

Learning definitions of complex notions – the case of Avian Influenza Virus

Knowledge-based Systems in Bioinformatics
Fall 2019
Prof. Jan Komorowski

Slides contributed by Zeeshan Khaliq

RESEARCH ARTICLE

Open Access



A complete map of potential pathogenicity markers of avian influenza virus subtype H5 predicted from 11 expressed proteins

Zeeshan Khalil¹, Mikael Leijon^{2,3}, Sándor Belák^{3,4} and Jan Komorowski^{1,5*}

Abstract

Background: Polybasic cleavage sites of the hemagglutinin (HA) proteins are considered to be the most important determinants indicating virulence of the avian influenza viruses (AIV). However, evidence is accumulating that these sites alone are not sufficient to establish high pathogenicity. There need to exist other sites located on the HA protein outside the cleavage site or on the other proteins expressed by AIV that contribute to the pathogenicity.

Results: We employed rule-based computational modeling to construct a map, with high statistical significance, of amino acid (AA) residues associated to pathogenicity in 11 proteins of the H5 type viruses. We found potential markers of pathogenicity in all of the 11 proteins expressed by the H5 type of AIV. AA mutations S-43^{HA1}-D, D-83^{HA1}-A in HA; S-269-D, E-41-H in NA; S-48-N, K-212-N in NS1; V-166-A in M1; G-14-E in M2; K-77-R, S-377-N in NP; and Q-48-P in PB1-F2 were identified as having a potential to shift the pathogenicity from low to high. Our results suggest that the low pathogenicity is common to most of the subtypes of the H5 AIV while the high pathogenicity is specific to each subtype. The models were developed using public data and validated on new, unseen sequences.

Conclusions: Our models explicitly define a viral genetic background required for the virus to be highly pathogenic and thus confirm the hypothesis of the presence of pathogenicity markers beyond the cleavage site.

Keywords: Avian Influenza virus, Pathogenicity, Virulence, MCFS, Rosetta, Rough sets

RESEARCH ARTICLE

Open Access



Identification of combinatorial host-specific signatures with a potential to affect host adaptation in influenza A H1N1 and H3N2 subtypes

Zeeshan Khaliq¹, Mikael Leijon^{2,3}, Sándor Belák^{3,4} and Jan Komorowski^{1,5*} 

Abstract

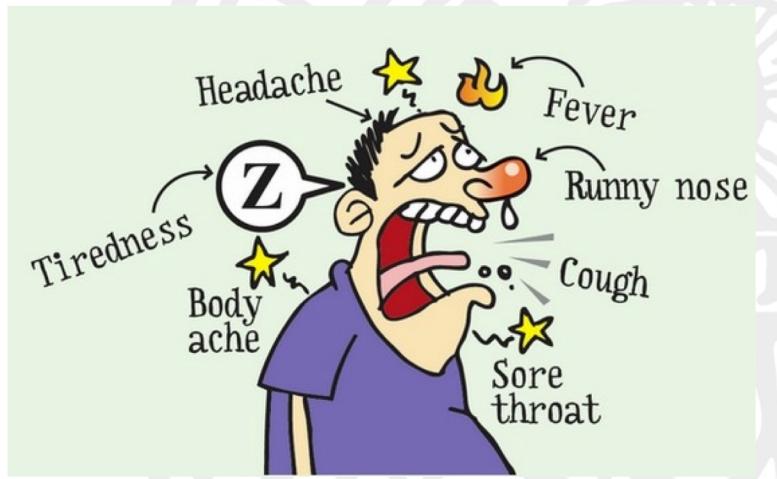
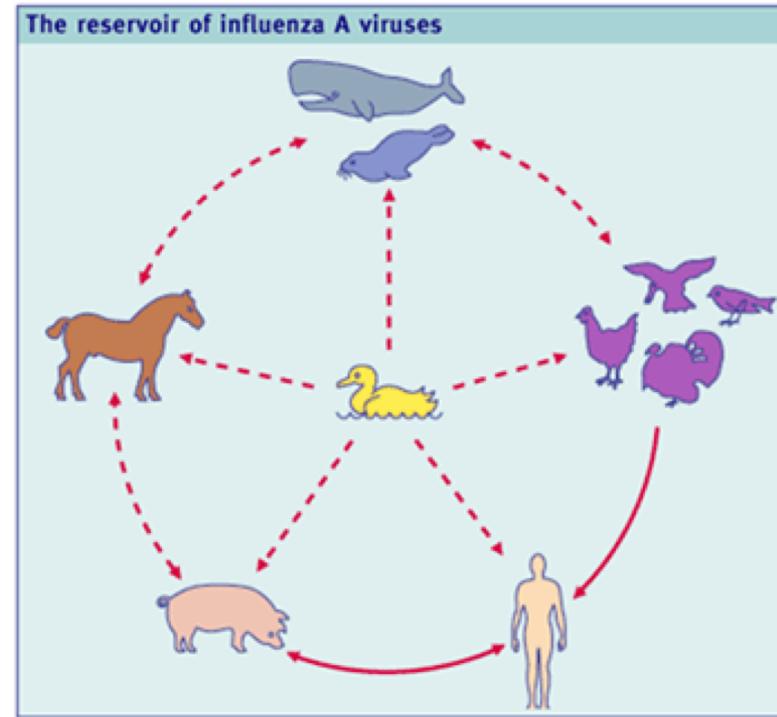
Background: The underlying strategies used by influenza A viruses (IAVs) to adapt to new hosts while crossing the species barrier are complex and yet to be understood completely. Several studies have been published identifying singular genomic signatures that indicate such a host switch. The complexity of the problem suggested that in addition to the singular signatures, there might be a combinatorial use of such genomic features, in nature, defining adaptation to hosts.

Results: We used computational rule-based modeling to identify combinatorial sets of interacting amino acid (aa) residues in 12 proteins of IAVs of H1N1 and H3N2 subtypes. We built highly accurate rule-based models for each protein that could differentiate between viral aa sequences coming from avian and human hosts. We found 68 host-specific combinations of aa residues, potentially associated to host adaptation on HA, M1, M2, NP, NS1, NEP, PA, PA-X, PB1 and PB2 proteins of the H1N1 subtype and 24 on M1, M2, NEP, PB1 and PB2 proteins of the H3N2 subtypes. In addition to these combinations, we found 132 novel singular aa signatures distributed among all proteins, including the newly discovered PA-X protein, of both subtypes. We showed that HA, NA, NP, NS1, NEP, PA-X and PA proteins of the H1N1 subtype carry H1N1-specific and HA, NA, PA-X, PA, PB1-F2 and PB1 of the H3N2 subtype carry H3N2-specific signatures. M1, M2, PB1-F2, PB1 and PB2 of H1N1 subtype, in addition to H1N1 signatures, also carry H3N2 signatures. Similarly M1, M2, NP, NS1, NEP and PB2 of H3N2 subtype were shown to carry both H3N2 and H1N1 host-specific signatures (HSSs).



Influenza A viruses (IAVs)

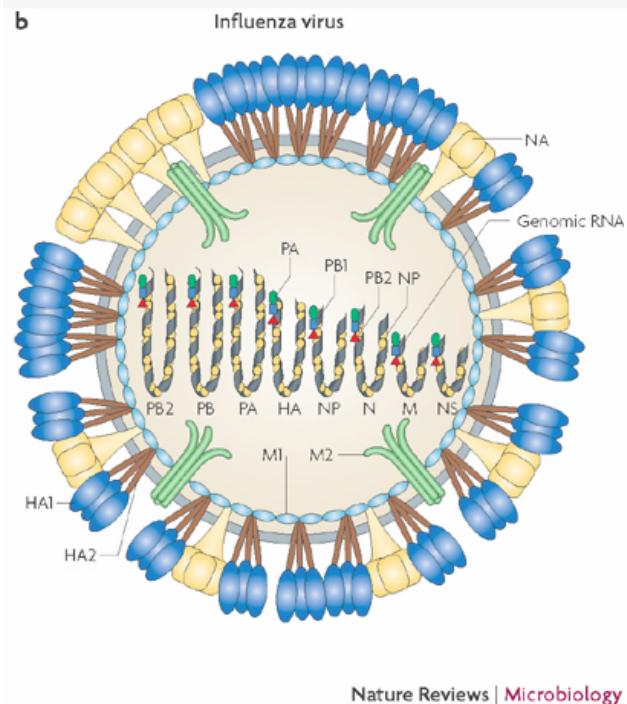
- Wild birds are natural reservoirs of IAVs
- Cross species transmission is rare but does happen
- IAVs can cause mild to severe disease in both humans and animals
 - Yearly epidemics prove fatal for at least 250,000-500,000 humans (CDC. Influenza (Seasonal) Fact Sheet. 2014)





Influenza A viruses (IAVs)

- 8 single RNA strands coding for 11 proteins
- Classification based on hemagglutinin (HA) and neuraminidase (NA)
- 18 subtypes of HA and 9 types of NA
- Further classified as high and low pathogenic
- Pathogenicity is related to insertions at the cleavage site on the surface glycoprotein HA



Nature Reviews | Microbiology

[Hedestem et. al, 2008]

What makes the virus highly pathogenic?

- How would you approach answering this question?
- If classification is the answer, what do you need in terms of data?
- Once you find a classifier, how would you prove it is correct?



Known markers of HP

1. Elongated cleavage site
2. Last four aa's must be R/K XX R/K

Leijon M et. al., 2011

	P (X) R/K XX R/K	High pathogenic
AAA92245	LATGMKNVP-E	TPKKKKK-K-R-GLF
AEZ68704	LATGMKNVP-E	IPKKKKK-K-R-GLF
AAA92246	LATGMKNVP-E	IP-KKRK-K-R-GLF
AAA92244	LATGMKNVP-E	IP-KRK-K-R-GLF
AAC54377	LATGMKNVP-E	IP-KKK-K-R-GLF
ADI34057	LATGMKNVP-E	IP-K-G-R-GLF
AAG10676	LATGMKNVP-E	IP-K-G-R-GLF
ACJ03944	LATGMKNVP-E	IP-K-G-R-GLF

Low pathogenic

Previous studies

Experimentally proven: Insertion in the cleavage site is a necessary but not a sufficient condition for a virus to become HP

- Bogs J, et al., 2010: **Highly Pathogenic H5N1 Influenza Viruses Carry Virulence Determinants beyond the Polybasic Hemagglutinin Cleavage Site.**
- Veits J, et al. 2012: **Avian influenza virus hemagglutinins H2, H4, H8, and H14 support a highly pathogenic phenotype**
- Gohrbandt S et al. 2011: **H9 avian influenza reassortant with engineered polybasic cleavage site displays a highly pathogenic phenotype in chicken.**

Conclusion: For HP, a proper viral background is needed which is suggested but not identified in previous studies



UPPSALA
UNIVERSITET

Goals

Identify the viral background necessary for high pathogenicity

RESEARCH ARTICLE

Open Access



A complete map of potential pathogenicity markers of avian influenza virus subtype H5 predicted from 11 expressed proteins

Zeeshan Khaliq¹, Mikael Leijon^{2,3}, Sándor Belák^{3,4} and Jan Komorowski^{1,5*}

Abstract

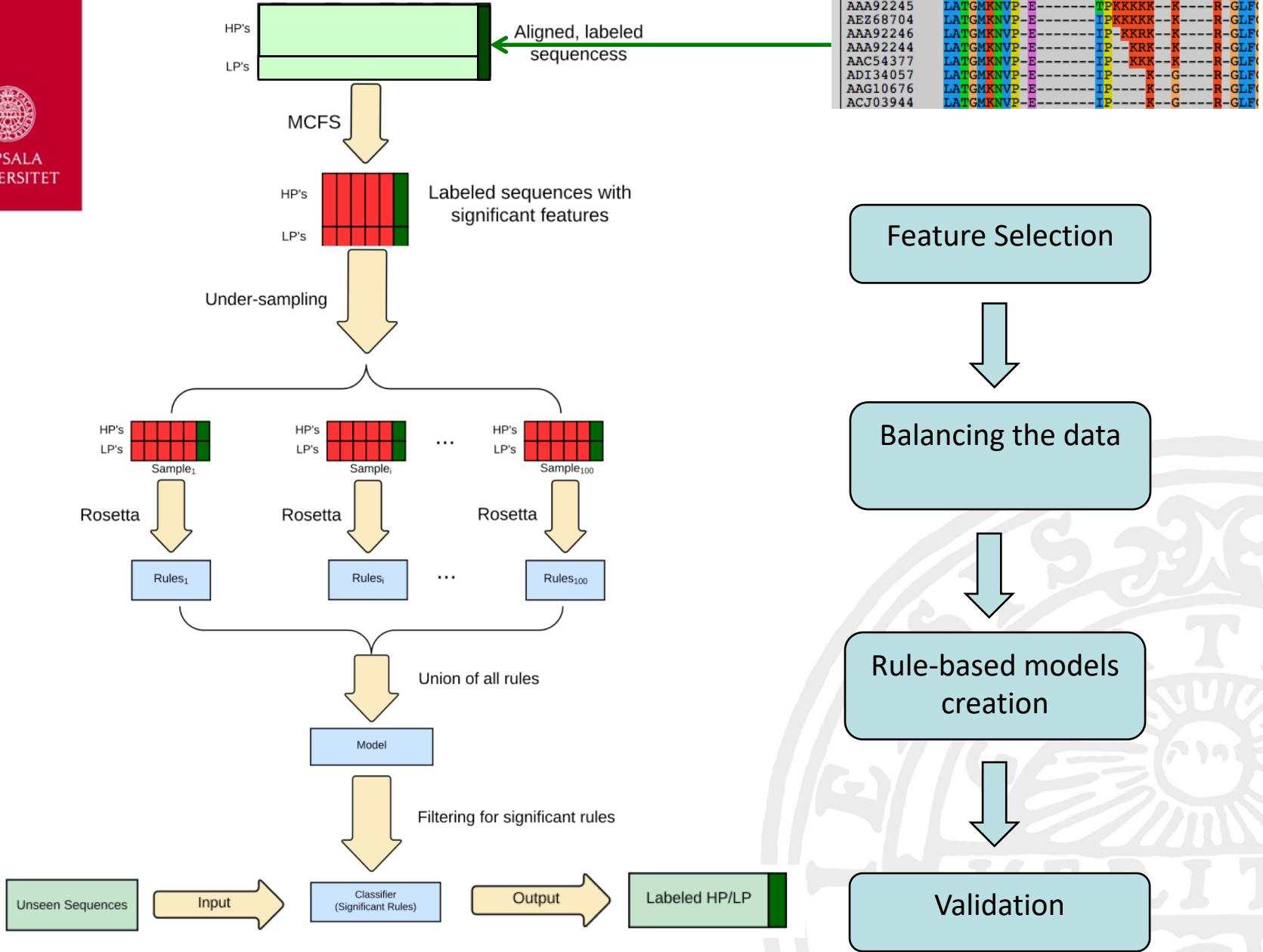
Background: Polybasic cleavage sites of the hemagglutinin (HA) proteins are considered to be the most important determinants indicating virulence of the avian influenza viruses (AIV). However, evidence is accumulating that these sites alone are not sufficient to establish high pathogenicity. There need to exist other sites located on the HA protein outside the cleavage site or on the other proteins expressed by AIV that contribute to the pathogenicity.

Results: We employed rule-based computational modeling to construct a map, with high statistical significance, of amino acid (AA) residues associated to pathogenicity in 11 proteins of the H5 type viruses. We found potential markers of pathogenicity in all of the 11 proteins expressed by the H5 type of AIV. AA mutations S-43^{HA1}-D, D-83^{HA1}-A in HA; S-269-D, E-41-H in NA; S-48-N, K-212-N in NS1; V-166-A in M1; G-14-E in M2; K-77-R, S-377-N in NP; and Q-48-P in PB1-F2 were identified as having a potential to shift the pathogenicity from low to high. Our results suggest that the low pathogenicity is common to most of the subtypes of the H5 AIV while the high pathogenicity is specific to each subtype. The models were developed using public data and validated on new, unseen sequences.

Conclusions: Our models explicitly define a viral genetic background required for the virus to be highly pathogenic and thus confirm the hypothesis of the presence of pathogenicity markers beyond the cleavage site.

Keywords: Avian Influenza virus, Pathogenicity, Virulence, MCFS, Rosetta, Rough sets

Methodology





UPPSALA
UNIVERSITET

Results and Discussion ...

The Strongest Rules from HA classifier

Table 3: The strongest rules for highly and low pathogenic viruses from the HA classifier.

	Rule	Accuracy ^a (%)	Support ^b	Class-Specific-Coverage ^c (%)
HP-Rules	IF P43(HA1)=D THEN virus=HP	99.8	1225	86
	IF P83(HA1)=A THEN virus=HP	100	807	57
	IF P71(HA1)=I THEN virus=HP	100	759	53
LP-Rules	IF P43(HA1)=S THEN virus=LP	95.2	589	99
	IF P83(HA1)=D THEN virus=LP	94.6	571	95
	IF P107(HA1)=S THEN virus=LP	95.8	552	93
	IF P138(HA1)=N THEN virus=LP	92.7	536	88
	IF P309(HA1)=D THEN virus=LP	94.9	533	89
	IF P320(HA1)=V THEN virus=LP	95.7	532	90
	IF P195(HA1)=N THEN virus=LP	88.8	400	63
	IF P16(SP)=G THEN virus=LP	89.3	392	62
	IF P203(HA2)=I THEN virus=LP	82.4	380	55
	IF P6(SP)=I THEN virus=LP	97.5	354	61
	IF P7(SP)=A THEN virus=LP	98	352	61
	IF P3(SP)=R THEN virus=LP	94.1	341	57
	IF P240(HA1)=S THEN virus=LP	95.2	332	56
	IF P275(HA1)=D THEN virus=LP	97.3	300	52

^a Accuracy is the percentage of the sequences in the support set classified correctly by the rule.

^b Support is the number of sequences that satisfy the “IF” conditions of the rule.

^c Class-Specific-Coverage is the percentage per class of the sequences that support the rule and are correctly classified by the rule. For instance, if a rule is an HP class rule then the Class-Specific-Coverage would mean percentage of the HP sequences classified correctly by this rule.

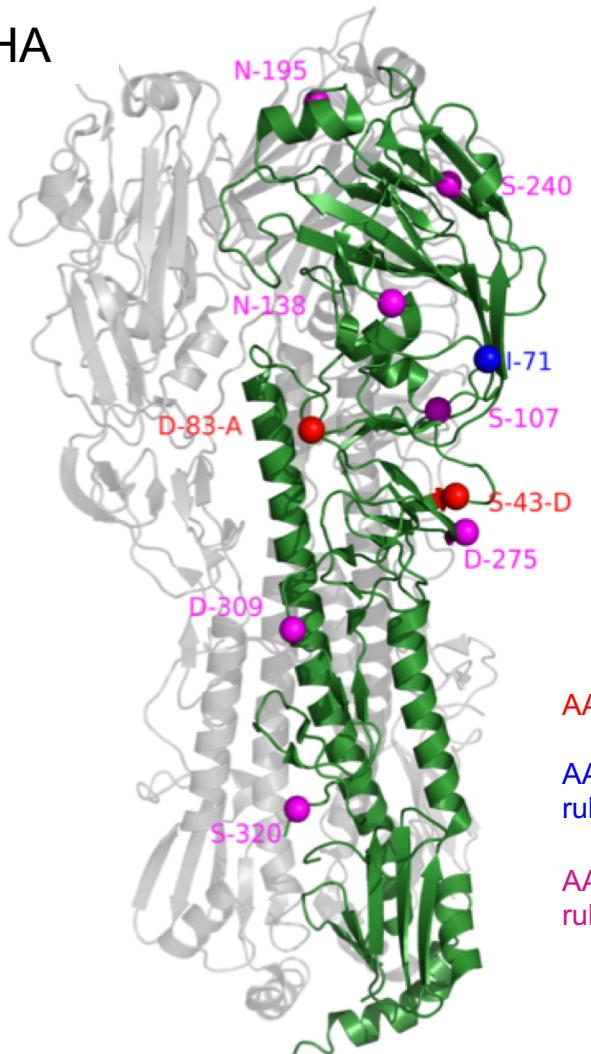
Mutations associated with shifts of pathogenicity from low to high

Table 5: Amino acid mutations associated with a shift of pathogenicity from low to high

AA changes associated with a change in pathogenicity from low to high	
HA	S-43 ^{HA1} -D, D-83 ^{HA1} -A
NA	S-269-D, E-41-H
NS1	S-48-N, K-212-N
NS2	-
M1	V-166-A
M2	G-14-E
NP	K-77-R, S-377-N
PA	-
PB1	-
PB2	-
PB1-F2	Q-48-P

AA's Associated to pathogenicity

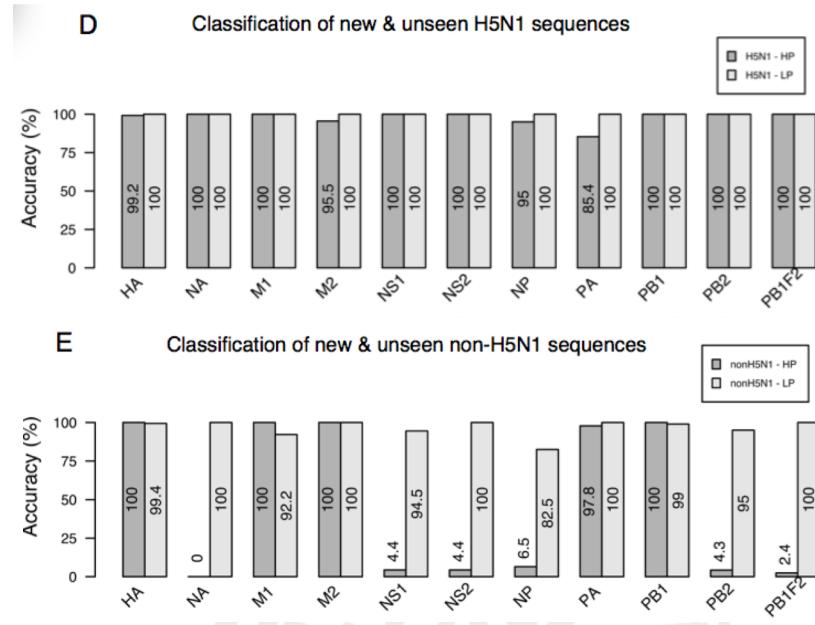
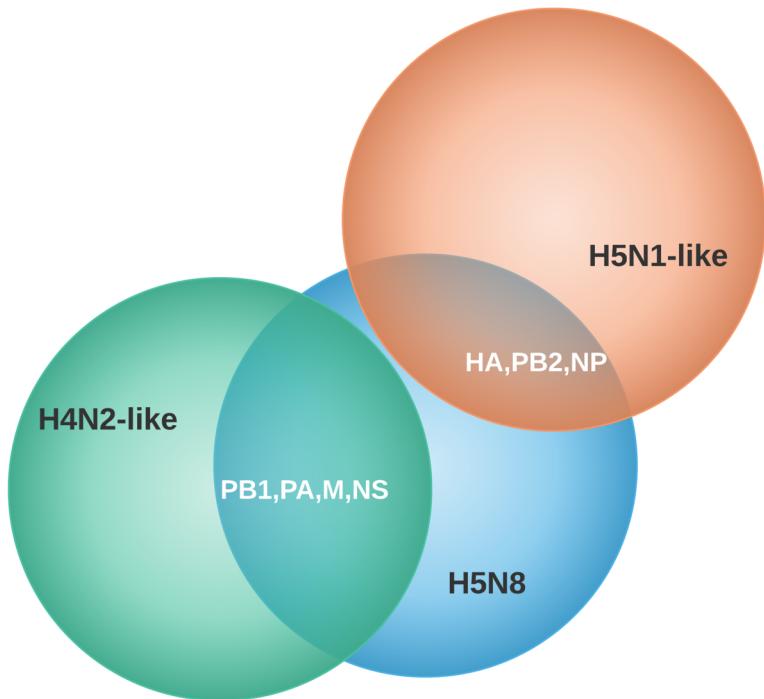
HA



Protein	AA's associated with high pathogenicity	AA's associated with low pathogenicity
HA	D-43 ^{HA1} , A-83 ^{HA1} , I-71 ^{HA1}	S-43 ^{HA1} , D-83 ^{HA1} , S-107 ^{HA1} , N-138 ^{HA1} , D-309 ^{HA1} , V-302 ^{HA1} , A-7 ^{sp} , I-6 ^{sp} , D-275 ^{HA1} , N-195 ^{HA1} , S-240 ^{HA1} , R-3 ^{SP} , S-194 ^{HA1}
NA	N-369, G-386, T-288, H-100, D-269, H-41	N-400, K-38, V-192, P-90, I-73, I-262, L-255, M-24, S-14, E-41, S-269, K-187, T-434, E-74, S-43

High pathogenicity specific to subtype

- Mechanism of HP is subtype specific.
- Mechanism of LP is similar across subtypes.



Mostly H5N8, classic reassortant viruses.



UPPSALA
UNIVERSITET

Genomic signatures of Host Specificity in Influenza A viruses (H1N1 and H3N2 types)

Slides contributed by Zeeshan Khaliq

RESEARCH ARTICLE

Open Access



Identification of combinatorial host-specific signatures with a potential to affect host adaptation in influenza A H1N1 and H3N2 subtypes

Zeeshan Khaliq¹, Mikael Leijon^{2,3}, Sándor Belák^{3,4} and Jan Komorowski^{1,5*} 

Abstract

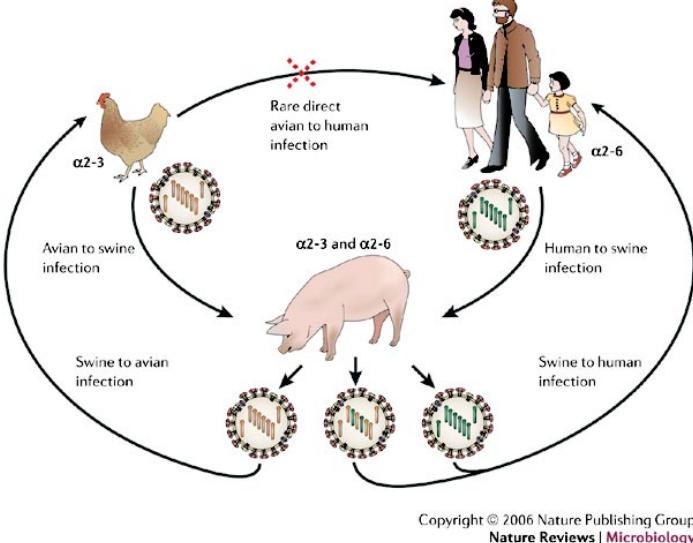
Background: The underlying strategies used by influenza A viruses (IAVs) to adapt to new hosts while crossing the species barrier are complex and yet to be understood completely. Several studies have been published identifying singular genomic signatures that indicate such a host switch. The complexity of the problem suggested that in addition to the singular signatures, there might be a combinatorial use of such genomic features, in nature, defining adaptation to hosts.

Results: We used computational rule-based modeling to identify combinatorial sets of interacting amino acid (aa) residues in 12 proteins of IAVs of H1N1 and H3N2 subtypes. We built highly accurate rule-based models for each protein that could differentiate between viral aa sequences coming from avian and human hosts. We found 68 host-specific combinations of aa residues, potentially associated to host adaptation on HA, M1, M2, NP, NS1, NEP, PA, PA-X, PB1 and PB2 proteins of the H1N1 subtype and 24 on M1, M2, NEP, PB1 and PB2 proteins of the H3N2 subtypes. In addition to these combinations, we found 132 novel singular aa signatures distributed among all proteins, including the newly discovered PA-X protein, of both subtypes. We showed that HA, NA, NP, NS1, NEP, PA-X and PA proteins of the H1N1 subtype carry H1N1-specific and HA, NA, PA-X, PA, PB1-F2 and PB1 of the H3N2 subtype carry H3N2-specific signatures. M1, M2, PB1-F2, PB1 and PB2 of H1N1 subtype, in addition to H1N1 signatures, also carry H3N2 signatures. Similarly M1, M2, NP, NS1, NEP and PB2 of H3N2 subtype were shown to carry both H3N2 and H1N1 host-specific signatures (HSSs).

Zoonosis - Transmission of Avian Influenza A viruses between animals and people

- Influenza A viruses have infected many different animals, including ducks, chickens, pigs, whales, horses, and seals.
- Certain subtypes of influenza A virus are specific to certain species, except for birds, which are hosts to all known subtypes of influenza A viruses.
- Avian influenza A viruses may be transmitted from animals to humans in two main ways:
 - Directly from birds or from avian influenza A virus-contaminated environments to people.
 - Through an intermediate host, such as a pig.
- From: Centers for Disease Control and Prevention

Background



Stevens et al. *Nature Reviews Microbiology* 4, 857-864 (November 2006)
| doi:10.1038/nrmicro1530

nature
REVIEWS MICROBIOLOGY

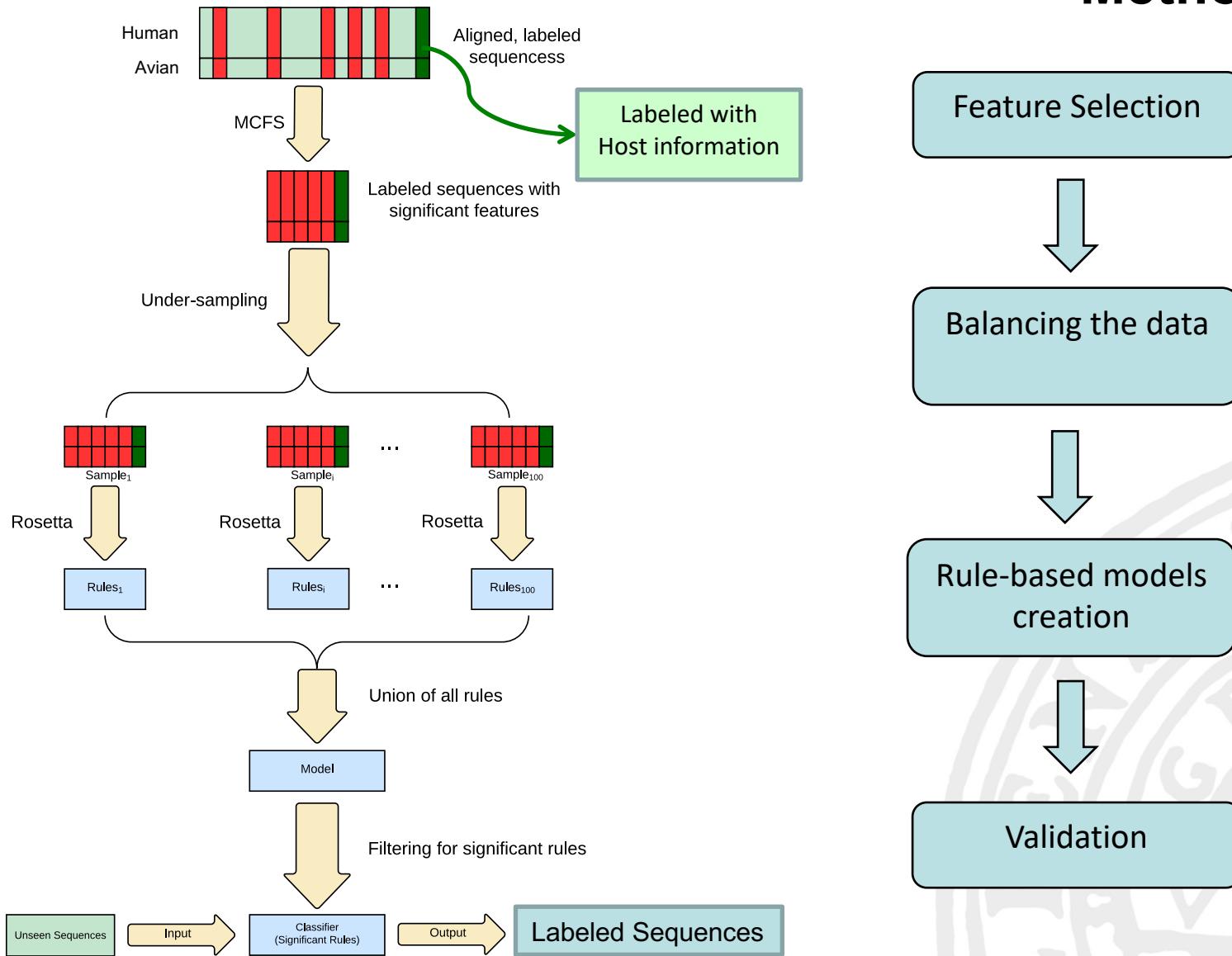
- Species barrier crosses are complex and not fully understood as of yet
- There are genomic markers facilitating crosses of the species barrier
- Several studies have produced lists of amino acids (aa) signatures associated with host specificity
- Given the complexity of the problem, we suspect that there might be a combinatorial use of aa signatures to specify host specificity.

Goals

Identify combinatorial markers of host specificity in addition to singular ones ...

Show if the markers are valid across subtypes ...

Methodology





UPPSALA
UNIVERSITET

Results and Discussion...

Predicted combinatorial signatures in H1N1 and H3N2 subtypes

Subtype	Protein	Combination	Host	Comb. Accuracy	Conditional Accuracy	Combinatorial Accuracy Gain
H1N1	HA	P200=P,P222=K	Avian	91.3%	(P200=P):Acc=15.9% (P222=K):Acc=54.0%	56.3%
H1N1	HA	P2=E,P222=K	Avian	96.2%	(P2=E):Acc=66.2% (P222=K):Acc=54.0%	36.1%
H1N1	HA	P137=A,P544=L	Avian	96.1%	(P137=A):Acc=83.8% (P544=L):Acc=14.7%	46.9%
H1N1	HA	P78=L,P435=V	Avian	97.1%	(P78=L):Acc=72.1% (P435=V):Acc=71.9%	25.1%
H1N1	M1	P121=T,P142=V	Avian	97.4%	(P121=T):Acc=11.7% (P142=V):Acc=48.8%	67.2%
H1N1	M1	P116=A,P121=T	Avian	94.9%	(P116=A):Acc=21.7% (P121=T):Acc=11.7%	78.2%
H1N1	M1	P101=R,P121=T	Avian	92.3%	(P101=R):Acc=20.1% (P121=T):Acc=11.7%	76.4%
H1N1	M1	P15=V,P121=T	Avian	94.7%	(P15=V):Acc=33.3% (P121=T):Acc=11.7%	72.2%
H1N1	M1	P30=D,P121=T	Avian	92.9%	(P30=D):Acc=22.1% (P121=T):Acc=11.7%	76.0%
H1N1	M1	P121=T,P207=S	Avian	92.5%	(P121=T):Acc=11.7% (P207=S):Acc=21.6%	75.8%
H1N1	M1	P15=I,P227=A	Human	98.7%	(P15=I):Acc=98.7% (P227=A):Acc=89.1%	4.8%
H1N1	M1	P166=A,P214=H	Human	99.0%	(P166=A):Acc=98.7% (P214=H):Acc=98.7%	0.3%

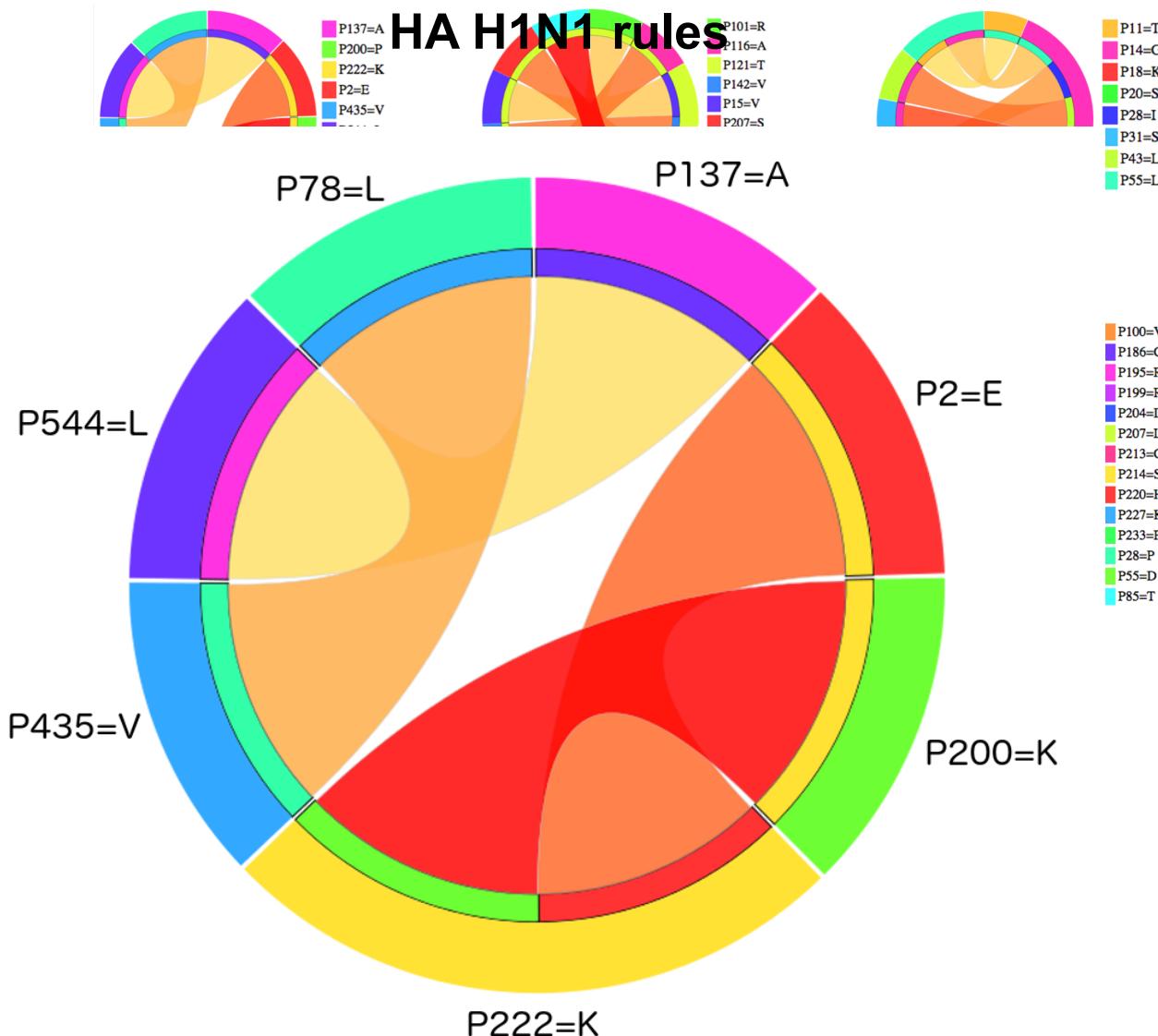
P200=P & P222=K : Avian=209, Human = 20., Acc=209 ÷ (209+20)=91.3%

P200=P: Avian=212, Human = 1119., Acc=212 ÷ (212+1119)=15.9%

P222=K: Avian=211, Human = 180., Acc=211 ÷ (211+180)=54%

Combinatorial Gain = Comb.Accurcay – averaged conditional accuracy
 $= 91.3\% - ((54+15.9) \div 2) = 56.3\%$

Combinations from H1N1 rules



Validity of host specificity signatures across H1N1 and H3N2 subtypes

H3N2 data – H1N1 models

Protein	Sensitivity	Specificity	MCC
HA	1	0	na
M1	1	0.895	0.941
M2	1	0.742	0.848
NA	1	0	na
NP	1	0.891	0.938
NS1	1	0.745	0.849
NEP	1	0.642	0.776
PA-X	0	1	na
PA	0.021	0.93	-0.11
PB1-F2	0.023	1	0.056
PB1	0.563	0.909	0.302
PB2	0.979	0.949	0.873

- HA, NA, PA-X, PA, PB1-F2 and PB1 proteins of the H3N2 viruses carry only H3N2 specific markers.
- M1, M2, NP, NS1, NEP(NS2) and PB2 protein of the H3N2 viruses, in addition to carrying H3N2 markers, also carry H1N1 markers.

Validity of host specificity signatures across H1N1 and H3N2 subtypes

H1N1 data – H3N2 models

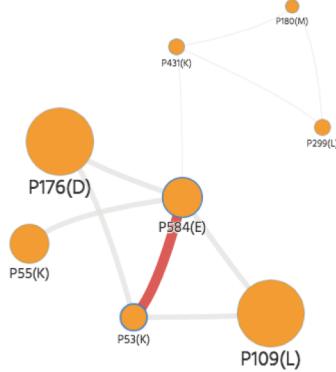
Protein	Sensitivity	Specificity	MCC
HA	0	na	na
M1	0.957	0.975	0.885
M2	0.987	0.766	0.804
NA	1	0	-0.004
NP	0.363	0.984	0.251
NS1	0.365	0.993	0.236
NEP	0.027	1	0.06
PA-X	0.202	0.982	0.224
PA	0.247	0.995	0.177
PB1-F2	0.991	0.804	0.831
PB1	0.992	0.877	0.888
PB2	0.956	0.951	0.788

- HA, NA, NP, NS1, NEP, PA-X and PA proteins of the H1N1 viruses carry only H1N1 specific markers.
- M1, M2, PB1-F2, PB2 and PB1 proteins of the H1N1 viruses, in addition to carrying H1N1 markers, also carry H3N2 markers.
- It follows, the surface proteins (HA and NA) are specifically tailored for each subtype
- M1, M2 and PB2 are the most conserved proteins from the host specificity markers point of view

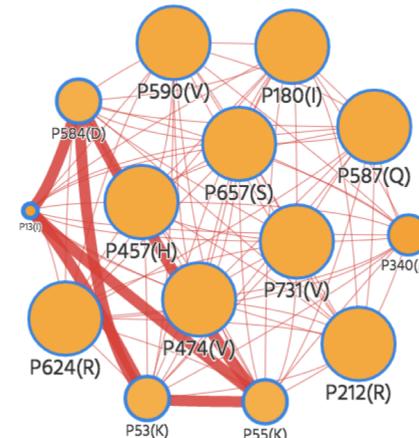
RNs for AIV adaptation to human host

Rule networks generated from the rule models:

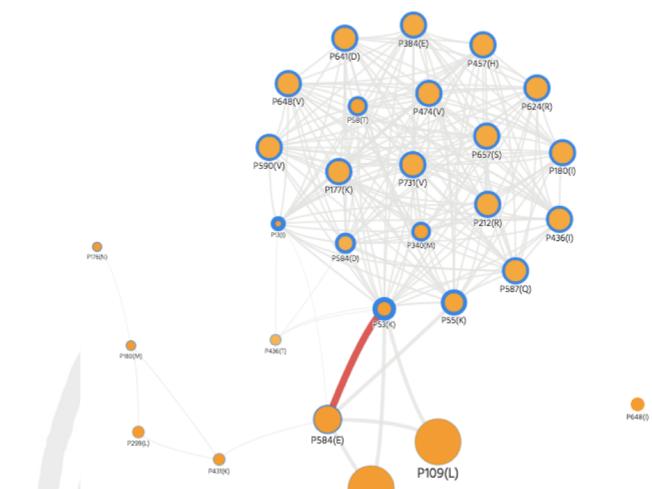
Avian host



Human-adapted



All classes



- How would you explain the remarkable difference in the complexity of the networks for avian versus human host?
 - Hint: evolution of the virus.

Conclusions

- Use of a combined MCFS-ROSETTA (feature selection – rule-based models) methodology
- Computationally identified the anticipated viral background necessary for high pathogenicity of IAVs
- Showed that high pathogenicity is subtype-specific while low pathogenicity is common across subtypes
- Identified combinatorial markers of host specificity in H1N1 and H3N2 viruses (in addition to singular markers)
- Showed the surface proteins (HA and NA) are highly customized for each subtype from the host specificity point of view.
- The matrix proteins (M1 and M2) and PB2 are the most conserved ones from the host specificity point of view.