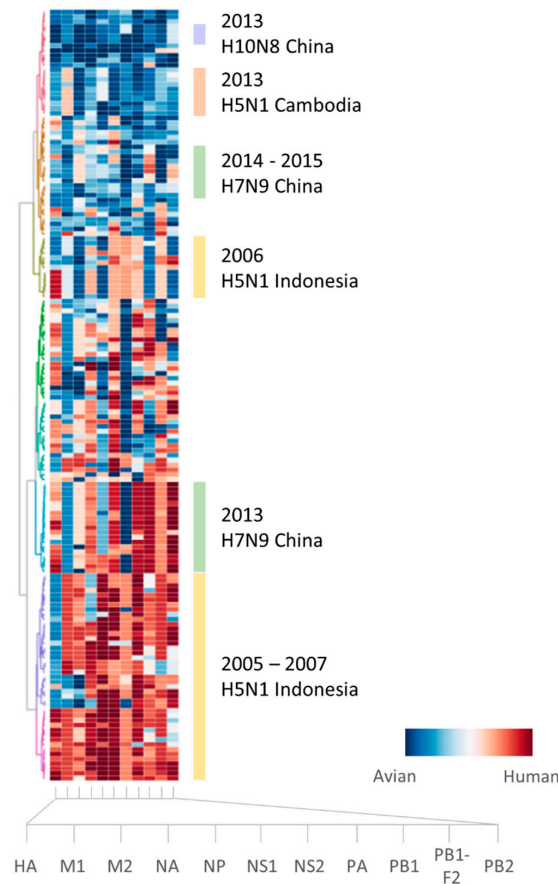


Figure A1. Hierarchical clustering of host tropism protein signatures of 160 confirmed zoonotic strains. Zoonotic strains from the same influenza outbreaks typically share similar signatures. Each row depicts the signature of a zoonotic strain, with host tropism predictions of 11 proteins shown in each column (HA, M1, M2, NA, NP, NS1, NS2, PA, PB1, PB1-F2, and PB2). Avian protein predictions are illustrated in blue while human proteins are in red. The confidence of the avian or human host tropism prediction is expressed by the intensity of the color, based on the prediction probability estimates found in Supplementary S1 and S2 datasets.

Table A1. Total number of suspected and confirmed zoonotic strains with complete proteome identified from zoonotic influenza outbreaks.

Year	Subtype	Country	Avian-isolated Suspected Zoonotic Samples	Human-Isolated Confirmed Zoonotic Samples
1997	H5N1	Hong Kong	2	
	H9N2	Hong Kong		
1998	H9N2	China	1	
1999	H9N2	China	2	
2003	H5N1	China	32	
	H7N7	Vietnam	1	
	H9N2	Netherlands Hong Kong	2	
2004	H5N1	Thailand	10	
		Vietnam	14	
	H7N3	Canada	1	1

Table A1. Cont.

Year	Subtype	Country	Avian-isolated Suspected Zoonotic Samples	Human-Isolated Confirmed Zoonotic Samples	
2005	H5N1	Cambodia			
		China	2		
		Indonesia	11	8	
		Thailand	22		
		Vietnam	46		
2006	H5N1	Azerbaijan			
		Cambodia			
		China	19		
		Djibouti			
		Egypt	4	2	
		Indonesia	11	46	
		Iraq		1	
		Thailand	7	1	
		Turkey		3	
2007	H3N8	Laos	1		
	H5N1	Cambodia			
		China	7		
		Egypt	2		
		Indonesia	4	10	
		Laos	16		
		Myanmar			
		Nigeria	20		
		Pakistan			
		Vietnam	54		
		United Kingdom			
2008	H3N8	Vietnam	1		
	H5N1	Bangladesh		1	
		Cambodia			
		China	5		
		Egypt	18		
		Indonesia	1		
		Vietnam	4		
		H9N2	Hong Kong	9	
		H11N9	Vietnam	2	
2009	H5N1	Cambodia	1		
		China	6		
		Egypt	6		
		Indonesia	1		
		Vietnam			
H9N2	Hong Kong	4			
2010	H5N1	Cambodia	5	1	
		China	3	1	
		Egypt	25		
		Indonesia			
		Vietnam	3		
2011	H5N1	Bangladesh	1	1	
		Cambodia	6	4	
		China	11		
		Egypt	13		
		Indonesia			
2012	H5N1	Bangladesh	18		
		Cambodia	1	3	
		China	3		
		Egypt	6		
		Indonesia			
	Vietnam	30			
H7N3	Mexico		1		

Table A1. Cont.

Year	Subtype	Country	Avian-isolated Suspected Zoonotic Samples	Human-Isolated Confirmed Zoonotic Samples
2013	H5N1	Bangladesh	5	10
		Cambodia	8	
		Canada		
		China	3	
		Egypt	3	
	H7N9	Indonesia		35
		Vietnam	19	
		China	114	
		Hong Kong		
		Taiwan		
H9N2	China	135	1	
	H10N8	China	26	4
2014	H5N1	Cambodia		1
		China	7	
		Egypt	3	
		Indonesia		
	Vietnam	24	19	
H7N9	China	226		
2015	H5N1	China		1
		Egypt		
	Indonesia			
	H5N6	China		
	H7N9	Canada		
		China		2
		Total	1047	160

References

- Centers for Disease Control and Prevention. Isolation of avian influenza A (H5N1) viruses from humans—Hong Kong, May–December 1997. *Morb. Mortal. Wkly. Rep.* **1997**, *46*, 1204–1207.
- Peiris, J.S.; de Jong, M.D.; Guan, Y. Avian influenza virus (H5N1): A threat to human health. *Clin. Microbiol. Rev.* **2007**, *20*, 243–267. [[CrossRef](#)] [[PubMed](#)]
- World Health Organization. Cumulative Number of Confirmed Human Cases for Avian Influenza A (H5N1) Reported to WHO, 2003–2016. Available online: http://www.who.int/influenza/human_animal_interface/EN_GIP_20160404cumulativenumberH5N1cases.pdf (accessed on 18 December 2016).
- Fouchier, R.A.; Schneeberger, P.M.; Rozendaal, F.W.; Broekman, J.M.; Kemink, S.A.; Munster, V.; Kuiken, T.; Rimmelzwaan, G.F.; Schutten, M.; van Doornum, G.J.; et al. Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1356–1361. [[CrossRef](#)] [[PubMed](#)]
- Koopmans, M.; Wilbrink, B.; Conyn, M.; Natrop, G.; van der Nat, H.; Vennema, H.; Meijer, A.; van Steenbergen, J.; Fouchier, R.; Osterhaus, A.; et al. Transmission of H7N7 avian influenza A virus to human beings during a large outbreak in commercial poultry farms in the Netherlands. *Lancet* **2004**, *363*, 587–593. [[CrossRef](#)]
- Kurtz, J.; Manvell, R.J.; Banks, J. Avian influenza virus isolated from a woman with conjunctivitis. *Lancet* **1996**, *348*, 901–902. [[CrossRef](#)]
- Butt, K.M.; Smith, G.J.; Chen, H.; Zhang, L.J.; Leung, Y.H.; Xu, K.M.; Lim, W.; Webster, R.G.; Yuen, K.Y.; Peiris, J.S.; et al. Human infection with an avian H9N2 influenza A virus in Hong Kong in 2003. *J. Clin. Microbiol.* **2005**, *43*, 5760–5767. [[CrossRef](#)] [[PubMed](#)]
- Peiris, M.; Yuen, K.Y.; Leung, C.W.; Chan, K.H.; Ip, P.L.; Lai, R.W.; Orr, W.K.; Shortridge, K.F. Human infection with influenza H9N2. *Lancet* **1999**, *354*, 916–917. [[CrossRef](#)]
- Chen, Y.; Liang, W.; Yang, S.; Wu, N.; Gao, H.; Sheng, J.; Yao, H.; Wo, J.; Fang, Q.; Cui, D.; et al. Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: Clinical analysis and characterisation of viral genome. *Lancet* **2013**, *381*, 1916–1925. [[CrossRef](#)]

10. Gao, R.; Cao, B.; Hu, Y.; Feng, Z.; Wang, D.; Hu, W.; Chen, J.; Jie, Z.; Qiu, H.; Xu, K.; et al. Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* **2013**, *368*, 1888–1897. [[CrossRef](#)] [[PubMed](#)]
11. Richard, M.; de Graaf, M.; Herfst, S. Avian influenza A viruses: From zoonosis to pandemic. *Future Virol.* **2014**, *9*, 513–524. [[CrossRef](#)] [[PubMed](#)]
12. Yu, H.; Wu, J.T.; Cowling, B.J.; Liao, Q.; Fang, V.J.; Zhou, S.; Wu, P.; Zhou, H.; Lau, E.H.; Guo, D.; et al. Effect of closure of live poultry markets on poultry-to-person transmission of avian influenza A H7N9 virus: An ecological study. *Lancet* **2014**, *383*, 541–548. [[CrossRef](#)]
13. World Health Organization. H5N1 Research Issues. Available online: http://www.who.int/influenza/human_animal_interface/avian_influenza/h5n1_research/en/ (accessed on 14 October 2014).
14. Heymann, D.L.; Dixon, M. Infections at the animal/human interface: Shifting the paradigm from emergency response to prevention at source. *Curr. Top. Microbiol. Immunol.* **2013**, *366*, 207–215. [[CrossRef](#)] [[PubMed](#)]
15. Reid, A.H.; Taubenberger, J.K. The origin of the 1918 pandemic influenza virus: A continuing enigma. *J. Gen. Virol.* **2003**, *84*, 2285–2292. [[CrossRef](#)] [[PubMed](#)]
16. Peiris, M.; Yen, H.L. Animal and human influenzas. *Rev. Sci. Tech.* **2014**, *33*, 539–553. [[CrossRef](#)] [[PubMed](#)]
17. World Health Organization. Antigenic and Genetic Characteristics of Influenza A(H5N1) and Influenza A(H9N2) Viruses and Candidate Vaccine Viruses Developed for Potential Use in Human Vaccines. Available online: http://www.who.int/influenza/resources/documents/201009_H5_H9_VaccineVirusUpdate.pdf (accessed on 8 April 2014).
18. Chien, Y.J. How did international agencies perceive the avian influenza problem? The adoption and manufacture of the “one world, one health” framework. *Soc. Health Illn.* **2013**, *35*, 213–226. [[CrossRef](#)] [[PubMed](#)]
19. Hartaningsih, N.; Wibawa, H.; Pudjiatmoko; Rasa, F.S.; Irianingsih, S.H.; Dharmawan, R.; Azhar, M.; Siregar, E.S.; McGrane, J.; Wong, F.; et al. Surveillance at the molecular level: Developing an integrated network for detecting variation in avian influenza viruses in indonesia. *Prev. Vet. Med.* **2015**, *120*, 96–105. [[CrossRef](#)] [[PubMed](#)]
20. Russell, C.A.; Kasson, P.M.; Donis, R.O.; Riley, S.; Dunbar, J.; Rambaut, A.; Asher, J.; Burke, S.; Davis, C.T.; Garten, R.J.; et al. Improving pandemic influenza risk assessment. *eLife* **2014**, *3*, e03883. [[CrossRef](#)] [[PubMed](#)]
21. Chen, G.W.; Chang, S.C.; Mok, C.K.; Lo, Y.L.; Kung, Y.N.; Huang, J.H.; Shih, Y.H.; Wang, J.Y.; Chiang, C.; Chen, C.J.; et al. Genomic signatures of human versus avian influenza a viruses. *Emerg. Infect. Dis.* **2006**, *12*, 1353–1360. [[CrossRef](#)] [[PubMed](#)]
22. Finkelstein, D.B.; Mukatira, S.; Mehta, P.K.; Obenauer, J.C.; Su, X.; Webster, R.G.; Naeve, C.W. Persistent host markers in pandemic and H5N1 influenza viruses. *J. Virol.* **2007**, *81*, 10292–10299. [[CrossRef](#)] [[PubMed](#)]
23. Steel, J.; Lowen, A.C.; Mubareka, S.; Palese, P. Transmission of influenza virus in a mammalian host is increased by PB2 amino acids 627K or 627E/701N. *PLoS Pathog.* **2009**, *5*, e1000252. [[CrossRef](#)] [[PubMed](#)]
24. Subbarao, E.K.; London, W.; Murphy, B.R. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J. Virol.* **1993**, *67*, 1761–1764. [[PubMed](#)]
25. Khaliq, Z.; Leijon, M.; Belak, S.; Komorowski, J. Identification of combinatorial host-specific signatures with a potential to affect host adaptation in influenza A H1N1 and H3N2 subtypes. *BMC Genom.* **2016**, *17*, 529. [[CrossRef](#)] [[PubMed](#)]
26. Sjaugi, M.F.; Tan, S.; Abd Raman, H.S.; Lim, W.C.; Nik Mohamed, N.E.; August, J.; Khan, A.M. g-FLUA2H: A web-based application to study the dynamics of animal-to-human mutation transmission for influenza viruses. *BMC Med. Genom.* **2015**, *8*, S5. [[CrossRef](#)] [[PubMed](#)]
27. Taubenberger, J.K.; Kash, J.C. Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host Microbe* **2010**, *7*, 440–451. [[CrossRef](#)] [[PubMed](#)]
28. Global Influenza Programme. Tool for Influenza Pandemic Risk Assessment (TIPRA). Available online: http://www.who.int/influenza/publications/TIPRA_manual_v1/en/ (accessed on 14 March 2017).
29. Trock, S.C.; Burke, S.A.; Cox, N.J. Development of framework for assessing influenza virus pandemic risk. *Emerg. Infect. Dis.* **2015**, *21*, 1372–1378. [[CrossRef](#)] [[PubMed](#)]
30. Qiang, X.; Kou, Z. Prediction of interspecies transmission for avian influenza A virus based on a back-propagation neural network. *Math. Comput. Model.* **2010**, *52*, 2060–2065. [[CrossRef](#)]

31. Wang, J.; Ma, C.; Kou, Z.; Zhou, Y.; Liu, H. Predicting transmission of avian influenza A viruses from avian to human by using informative physicochemical properties. *Int. J. Data Min. Bioinform.* **2013**, *7*, 166–179. [[CrossRef](#)] [[PubMed](#)]
32. Eng, C.L.; Tong, J.C.; Tan, T.W. Predicting host tropism of influenza A virus proteins using random forest. *BMC Med. Genom.* **2014**, *7*, S1. [[CrossRef](#)] [[PubMed](#)]
33. Eng, C.L.; Tong, J.C.; Tan, T.W. Distinct host tropism protein signatures to identify possible zoonotic influenza A viruses. *PLoS ONE* **2016**, *11*, e0150173. [[CrossRef](#)] [[PubMed](#)]
34. Kageyama, T.; Fujisaki, S.; Takashita, E.; Xu, H.; Yamada, S.; Uchida, Y.; Neumann, G.; Saito, T.; Kawaoka, Y.; Tashiro, M. Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China, February to April 2013. *Euro Surveill.* **2013**, *18*, 20453. [[PubMed](#)]
35. World Health Organization. Overview of the Emergence and Characteristics of the Avian Influenza A(H7N9) Virus. Available online: [Http://www.who.int/influenza/human_animal_interface/influenza_h7n9/WHO_H7N9_review_31May13.pdf?ua=1](http://www.who.int/influenza/human_animal_interface/influenza_h7n9/WHO_H7N9_review_31May13.pdf?ua=1) (accessed on 8 April 2014).
36. Pantin-Jackwood, M.J.; Miller, P.J.; Spackman, E.; Swayne, D.E.; Susta, L.; Costa-Hurtado, M.; Suarez, D.L. Role of poultry in the spread of novel H7N9 influenza virus in china. *J. Virol.* **2014**, *88*, 5381–5390. [[CrossRef](#)] [[PubMed](#)]
37. Yang, J.R.; Kuo, C.Y.; Huang, H.Y.; Wu, F.T.; Huang, Y.L.; Cheng, C.Y.; Su, Y.T.; Wu, H.S.; Liu, M.T. Characterization of influenza A (H7N9) viruses isolated from human cases imported into Taiwan. *PLoS ONE* **2015**, *10*, e0119792. [[CrossRef](#)] [[PubMed](#)]
38. Rith, S.; Davis, C.T.; Duong, V.; Sar, B.; Horm, S.V.; Chin, S.; Ly, S.; Laurent, D.; Richner, B.; Oboho, I.; et al. Identification of molecular markers associated with alteration of receptor-binding specificity in a novel genotype of highly pathogenic avian influenza A(H5N1) viruses detected in Cambodia in 2013. *J. Virol.* **2014**, *88*, 13897–13909. [[CrossRef](#)] [[PubMed](#)]
39. Setiawaty, V.; Dharmayanti, N.L.; Misriyah; Pawestri, H.A.; Azhar, M.; Tallis, G.; Schoonman, L.; Samaan, G. Avian influenza A(H5N1) virus outbreak investigation: Application of the FAO-OIE-WHO four-way linking framework in Indonesia. *Zoonoses Public Health* **2015**, *62*, 381–387. [[CrossRef](#)] [[PubMed](#)]
40. Wibawa, H.; Henning, J.; Wong, F.; Selleck, P.; Junaidi, A.; Bingham, J.; Daniels, P.; Meers, J. A molecular and antigenic survey of H5N1 highly pathogenic avian influenza virus isolates from smallholder duck farms in Central Java, Indonesia during 2007–2008. *J. Virol.* **2011**, *8*, 425. [[CrossRef](#)] [[PubMed](#)]
41. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
42. Squires, R.B.; Noronha, J.; Hunt, V.; Garcia-Sastre, A.; Macken, C.; Baumgarth, N.; Suarez, D.; Pickett, B.E.; Zhang, Y.; Larsen, C.N.; et al. Influenza research database: An integrated bioinformatics resource for influenza research and surveillance. *Influenza Respir. Viruses* **2012**, *6*, 404–416. [[CrossRef](#)] [[PubMed](#)]
43. Ahmed, S.S.; Themudo, G.E.; Christensen, J.P.; Biswas, P.K.; Giasuddin, M.; Samad, M.A.; Toft, N.; Ersboll, A.K. Molecular epidemiology of circulating highly pathogenic avian influenza (H5N1) virus in chickens, in Bangladesh, 2007–2010. *Vaccine* **2012**, *30*, 7381–7390. [[CrossRef](#)] [[PubMed](#)]
44. Centers for Disease Control and Prevention. Update: Influenza activity—United States and worldwide, 2005–06 season, and composition of the 2006–07 influenza vaccine. *Morb. Mortal. Wkly. Rep.* **2006**, *55*, 648–653.
45. Pabbaraju, K.; Tellier, R.; Wong, S.; Li, Y.; Bastien, N.; Tang, J.W.; Drews, S.J.; Jang, Y.; Davis, C.T.; Fonseca, K.; et al. Full-genome analysis of avian influenza A(H5N1) virus from a human, North America, 2013. *Emerg. Infect. Dis.* **2014**, *20*, 887–891. [[CrossRef](#)] [[PubMed](#)]
46. Rabinowitz, P.M.; Galusha, D.; Vegso, S.; Michalove, J.; Rinne, S.; Scotch, M.; Kane, M. Comparison of human and animal surveillance data for H5N1 influenza A in Egypt 2006–2011. *PLoS ONE* **2012**, *7*, e43851. [[CrossRef](#)] [[PubMed](#)]
47. Claas, E.C.; de Jong, J.C.; van Beek, R.; Rimmelzwaan, G.F.; Osterhaus, A.D. Human influenza virus A/HongKong/156/97 (H5N1) infection. *Vaccine* **1998**, *16*, 977–978. [[CrossRef](#)]
48. Kandun, I.N.; Wibisono, H.; Sedyaningsih, E.R.; Yusharmen; Hadisoedarsuno, W.; Purba, W.; Santoso, H.; Septiawati, C.; Tresnaningsih, E.; Heriyanto, B.; et al. Three Indonesian clusters of H5N1 virus infection in 2005. *N. Engl. J. Med.* **2006**, *355*, 2186–2194. [[CrossRef](#)] [[PubMed](#)]
49. Centers for Disease Control and Prevention. Update: Influenza activity—United States and worldwide, 2006–07 season, and composition of the 2007–08 influenza vaccine. *Morb. Mortal. Wkly. Rep.* **2007**, *56*, 789–794.

50. Zaman, M.; Ashraf, S.; Dreyer, N.A.; Toovey, S. Human infection with avian influenza virus, Pakistan, 2007. *Emerg. Infect. Dis.* **2011**, *17*, 1056–1059. [[CrossRef](#)] [[PubMed](#)]
51. Tiensin, T.; Chaitaweesub, P.; Songserm, T.; Chaisingh, A.; Hoonsuwan, W.; Buranathai, C.; Parakamawongsa, T.; Premashthira, S.; Amonsin, A.; Gilbert, M.; et al. Highly pathogenic avian influenza H5N1, Thailand, 2004. *Emerg. Infect. Dis.* **2005**, *11*, 1664–1672. [[CrossRef](#)] [[PubMed](#)]
52. Oner, A.F.; Bay, A.; Arslan, S.; Akdeniz, H.; Sahin, H.A.; Cesur, Y.; Epcacan, S.; Yilmaz, N.; Deger, I.; Kizilyildiz, B.; et al. Avian influenza A (H5N1) infection in Eastern Turkey in 2006. *N. Engl. J. Med.* **2006**, *355*, 2179–2185. [[CrossRef](#)] [[PubMed](#)]
53. Manabe, T.; Yamaoka, K.; Tango, T.; Binh, N.G.; Co, D.X.; Tuan, N.D.; Izumi, S.; Takasaki, J.; Chau, N.Q.; Kudo, K. Chronological, geographical, and seasonal trends of human cases of avian influenza A (H5N1) in Vietnam, 2003–2014: A spatial analysis. *BMC Infect. Dis.* **2016**, *16*, 64. [[CrossRef](#)] [[PubMed](#)]
54. Qi, X.; Cui, L.; Yu, H.; Ge, Y.; Tang, F. Whole-genome sequence of a reassortant H5N6 avian influenza virus isolated from a live poultry market in China, 2013. *Genome Announc.* **2014**, *2*. [[CrossRef](#)] [[PubMed](#)]
55. Yuan, J.; Zhang, L.; Kan, X.; Jiang, L.; Yang, J.; Guo, Z.; Ren, Q. Origin and molecular characteristics of a novel 2013 avian influenza A(H6N1) virus causing human infection in Taiwan. *Clin. Infect. Dis.* **2013**, *57*, 1367–1368. [[CrossRef](#)] [[PubMed](#)]
56. Editorial Team. Avian influenza A/(H7N2) outbreak in the United Kingdom. *Euro Surveill.* **2007**, *12*, E070531. [[PubMed](#)]
57. Editorial team. Avian influenza H5N1 detected in poultry in Nigeria, further human cases reported in Iraq, Indonesia and China. *Euro Surveill.* **2006**, *11*, E060209 060201. [[PubMed](#)]
58. Ostrowsky, B.; Huang, A.; Terry, W.; Anton, D.; Brunagel, B.; Traynor, L.; Abid, S.; Johnson, G.; Kacica, M.; Katz, J.; et al. Low pathogenic avian influenza A (H7N2) virus infection in immunocompromised adult, New York, USA, 2003. *Emerg. Infect. Dis.* **2012**, *18*, 1128–1131. [[CrossRef](#)] [[PubMed](#)]
59. Tweed, S.A.; Skowronski, D.M.; Davies, T.M.; Larder, A.; Petric, M.; Lees, W.; Li, Y.; Katz, J.; Krajdien, M.; Tellier, R.; et al. Human illness from avian influenza H7N3, British Columbia. *Emerg. Infect. Dis.* **2004**, *10*, 2196–2199. [[CrossRef](#)] [[PubMed](#)]
60. Puzelli, S.; di Trani, L.; Fabiani, C.; Campitelli, L.; de Marco, M.A.; Capua, I.; Aguilera, J.F.; Zamboni, M.; Donatelli, I. Serological analysis of serum samples from humans exposed to avian H7 influenza viruses in Italy between 1999 and 2003. *J. Infect. Dis.* **2005**, *192*, 1318–1322. [[CrossRef](#)] [[PubMed](#)]
61. Lopez-Martinez, I.; Balish, A.; Barrera-Badillo, G.; Jones, J.; Nunez-Garcia, T.E.; Jang, Y.; Aparicio-Antonio, R.; Azziz-Baumgartner, E.; Belser, J.A.; Ramirez-Gonzalez, J.E.; et al. Highly pathogenic avian influenza A(H7N3) virus in poultry workers, Mexico, 2012. *Emerg. Infect. Dis.* **2013**, *19*, 1531–1534. [[CrossRef](#)] [[PubMed](#)]
62. Nguyen-Van-Tam, J.S.; Nair, P.; Acheson, P.; Baker, A.; Barker, M.; Bracebridge, S.; Croft, J.; Ellis, J.; Gelletlie, R.; Gent, N.; et al. Outbreak of low pathogenicity H7N3 avian influenza in UK, including associated case of human conjunctivitis. *Euro Surveill.* **2006**, *11*, E060504 060502. [[PubMed](#)]
63. Puzelli, S.; Rossini, G.; Facchini, M.; Vaccari, G.; di Trani, L.; di Martino, A.; Gaibani, P.; Vocale, C.; Cattoli, G.; Bennett, M.; et al. Human infection with highly pathogenic A(H7N7) avian influenza virus, Italy, 2013. *Emerg. Infect. Dis.* **2014**, *20*, 1745–1749. [[CrossRef](#)] [[PubMed](#)]
64. World Health Organization. Background and Summary of Human Infection with Avian Influenza A(H7N9) Virus—As of 31 January 2014. Available online: http://www.who.int/influenza/human_animal_interface/20140131_background_and_summary_H7N9_v1.pdf?ua=1 (accessed on 21 December 2016).
65. Huang, Y.; Li, X.; Zhang, H.; Chen, B.; Jiang, Y.; Yang, L.; Zhu, W.; Hu, S.; Zhou, S.; Tang, Y.; et al. Human infection with an avian influenza A (H9N2) virus in the middle region of China. *J. Med. Virol.* **2015**, *87*, 1641–1648. [[CrossRef](#)] [[PubMed](#)]
66. Pan American Health Organization. Avian Influenza Virus A(H10N7) Circulating among Humans in Egypt. Available online: http://www1.paho.org/hq/dmdocuments/2010/Avian_Influenza_Egypt_070503.pdf (accessed on 17 December 2016).
67. Arzey, G.G.; Kirkland, P.D.; Arzey, K.E.; Frost, M.; Maywood, P.; Conaty, S.; Hurt, A.C.; Deng, Y.M.; Iannello, P.; Barr, I.; et al. Influenza virus A (H10N7) in chickens and poultry abattoir workers, Australia. *Emerg. Infect. Dis.* **2012**, *18*, 814–816. [[CrossRef](#)] [[PubMed](#)]
68. Chen, H.; Yuan, H.; Gao, R.; Zhang, J.; Wang, D.; Xiong, Y.; Fan, G.; Yang, F.; Li, X.; Zhou, J.; et al. Clinical and epidemiological characteristics of a fatal case of avian influenza A H10N8 virus infection: A descriptive study. *Lancet* **2014**, *383*, 714–721. [[CrossRef](#)]

69. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *SIGKDD Explor.* **2009**, *11*, 10–18. [[CrossRef](#)]
70. Fawcett, T. An introduction to ROC analysis. *Pattern Recogn. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
71. Linden, A. Measuring diagnostic and predictive accuracy in disease management: An introduction to receiver operating characteristic (ROC) analysis. *J. Eval. Clin. Pract.* **2006**, *12*, 132–139. [[CrossRef](#)] [[PubMed](#)]
72. Herrick, K.A.; Huettmann, F.; Lindgren, M.A. A global model of avian influenza prediction in wild birds: The importance of northern regions. *BMC Vet. Res.* **2013**, *44*, 42. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).