SUPPLEMENTARY INFORMATION

Survival analysis of infected mice reveals pathogenic variations in the genome of avian H1N1 viruses

Zeynep A. Koçer[1], Yiping Fan[2], Robert Huether[2,†], John Obenauer[2, ‡], Richard J. Webby[1], Jinghui Zhang[2], Robert G. Webster[1*], Gang Wu[2*]

[1]Department of Infectious Diseases, Division of Virology, and [2]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, Tennessee, 38105, United States.

Current addresses:

[†] Ambry Genetics, 15 Argonaut, Aliso Viejo, California 92656, [‡] Digital Genomics LLC 2151 Dickens Place Drive, Southaven, Mississippi, 38672, United States

[*]Corresponding authors at: Department of Computational Biology, MS 1135, and Department of Infectious Diseases, Division of Virology, MS 330, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, United States

Phone: 901-595-8126 and 901-595-3414; FAX: 901-595-7100 and 901-595-8559; Email: gang.wu@stjude.org; robert.webster@stjude.org

Table of Contents

Supplementary Results 4

Polymerase proteins 4

Surface glycoproteins 7

Nucleoprotein 9

Matrix and nonstructural proteins 10

Supplementary Figures 12

Supplementary Figure S1. Polymorphic residues in the PB2-coding region. 12

Supplementary Figure S2. Polymorphic residues in the PB1-coding region. 13

Supplementary Figure S3. Polymorphic residues in the PB1-F2–coding region. 14

Supplementary Figure S4. Polymorphic residues in the PA-coding region. 15

Supplementary Figure S5. Polymorphic residues in the 61-aa C terminus of the PA-X–coding region. 16

Supplementary Figure S6. Polymorphic residues in the HA-coding region. 17

Supplementary Figure S7. Polymorphic residues in the NP-coding region. 18

Supplementary Figure S8. Polymorphic residues in the NA-coding region. 19

Supplementary Figure S9. Polymorphic residues in the (a) M1- and (b) M2-coding regions. 20

Supplementary Figure S10. Polymorphic residues in the NS1-coding region. 21

Supplementary Figure S11. Polymorphic residues in the NEP-coding region. 22

Supplementary Figure S12. Variants in the cap-binding domain of PB2 in avian H1N1 IAVs. 23

Supplementary Figure S13. Variants in the PA-PB1–binding site of avian H1N1 IAVs. 24

Supplementary Figure S14. Variants in the PA N terminus of avian H1N1 IAVs. 25

Supplementary Figure S15. The ribbon model of the structure of avian H1 hemagglutinin (HA). 26

Supplementary Figure S16. HA variants at the antigenic sites of avian H1N1 IAVs of Anseriformes or Charadriiformes origin. 27

Supplementary Figure S17. Observed variants mapped on the NA structure. 28

Supplementary Figure S18. Variants mapped on the NP structure. 29

Supplementary Figure S19. The NS1-PI3K interface. 30

Supplementary Tables 31

Supplementary Table S1. Pathogenicity and mortality rate of avian H1N1 IAVs with accompanying full-genome accession numbers 31

Supplementary Table S2. Polymorphic sites in the proteome of North American avian H1N1 IAVs in association with their pathogenicity in DBA/2J mice by residue effect after adjusting for host effect 32

Supplementary Table S3. Polymorphic sites in the proteome of North American avian H1N1 IAVs in association with their pathogenicity in DBA/2J mice by host effect after adjusting for residue effect 36

Supplementary Table S4. Polymorphic sites in the proteome of North American avian H1N1 IAVs affecting the pathogenicity of viruses in DBA/2J mice by host-residue interactions 37

Supplementary Table S5. Polymorphic residues at HA antigenic sites of avian H1N1 IAVs compared to human and pandemic H1N1 viruses 38

Supplementary Table S6. Distribution of the GSEV PDZ-binding motif among various IAV subtypes from different hosts in publicly available NS1 sequences 39

Supplementary Table S7. Distribution of the GSEV motif in publicly available NS1 sequences in terms of isolation year, host, region, and subtype 40

Supplementary References 42

Supplementary Results

For statistical robustness, we excluded the genetic variants appearing as a singleton in our dataset. Therefore, in addition to the polymorphic sites that are identified associated with pathogenicity by Cox proportional hazard model, we also investigated the structural and functional importance of all observed variants in the proteome of avian H1N1 IAVs. We also manually checked the genomes of avian H1N1 IAVs for the presence of previously known adaptation, virulence, transmission, and drug resistance markers, as well as the antigenic similarities between avian and pandemic H1N1 IAVs.

Polymerase proteins

We identified 12 variants in the cap-binding domain of PB2 (residues 318-483) (Supplementary Fig. S12a), six of which (V338I, R340K, E342D, R355K, V356I, and E358V) were clustered within 10 Å of m^7GTP (pre-mRNA cap analog) binding (Supplementary Fig. S12b). Interestingly, none of those six variants was observed in low-pathogenic strains (survival score ≥ 0.457, Supplementary Table 1). Variants surrounding m^7GTP are not anticipated to drastically affect protein stability or solubility, because they were of similar physicochemical properties, surface exposed, and did not directly interact with the m^7GTP. However, their positions in the structure indicated that they could slightly perturb and influence residues that line the m^7GTP-binding pocket. Four of the 10 most pathogenic strains (A/mallard/ALB/119/1998, A/mallard/MN/AI07-3100/2007, A/mallard/ALB/201/1998, and A/mallard/ALB/88/2004) (Supplementary Table S1) contain variants in their cap-binding domains at their methyl guanylyltransferase (MGT)-binding pocket. Additionally, all viruses with the variants in the cap-binding domain belong to avian IAVs of the Anseriformes origin. Previously, it was shown that arginine and lysine variants at residue 355 at the cap-binding

domain of PB2 are related to high pathogenicity of H5N1 IAVs in mice models, and Q355 is associated with low pathogenicity [1-3]. Therefore, the arginine or lysine variants at residue 355 of 31 avian H1N1 viruses may increase the virulence of the viruses in mice, although we cannot support this notion statistically due to the lack of statistical power (K355; n=1 and R355; n=30) variants in our sample set (Supplementary Fig. S1).

We detected six variants in the host-specific domain of PB2 (538-678 aa) of avian H1N1 viruses: four were surface exposed (D567A, G590S, V649I, and K660R), and two were buried (I553V and V667I). The variant V649I was identified in relation to increased pathogenicity in mice by residue effect using Cox proportional hazard model (Supplementary Table S2). Amino acid substitutions at residue 627 (glutamic acid to lysine) and 701 (aspartic acid to asparagine) are important in terms of mammalian adaptation of an avian virus [4,5]. All 31 North American avian H1N1 viruses carry avian-like E627 and D701 at those sites. However, the PB2 protein of the 2009 pandemic H1N1 did not acquire the E627K or D701N mutation. Instead, the acquisition of mutations at other sites, such as G590S and Q591R (SR polymorphism), have been identified as compensatory of E627K substitution that enhances the viral replication and polymerase activity in mammalian cells [6,7]. In our dataset of 31 avian H1N1 IAVs, we observed the avian-like GQ motif at positions 590-591 of PB2 in 27 viruses; however, we detected SQ motif in five of 31 viruses (A/mallard/ALB/88/2004, A/mallard/MN/Sg-00121/2007, A/mallard/ALB/267/1996, A/mallard/MN/AI07-3018/2007, and A/blue winged teal/LA/B228/1986) due to the G590S substitution. Thus, over the years, some avian H1N1 viruses have been one step closer to acquiring SR polymorphism, as seen in the 2009 pandemic virus.

We detected 16 polymorphic sites at the C-terminal region of the PB1-F2 transcript of North American avian H1N1 viruses, but we could not extrapolate the functional importance of those residues. Notably, the N66S substitution on PB1-F2 increases the virulence of H5N1 and 1918 H1N1 viruses [8]. Among the North American avian H1N1 IAVs, 29 of 31 carry S66; only

two (A/blue winged teal/LA/B228/1986 and A/shorebird/DE/274/2009) carry N66 (Supplementary Fig. S3). The latter two viruses were among the low pathogenic viruses with survival scores 0.457 and 0.571, respectively (Supplementary Table S1). So, S66 in the PB1-F2 of avian H1N1 IAVs might explain the increased virulence of most of the viruses in the mouse model.

Using the available crystal structures of PA, we structurally investigated the variants for importance, with the majority of variants (26 of 33) being detected in the C-terminal PA-PB1–binding site (Supplementary Figs. S4, S13). Most of the residues were present on the protein surface, and seven variants (V323I, P332S, A337T, V542I, L543I, I545V, and V565I) clustered within 10Å of each other (Supplementary Fig. S13a). Variants V323I and I545V in PA coevolved in most of the viruses, and given their location in the core of the protein might compensate for each other to stabilize the molecule (Table 2 and Supplementary Fig. 13a). Human-like signatures D382 and N409 were observed in several avian H1N1 IAVs [9]. Among the 31 avian H1N1 viruses, four (A/mallard/MN/AI07-3127/2007, A/mallard/MO/466554-14/2008, A/mallard/MN/AI07-3140/2007, and A/shorebird/DE/274/2009) carried the human-like D382 on the PA protein, and four others (A/shorebird/DE/300/2009, A/shorebird/DE/324/2009, A/shorebird/DE/170/2009, and A/shorebird/DE/318/2009) carried human-like N409. The increased length of N409 might interact weakly with PB1, which would not occur with the smaller S409 (Supplementary Fig. S13b). Therefore, S409N might provide an additional stabilizing contact between PA and PB1. Overall, five of the six avian H1N1 viruses of Charadriiformes origin had one or the other human-like residue in their PA that might increase their pathogenicity in mammalian models.

Seven variants were detected in the N-terminal endonuclease domain of PA (residues 1-197) (Supplementary Fig. S4). None of the variants were located directly in the catalytic site, but two were adjacent (I118V and T162I). Structurally, these variants were anticipated to affect the

stability of the secondary shell, but they did not appear to drastically disrupt the orientation of residues lining the binding pocket (Supplementary Fig. S14).

Surface glycoproteins

Avian H1N1 IAVs have seven conserved glycosylation sites at residues 22, 23, 40, 104, 304, 498, and 557 in their HA. It is not clear whether the viruses use both glycosylation sites at residues 22 and 23, or if residue 23 is only used alternatively. Using the available crystal structures, we modeled 40 variants: 32 in the head domain and eight in the stalk domain (Supplementary Fig. S15a). Ten of the 25 variants in the globular head (residues 59-292) were observed at antigenic sites, and 15 were at nonantigenic sites. All H1N1 viruses isolated from Charadriiformes displayed four variants (E132Q, A137S, T169I, and K236Q) that co-occurred. The T169I substitution was also observed in eight viruses of Anseriformes origin. Four of the 25 variants in the globular head were in the receptor-binding site (I168L, T169I, L208I, and I266V) (Supplementary Fig. S15b). Residues 169, 208, and 266 would influence carbohydrate interactions, and residue 168 would influence residues on the surface of the Sa antigenic site. The T169I variant is the only one seen in almost half of the viruses. The side chain faces the binding site altering carbohydrate-protein interactions.

We determined the distribution of the variants at the antigenic sites (Supplementary Fig. S16) for several selected viruses from the most pathogenic group in DBA/2J mice. Eleven of the variants at the nonantigenic sites (60, 132, 137, 138, 169, 188, 213, 228, 236, 279, and 286) were surface exposed, and four (125, 168, 266, and 289) were buried. We did not observe a defined pattern in the distribution of the variants at nonantigenic sites of the H1N1 viruses of Anseriformes origin. However, we observed localized and enriched variants at the nonantigenic sites of H1N1 viruses of Charadriiformes origin (Supplementary Fig. S16f).

We examined the residues at the HA antigenic sites based on the variations observed between human H1N1 IAVs throughout the years and pandemic H1N1 viruses (1918 and 2009) [10] and compared those with avian H1N1 IAVs. Of the 50 residues found at the antigenic sites of HA (Sa, Sb, Ca, and Cb), 21 residues appeared to be conserved among human, avian, and pandemic H1N1 IAVs. North American avian H1N1 IAVs were similar to both 1918 and 2009 pandemic viruses at 13 residues (K170, G172, S174, and K177 in Sa; G187 in Ca; T201, Q206, S207, L208, Q210, and N211 in Sb; and T89 and A90 in Cb), while human H1N1 viruses showed variations throughout the years. We observed similarities to either pandemic virus at some other positions. For instance, avian H1N1 IAVs carry five additional residues identical to those in the 1918 virus (Y155, S159, N185, and K222 in Ca; and L88 in Cb) but not those in human H1N1 or 2009 pandemic virus. Similarly, these viruses carry residues identical to those of the 2009 pandemic virus (N173 in the Sa antigenic site in H1N1 IAVs of Anseriformes and Charadriiformes origin, and P154 in the Ca antigenic site only in H1N1 IAVs of Charadriiformes origin). On the basis of 10 polymorphic residues identified at the HA antigenic sites of avian H1N1 IAVs, we did not observe any significant difference between H1N1 viruses of Anseriformes and Charadriiformes origin in terms of their antigenic similarity to either pandemic virus (Supplementary Table S5).

Seven glycosylation sites were detected in the NA protein of North American avian H1N1 viruses at residues 50, 58, 63, 68, 88, 146, and 235. Six of these were observed in all 31 H1N1 viruses, but the glycosylation site at residue 68 was lost in two viruses (A/mallard/MN/AI07-3127/2007 and A/mallard/MN/AI07-3140/2007) due to a S70N substitution. Five of the polymorphic sites (260, 262, 416, 449, and 450) were surface exposed and observed near the interface between monomers (Supplementary Fig. S17b).

The effect of the variants on oligomerization was most likely small due to the variable nature of the positions. Two variable residues (385 and 386) were positioned in a loop that

coordinates Ca^{+2} binding (Supplementary Fig. S17c) which is important for stabilizing the NA protein [11]. Residues in position 386 help coordinate Ca^{+2} via backbone interactions, indicating that small changes in the side chain can be tolerated. One avian H1N1 virus, A/pintail/ALB/210/2002, contained the R430K variant. This variant was not a drastic electrostatic or stereochemical change, but it does occur adjacent to the residues on the "150 loop", which forms a cavity adjacent to the catalytic pocket and is being explored as a target for antiviral inhibitors [11,12]. All the other variants (surface-exposed or buried) were similar in their physiochemical changes and did not appear to affect the function of the protein. We identified three variable sites (220, 221, and 430) surrounding the active site; however, none of those positions were directly involved in the formation of the active-site pocket.

All 31 North American avian H1N1 viruses in our study displayed H275 (274 in N2 numbering) and N295 (294 in N2 numbering); therefore, these residues are probably sensitive to NA inhibitors. Stalk deletion in NA is known to facilitate the adaptation of avian IAVs to mammals. However, we did not detect deletion in the NA stalk of avian H1N1 IAVs.

Nucleoprotein

We detected 22 polymorphic sites in the NP-coding region (Supplementary Fig. S7). Of those, only two (143 and 478) were associated with decreased pathogenicity in two H1N1 viruses of Anseriformes origin after adjusting the residue effect (adjusted *P* (FDR) = 9.50E-03; Supplementary Table S3).

The NP protein contains a perpendicular helix that divides the protein into two regions, the head and body. RNA is thought to bind along the perpendicular helix of NP (Supplementary Fig. S18), which contains five (23%) of the NP variants (D128E, L133I, L136M, L143M, and A146S) that are in viruses with lower pathogenicity. Substitutions at two of these residues

(L136M in A/blue winged teal/LA/B228/1986 and A/mallard/MN/AI07-3140/2007 and L143M in A/mallard/MN/Sg-00627/2008 and A/mallard/MN/Sg-00628) appeared to enlarge the side chain slightly, which might affect the stability of the helix and RNA binding. The variant S478P is surface exposed and present at the C terminus of the protein (data not shown).

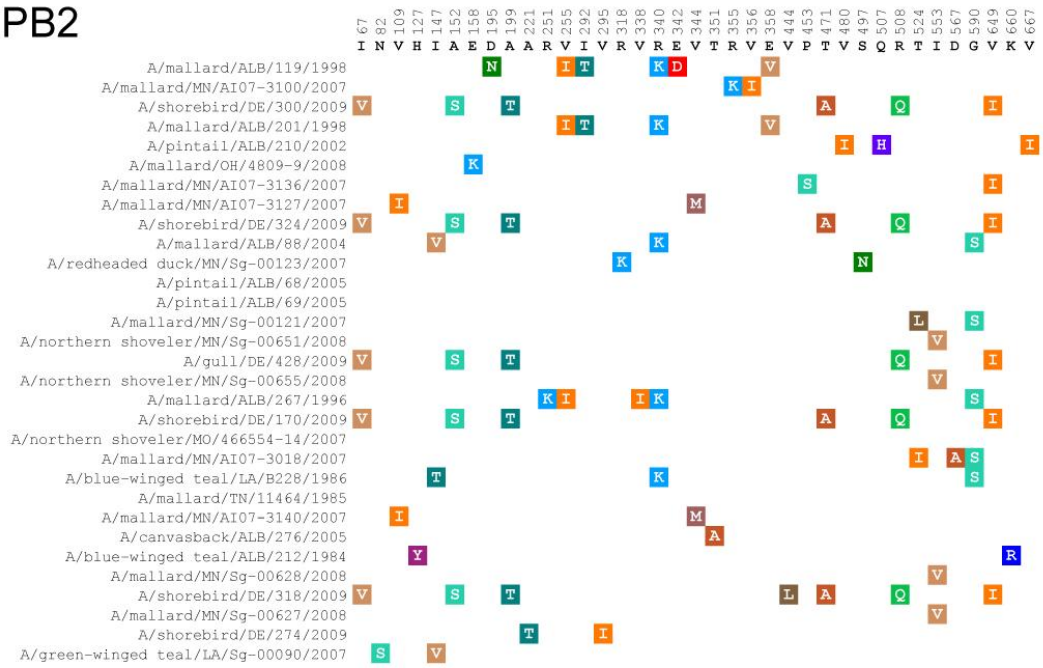
Matrix and nonstructural proteins

The variants in M1 (S70N and N85D) were not located on the nuclear-localization signal (residues 101-105) or on the NEP-interaction interface. Both were solvent exposed, and their effect on function is unknown. V27A, A30S, and S31N substitutions were previously described as amantadine resistance markers in M2 [13]. None of the 31 North American avian H1N1 viruses in our study contained these markers; therefore, they are predicted to be sensitive to amantadine.

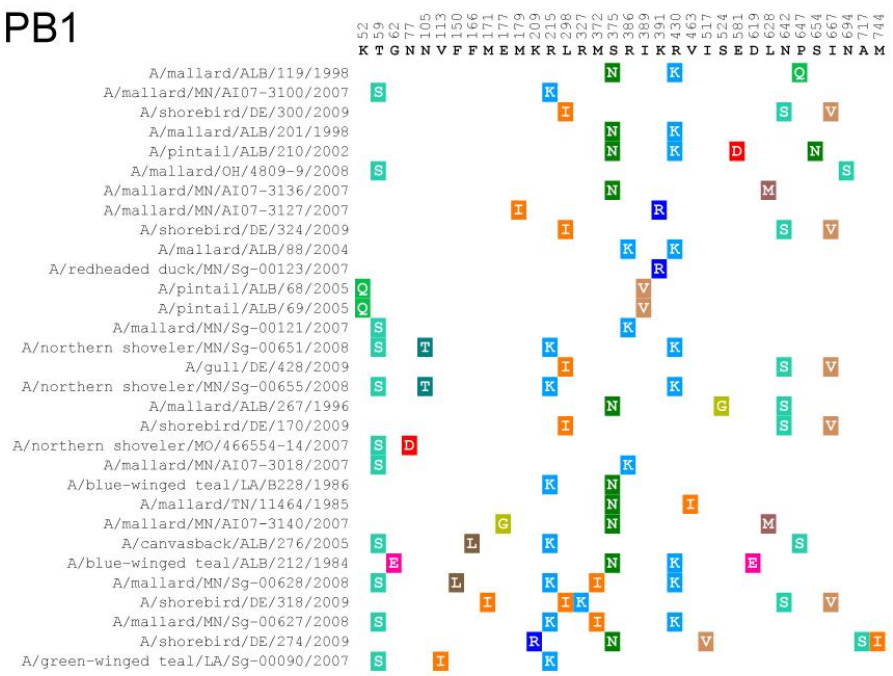
In terms of the structural and functional importance of each substitution in NS1 on the pathogenicity of viruses, our main focus was on the effector domain (residues 74-230), because it affects various key functions by binding the host cell's proteins [12]. Key variants were identified at the PDZ-ligand motif, the nuclear-export signal, and the dimer interface. On the basis of interface interactions between NS1 and PI3K domain, we identified six interactions (Y89, R96, S99, D101, R118, and E159) that participate in hydrogen-bond interactions with the PI3K-inhibitory domain (Supplementary Fig. S19a), two of which (D101E and R118K) were present in all pathogenic viruses. Six of the remaining 11 hydrophobic interactions were observed variants in avian H1N1 IAVs (S87, L95, M98, I145, L146, and S161). The variant R118K might allow closer packing of the PI3K helix to the NS1 protein (Table 2 and Supplementary Fig. S18) possibly by a decrease in the steric size at residue 118 due to the arginine-to-lysine substitution.

We observed E227G substitution in NS1 of three pathogenic H1N1 IAVs of Charadriiformes origin. We then assessed all 195 full-length NS1 sequences with a GSEV motif that were on the NCBI Influenza Virus Resources database to identify the hosts and virus subtypes in which the GSEV motifs were located at the 3' end of the ORF. Of the NS1 sequences with a GSEV motif, 78.5% were detected in avian IAVs; 35.9% belonged to the H5N1 subtype (Supplementary Table S6); and 29.4% were isolated from highly pathogenic avian H5N1 viruses. The circulation of a GSEV PDZ-binding domain in IAVs of various regions and hosts over the years (Supplementary Table S7) allows us to speculate that the GSEV motif appeared in the early 1980s in Eurasian swine with transitory occurrence in avian subtypes through the years until 2006. That year, GSEV was acquired mainly by Charadriiformes from Eurasia and North America, as well as by a Eurasian H5N1 virus that was highly pathogenic. In 2007, GSEV was circulating in highly pathogenic H5N1 viruses that were isolated from birds, humans, and pikas. In 2009, GSEV circulation peaked in the IAVs of Charadriiformes origin, including the three H1N1 viruses with a GSEV motif investigated in this study. Today, GSEV still circulates in IAVs of Charadriiformes origin and in some H9N2 viruses from Eurasia [14].

Supplementary Figures



Supplementary Figure S1. Polymorphic residues in the PB2-coding region. These positions were based on the sequences used in this study.

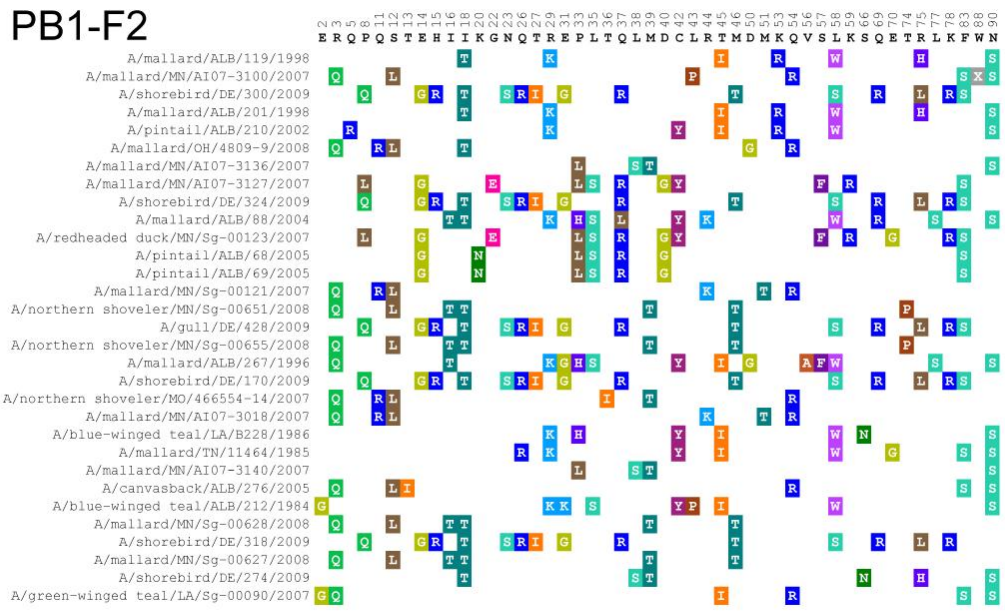


261

262 **Supplementary Figure S2.** Polymorphic residues in the PB1-coding region. These positions

263 were based on the sequences used in this study.

264

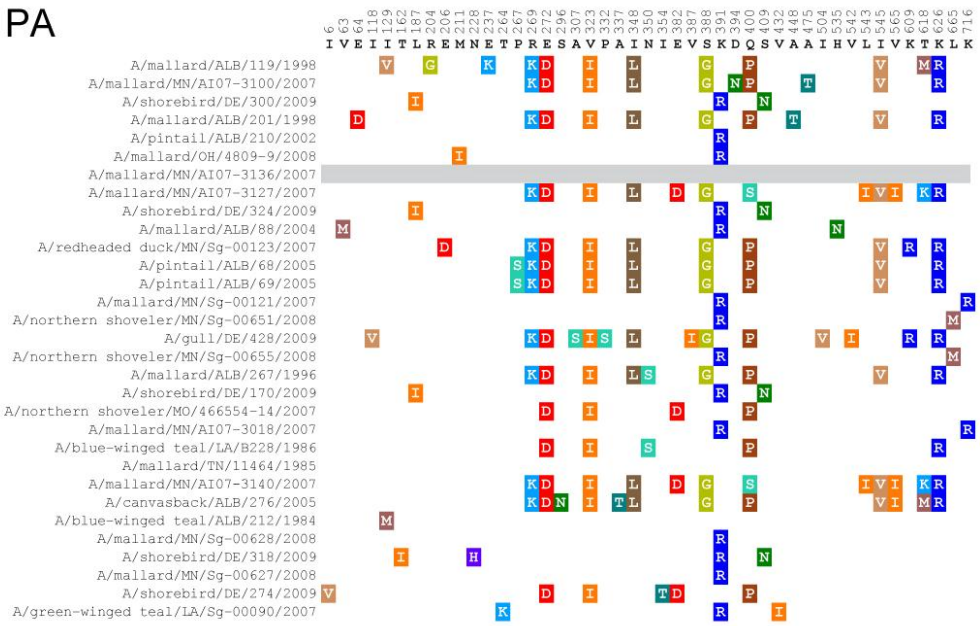


265

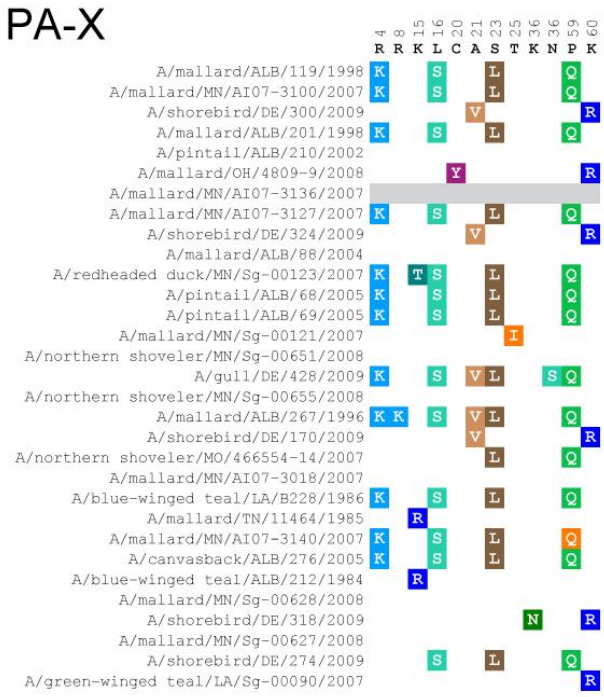
266 **Supplementary Figure S3.** Polymorphic residues in the PB1-F2–coding region. These

267 positions were based on the sequences used in this study.

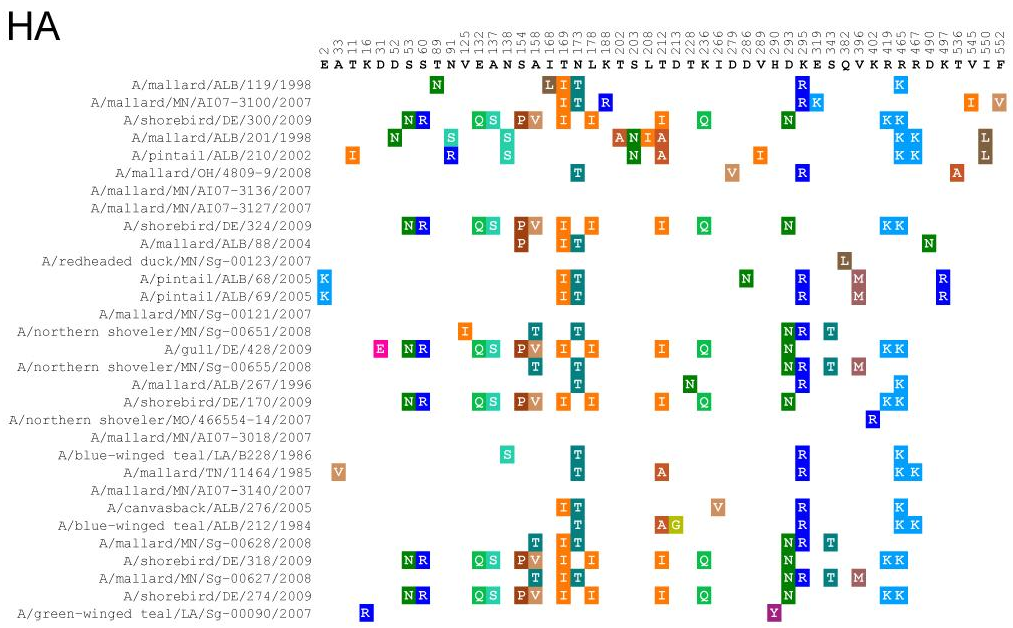
268



Supplementary Figure S4. Polymorphic residues in the PA-coding region. The PA sequence for A/mallard/MN/AI07-3136/2007 (shown in gray) is excluded due to high number of ambiguous positions. These positions were based on the sequences used in this study.



Supplementary Figure S5. Polymorphic residues in the 61-aa C terminus of the PA-X–coding region. The PA sequence for A/mallard/MN/AI07-3136/2007 (shown in gray) is excluded due to high number of ambiguous positions. These positions were based on the sequences used in this study.

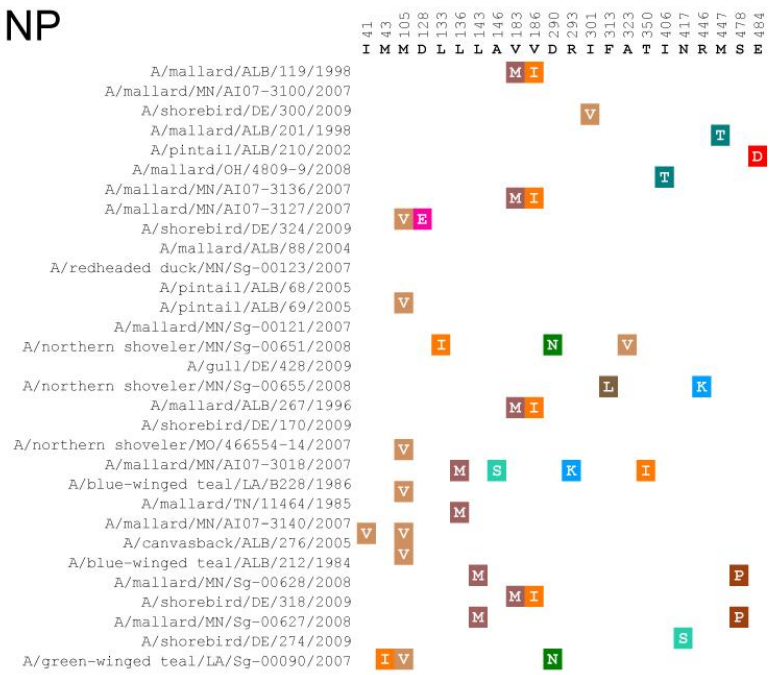


280

281 **Supplementary Figure S6.** Polymorphic residues in the HA-coding region. These positions

282 were based on the sequences used in this study.

283

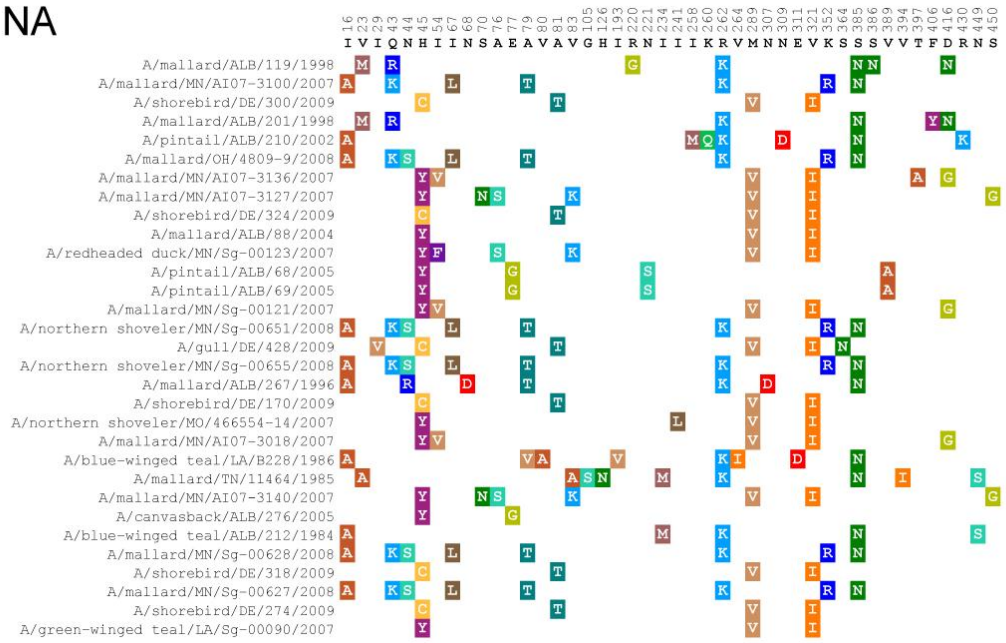


284

285 **Supplementary Figure S7.** Polymorphic residues in the NP-coding region. These positions

286 were based on the sequences used in this study.

287

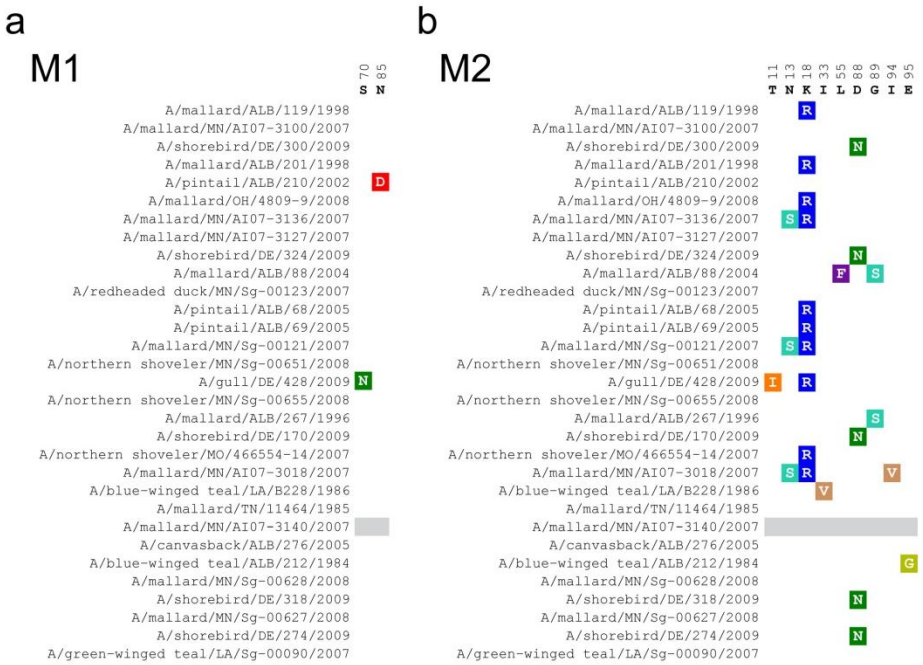


288

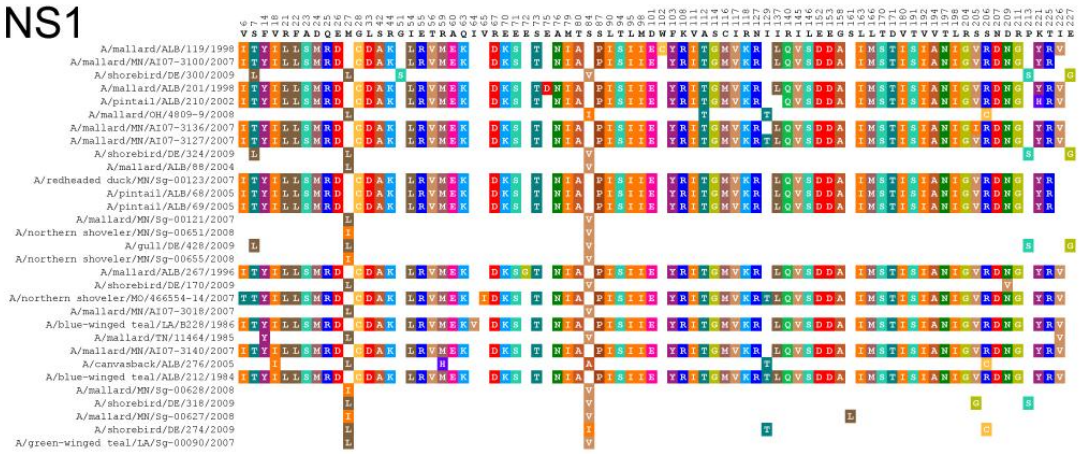
289 **Supplementary Figure S8.** Polymorphic residues in the NA-coding region. These positions

290 were based on the sequences used in this study.

291



Supplementary Figure S9. Polymorphic residues in the **(a)** M1- and **(b)** M2-coding regions. The M sequence for A/mallard/MN/AI07-3140/2007 (shown in gray) is excluded due to high number of ambiguous positions. These positions were based on the sequences used in this study.

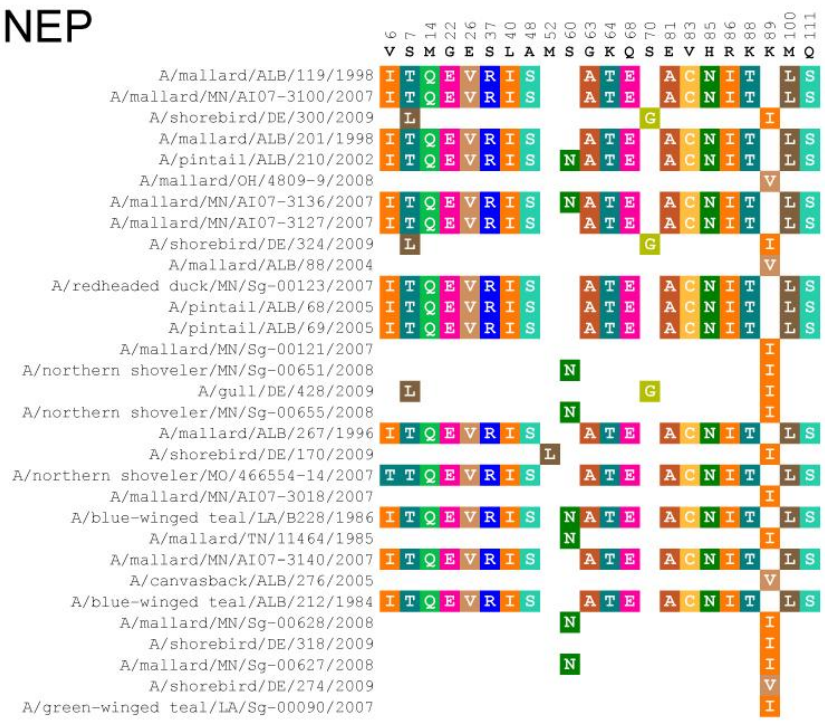


299

300 **Supplementary Figure S10.** Polymorphic residues in the NS1-coding region. These positions

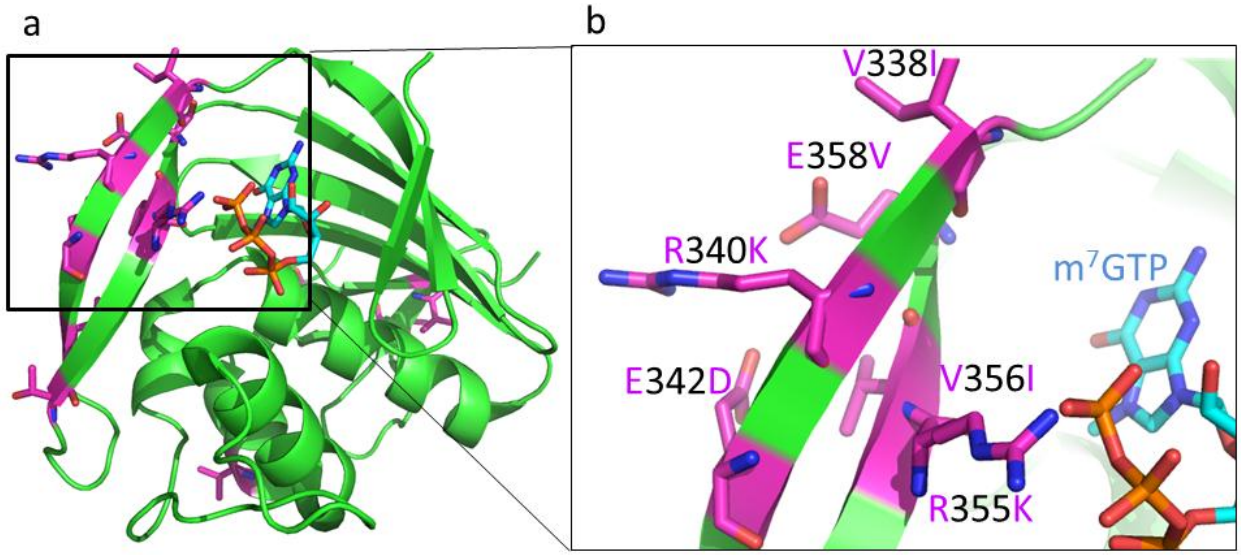
301 were based on the sequences used in this study.

302

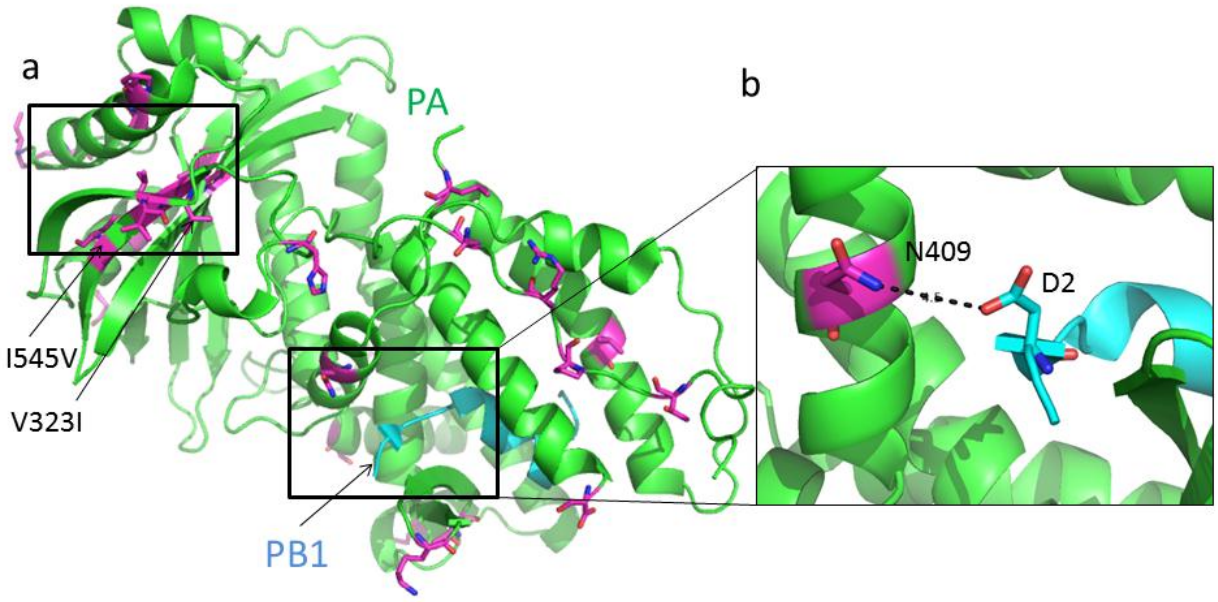


303

304 **Supplementary Figure S11.** Polymorphic residues in the NEP-coding region. These positions
305 were based on the sequences used in this study.
306



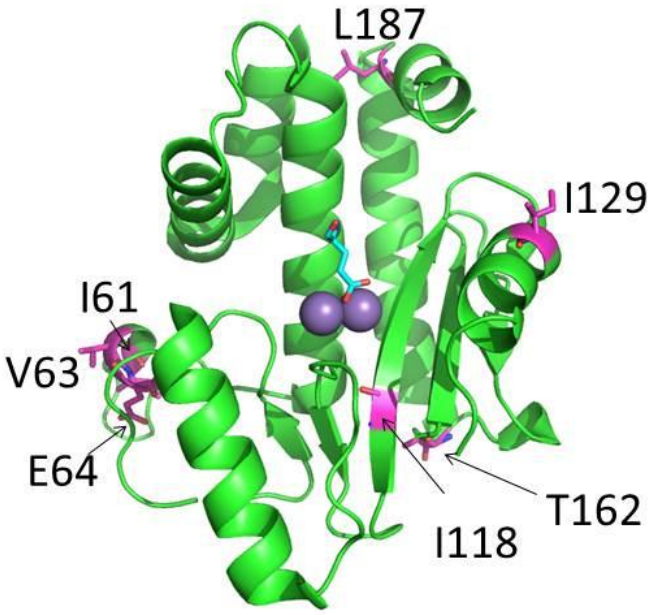
Supplementary Figure S12. Variants in the cap-binding domain of PB2 in avian H1N1 IAVs. **(a)** Ribbon models of the structure of PB2 (green; [PDB: 2VQZ]) with 12 variants (magenta sticks) located in the cap-binding domain and bound to the pre-mRNA–cap analog (teal stick). **(b)** Six variants (I338, E358, R340, E342, R355, and V356) in the PB2 protein that cluster within approximately 10 Å of the bound pre-mRNA–cap analog (m^7GTP). These positions were based on the sequences used in this study.



315

316 **Supplementary Figure S13.** Variants in the PA-PB1–binding site of avian H1N1 IAVs. **(a)**
317 Model of the C-terminal domain of PA (green; [PDB: 3CM8]) bound to the N–terminal peptide
318 (teal) of PB1. Observed variants are shown (magenta sticks). Of note, a cluster (<10 Å) of
319 seven variants (V323I, P332S, A337T, V542I, L543I, I545V, and V565I) are observed (black
320 boxes). **(b)** Region around variant N409 on PA and D2 on PB1. Potential interaction is shown as
321 a black dotted line. These positions were based on the sequences used in this study.

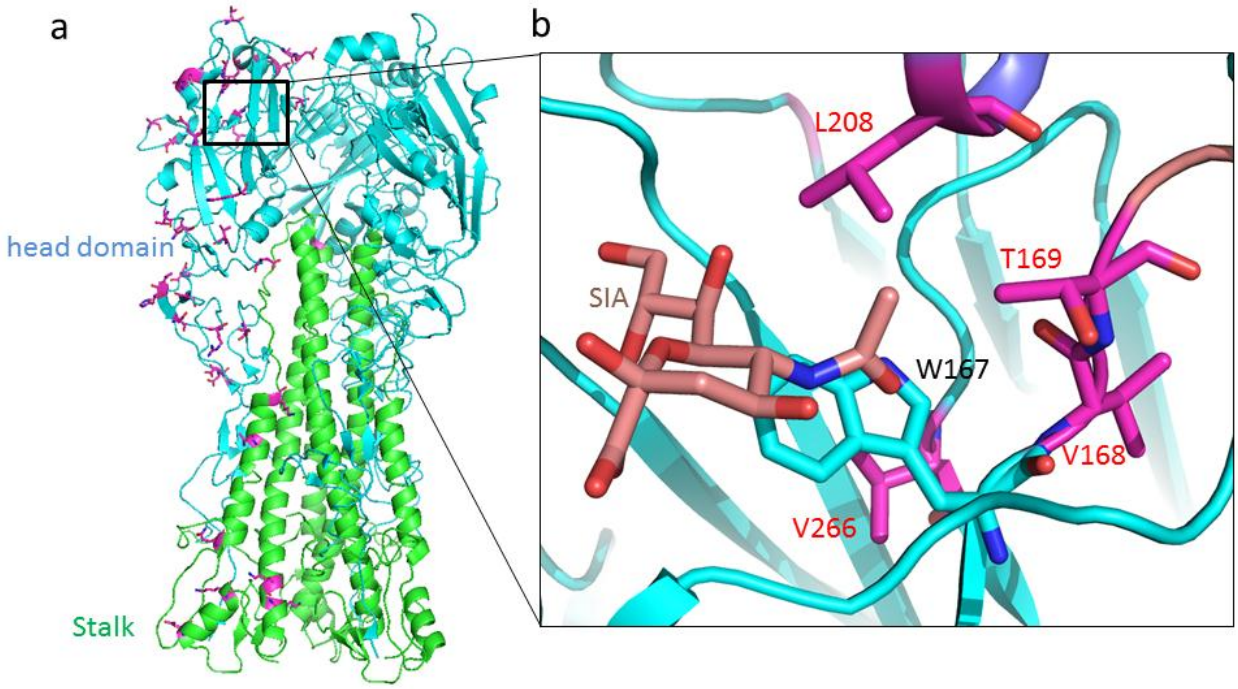
322



323

324 **Supplementary Figure S14.** Variants in the PA N terminus of avian H1N1 IAVs. The seven
325 variants are indicated (magenta sticks) and labeled within the PA endonuclease domain (green;
326 [PDB: 2W69]). Manganese ions are shown as purple spheres, and the bound substrate is
327 shown as a teal stick. These positions were based on the sequences used in this study.

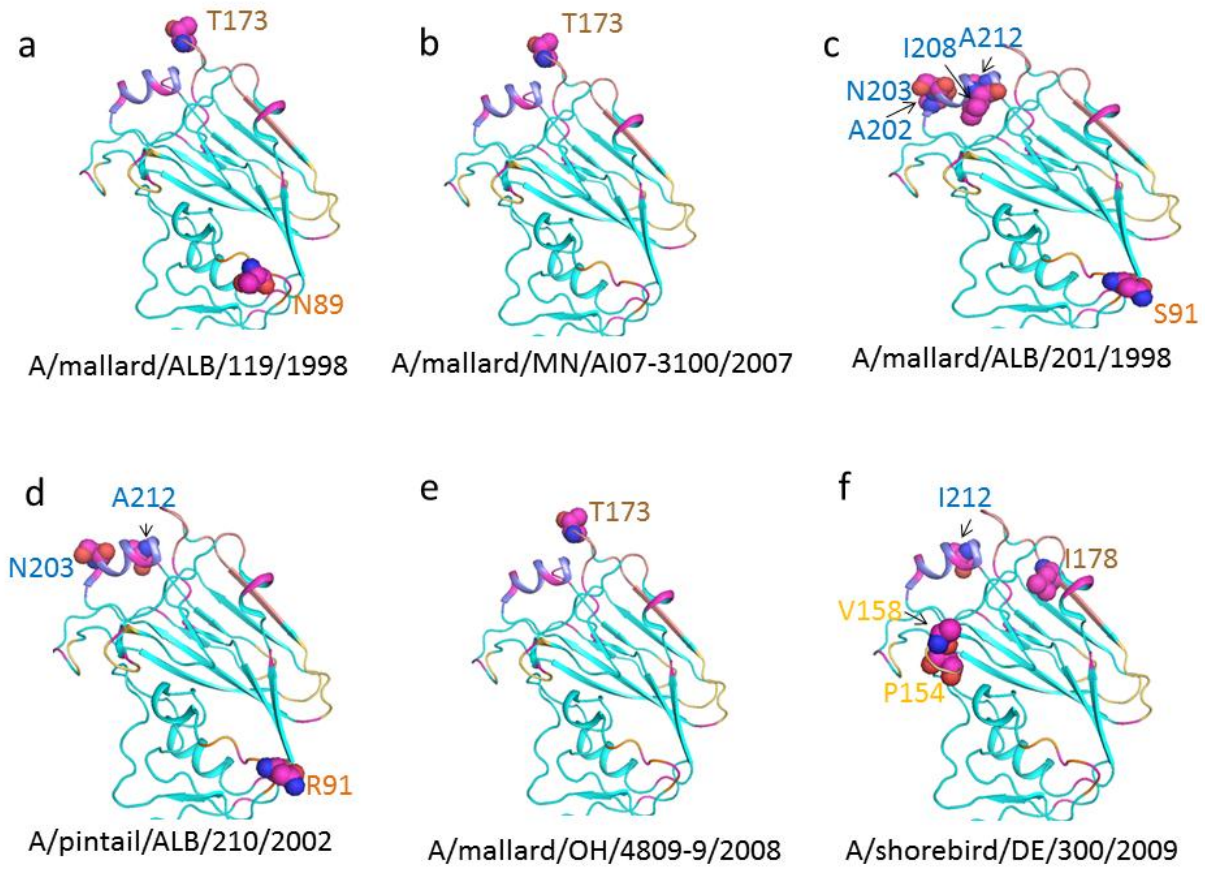
328



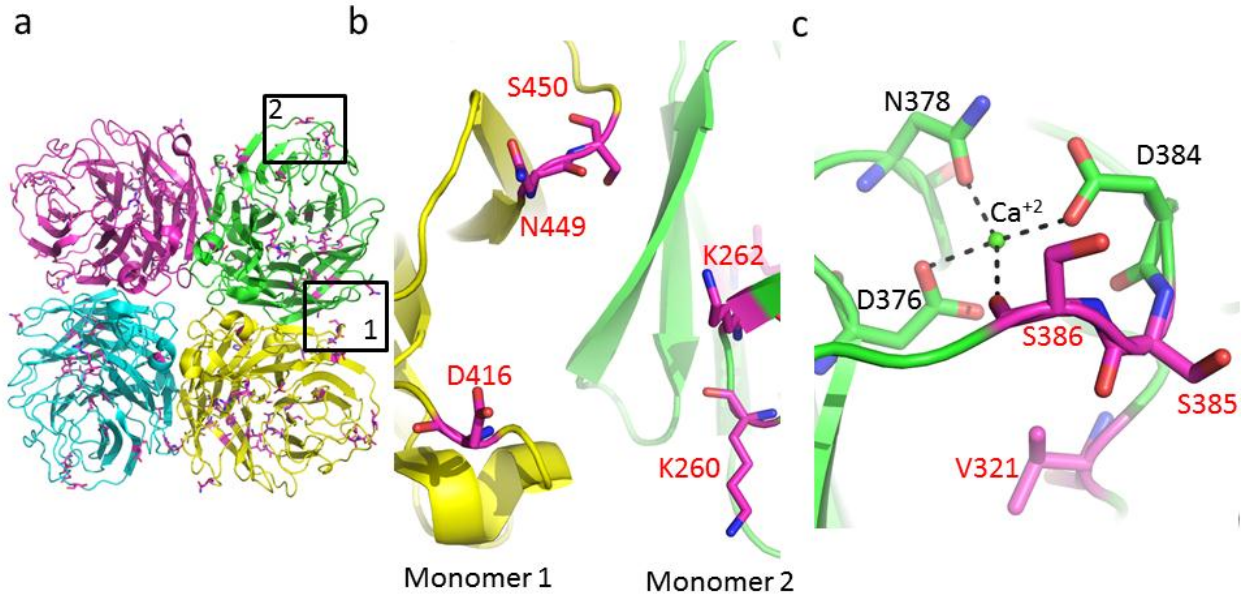
329

330 **Supplementary Figure S15.** The ribbon model of the structure of avian H1 hemagglutinin (HA).
331 **(a)** Structural tetramer [PDB: 2WRH] of the HA protein in which the head (teal) and stalk (green)
332 domains are labeled. Observed variants are indicated (magenta sticks) on only the monomer of
333 the head and stalk domains. The carbohydrate receptor–binding site is indicated by a black box.
334 **(b)** Region surrounding the receptor-binding site includes variant L208 in the Sb antigenic site,
335 and variants T169, V168, and V266 are nonantigenic. Both carbohydrates (SIA) are visualized
336 as a peach stick. The locations of variants discussed in the text are labeled in red. These
337 positions were based on the sequences used in this study.

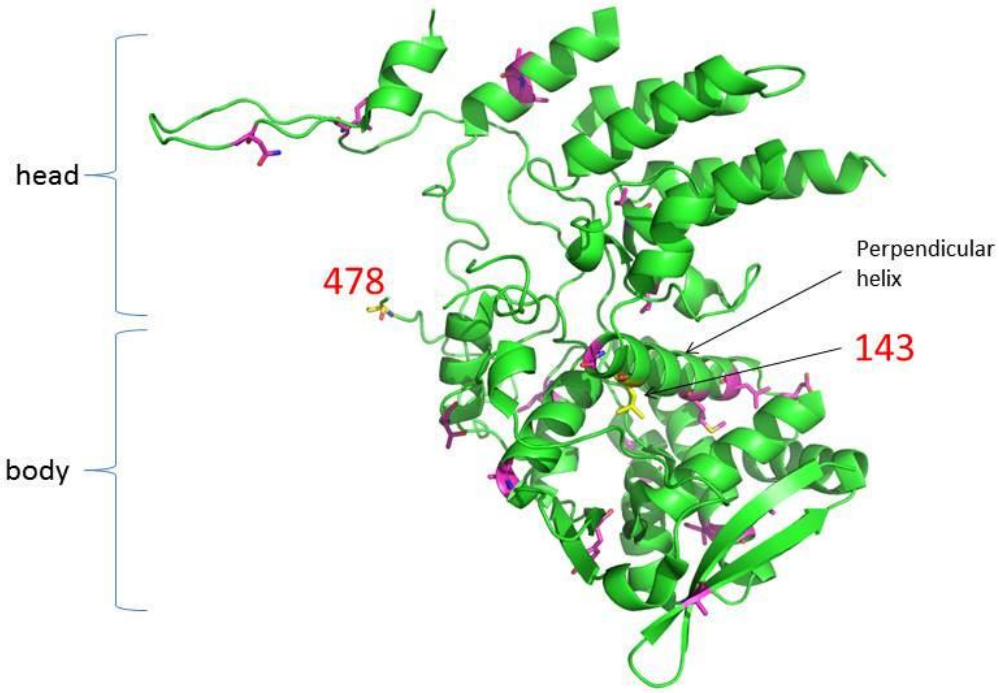
338



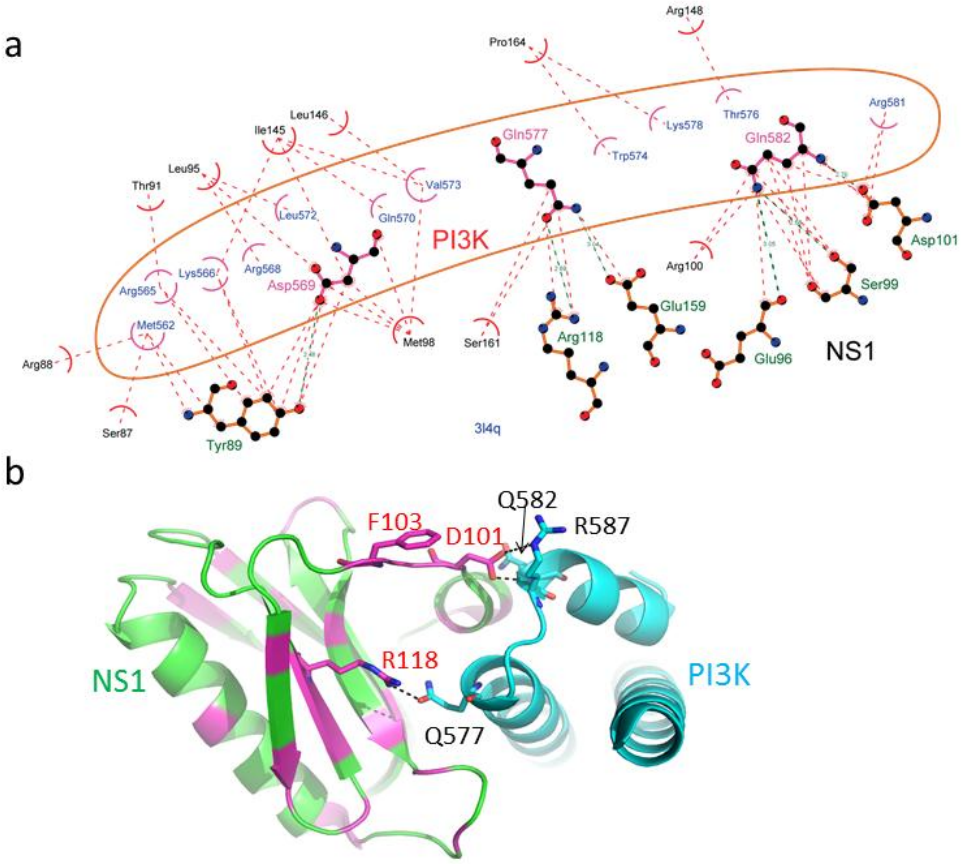
Supplementary Figure S16. HA variants at the antigenic sites of avian H1N1 IAVs of Anseriformes or Charadriiformes origin. Ribbon models indicate the variant epitopes in HA antigenic sites (colors and structures are the same as those in Fig. 3). The residues associated with increased pathogenicity are labeled (magenta space-filled spheres) on the HA of the most pathogenic avian H1N1 IAVs of Anseriformes origin: **(a)** A/mallard/ALB/119/1998, **(b)** A/mallard/MN/AI07-3100/2007, **(c)** A/mallard/ALB/201/1998, **(d)** A/pintail/ALB/210/2002, and **(e)** A/mallard/OH/4809-9/2008. **(f)** Those associated with increased pathogenicity on the HA of the most pathogenic H1N1 IAV, of Charadriiformes origin, A/shorebird/DE/300/2009, are also noted. A small difference in variant epitopes was observed between the Ans and Char IAVs. These positions were based on the sequences used in this study.



Supplementary Figure S17. Observed variants mapped on the NA structure. **(a)** Variants observed within the biological tetramer of protein NA [PDB: 2HTY]. Interesting variant locations are labeled (1: dimer interface, 2: calcium-binding site) and boxed in black. **(b)** Variants on the protein-protein interface in the tetramer between monomers 1 (yellow) and 2 (green) are indicated by purple sticks and red text. **(c)** Variants observed in a Ca^{2+}-binding site are indicated by green sticks and black text [PDB: 3B7E]. Coordinating interactions with Ca^{2+} are represented as black dotted lines. These positions were based on the sequences used in this study.



Supplementary Figure S18. Variants mapped on the NP structure. Crystal structure of the NP protein (green; [PDB: 3RO5]) with variant positions (magenta sticks). The protein can be subdivided into two domains (the head and body) separated by a perpendicular helix. Variants at positions 143 and 478 are associated with decreased pathogenicity (yellow sticks). These positions were based on the sequences used in this study.



370

371 **Supplementary Figure S19.** The NS1-PI3K interface. **(a)** Interface interactions between NS1
372 (outside the circle) bound to the PI3K inhibitory domain (inside the circle; [PDB: 3L4Q]). Amino
373 acids shown as whole atoms are involved in hydrogen bonding, and those shown as spiked
374 semicircles participate in hydrophobic interactions. Green dotted lines connect amino acids
375 involved in hydrogen bonding, and red dotted lines connect those involved in hydrophobic
376 interactions. The six residues involved in hydrogen bonding are Y89, R96, S99, D101, R118,
377 and E159. **(b)** Ribbon model of the interaction between NS1 and the PI3K inhibitory domain.
378 Variants observed in NS1 (magenta) and those in NS1 that are discussed in the text (red –sticks
379 and text) are labeled. Variants in PI3K are indicated in black text, and interactions between the
380 residues are indicated in black-dotted lines. These positions were based on the sequences used
381 in this study.

Supplementary Tables

Supplementary Table S1. Pathogenicity and mortality rate of avian H1N1 IAVs with accompanying full-genome accession numbers

Influenza A virus isolate	Mortality (%)	Survival score[a]	Pathogenicity Index (PI)[b]	Accession numbers[c]
A/mallard/ALB/119/1998	100	0.286	4	KF424175-182
A/mallard/MN/AI07-3100/2007	100	0.286	4	KF424015-022
A/shorebird/DE/300/2009	100	0.309	4	KF424127-134
A/mallard/ALB/201/1998	100	0.320	4	KF424047-054
A/pintail/ALB/210/2002	100	0.343	4	KF424111-118
A/mallard/OH/4809-9/2008	100	0.343	4	KF424151-158
A/mallard/MN/AI07-3136/2007	100	0.343	4	KF424167-174
A/mallard/MN/AI07-3127/2007[d]	100	0.343	4	KF424207-214
A/shorebird/DE/324/2009	100	0.377	4	KF424079-086
A/mallard/ALB/88/2004	100	0.366	4	KF424039-046
A/redheaded duck/MN/Sg-00123/2007	100	0.400	3	KF424191-198
A/pintail/ALB/68/2005[d]	100	0.400	3	KF424215-222
A/pintail/ALB/69/2005	100	0.400	3	KF424103-110
A/mallard/MN/Sg-00121/2007[d]	100	0.400	3	KF424239-246
A/northern shoveler/MN/Sg-00651/2008	100	0.400	3	KF424031-038
A/gull/DE/428/2009	100	0.400	3	KF424023-030
A/northern shoveler/MN/Sg-00655/2008[d]	100	0.400	3	KF424247-254
A/mallard/ALB/267/1996	100	0.400	3	KF424095-102
A/shorebird/DE/170/2009	100	0.400	3	KF424071-078
A/northern shoveler/MO/466554-14/2007	100	0.411	3	KF424119-126
A/mallard/MN/AI07-3018/2007[d]	100	0.423	3	KF424199-206
A/blue-winged teal/LA/B228/1986	100	0.457	3	KF424183-190
A/mallard/TN/11464/1985	100	0.457	3	KF424143-150
A/mallard/MN/AI07-3140/2007[d]	100	0.457	3	KF424255-262
A/canvasback/ALB/276/2005	100	0.457	3	KF424159-166
A/blue-winged teal/ALB/212/1984[d]	100	0.457	3	KF424231-238
A/mallard/MN/Sg-00628/2008	100	0.457	3	KF424135-142
A/shorebird/DE/318/2009[d]	100	0.514	2	KF424223-230
A/mallard/MN/Sg-00627/2008	100	0.571	2	KF424055-062
A/shorebird/DE/274/2009	100	0.571	2	KF424063-070
A/green-winged teal/LA/Sg-00090/2007	20	0.766	1	KF424087-094

[a]The survival score is calculated as a function of percent survival by time [0.8 × (survival AUC/maximum AUC) [15]. Individual mortality by time data for each mouse in each group is available upon request.

[b]PIs were classified as follows: 4, most pathogenic; 3, moderately pathogenic; 2, low pathogenic; 1, least pathogenic [15].

[c]The GenBank accession numbers for the whole genome of each avian H1N1 IAVs (PB2, PB1, PA, HA, NP, NA, M, NS, respectively).

Supplementary Table S2. Polymorphic sites in the proteome of North American avian H1N1 IAVs in association with their pathogenicity in DBA/2J mice by residue effect after adjusting for host effect

coding region	position	aa distribution in *Ans*Char	totalGroup	nonreferenceResidue	AAcoef	exp.AAcoef.	AApvalue	aaFDR
NA	416	D*19*6,G*3*0,N*2*0	3	N	2.506931	12.26722	4.20E-11	4.24E-09
PB1-F2	53	K*21*6,R*3*0	2	R	2.201197	9.035823	1.87E-11	4.24E-09
PB1-F2	75	H*2*1,L*0*5,R*22*0	3	R	-2.47556	0.0841161	5.11E-11	4.24E-09
PB2	292	I*22*6,T*2*0	2	T	2.424705	11.2989	1.13E-10	5.63E-09
PB2	358	E*22*6,V*2*0	2	V	2.424705	11.2989	1.13E-10	5.63E-09
NA	43	K*6*0,Q*16*6,R*2*0	3	R	2.632694	13.9112	2.27E-10	9.41E-09
NEP	7	L*0*3,S*10*3,T*14*0	3	S	-2.04118	0.1298748	4.40E-07	1.37E-05
NS1	7	L*0*3,S*10*3,T*14*0	3	S	-2.04118	0.1298748	4.40E-07	1.37E-05
HA	203	N*2*0,S*22*6	2	S	-1.72705	0.177809	2.25E-06	4.32E-05
HA	550	I*22*6,L*2*0	2	L	1.727045	5.624013	2.25E-06	4.32E-05
NEP	70	G*0*3,S*24*3	2	S	-1.87937	0.1526867	2.04E-06	4.32E-05
NS1	227	E*24*3,G*0*3	2	G	1.879367	6.549359	2.04E-06	4.32E-05
PA	187	I*0*3,L*24*3	2	L	-1.87937	0.1526867	2.04E-06	4.32E-05
PB2	255	I*3*0,V*21*6	2	V	-1.24868	0.2868838	1.37E-05	2.44E-04
M2	18	K*15*5,R*9*1	2	R	0.826066	2.284315	3.06E-05	4.23E-04
NS1	112	A*9*6,T*15*0	2	T	0.794422	2.213162	5.49E-05	7.19E-04
NEP	14	M*10*6,Q*14*0	2	Q	0.647753	1.911241	7.52E-04	1.87E-03
NEP	22	E*14*0,G*10*6	2	G	-0.64775	0.5232204	7.52E-04	1.87E-03
NEP	26	E*10*6,V*14*0	2	V	0.647753	1.911241	7.52E-04	1.87E-03
NEP	37	R*14*0,S*10*6	2	S	-0.64775	0.5232204	7.52E-04	1.87E-03
NEP	40	I*14*0,L*10*6	2	L	-0.64775	0.5232204	7.52E-04	1.87E-03
NEP	48	A*10*6,S*14*0	2	S	0.647753	1.911241	7.52E-04	1.87E-03
NEP	63	A*14*0,G*10*6	2	G	-0.64775	0.5232204	7.52E-04	1.87E-03
NEP	64	K*10*6,T*14*0	2	T	0.647753	1.911241	7.52E-04	1.87E-03
NEP	68	E*14*0,Q*10*6	2	Q	-0.64775	0.5232204	7.52E-04	1.87E-03
NEP	81	A*14*0,E*10*6	2	E	-0.64775	0.5232204	7.52E-04	1.87E-03

NEP	83	C*14*0,V*10*6	2	V	-0.64775	0.5232204	7.52E-04	1.87E-03
NEP	85	H*10*6,N*14*0	2	N	0.647753	1.911241	7.52E-04	1.87E-03
NEP	86	I*14*0,R*10*6	2	R	-0.64775	0.5232204	7.52E-04	1.87E-03
NEP	88	K*10*6,T*14*0	2	T	0.647753	1.911241	7.52E-04	1.87E-03
NEP	100	L*14*0,M*10*6	2	M	-0.64775	0.5232204	7.52E-04	1.87E-03
NEP	111	Q*10*6,S*14*0	2	S	0.647753	1.911241	7.52E-04	1.87E-03
NS1	21	L*14*0,R*10*6	2	R	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	22	F*10*6,L*14*0	2	L	0.647753	1.911241	7.52E-04	1.87E-03
NS1	23	A*10*6,S*14*0	2	S	0.647753	1.911241	7.52E-04	1.87E-03
NS1	24	D*10*6,M*14*0	2	M	0.647753	1.911241	7.52E-04	1.87E-03
NS1	25	Q*10*6,R*14*0	2	R	0.647753	1.911241	7.52E-04	1.87E-03
NS1	26	D*14*0,E*10*6	2	E	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	27	I*4*0,L*6*6,M*14*0	3	M	1.024647	2.786111	1.74E-04	1.87E-03
NS1	28	C*14*0,G*10*6	2	G	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	33	D*14*0,L*10*6	2	L	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	42	A*14*0,S*10*6	2	S	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	44	K*14*0,R*10*6	2	R	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	54	I*10*6,L*14*0	2	L	0.647753	1.911241	7.52E-04	1.87E-03
NS1	55	E*10*6,R*14*0	2	R	0.647753	1.911241	7.52E-04	1.87E-03
NS1	56	T*10*6,V*14*0	2	V	0.647753	1.911241	7.52E-04	1.87E-03
NS1	60	A*10*6,E*14*0	2	E	0.647753	1.911241	7.52E-04	1.87E-03
NS1	63	K*14*0,Q*10*6	2	Q	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	67	D*14*0,R*10*6	2	R	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	70	E*10*6,K*14*0	2	K	0.647753	1.911241	7.52E-04	1.87E-03
NS1	71	E*10*6,S*14*0	2	S	0.647753	1.911241	7.52E-04	1.87E-03
NS1	73	S*10*6,T*14*0	2	T	0.647753	1.911241	7.52E-04	1.87E-03
NS1	76	A*10*6,N*14*0	2	N	0.647753	1.911241	7.52E-04	1.87E-03
NS1	79	I*14*0,M*10*6	2	M	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	80	A*14*0,T*10*6	2	T	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	87	P*14*0,S*10*6	2	S	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	90	I*14*0,L*10*6	2	L	-0.64775	0.5232204	7.52E-04	1.87E-03

NS1	94	S*14*0,T*10*6	2	T	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	95	I*14*0,L*10*6	2	L	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	98	I*14*0,M*10*6	2	M	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	101	D*10*6,E*14*0	2	E	0.647753	1.911241	7.52E-04	1.87E-03
NS1	103	F*10*6,Y*14*0	2	Y	0.647753	1.911241	7.52E-04	1.87E-03
NS1	108	K*10*6,R*14*0	2	R	0.647753	1.911241	7.52E-04	1.87E-03
NS1	111	I*14*0,V*10*6	2	V	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	114	G*14*0,S*10*6	2	S	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	116	C*10*6,M*14*0	2	M	0.647753	1.911241	7.52E-04	1.87E-03
NS1	117	I*10*6,V*14*0	2	V	0.647753	1.911241	7.52E-04	1.87E-03
NS1	118	K*14*0,R*10*6	2	R	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	127	N*10*6,R*14*0	2	R	0.647753	1.911241	7.52E-04	1.87E-03
NS1	140	Q*14*0,R*10*6	2	R	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	145	I*10*6,V*14*0	2	V	0.647753	1.911241	7.52E-04	1.87E-03
NS1	146	L*10*6,S*14*0	2	S	0.647753	1.911241	7.52E-04	1.87E-03
NS1	152	D*14*0,E*10*6	2	E	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	153	D*14*0,E*10*6	2	E	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	158	A*14*0,G*10*6	2	G	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	163	I*14*0,L*10*6	2	L	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	166	L*10*6,M*14*0	2	M	0.647753	1.911241	7.52E-04	1.87E-03
NS1	170	S*14*0,T*10*6	2	T	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	171	D*10*6,T*14*0	2	T	0.647753	1.911241	7.52E-04	1.87E-03
NS1	180	I*14*0,V*10*6	2	V	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	191	S*14*0,T*10*6	2	T	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	192	I*14*0,V*10*6	2	V	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	194	A*14*0,V*10*6	2	V	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	197	N*14*0,T*10*6	2	T	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	198	I*14*0,L*10*6	2	L	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	204	G*14*0,R*10*6	2	R	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	207	D*14*0,N*10*6	2	N	-0.64775	0.5232204	7.52E-04	1.87E-03
NS1	211	G*14*0,R*10*6	2	R	-0.64775	0.5232204	7.52E-04	1.87E-03

NS1	225	R*14*0,T*10*6	2	T	-0.64775	0.5232204	7.52E-04	1.87E-03
PB1	215	K*7*0,R*17*6	2	R	0.748074	2.112925	3.89E-04	1.87E-03
PB1-F2	75	H*2*1,L*0*5,R*22*0	3	L	1.750964	5.760152	1.68E-03	4.14E-03
PA	269	K*10*1,R*14*5	2	R	-0.59046	0.5540735	2.00E-03	4.74E-03
PA	348	I*14*5,L*10*1	2	L	0.590458	1.804815	2.00E-03	4.74E-03
PA	388	G*10*1,S*14*5	2	S	-0.59046	0.5540735	2.00E-03	4.74E-03
NEP	7	L*0*3,S*10*3,T*14*0	3	T	-1.30616	0.2708583	2.23E-03	4.77E-03
NS1	7	L*0*3,S*10*3,T*14*0	3	T	-1.30616	0.2708583	2.23E-03	4.77E-03
PB1	298	I*0*5,L*24*1	2	L	-1.69903	0.1828609	2.26E-03	4.77E-03
PB1	667	I*24*1,V*0*5	2	V	1.699029	5.468637	2.26E-03	4.77E-03
PB1-F2	15	H*24*1,R*0*5	2	R	1.699029	5.468637	2.26E-03	4.77E-03
PB1-F2	23	N*24*1,S*0*5	2	S	1.699029	5.468637	2.26E-03	4.77E-03
PB1-F2	27	I*0*5,T*24*1	2	T	-1.69903	0.1828609	2.26E-03	4.77E-03
PB1-F2	58	L*16*1,S*0*5,W*8*0	3	S	1.7196	5.582298	2.06E-03	4.77E-03
PB2	67	I*24*1,V*0*5	2	V	1.699029	5.468637	2.26E-03	4.77E-03
PB2	152	A*24*1,S*0*5	2	S	1.699029	5.468637	2.26E-03	4.77E-03
PB2	199	A*24*1,T*0*5	2	T	1.699029	5.468637	2.26E-03	4.77E-03
PB2	508	Q*0*5,R*24*1	2	R	-1.69903	0.1828609	2.26E-03	4.77E-03
PB2	649	I*1*5,V*23*1	2	V	-1.47445	0.2289058	2.08E-03	4.77E-03
PA	545	I*14*6,V*10*0	2	V	0.575901	1.778733	2.49E-03	5.17E-03

Note: The sites 75 in PB1-F2, 7 in NS1 and 7 in NEP have multiple residue variations associated with pathogenicity.
These positions were based on the sequences used in this study.

Supplementary Table S3. Polymorphic sites in the proteome of North American avian H1N1 IAVs in association with their pathogenicity in DBA/2J mice by host effect after adjusting for residue effect

coding region	position	aa distribution in*Ans*Char	Hostcoef	exp.HostCoef.	HostPvalue	HostFDR
PB1-F2	75	H*2*1	-4.1433	0.01587038	1.01E-10	1.18E-08
NEP	70	S*24*3	-1.07086	0.3427139	2.52E-04	7.39E-03
NS1	227	E*24*3	-1.07086	0.3427139	2.52E-04	7.39E-03
PA	187	L*24*3	-1.07086	0.3427139	2.52E-04	7.39E-03
PB1	642	N*23*1,S*1*5	-1.69259	0.1840418	1.24E-03	9.50E-03
PB1-F2	78	K*23*1,R*1*5	-1.69259	0.1840418	1.24E-03	9.50E-03
HA	295	K*10*6	-0.79213	0.4528804	1.25E-03	9.50E-03
NP	143	L*22*6	-0.83238	0.4350107	5.19E-04	9.50E-03
NP	478	S*22*6	-0.83238	0.4350107	5.19E-04	9.50E-03
PB1	215	R*17*6	-0.76757	0.4641406	9.58E-04	9.50E-03
PB1	298	L*24*1	-1.69822	0.1830083	1.19E-03	9.50E-03
PB1	372	M*22*6	-0.83238	0.4350107	5.19E-04	9.50E-03
PB1	667	I*24*1	-1.69822	0.1830083	1.19E-03	9.50E-03
PB1-F2	15	H*24*1	-1.69822	0.1830083	1.19E-03	9.50E-03
PB1-F2	23	N*24*1	-1.69822	0.1830083	1.19E-03	9.50E-03
PB1-F2	27	T*24*1	-1.69822	0.1830083	1.19E-03	9.50E-03
PB2	67	I*24*1	-1.69822	0.1830083	1.19E-03	9.50E-03
PB2	152	A*24*1	-1.69822	0.1830083	1.19E-03	9.50E-03
PB2	199	A*24*1	-1.69822	0.1830083	1.19E-03	9.50E-03
PB2	508	R*24*1	-1.69822	0.1830083	1.19E-03	9.50E-03
PB1-F2	69	Q*23*1,R*1*5	-1.6865	0.1851662	1.30E-03	9.57E-03
PB1-F2	8	P*22*1	-1.67741	0.1868569	1.39E-03	9.62E-03
HA	173	N*9*6	-0.80132	0.4487382	1.48E-03	9.96E-03

These positions were based on the sequences used in this study.

Supplementary Table S4. Polymorphic sites in the proteome of North American avian H1N1 IAVs affecting the pathogenicity of viruses in DBA/2J mice by host-residue interactions

coding region	position	aa distribution in *Ans*Char	residue.host	AA.HostCoef	exp.AA.HostCoef.	AA.HostPvalue	residueHostFDR
PB1-F2	46*M	M*21,T*9	T:host_Charadriiformes	3.1599	23.5685	7.21E-06	2.38E-04
PB1-F2	83*F	F*19,S*11	S:host_Charadriiformes	1.9405	6.9621	2.93E-05	4.84E-04
NEP	89*K	I*12,K*14,V*4	V:host_Charadriiformes	-2.5932	0.0748	1.23E-04	1.35E-03
NS1	206*S	C*3,R*14,S*13	S:host_Charadriiformes	2.4062	11.0921	5.08E-04	4.19E-03
PB1-F2	90*N	N*17,S*13	S:host_Charadriiformes	-1.9701	0.1394	1.06E-03	6.98E-03
PB1	375*S	N*10,S*20	S:host_Charadriiformes	1.8663	6.4642	1.84E-03	8.66E-03
PB1-F2	26*Q	Q*24,R*6	R:host_Charadriiformes	2.2504	9.4916	1.70E-03	8.66E-03
PA	272*E	D*14,E*16	E:host_Charadriiformes	1.3672	3.9242	3.48E-03	8.83E-03
PA	323*V	I*14,V*16	V:host_Charadriiformes	1.3672	3.9242	3.48E-03	8.83E-03
PA	400*Q	P*12,Q*16,S*2	Q:host_Charadriiformes	1.4109	4.0995	2.78E-03	8.83E-03
PA-X	16*L	L*17,S*13	S:host_Charadriiformes	-1.3964	0.2475	2.79E-03	8.83E-03
PA-X	23*S	L*14,S*16	S:host_Charadriiformes	1.3672	3.9242	3.48E-03	8.83E-03
PA-X	59*P	P*16,Q*14	Q:host_Charadriiformes	-1.3672	0.2548	3.48E-03	8.83E-03
PA-X	21*A	A*25,V*5	V:host_Charadriiformes	1.7787	5.9223	3.83E-03	9.02E-03

These positions were based on the sequences used in this study.

Supplementary Table S5. Polymorphic residues at HA antigenic sites of avian H1N1 IAVs compared to human and pandemic H1N1 viruses

	Antigenic sites and residues									
	Sa		Ca		Sb				Cb	
Year	173	178	154	158	202	203	208	212	89	91
1918pdm	S	L	S	**A**	G	T	**L**	**A**	**T**	S
1934	G	L	S	K	S	K	I	E	P	R
1942	G	L	S	K	I	K	L	E	S	R
1950	G	L	S	K	I	E	L	E	S	K
1977	G	L	S	K	I	E	I	E	S	K
1986	G	L	S	R	I	G	I	E	S	K
1991	G	V	S	K	I	G	I	E	S	E
1998	G	L	S	K	I	G	I	E	S	E
2007	G	L	S	E	I	G	L	E	S	E
2009pdm	**N**	L	**P**	**A**	S	A	**L**	**A**	**T**	S
H1N1_Charadriiformes	**N**	I	**P**	V	T	S	**L**	I	**T**	N
H1N1_Anseriformes	T/**N**	L	S	**A**	T	S/N	**L**	T/**A**	**T**	N

Abbreviations: pdm, pandemic; A, alanine; E, glutamic acid; G, glycine; I, isoleucine; K, lysine; L, leucine; N, asparagine; P, proline; R, arginine; S, serine; T, threonine; V, valine

406 **Supplementary Table S6.** Distribution of the GSEV PDZ-binding motif among various IAV

407 subtypes from different hosts in publicly available NS1 sequences

Virus subtype	Human	Swine	Avian	Pika[408]
H1N1	1	16	6	
H1N2		3		
H1N7			2	
H1N8			5	
H2N1			1	
H3N2	2	4		
H4N8			4	
H5N1[a]	12 (11)	na	54 (45)	4 (4)
H5N2			4	
H6N1			2	
H6N2			4	
H6N8			1	
H6N9			1	
H7N2			1	
H7N6			5	
H9N2			12	
H13N2			1	
H13N6			20	
H13N8			1	
H13N9			6	
H16N3			15	
H16N9			1	
H13			1	
N6			1	
Mixed			5	
TOTAL	**15**	**23**	**153**	**4**

409 [a]The number of highly pathogenic H5N1 viruses is shown in parentheses.

Supplementary Table S7. Distribution of the GSEV motif in publicly available NS1 sequences in terms of isolation year, host, region, and subtype

Isolation year	Host	Region	Virus subtypes (n) [a]	Number of strains
1981	swine	Eurasia	H1N1	1
1982	swine	Eurasia	H1N1	1
1983	swine	Eurasia	H1N1	1
1984	swine	Eurasia	H1N1 (1), H3N2 (2)	3
1985	swine	Eurasia	H1N1 (2), H3N2 (1)	3
1989	swine	Eurasia	H1N1	1
1990	avian (Char)	S. America	H2N1	1
1992	swine	Eurasia	H1N1	1
1993	swine	Eurasia	H3N2	1
	avian (Galli)	N. America	H5N2	3
1995	swine	Eurasia	H1N1	3
1997	avian	Eurasia	H6N9	1
1998	avian (Galli)	N. America	H6N8, H7N2	2
1999	avian (Char)	Eurasia	H16N3	2
2000	swine	Eurasia	H1N1, H1N2	2
	avian (Galli)	Eurasia	H9N2	4
2001	swine	Eurasia	H1N1, H1N2	2
	human	N. America	H1N1	1
	avian (Galli)	N. America	H6N2	1
	avian	Eurasia	H5N1	2
2002	avian (Galli)	Eurasia	H9N2	1
2003	swine	Eurasia	H1N1	1
2004	swine	Eurasia	H1N1	1
	avian (Galli)	N. America	H6N2	2
2005	avian (Char)	Eurasia	H13	1
2006	swine	Eurasia	H1N2	1
	avian	Eurasia	H5N1 (1), H9N2 (3)	4
	avian (Char)	Eurasia	H4N8 (4), H16N3 (2)	6
	avian (Char)	N. America	H16N3(3), H13N9 (2)	5
2007	pika	Eurasia	H5N1	4
	human	Eurasia	H5N1	8
	avian (Galli)	Eurasia	H5N1	6
	avian	Eurasia	H5N1	31
	avian (Char)	N. America	H13N9	1
	avian (Passer)	Eurasia	H5N1	7

	avian (Falcon)	Eurasia	H5N1	1
2008	swine	Eurasia	H1N1	1
	human	Eurasia	H5N1	4
	avian	Eurasia	H5N1	1
		N. America	H6N1	1
	avian (Galli)	Eurasia	H5N1	1
	avian (Char)	Eurasia	H13N8	1
		N. America	H13N9 (3), H13N2 (1)	4
	avian (Falcon)	Eurasia	H5N1	1
2009	avian	Eurasia	H5N1	1
		N. America	H13N6 (2), H16N3(1), mixed (1)	4
	avian (Char)	N. America	H13N6 (10), H16N3 (1), H1N1 (6), H1N7 (2), H1N8 (5), H16N9 (1), H6N1 (1), mixed (3)	29
	avian (Galli)	Eurasia	H7N6	5
2010	avian (Char)	Eurasia	H16N3	1
		N. America	H16N3(4), H13N6 (2), unk (1)	7
	avian (Galli)	Eurasia	H5N2, H9N2	2
2011	avian (Galli)	Eurasia	H6N2, H9N2 (2)	3
	avian	Eurasia	H5N1, H9N2	2
	avian (Char)	N. America	H13N6 (6), H16N3 (1), N6 (1)	8
2012	human	N. America	H3N2	1
	avian	Eurasia	H5N1	1
2013	human	N. America	H3N2	1
TOTAL				**195**

Abbreviations: Char, Charadriiformes; Galli, Galliformes; Passer, Passeriformes; Falcon, Falconiformes

[a] n indicates the number of viruses from each subtype

Supplementary References

1. Katz, J. M. *et al.* Molecular correlates of influenza A H5N1 virus pathogenesis in mice. *J. Virol.* **74**, 10807-10810 (2000).
2. Lee, M. S. *et al.* Characterization of an H5N1 avian influenza virus from Taiwan. *Vet. Microbiol.* **124**, 193-201, doi:10.1016/j.vetmic.2007.04.021 (2007).
3. Chen, H. *et al.* Polygenic virulence factors involved in pathogenesis of 1997 Hong Kong H5N1 influenza viruses in mice. *Virus Res.* **128**, 159-163, doi:10.1016/j.virusres.2007.04.017 (2007).
4. 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.* **67**, 1761-1764 (1993).
5. Gao, Y. *et al.* Identification of amino acids in HA and PB2 critical for the transmission of H5N1 avian influenza viruses in a mammalian host. *PLoS Pathog.* **5**, e1000709, doi:10.1371/journal.ppat.1000709 (2009).
6. Yamada, S. *et al.* Biological and structural characterization of a host-adapting amino acid in influenza virus. *PLoS Pathog.* **6**, e1001034, doi:10.1371/journal.ppat.1001034 (2010).
7. Mehle, A. & Doudna, J. A. Adaptive strategies of the influenza virus polymerase for replication in humans. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21312-21316, doi:10.1073/pnas.0911915106 (2009).
8. Conenello, G. M., Zamarin, D., Perrone, L. A., Tumpey, T. & Palese, P. A single mutation in the PB1-F2 of H5N1 (HK/97) and 1918 influenza A viruses contributes to increased virulence. *PLoS Pathog.* **3**, 1414-1421, doi:10.1371/journal.ppat.0030141 (2007).
9. Chen, G. W. *et al.* Genomic signatures of human versus avian influenza A viruses. *Emerg. Infect. Dis.* **12**, 1353-1360, doi:10.3201/eid1209.060276 (2006).
10. Igarashi, M. *et al.* Predicting the antigenic structure of the pandemic (H1N1) 2009 influenza virus hemagglutinin. *PLoS One* **5**, e8553, doi:10.1371/journal.pone.0008553 (2010).
11. Xu, X., Zhu, X., Dwek, R. A., Stevens, J. & Wilson, I. A. Structural characterization of the 1918 influenza virus H1N1 neuraminidase. *J. Virol.* **82**, 10493-10501, doi:10.1128/JVI.00959-08 (2008).
12. Hale, B. G., Barclay, W. S., Randall, R. E. & Russell, R. J. Structure of an avian influenza A virus NS1 protein effector domain. *Virology* **378**, 1-5, doi:10.1016/j.virol.2008.05.026 (2008).
13. Hay, A. J., Wolstenholme, A. J., Skehel, J. J. & Smith, M. H. The molecular basis of the specific anti-influenza action of amantadine. *EMBO J.* **4**, 3021-3024 (1985).
14. Tombari, W., Nsiri, J., Larbi, I., Guerin, J. L. & Ghram, A. Genetic evolution of low pathogenecity H9N2 avian influenza viruses in Tunisia: acquisition of new mutations. *Virol. J.* **8**, 467, doi:10.1186/1743-422X-8-467 (2011).
15. Koçer, Z. A., Krauss, S., Stallknecht, D. E., Rehg, J. E. & Webster, R. G. The potential of avian H1N1 influenza A viruses to replicate and cause disease in mammalian models. *PLoS One* **7**, e41609, doi:10.1371/journal.pone.0041609 (2012).