

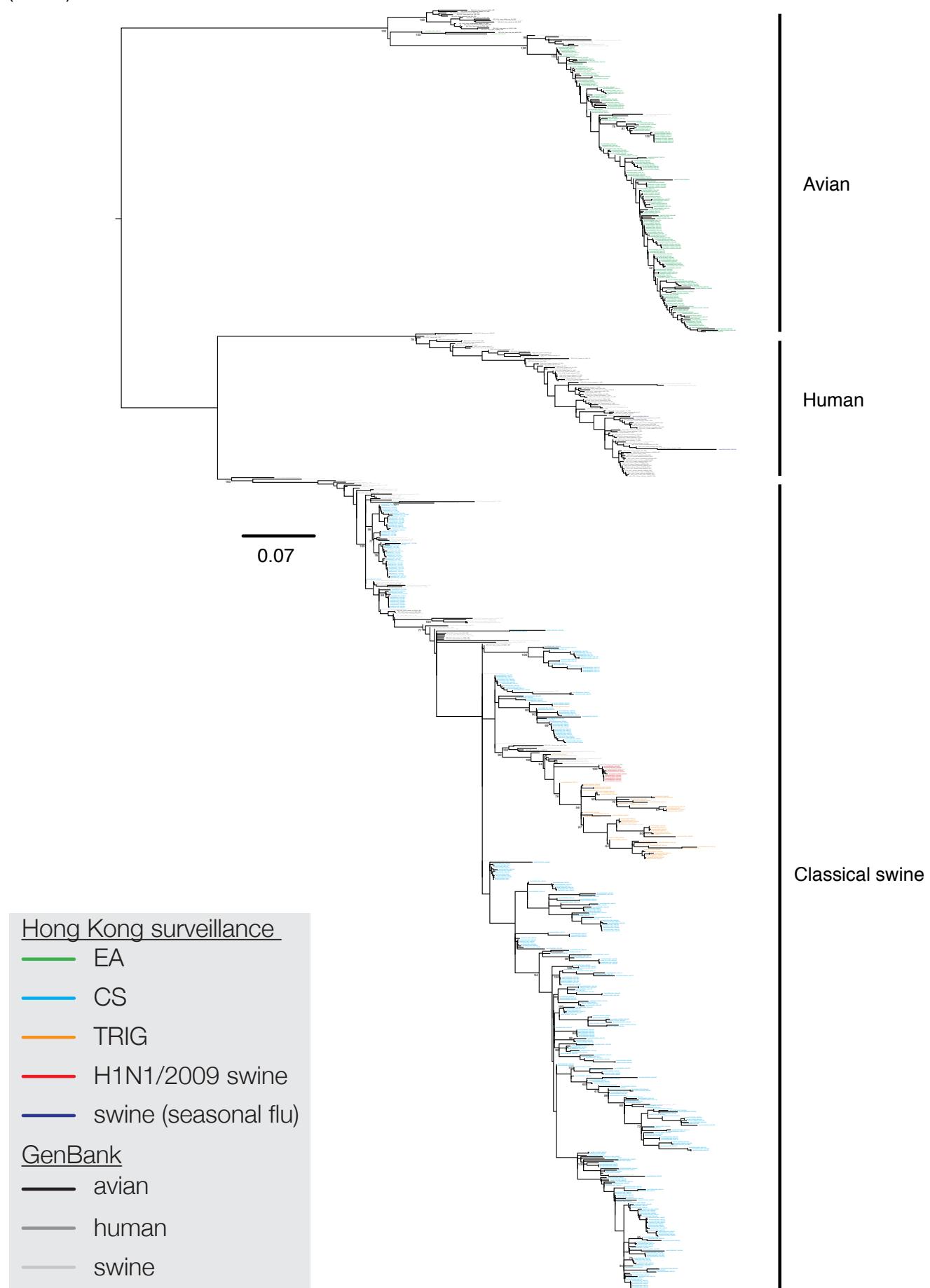
1. Supplementary Figures



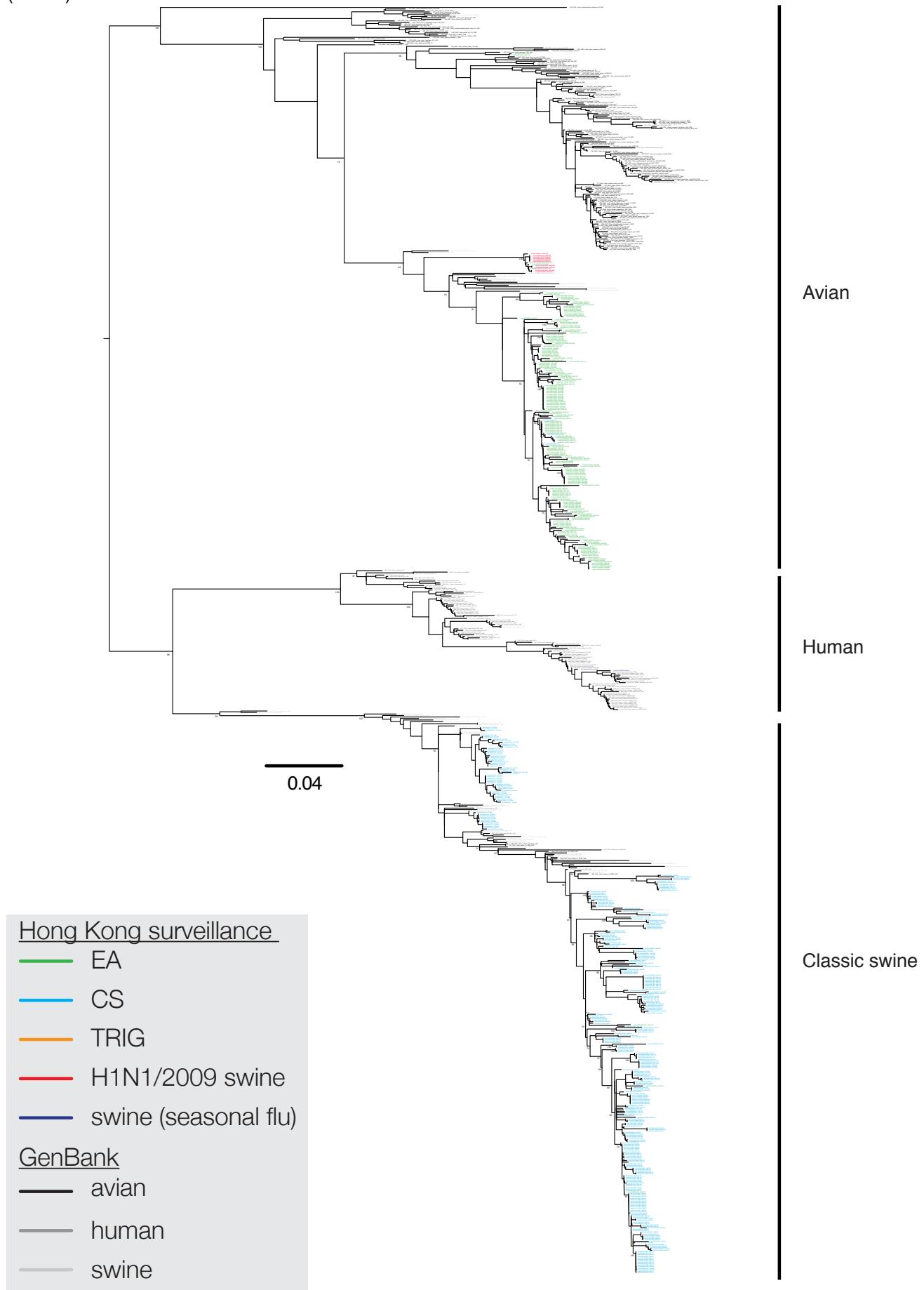
Figure S1. Map of PR China showing provinces (dark grey) from which pigs are imported into Hong Kong. Percentage imports for each province during 2008 are shown here. An annual breakdown of pig imports from 1999 to 2010 is provided in Table S1.

Figure S2. Phylogenetic relationships of sequences from each swine influenza virus (SwIV) gene segment (H1 (a), N1 (b), N2 (c), PB2 (d), PB1 (e), PA (f), NP (g), M (h) & NS (i)). Phylogenetic trees were estimated using the neighbour-joining distance method using genetic distances calculated by maximum likelihood under the HKY model with gamma distributed rates amongst sites (HKY+ γ). Clade labels indicate major influenza A virus lineages (Avian, Human and Classical Swine). SwIV isolated in this study are coloured according to the phylogenetic classification of the HA gene. Sequence colours (see key) represent viruses isolated in this study and representative viruses isolated in GenBank. Scale bar is in units of nucleotide substitutions per site. Node labels represent neighbour-joining bootstrap values.

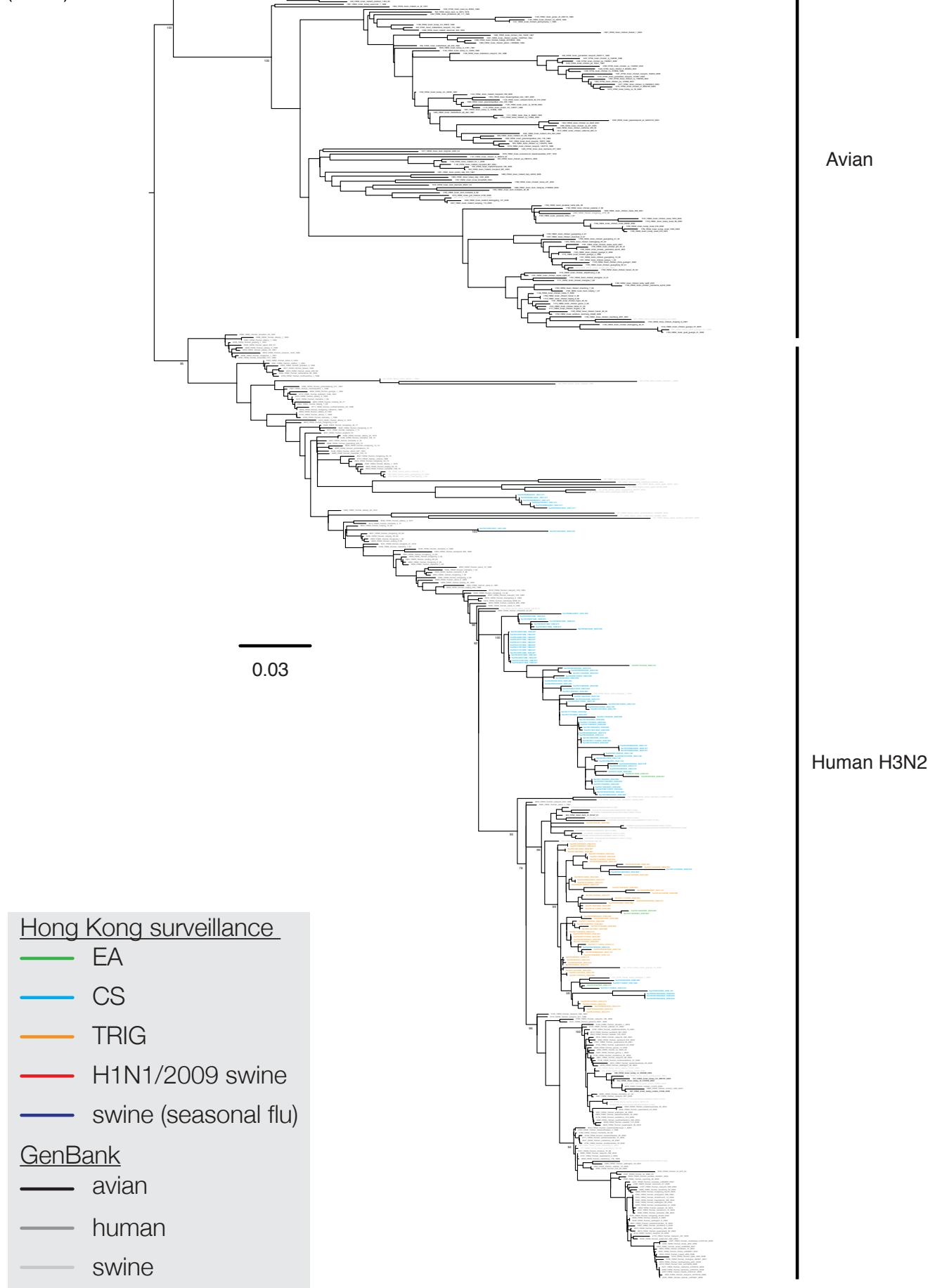
(a - H1)



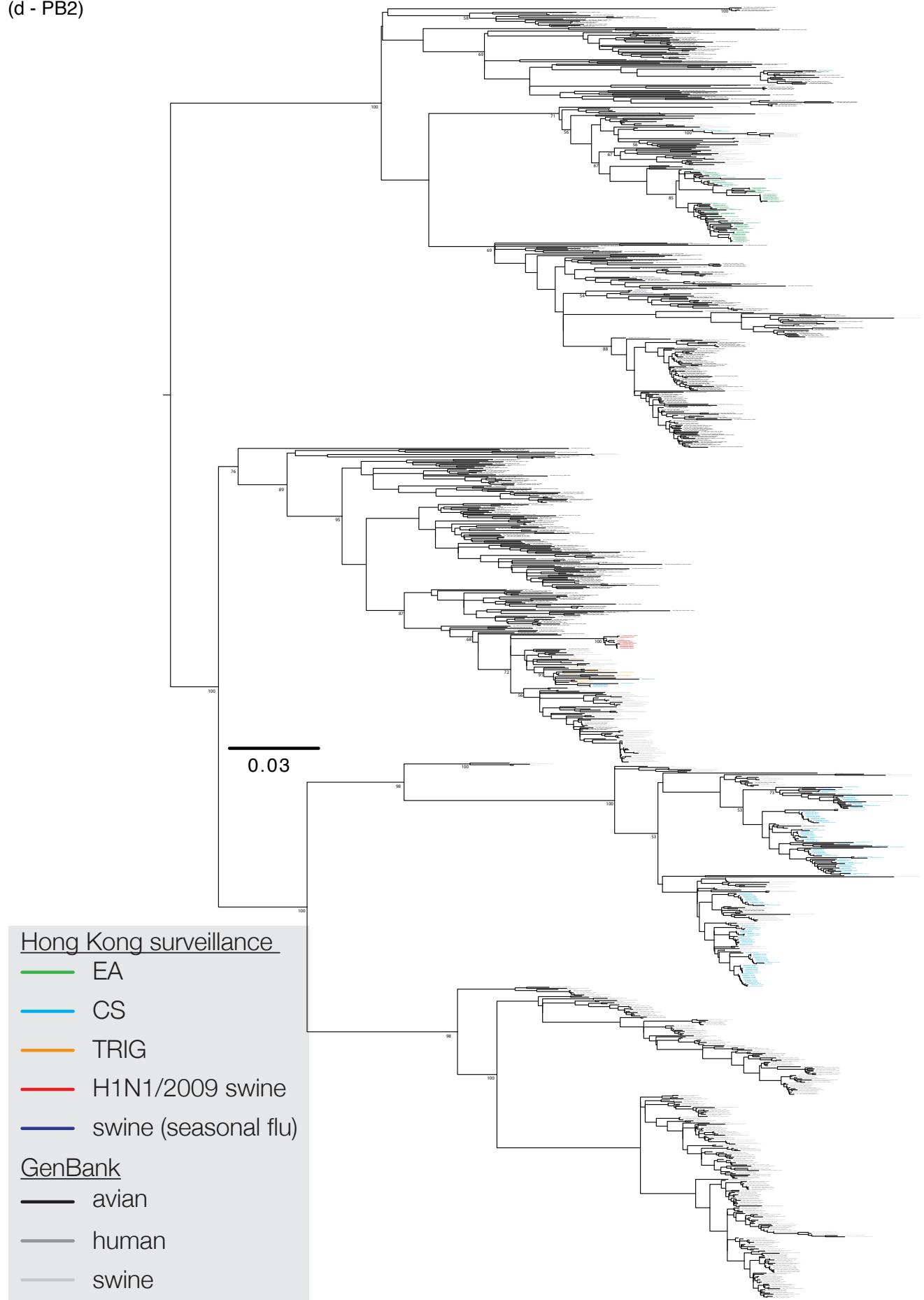
(b - N1)



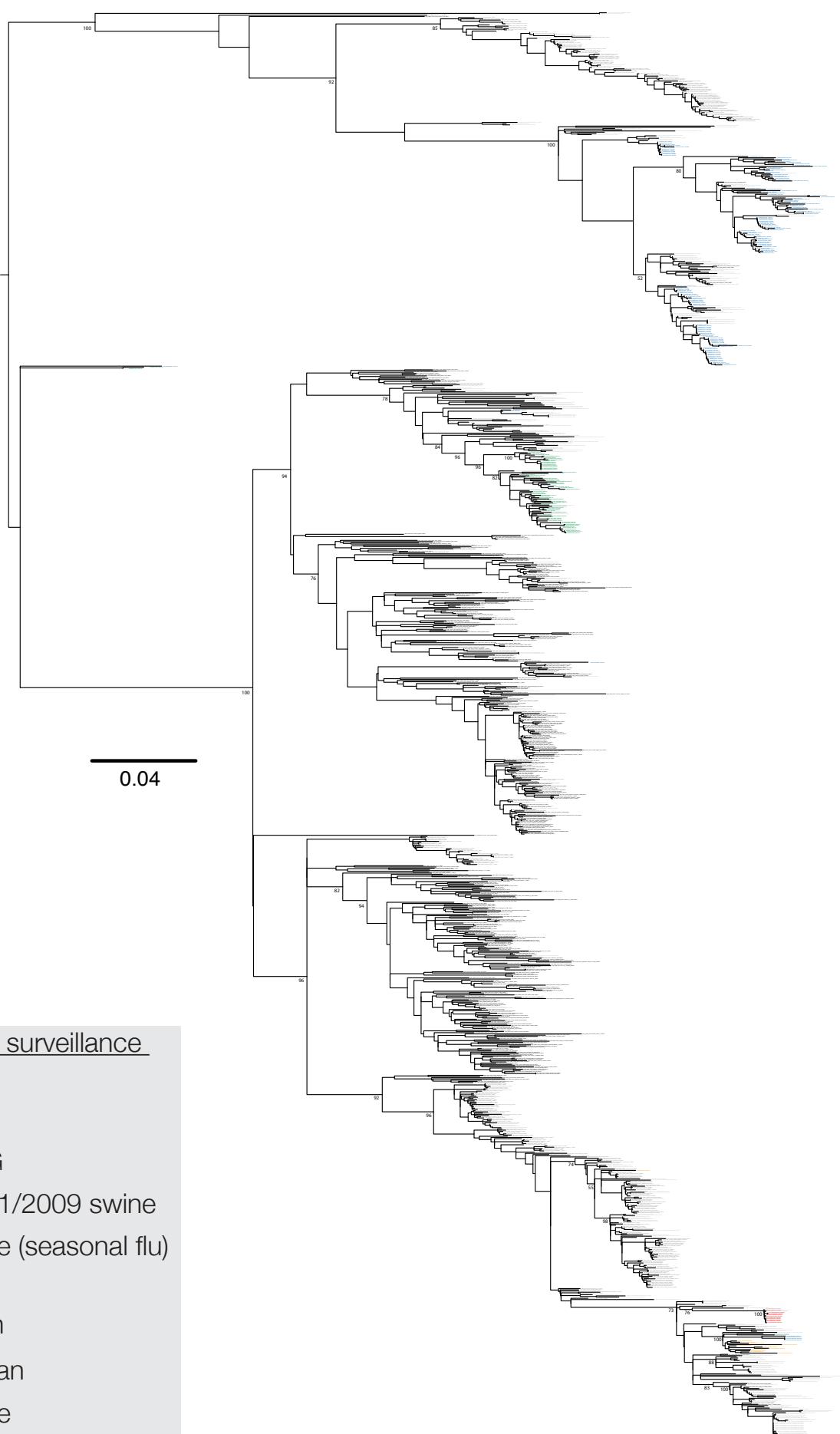
(c - N2)



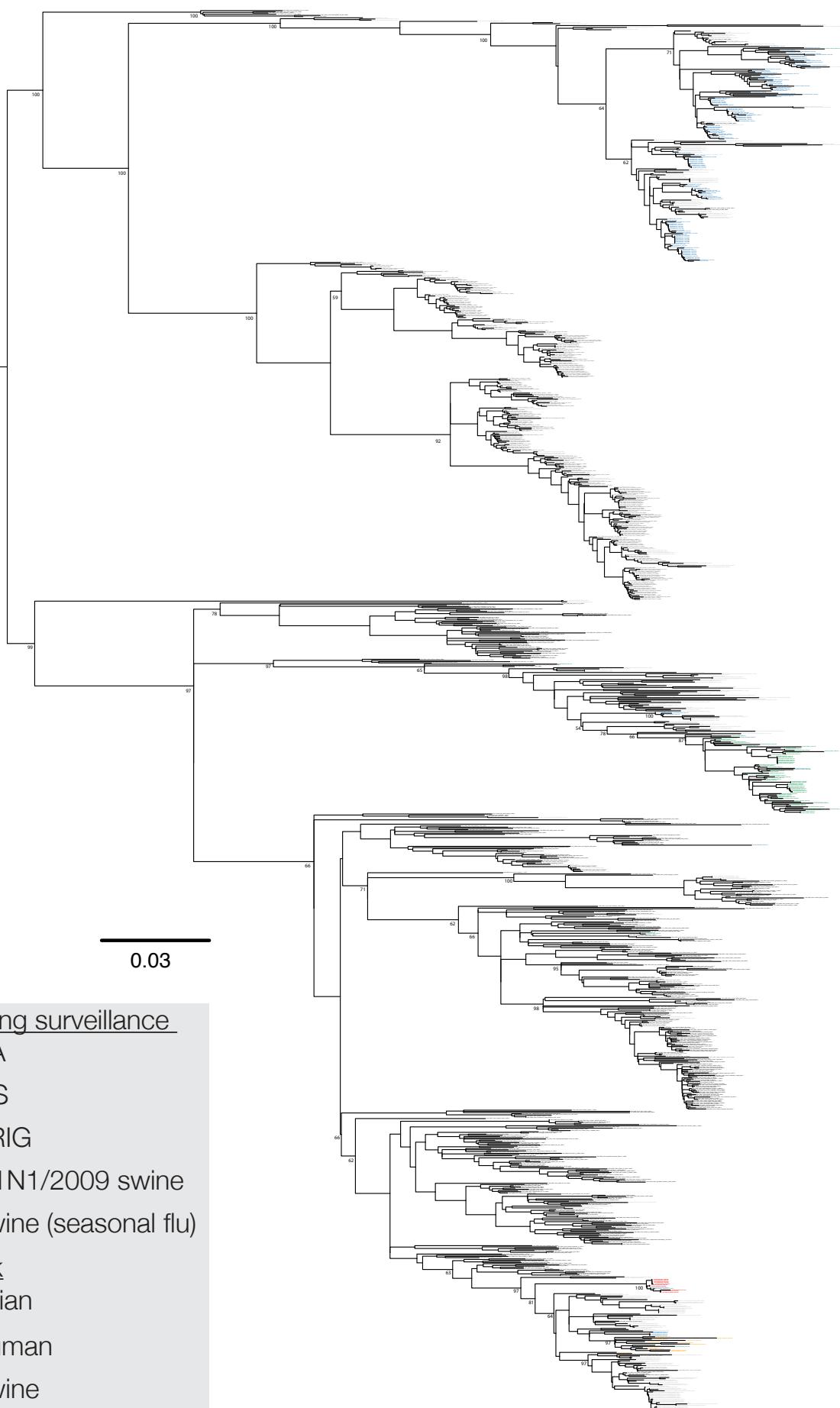
(d - PB2)

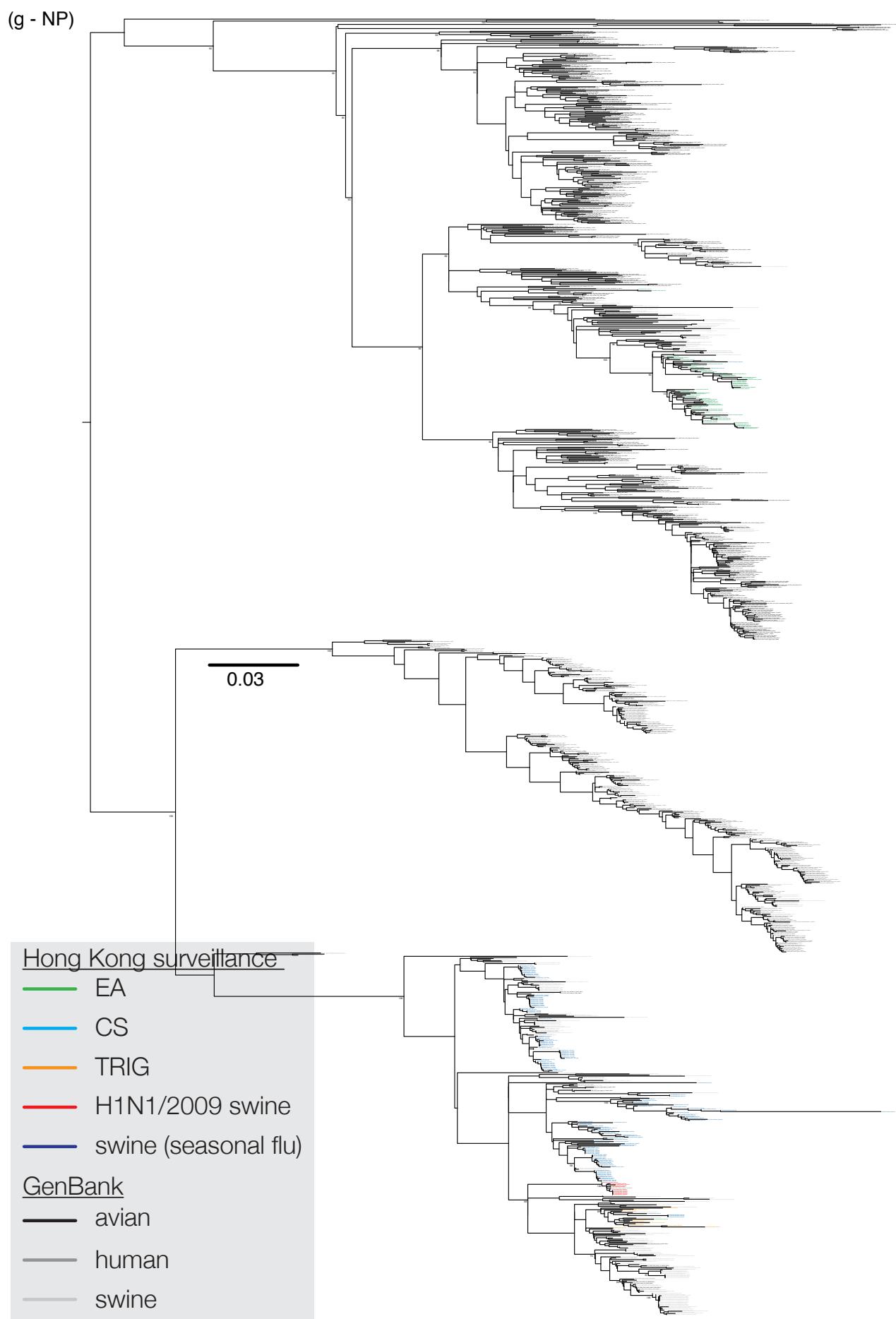


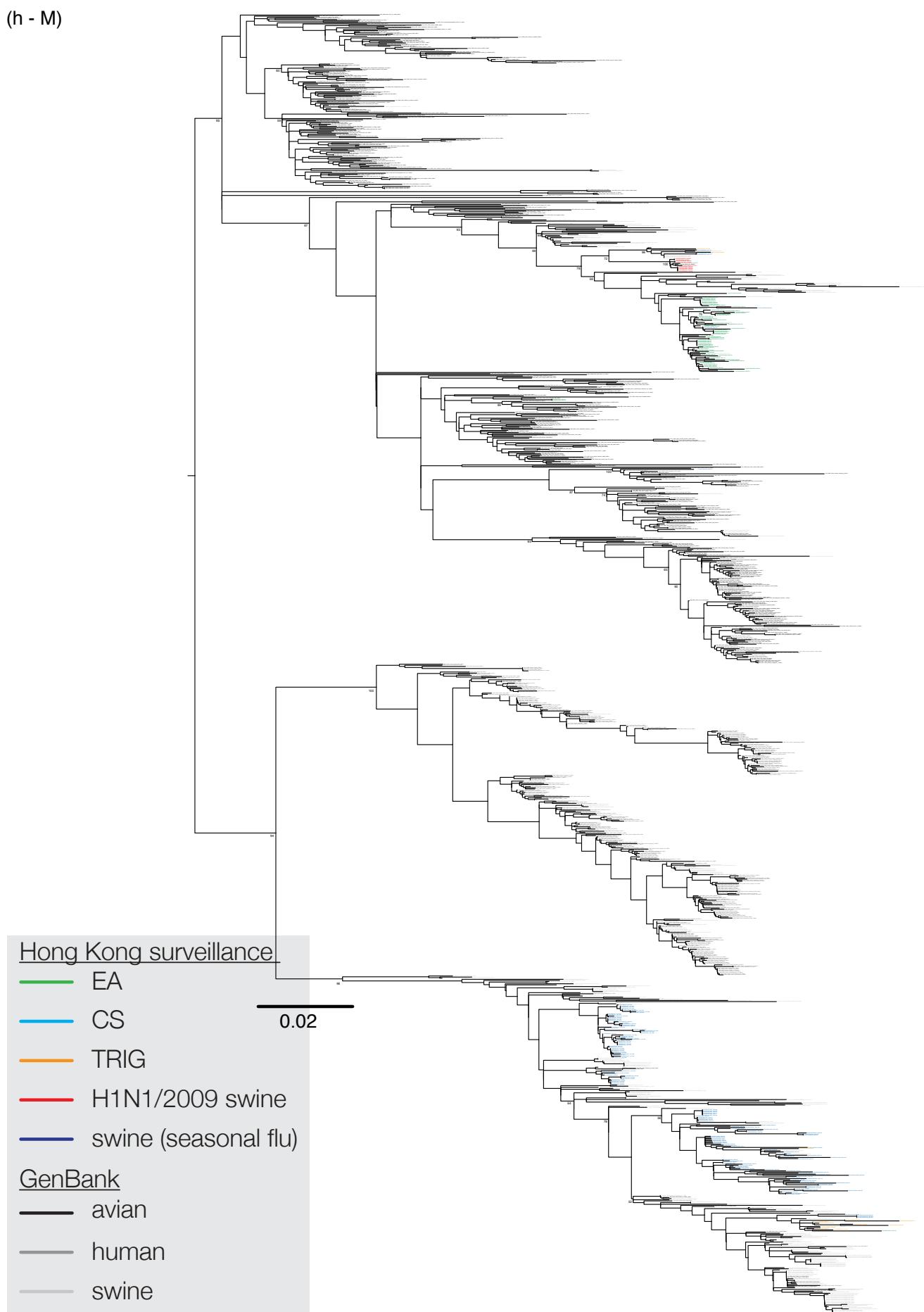
(e - PB1)



(f - PA)







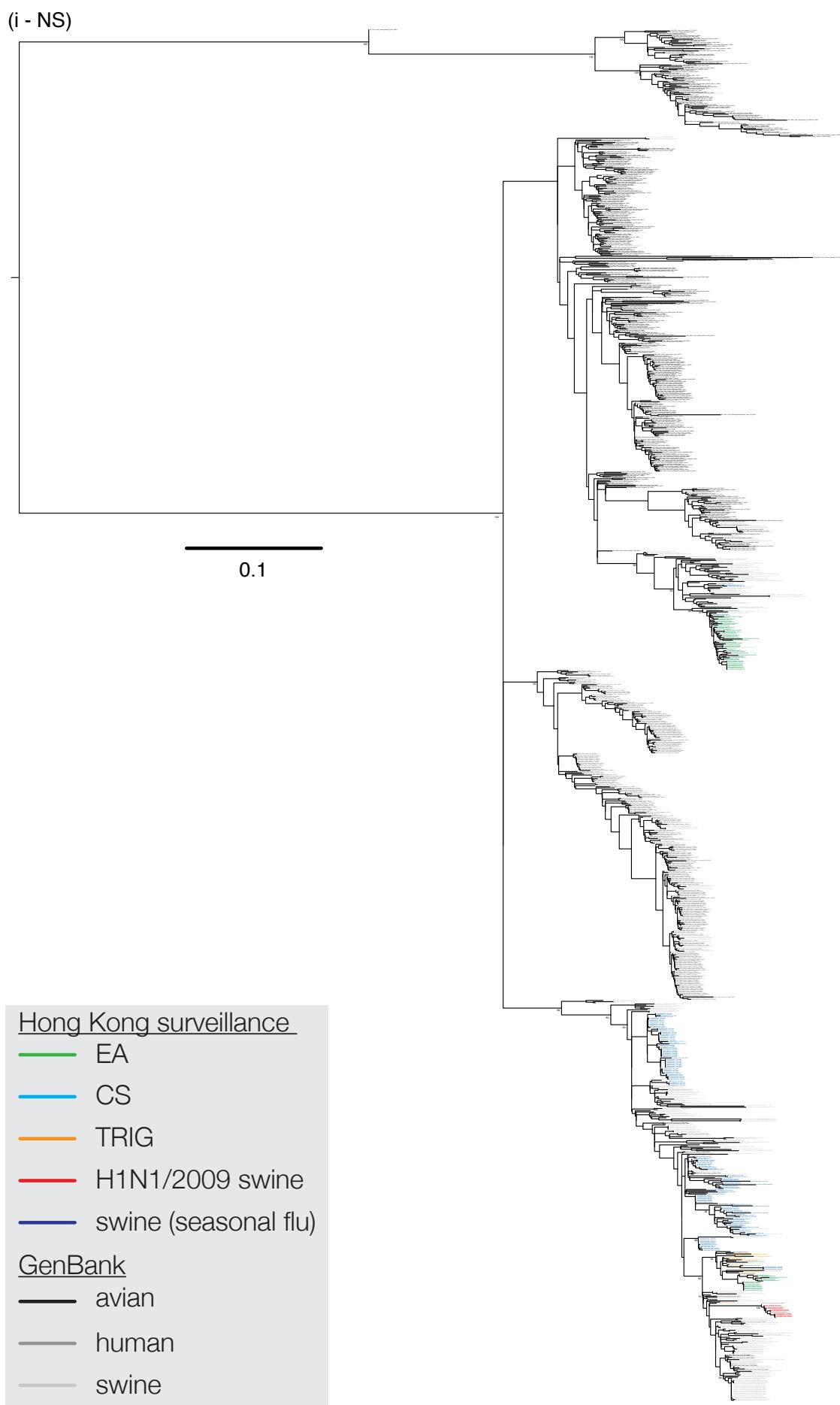
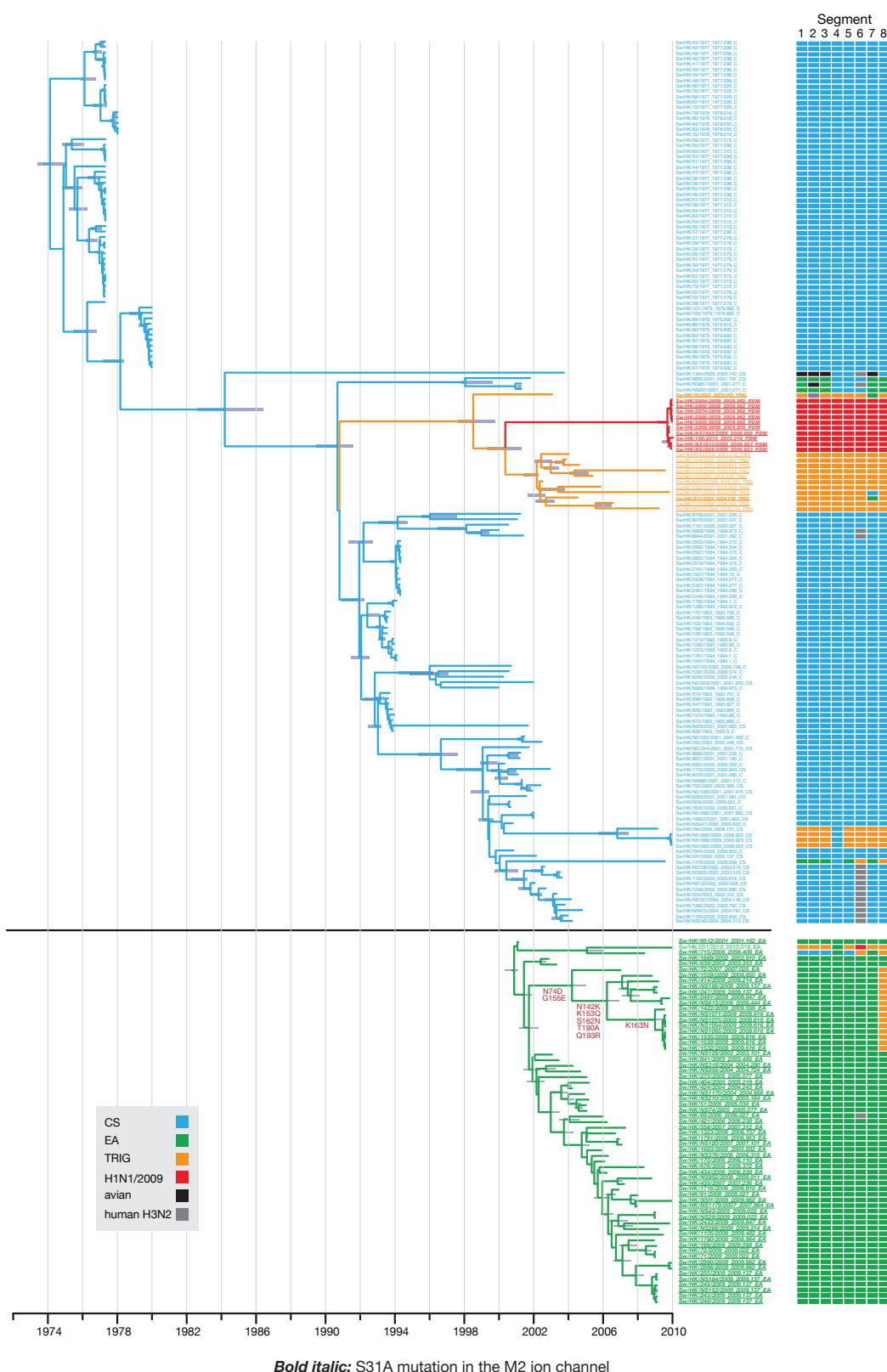
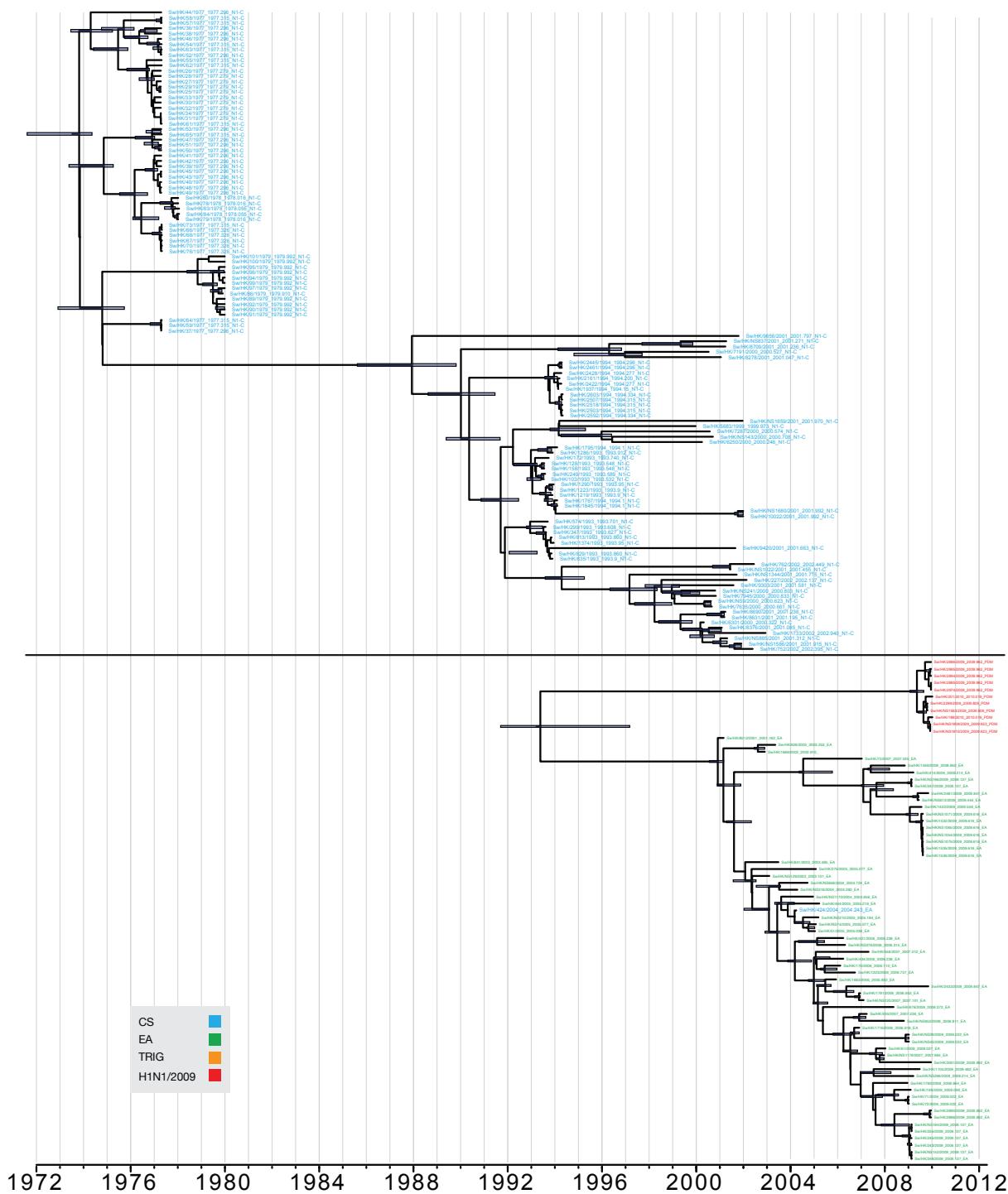


Figure S3. Phylogenetic relationships of fully sequenced SwIV from surveillance in Hong Kong, estimated using a molecular clock that places a timescale on virus evolution (H1 (a), N1 (b), N2 (c), PB2 (d), PB1 (e), PA (f), NP (g), M (h) & NS (i)). Phylogenies were estimated within a Bayesian Markov Chain Monte Carlo framework¹, using the GTR substitution model with gamma-distributed among site rate heterogeneity and a ‘strict molecular clock’ model. The 95% credible intervals of the age of internal nodes are represented by grey bars. Indicated on the HA phylogenies (a) are the antigenic changes observed in the novel EA reassortant (shown in Italics). The S31N mutation in the M gene associated with amantadine resistance is similarly indicated.

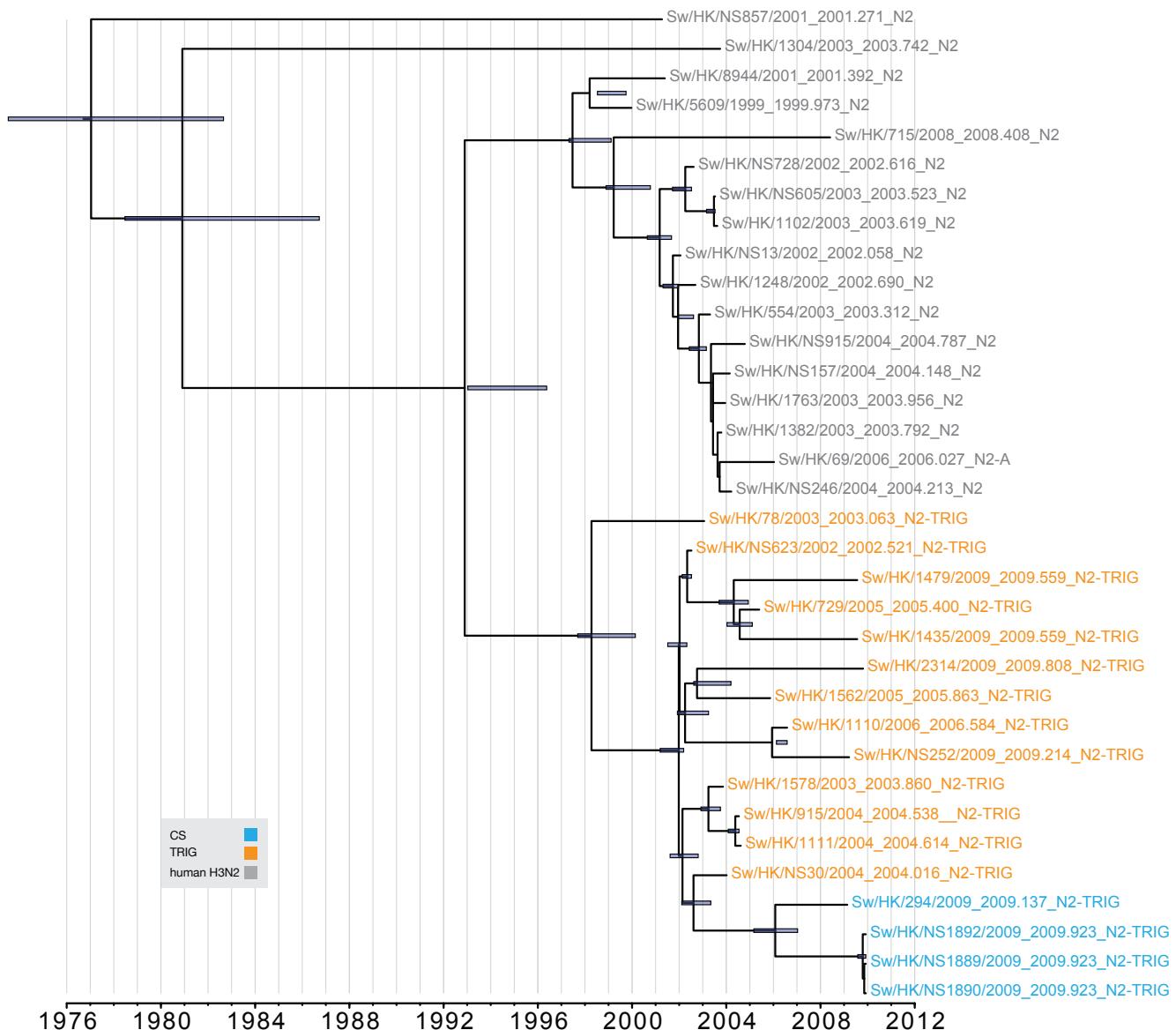
(a)



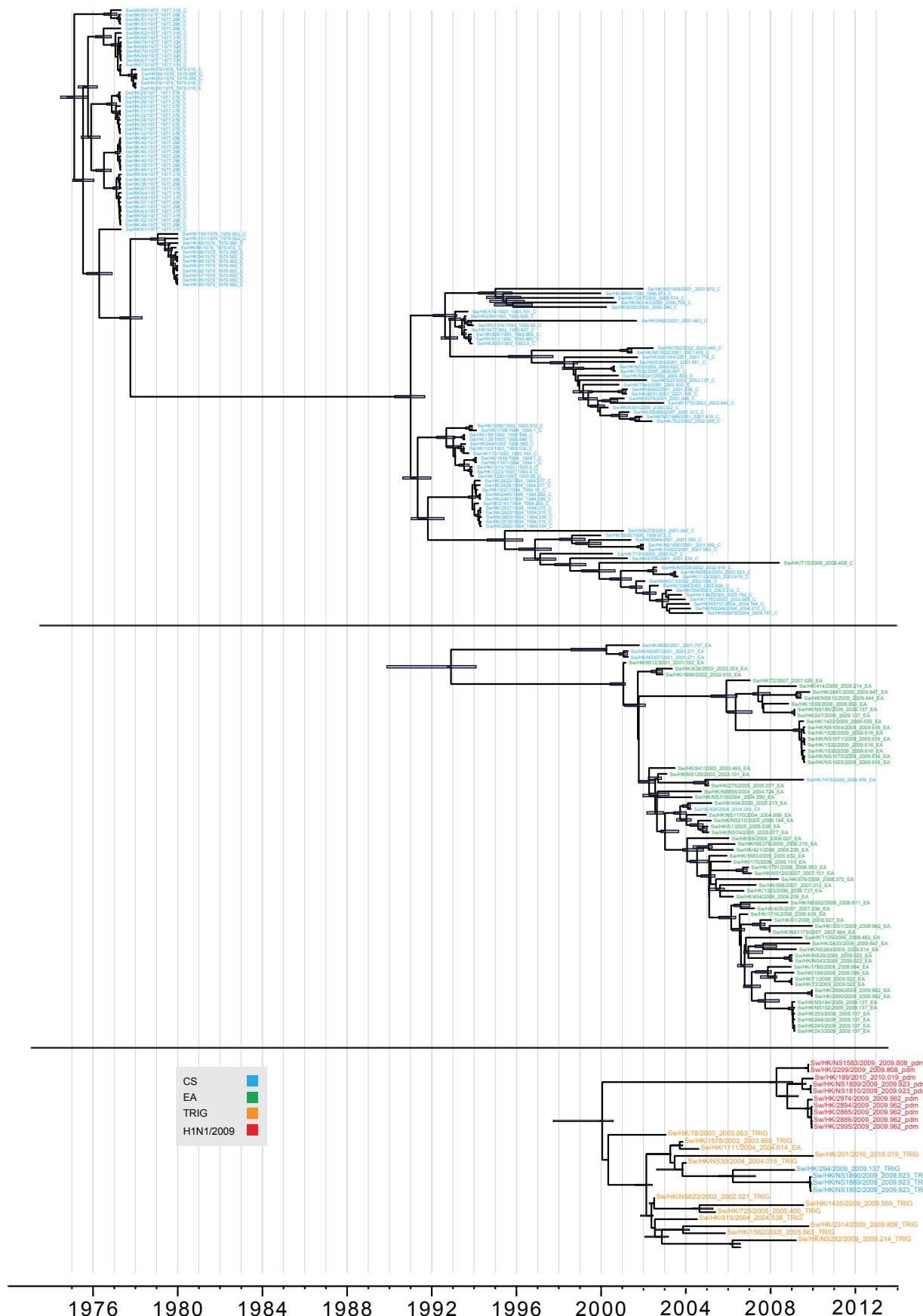
(b - N1)



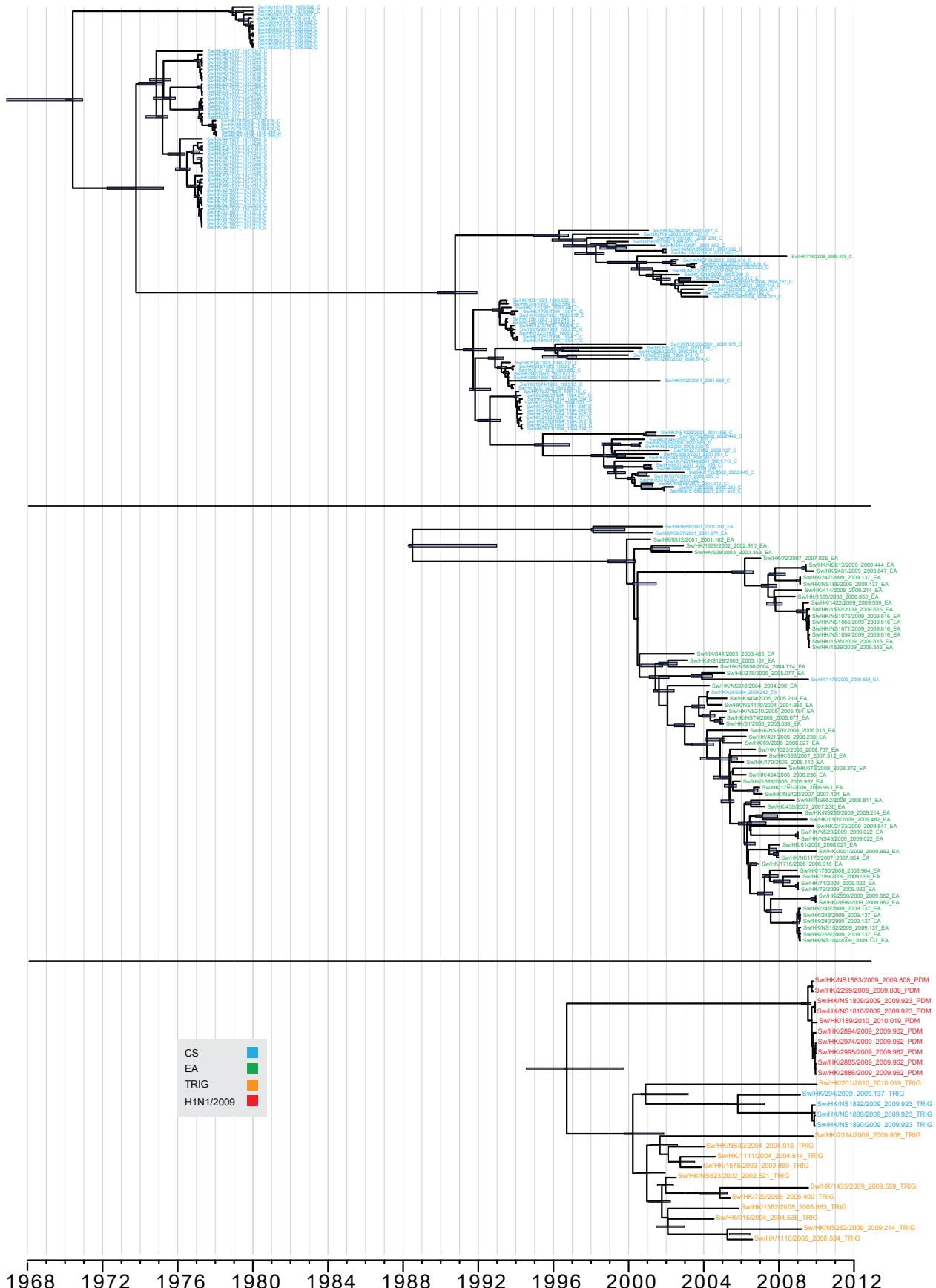
(c - N2)



(d - PB2)



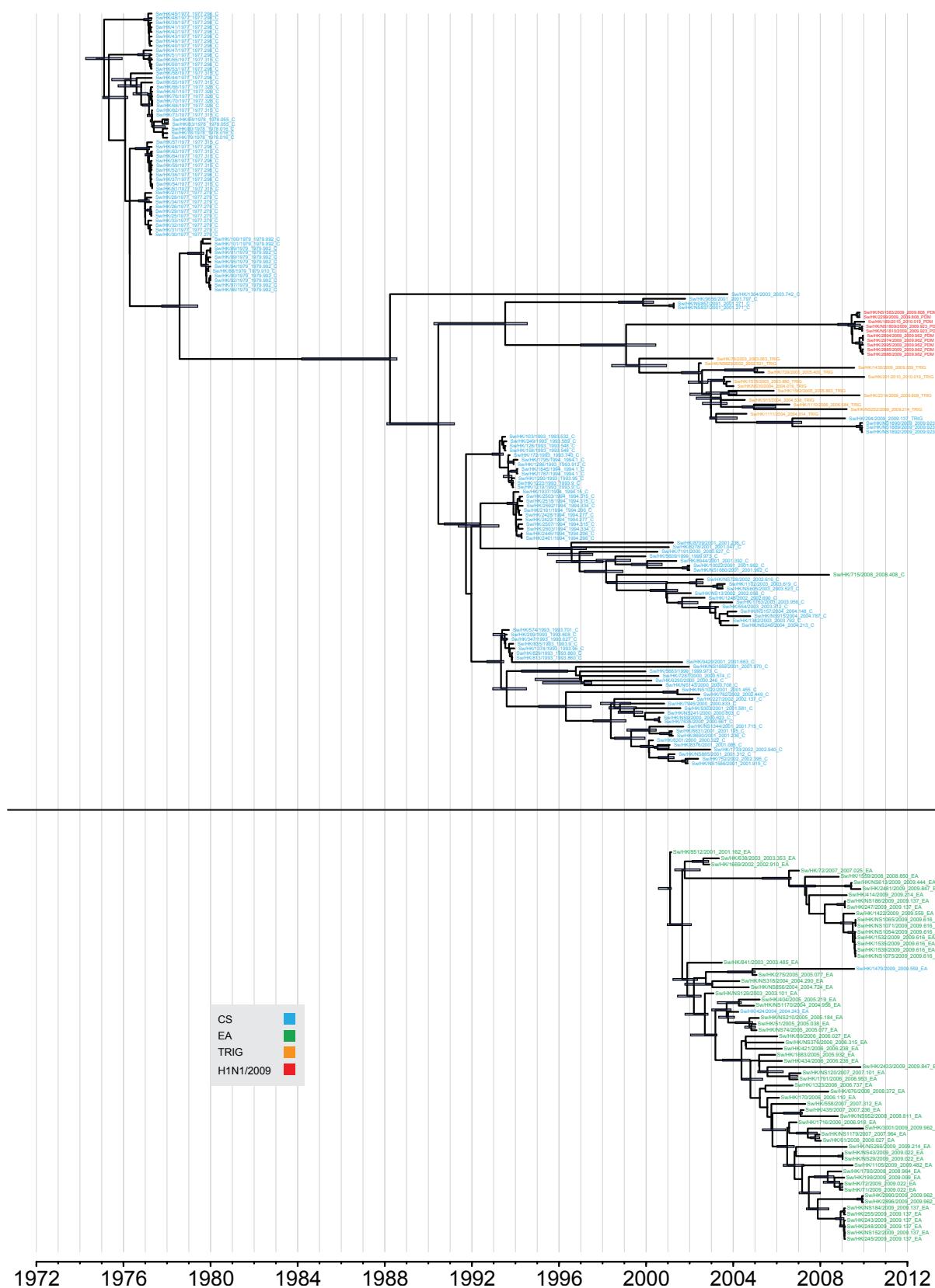
(e - PB1)

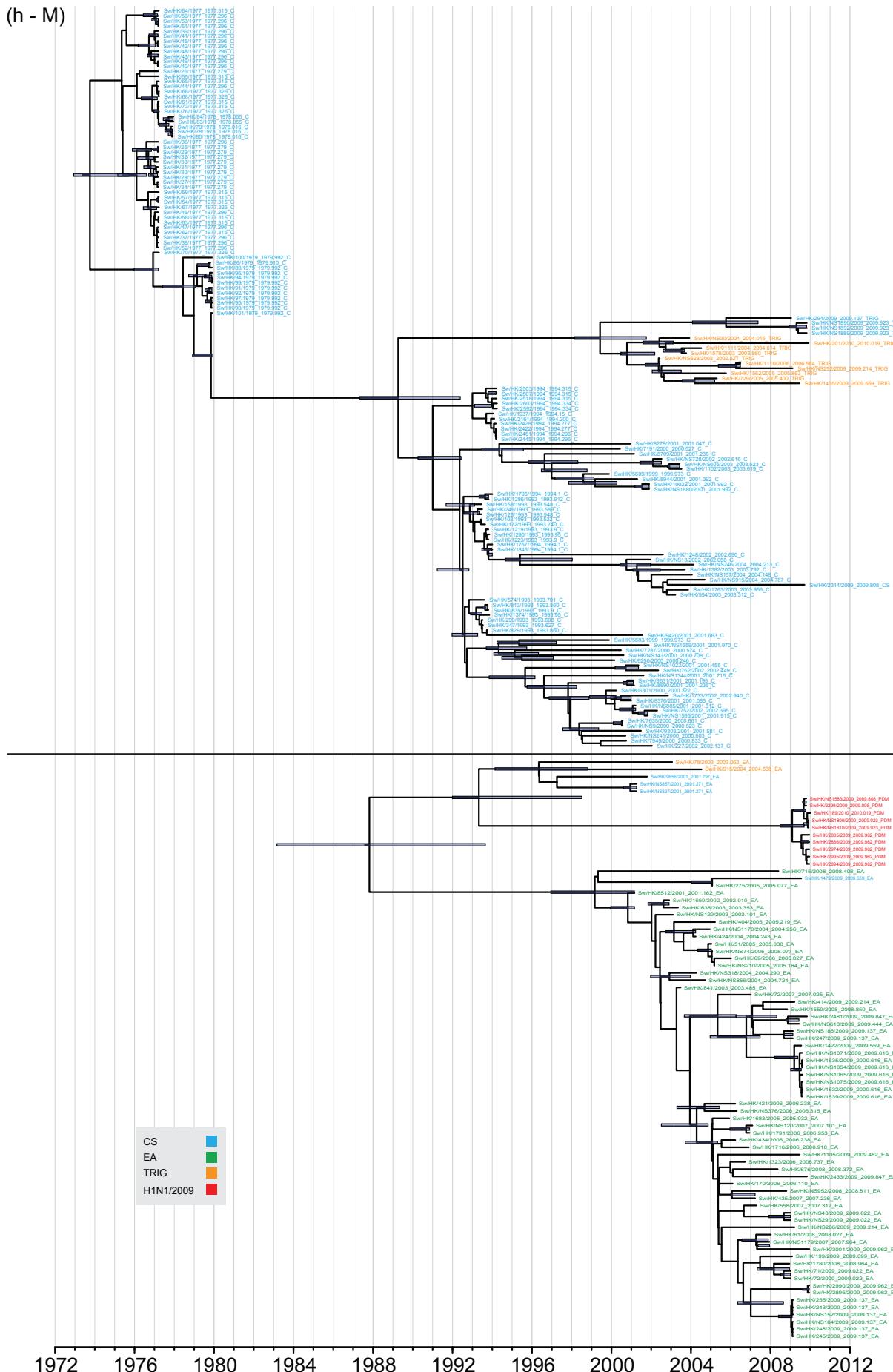


(f - PA)

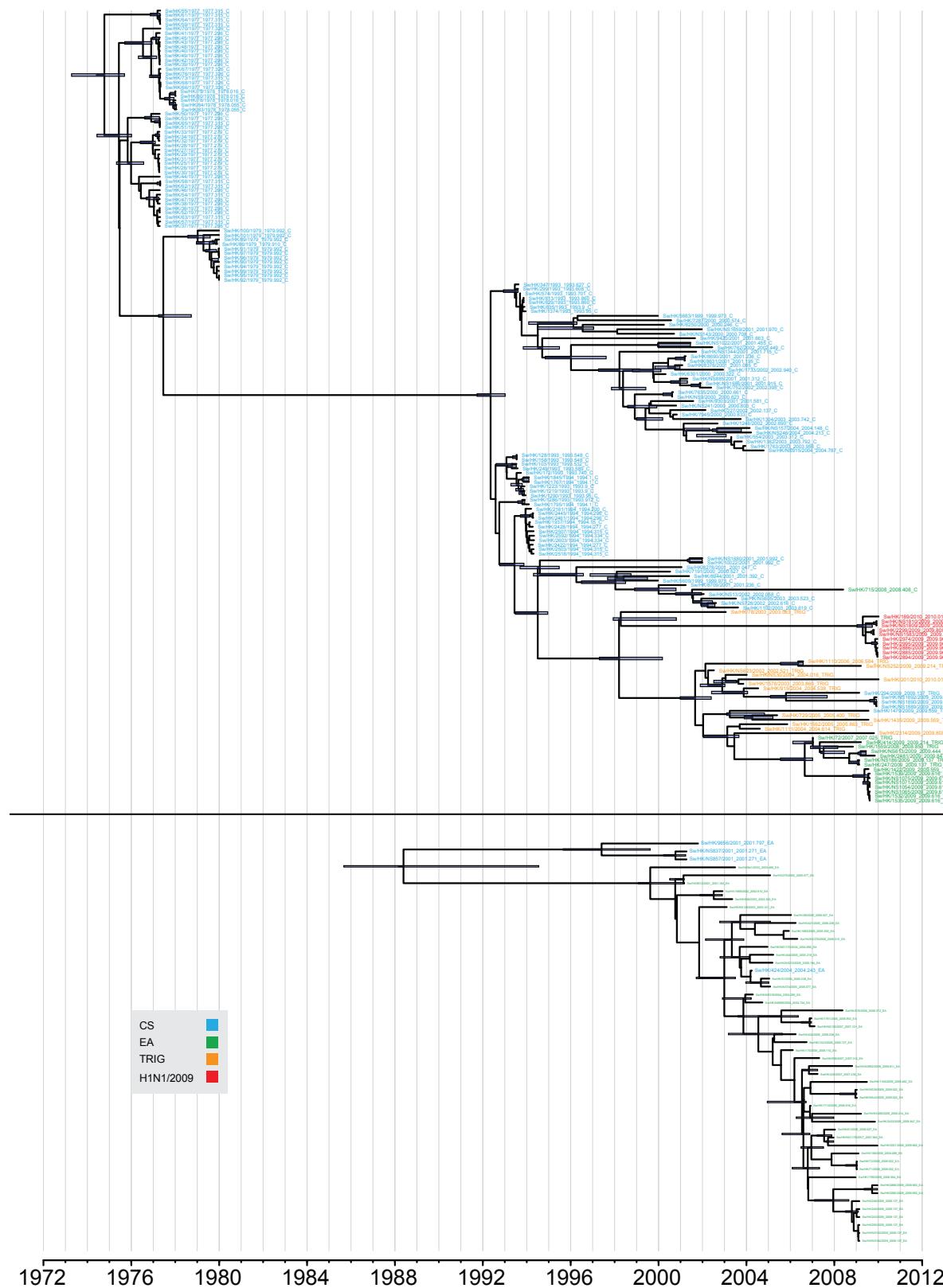


(g - NP)





(i - NS)



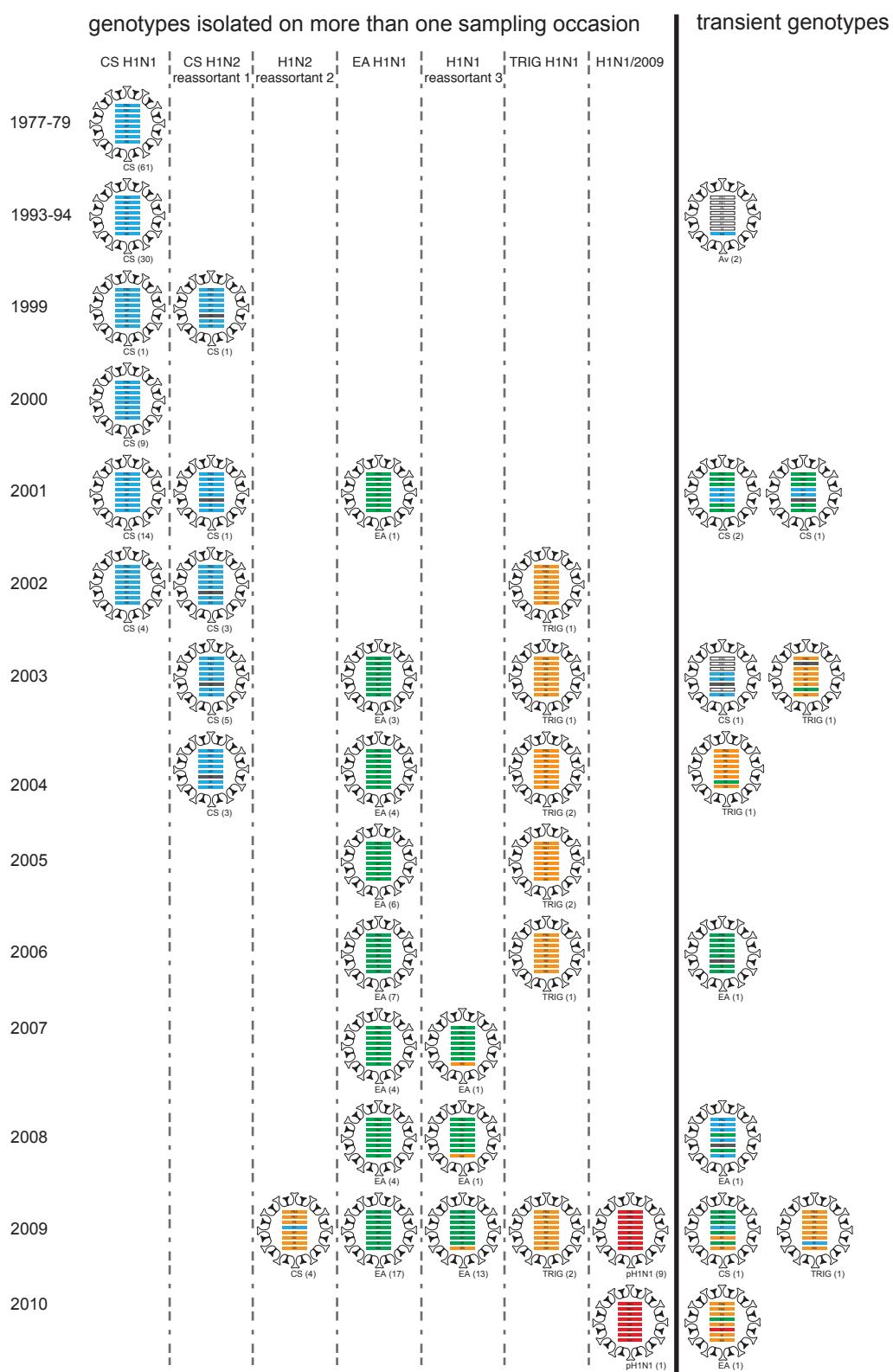


Figure S4. Diagram representing the emergence of multiple genotypes of influenza A viruses. Gene segments are ordered PB2, PB1, PA, HA, NP, NA, M and NS from top to bottom within each virus particle. Virus subtype and major genotypes are shown at the top. Genotypes that have been isolated on more than one sampling occasion are shown on the left, while other 'transient' genotypes are shown on the right. Numbers within brackets represents virus genotypes fully sequenced during the year.

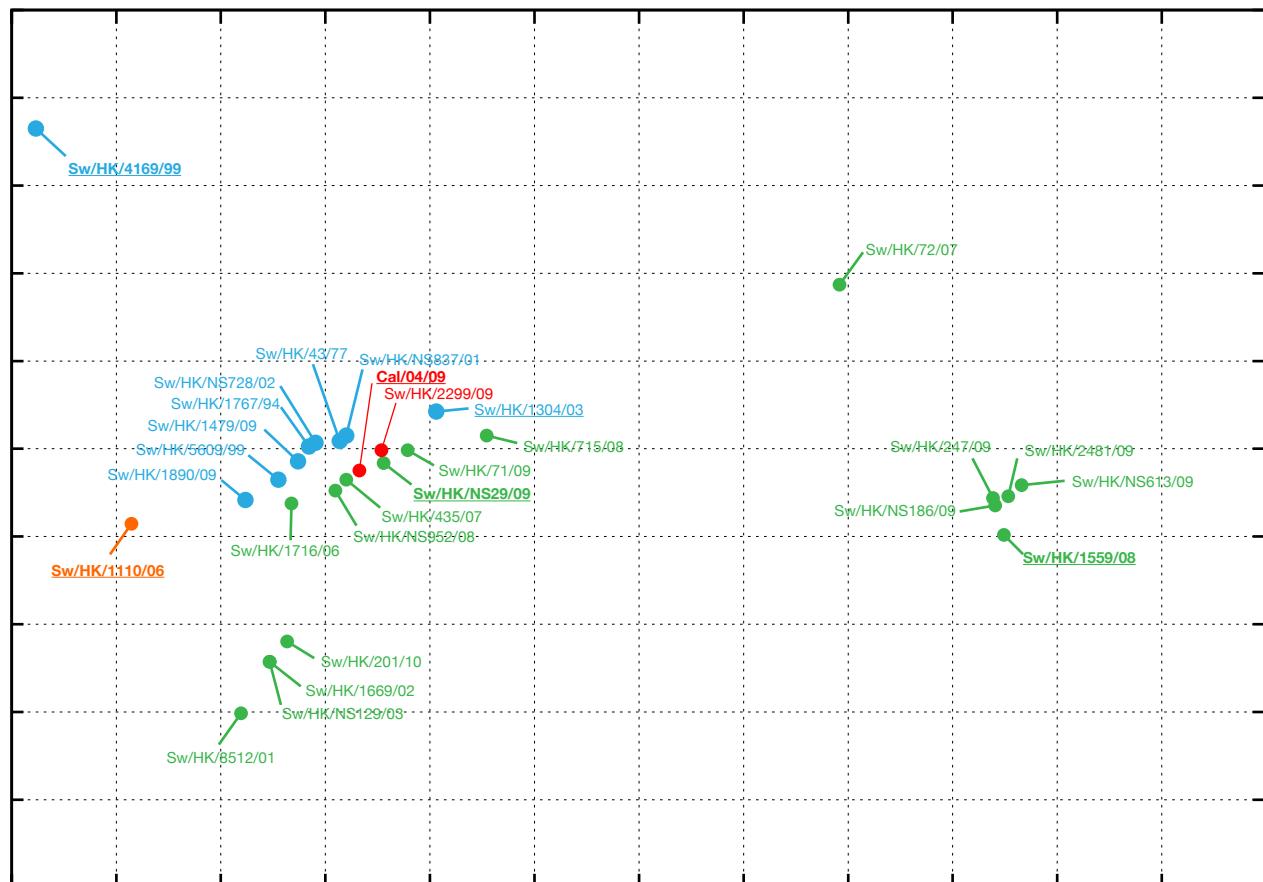


Figure S5. Antigenic map^{2,3}. Bold underlined text represents reference antigens of four major SwIV virus lineages, including CS, TRIG, EA and EA reassortants.

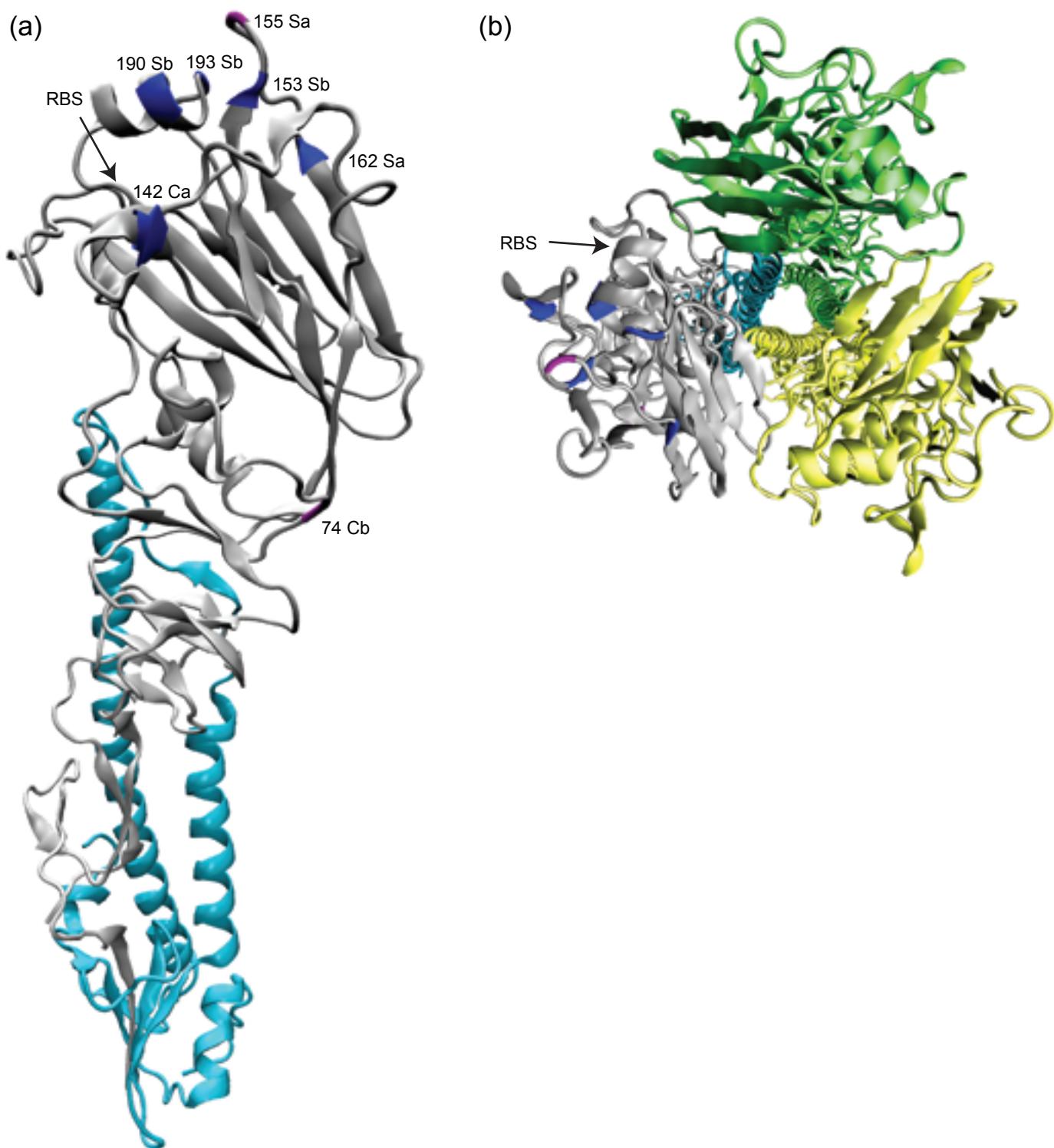


Figure S6. Crystal structure of HA0 (Protein DB: 3AL4) showing amino acid changes in EA reassortant (Sw/HK/72/2007-like) viruses. Changes at positions 74, 155 (purple) are observed in all EA reassortants viruses, while changes in amino acid positions 142, 153, 162, 190, 193 (blue) are not observed in the earliest EA reassortant virus (Sw/HK/72/2007), suggesting a progressive accumulation of amino acid changes after reassortment. Sa, Sb, Ca, Cb represent the predicted antigenic sites that these changes fall within. RBS denotes the receptor binding site. (b) HA trimer showing the abovementioned amino acid changes on the trimer's exposed region. The crystal structures were generated in VMD⁴ and rendered with Tachyon⁵.

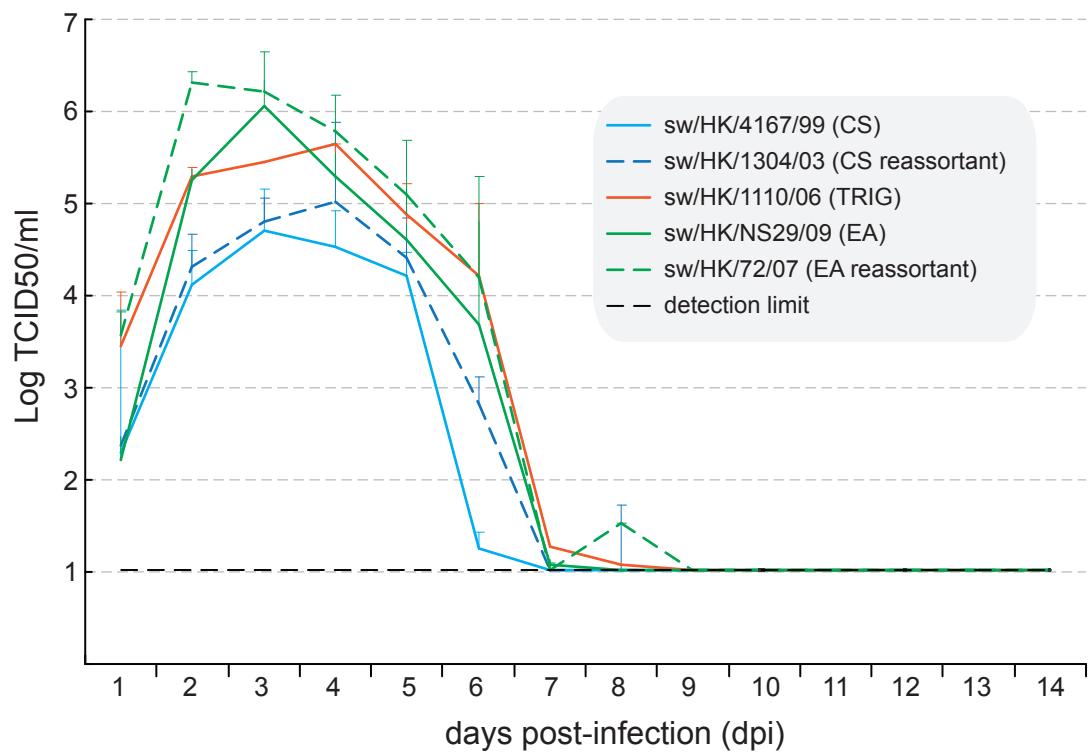


Figure S7. Virus shedding of representative SwIV in challenged pigs.

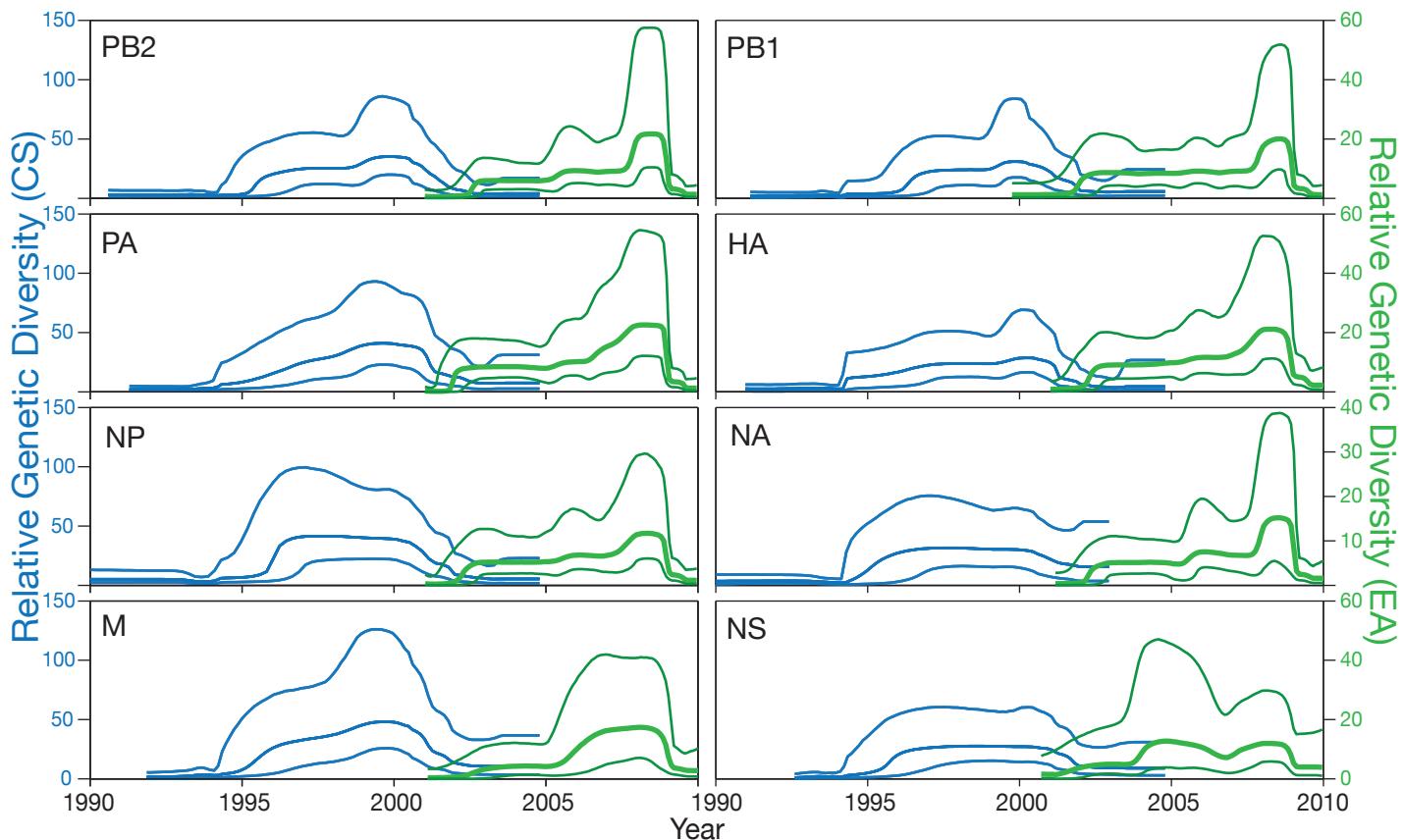


Figure S8. Population dynamics of SwIV genetic diversity . Bayesian skyline plots for each gene (PB2, PB1, PA, HA, NP, NA, M, NS) of the Classical swine and European “avian-like” SwIV lineages. The y-axes represent a measure of relative genetic diversity (see Methods for details). Analyses that included recent classic swine HA sequences exhibited very similar results (not shown).

Notes: Following recent studies of human influenza A, we reconstructed the dynamics of viral genomic diversity using a Bayesian coalescent approach^{6,7}. To meet the assumptions of this approach, we analysed the largest continuous monophyletic lineage of CS and EA viruses from Hong Kong, such that viruses sampled in the 1970s and reassortant genotypes (including pandemic H1N1/2009) were excluded. For both the CS and EA lineages, each genome segment gave similar results—viral diversity changes gradually over several years, in contrast to the strongly seasonal fluctuations observed for human influenza A⁷. This corroborates our observation of comparatively slow SwIV antigenic change: bottlenecks in human influenza A diversity are most likely caused by repeated antigenic selective sweeps^{2,7,8}. The plots clearly show a rise in CS genetic diversity from 1994–2000 and a decline from 2000–2004, coinciding with the introduction of the EA viruses (see main text Figures). EA lineage diversity gradually increases and peaks around 2008. The sharp drop in EA diversity after 2008 may be partly artificial, owing to a change in population sampling in 2009 (see Methods), but also reflects a drop in SwIV prevalence after the detection of H1N1/2009 virus. That issue excepted, the dynamics of genetic diversity are congruent with the longitudinal surveillance data and reveal the impact of SwIV lineage replacement (main text Fig. 1a), indicating that any potential bias arising from isolate selection procedures was not significant.

2. Supplementary tables

Table S1. Origin of pigs slaughtered in Hong Kong.

Location	Year											
	1999- 2000 ^a	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Hong Kong	15	19.3	20.2	20.5	18.1	16.9	18.4	19.3	15.2	5.6	5.0	5.8
Guangdong ^b	25				31.1	35.8	35.6	33.4	38.0	46.0	53.4	51.8
Other mainland provinces (breakdown below)	60				50.8	47.3	46.1	47.3	46.8	48.3	41.6	43.2
Fujian					0.6	0.3	0.3	0.5	0.2	0.3	0.3	0
Guangxi					2.8	2.8	2.7	2.7	2.6	1.9	1.8	1.5
Hainan					0.3	0.3	0.3	0.2	0.2	0.4	0.4	1.7
Hunan	80.7 ^c	79.8	79.5		11.8	12.9	12.7	12.3	12.1	9.6	5.7	13.4
Hubei					8.9	7.1	6.3	6.5	6.8	6.0	4.5	4.2
Zhejiang					7.4	5.4	4.9	5.0	3.7	4.0	2.5	2.5
Jiangxi					6.9	7.0	7.7	9.5	10.8	15.3	15.8	15.1
Henan					7.7	7.9	7.8	8.3	8.4	9.6	9.3	3.5
Shanxi					0.01	0.02	0.1	0.1	0.1	0.03	0	0
Shanghai					4.4	3.6	3.2	2.0	1.6	1.2	1.1	1.1
Chongqing					0	0	0	0.2	0.3	0.1	0.2	0.2

^a reference 9. ^bGuangdong also includes Shenzhen and Zhuhai prefecture cities.

^cBreakdown of mainland provinces is not available for years 2000–2002.

Table S2. Antigenic characterization of SwIV isolates measured by haemagglutinin inhibition assay.

Test antigens	Subtype	HA	Ferret antisera					
			4167/99	1304/03	1110/06	Cal/04/09	NS29/09	1559/08
			CS H1N1	CS H1N2	TRIG H1N2	H1N1/09	EA H1N1	EA* H1N1
Sw/HK/4167/99	H1N1	CS	<u>1:20480^a</u>	1:320	1:10240	1:2560	1:2560	<1:10
Sw/HK/1304/03	H1N2	CS	1:1280	<u>1:2560</u>	1:640	1:80	1:40	<1:10
Sw/HK/1110/06	H1N2	TRIG	1:40960	1:1280	<u>1:10240</u>	1:640	1:5120	<1:10
Cal/04/09	H1N1	pdm	1:640	1:640	1:2560	<u>1:1280</u>	160	<1:10
Sw/HK/NS29/09	H1N1	EA	1:640	1:160	1:1280	1:80	<u>1:10240</u>	<1:10
Sw/HK/1559/08	H1N1	EA	<1:10	<1:10	<1:10	<1:10	1:40	<u>1:5120</u>
Sw/HK/8512/01	H1N1	EA	1:10240	1:1280	1:5120	1:2560	1:10240	1:320
Sw/HK/1669/02	H1N1	EA	1:5120	1:1280	1:2560	1:1280	1:10240	1:160
Sw/HK/NS129/03	H1N1	EA	1:5120	1:1280	1:2560	1:1280	1:10240	1:160
Sw/HK/1716/06	H1N1	EA	1:2560	1:640	1:1280	1:640	1:2560	<1:10
Sw/HK/435/07	H1N1	EA	1:1280	1:320	1:640	1:320	1:5120	<1:10
Sw/HK/NS952/08	H1N1	EA	1:1280	1:640	1:640	1:320	1:5120	<1:10
Sw/HK/71/2009	H1N1	EA	1:640	1:160	1:640	1:320	1:1280	<1:10
Sw/HK/72/07	H1N1	EA* r ^b	<1:10	<1:10	<1:10	<1:10	1:40	<1:10
Sw/HK/247/09	H1N1	EA* r	<1:10	<1:10	<1:10	<1:10	1:40	1:2560
Sw/HK/NS613/09	H1N1	EA* r	<1:10	<1:10	<1:10	<1:10	1:10	1:2560
Sw/HK/2481/09	H1N1	EA* r	<1:10	<1:10	<1:10	<1:10	1:20	1:2560

Sw/HK/NS186/09	H1N1	EA* r	<1:10	<1:10	<1:10	<1:10	1:40	1:2560
Sw/HK/201/2010	H1N1	EA r	1:10240	1:640	1:2560	1:1280	1:10240	1:160
Sw/HK/715/2008	H1N1	EA r	1:320	1:1280	1:320	1:160	1:40	<1:10
<hr/>								
Sw/HK/43/77	H1N1	CS	1:5120	1:80	1:1280	1:640	1:640	<1:10
Sw/HK/1767/94	H1N1	CS	1:10240	1:160	1:1280	nd	1:640	<1:10
Sw/HK/5609/99	H1N2	CS r	1:10240	1:640	1:640	1:1280	1:640	<1:10
Sw/HK/NS837/01	H1N1	CS r	1:5120	1:80	1:1280	1:320	1:640	<1:10
Sw/HK/NS728/02	H1N2	CS r	1:10240	1:160	1:1280	1:320	1:640	<1:10
Sw/HK/1890/09	H1N1	CS r	1:5120	1:320	1:5120	1:1280	1:1280	<1:10
Sw/HK/1479/09	H1N2	CS r	1:5120	1:160	1:5120	1:320	1:640	<1:10
<hr/>								
Sw/HK/NS623/02	H1N2	TRIG	- ^c	-	1:1280	1:320	-	-
Sw/HK/1578/03	H1N2	TRIG	-	-	1:320	<1:10	-	-
Sw/HK/1435/09	H1N2	TRIG	-	-	1:2560	1:1280	-	-
Sw/HK/2314/09	H1N2	TRIG r	-	-	nd	nd	-	-
Sw/HK/1111/04	H1N2	TRIG r	-	-	1:1280	1:320	-	-
Sw/HK/78/03	H1N2	TRIG r	1:20	<1:10	<1:40	<1:10	<1:10	<1:10
<hr/>								
Sw/HK/2299/09	H1N1	pdm	1:2560	1:320	1:1:160	1:1280	1:320	1:10

^aUnderlined values represent homologous antigenic titres. ^br represents viruses that have a haemagglutinin gene derived from EA, CS or TRIG lineage, but are reassortants with one or more remaining genes derived from other lineages. EA* r denotes Sw/HK/72/2007-like EA reassortant viruses. See Table S5 (above) for gene sources of all viruses. ^cHI assay was not conducted against some antisera.

Table S3. Percentage of swine serum positive for representative SwIV antibodies during 2000, 2004 and 2009.

Virus antigen tested	Lineage ^a	% sero-prevalence HI antibody titre $\geq 40^b$			
		2000 ^c (n=221)	2004 (n=240)	2009 (n=360)	2010 (Jan-Jul) (n=170)
Sw/HK/PHK4167/1999 (H1N1)	CS	32.3	40.4	36.7	22.9
Sw/HK/PHK1110/2006 (H1N2)	TRIG	36.3	32.9	38.3	33.5
California/4/2009 (H1N1)	H1N1/2009	20.9	14.6	18.1	32.4
Sw/HK/NS29/2009 (H1N1)	EA	15.4	26.3	46.1	31.8
Sw/HK/1559/2008 (H1N1)	EA* r	5.0	12.2	27.8	31.8
Sw/HK/1128/2004 (H3N2)	Sydney-like human H3N2	19.0	- ^d	-	0
Sw/HK/5212/1999 (H3N2)	Victoria-like human H3N2	2.5	-	-	0
Japanese white-eye/HK/1038/2006 (H5N1)	Avian H5N1	-	-	0	-

^a Major HA lineages of SwIV. ^b includes sera that were positive for more than one antisera, hence total seroprevalence adds to more than 100% (also see below). ^cyear of serum collection. ^dnot tested. EA* r denotes Sw/HK/72/2007-like EA reassortant viruses.

Table S4. Reactivity profile of swine sera to H1 subtype influenza viruses of different lineages.

Virus	Virus lineage	% with homologous reaction pattern ^a in HI tests			
		2000	2004	2009	2010 (Jan-Jul)
		(n=221)	(n=240)	(n=360)	(n=170)
(Sw/HK/4167/1999 or Sw/HK/1110/2009 or California/4/2009)	CS, TRIG and pH1N1/2009	39.3	30.4	14.7 (0.3)	22.9 (5.3)
California/4/2009	pH1N1/2009	0	0	0.28	5.3
Sw/HK/NS29/2009	EA	0.5	7.1	23.1	12.9
Sw/HK/1559/2008	EA* r	0	0.8	2.2	7.1
Indeterminate		2.0	14.6	21.7	16.5
Negative		58.2	46.7	38.3	40.6

^a Percentages were calculated only for sera that show ≥ 4 fold higher titers to a specific virus (or virus group in the case of CS, TRIG and H1N1/09 viruses) compared to other viruses tested. For example serum showing ≥ 4 fold higher titers to EA reassortants virus (EA* r) (Sw/HK/1559/08) than other viruses are listed above.

Table S5. Isolation dates, gene sources and molecular markers of 230 fully sequenced SwIV isolates.

Table S6. Estimates of selective pressure in SwIV.

Segment	Clade	d_N/d_S			Num. of sites under positive selection ^a
		mean	95% lower	95% upper	
PB2	TRIG+PDM	0.085	0.067	0.105	
	EA	0.084	0.070	0.100	
	CS	0.078	0.068	0.090	
PB1	TRIG+PDM	0.068	0.054	0.085	
	EA	0.110	0.095	0.127	1
	CS	0.074	0.065	0.086	3
PA	TRIG+PDM	0.118	0.098	0.142	
	EA	0.090	0.076	0.104	
	CS	0.125	0.110	0.141	1
HA ^b	EA all (HK 2001-2009)	0.204	0.178	0.233	
	EA pure (HK 2001-2009)	0.197	0.168	0.228	
	EA reassortants (HK 2001-2009)	0.205	0.149	0.275	
	EA (EU 1980-2000)	0.243	0.217	0.271	2
	EA (EU 2001-2008)	0.171	0.142	0.203	
	CS+TRIG+PDM	0.185	0.170	0.202	
NP	EA	0.084	0.065	0.105	
	CS+TRIG+PDM	0.069	0.058	0.080	
N1	EA+PDM	0.189	0.163	0.217	
	CS	0.237	0.210	0.268	3
N2	N2+TRIG	0.172	0.151	0.196	
M1	EA+PDM	0.029	0.015	0.048	
	CS+TRIG	0.071	0.053	0.093	
M2	EA+PDM	0.527	0.346	0.763	
	CS+TRIG	0.334	0.256	0.427	
NS1	EA	0.300	0.228	0.385	
	CS+TRIG+PDM	0.235	0.201	0.273	
NS2	EA	0.183	0.084	0.342	
	CS+TRIG+PDM	0.331	0.246	0.434	1

^aNumber of positively selected amino acid positions ($p > 0.01$); ^b d_N/d_S values were also estimated for European EA viruses (**bold**) isolated from 1980–2000 and 2001–2009, in order to understand host adaptation mechanisms; All Hong Kong EA viruses, pure HK EA and HK EA reassortant (Sw/HK/72/2007-like) viruses were analysed independently.

Table S7. Seroprevalence of human sera to seasonal H1N1, pandemic H1N1 and swine influenza viruses by microneutralization tests.

% seropositive (titre≥40) to	lineage	Year of birth ^a					
		1992–09 (n=28)	1977–91 (n=52)	1968–76 (n=32)	1957–67 (n=91)	1944–56 (n=84)	Pre 1944 (n=46)
HK/400549/08	Seasonal H1N1	75	38.5	18.8	22	10.7	17.4
Cal/04/09	H1N1/2009	0	1.9	0	4.4	2.4	6.5
Sw/HK/4167/99	CS	0	3.8	9.4	5.5	6	13
Sw/HK/1304/03	CS reassortant	3.6	5.8	31.3	23.1	16.7	30.4
Sw/HK/1110/06	TRIG	7.1	9.6	12.5	6.6	7.1	17.4
Sw/HK/NS29/09	EA	0	0	0	5.5	2.4	10.9
Sw/HK/1559/08	EA reassortant	7.1	7.7	18.8	4.4	3.6	8.7
Sw/HK/201/10	EA-pH1N1	7.1	3.8	18.8	9.9	8.3	15.2

^aHuman sera were collected from community volunteers in 2009 prior to the spread of pandemic H1N1 in Hong Kong. Ages were categorised based on circulation of human influenza A virus. 1992–2009, pre influenza A H1N1/2009 emergence and post recent seasonal H1N1 circulation; 1977–1991, post re-emergence of seasonal H1N1; 1968–1976, post H3N2 pandemic emergence; 1957–1968, post H2N2 pandemic emergence.

Table S8: Seroprevalence of human sera to one or more swine influenza viruses by microneutralization tests^a.

% seropositive (titre ≥40) to	Year of birth ^b					
	1992–09 (n=28)	1977–91 (n=52)	1968–76 (n=32)	1957–67 (n=91)	1944–56 (n=84)	Pre 1944 (n=46)
	one or more viruses	25	21.1	50.0	29.7	27.3
two or more viruses	0	7.7	25.0	13.2	9.5	23.9
three or more viruses	0	1.9	12.5	5.5	4.8	19.6
four or more viruses	0	1.9	3.3	3.3	3.6	13.0

^aWhile cross reactive antibody to one SwIV is common, it is increasingly less common to have cross reactive antibody to multiple SwIVs. The prevalence of such broad cross reactivity increases with age and particularly so in the >65 year age group. ^bSee footnote^a of Table S7 for description of age categories.

3. Supplementary notes

3.1 Background of swine influenza A virus diversity

Currently available SwIV genomes from different locations suggest that multiple H1 and H3 subtype lineages co-circulate in swine, including strains with avian or human origins^{9–12}. One particular lineage of H1N1 viruses, termed classical swine (CS) influenza, has been isolated continuously from swine since the 1930s¹³ and is related to the H1N1/1918 pandemic virus¹⁴. Since 1918, avian H1N1 lineages have become established in swine on a number of occasions. Around 1979, a purely avian H1N1 virus (known as the European or Eurasian avian (EA) strain) was detected in swine in Europe. This virus, first reported in Belgium¹⁵, replaced the CS H1N1 strain previously endemic in parts of Europe and became widespread in European pigs¹⁶. During the mid-1980s, the EA virus reassorted with human H3N2 (A/Victoria/75-like) viruses in pigs¹⁷, generating viruses whose surface protein genes were of human origin, while the internal protein genes remained of avian origin. These H1N1, H1N2 and H3N2 subtype EA viruses continue to circulate in swine in Europe¹⁸, and there is evidence of continued generation of novel reassortants in that region¹⁹.

Selected antigenic-drift variants of human H3N2 viruses have established themselves in pigs (including A/Hong Kong/2/68-like, A/Victoria/75-like and A/Sydney/5/1997-like viruses)^{9,20,21} and have subsequently reassorted with other swine viruses. In 1998, an H3N2 virus whose surface proteins were derived from a contemporary human virus (A/Sydney/5/97-like) underwent reassortment in North America swine, first with a CS H1N1 virus, then subsequently with an avian influenza virus, giving rise to the triple reassortant H3N2 viruses, which have internal protein genes originating from North American avian (PA, PB2), human (PB1), and CS H1N1 (NP, M, NS) influenza viruses. The H3N2 triple-reassortant viruses caused outbreaks of clinically overt disease in North American swine and have been endemic in those

populations^{11,12}. The triple reassortant internal gene (TRIG) virus has proven adept at acquiring HA and NA genes from other sources, thus generating triple reassortant H1N1 and H1N2 swine viruses²², some of which have caused sporadic zoonotic human infections in North America²³. Subsequently, H1 and H3 TRIG SwIV have been detected in swine in various Asian countries^{9,24–26}. Co-infection of pigs with CS and TRIG viruses have shown TRIG to have a competitive advantage over other generated reassortants²⁷.

The 2009 pandemic virus was generated²⁴ by reassortment between TRIG viruses circulating in North American swine and EA viruses in European pigs. Hitherto, EA viruses had not been detected in the Americas, but the lack of data from Mexico, Central and South America means that their absence from the region is not certain.

3.2 Swine serology

Herd immunity may contribute to the establishment or decline of different viruses in swine. We used the haemagglutination assay to quantify the changes in seroprevalence in serum collected from swine during 2000, 2004, 2009 and 2010 to 5 representative viruses including the antigenically divergent Sw/Hong Kong/72/2007-like EA viruses (Table S3, S4). We also tested the 2000 and 2010 swine serum against two H3N2 human lineage viruses and 2009 swine serum against a highly pathogenic avian influenza H5N1 subtype virus (clade 2.3.4).

The low prevalence of antibody reactivity to EA viruses in 2000–2004 may explain the ease with which EA viruses established themselves in swine in this region, becoming progressively dominant in the early 2002–5 period. Similarly, the low seroprevalence to the antigenically-variant EA virus Sw/HK/1559/2008 with a TRIG NS gene; 5% in 2000, 12% in 2004 and 28% in 2009 may explain the emergence of this antigenically-variant sub-lineage in 2009. The 15–21% seroprevalence to pandemic

H1N1/2009 seen in sera collected in 2000 and 2004 likely represents serological cross-reactivity generated by the TRIG H1 viruses. Although pandemic H1N1/2009 has been repeatedly detected in pigs in this region since October 2009¹⁰, it was never detected prior to 2009. The observed low level of serologic cross-reactivity to pandemic H1N1/2009 suggests that this virus may become endemic in swine populations in the region. Serological analysis of swine sera collected in 2009/10 confirms the rarity of infection with human-like H3N2, Eurasian avian-like H3N2 and avian H5N1 viruses.

Sera often contained antibody to more than one virus, representing either dual infections or antigenic cross-reactivity across viruses (Table S4). In particular, sera that were positive against Sw/HK/4167/1999 (CS) were also often positive for Sw/HK/1110/2006 (TRIG), indicating a high level of cross-reactivity among these viruses. To attempt to identify the infecting virus more precisely, we analysed the serological reaction profiles to determine sera with ≥ 4 -fold higher titre to one virus compared to all other viruses (Table S4). Sera with an indeterminate reaction profile (i.e. with antibody titres to different viruses within 4-fold of each other) was low (2%) in 2000, a period when CS viruses were the only virus lineage detected by virus isolation. The indeterminate reaction profile increased from 2004 when multiple virus lineages were co-circulating in swine. A gradual decline in homologous reaction pattern to CS/TRIG/H1N1/2009 was observed until 2009 reflecting the decline in prevalence of CS viruses. There is however a sharp increase in seropositivity to this serogroup of viruses in 2010, most likely reflecting the introduction of pH1N1/2009 viruses into swine in the region since early 2010¹⁰. Sera with a homologous reaction to pandemic H1N1/2009 virus alone increased from 0% in 2000 and 2004 to 5.9% in 2010, supporting the contention that this virus has led to widespread infection of swine. In summary, 83 (41.8%) of 221 sera collected in 2000, 53.3% of 240 sera in 2004, 61.7% of 360 sera collected in 2009 and 59.4% of 170 sera collected during January–July 2010 have detectable antibody titres (≥ 40) to one or more swine viruses (Table S4). Thus,

although virus isolation rates have declined after the dominance of the EA lineage since 2005 (main text Fig. 1), these results show a gradual increase in seroprevalence to SwIV over the period of study.

3.3 Detailed molecular characterization

Full molecular characterization of amino acid residues at predicted antigenic sites, plus other host adaptation, drug resistance and virulence markers are listed in Table S5. The H1 and N1 numbering systems are used below for all haemagglutinin and neuraminidase residues, respectively.

Receptor binding domain and antigenic sites. Amino acid residues at receptor binding pocket of HA1 position Gln 226 (H1 numbering 223) and Gly 228 (H1 numbering 225) retain configurations (2,6-NeuAcGal linkages) predicted to have affinity for mammalian cell-surface receptors in all swine viruses fully sequenced from surveillance in Hong Kong²⁸.

Amino acid residues relevant to antigenic determinant domains²⁹ were conserved in the majority of CS viruses isolated from 1977-1993, however during 1998-2004 multiple variants of the antigenic determinant domain of CS viruses were detected. Three of these mutations (Gly 222 Asn/Asp (also receptor binding), Val 166 Ile and Val 73 Ala/Lys) had risen and became conserved in all CS viruses, including the TRIG viruses whose HA is derived from the CS lineage. This time corresponds with the dynamic period during which multiple antigenic variants of the CS lineage, including the TRIG, were identified in North American swine populations.

Two distinct groups of the antigenic determinant domains were observed for the EA viruses (Fig S5, Table S2). ‘Pure’ EA viruses (EA viruses with all gene segments derived from the avian gene pool) isolated from 2001 to 2009 were fairly conserved

with few random mutations observed. However the EA reassortant (Sw/Hong Kong/72/2007-like) viruses showed multiple non-synonymous substitutions in several antigenic sites, including antigenic site Sa: Gly155Glu, Ser162Asn, antigenic site Sb: Lys153Gln, Thr190Ala, Gln193Arg, antigenic site Ca: Asn142Lys, antigenic site Cb: Asn74Asp (Table S6), reflecting the lack of antigenic cross-reactivity towards other influenza viruses, including the ‘pure’ EA viruses. Progressive accumulation of amino acids in the EA reassortant viruses are marked on branches in Fig. S3a and marked on the crystal structure of the HA in Fig. S6. These reassortants form an independent phylogenetic lineage that is derived from the EA viruses suggesting that they represent a diversifying lineage.

Asp222Gly mutation. Mutation Asp222Gly in the HA protein has been reported in some patients with severe or fatal H1N1/2009 infection, however its significance is unclear³⁰. To visualize changes occurring at position 222 during the evolutionary history of this virus in swine, we mapped the amino acid changes at this residue on the HA phylogeny. Observed mutations at this position indicate that CS viruses isolated during 1977-1993 in Hong Kong all contained Gly at this position, whereas majority of CS isolated since 1998, and all TRIG and H1N1/2009 viruses have Asp at this position. It is interesting to note that changes Gly222Asn and Gly222Asp are associated with relatively longer branches on the phylogeny, indicating that these mutations may be associated with major antigenic and genetic change or reassortment that have occurred in this virus lineage.

The residues at position 222 on the EA lineage are specific to geographic areas. Viruses of the avian gene pool contain Gly-222, and viruses isolated from swine in Europe maintain this mutation, however the majority of EA viruses isolated from swine surveillance in Hong Kong have Glu-222. Interestingly swine viruses isolated in Europe that fall within the HK lineage (most likely derived from the Asian EA viruses) contain

Gly- and Lys-222. The independent emergence of mutations at position 222 in multiple swine influenza lineages suggests this may not be a host restriction factor, further highlighting the confusion regarding its importance. A thorough *in vitro* and *in silico* investigation of these mutations is warranted.

Drug resistance mutations. All CS and TRIG viruses exhibited an Asn31Ser mutation in the M2 protein that invariably confers resistance to the adamantanes, a group of antiviral drugs used for treatment of human influenza³¹. However all viruses that had an M gene derived from the EA lineage and one CS H1N2 virus (Sw/HK/NS915/04) maintained Asn at position 31, indicating susceptibility to amantadine. Other mutations that have been associated with amantadine resistance include positions 26, 27, 30 and 34 in the M2 ion channel. Lys, Ala and Glye were maintained at positions 26, 30 and 34, respectively, indicating susceptibility for amandatine. However a Val to Ile mutation was observed at position 27 in several viruses of both major lineages.

Amino acids His and Asn were maintained at positions 274 and 294 of the NA gene, respectively, for all swine influenza A viruses sequenced in this study, indicating susceptibility to Oseltamivir.

Virulence markers. The presence of a basic amino acid at position 591 of the PB2 protein has been shown to increase replication of A/California/04/2009 (H1N1/2009) in the ferret model³², and is considered to compensate for the lack of Lys 627 and/or Asn 701 in the PB2 protein, which are considered critical for mammalian adaptation of avian influenza viruses. Similar to human H1N1/2009 viruses, an Arg was present at position 591 in all TRIG and swine H1N1/2009, but only in 5 recent EA viruses (Sw/Hong Kong/72/2007-like) isolated in 2009, including the novel reassortant virus (Sw/Hong Kong/201/2010). The CS viruses had a Glu (non-basic) at this position. Lys-627 and Asn 701 were present in majority of the viruses whose PB2 gene were derived from the

CS lineage, while the TRIG and H1N1/2009 whose PB2 gene has a North American avian origin did not have Lys 627, but contained Arg at 591. The absence of any of these mutations in EA viruses combined with their apparently increased fitness in Eurasian swine suggests that the contributing factors affecting replicative fitness of swine viruses are complex.

Host specific markers. The PDZ-ligand at the 3' of the NS1 gene of EA viruses showed a clear pattern of change with time, indicating that it is involved in host adaptation as previously described³³. Early EA viruses from Europe had the avian ESEV motif, with a gradual shift to GSEV/GPEV motifs observed from multiple hosts, before the majority of viruses since 1999 had the motif GPKV that has previously been described from pigs. CS and TRIG viruses that contributed the NS gene for H1N1/2009 have a truncated NS gene, as does the antigenically variant Sw/Hong Kong/72/2007-like viruses.

Amino acid residues in the polymerase genes are presumed to distinguish host type (human, avian, swine and equine isolates)³⁴. We compared these amino acids across all swine influenza A viruses isolated in our study to identify any host association among these markers (Table S5). These include PB2 mutations 199, 475, 567, 591, 627, 701 and 702, PB1 mutation 375 and PA mutations 55, 100, 382 and 552. Residue analyses of SwIV sequences suggests that these previously-identified residues in PB2 and PA genes were highly conserved resulting in lineage specific residues. Importantly, these mutations were maintained even after multiple interspecies transmission events among birds, humans and swine, indicating that these residues may not be host specific. For example, EA viruses that were directly introduced to pigs from the Eurasian avian gene pool maintained purported avian residues despite circulating in pigs for over 30 years. Interestingly, amino acid changes at position 375 of the PB1 gene coincided with the phylogenetic relationship of the viruses HA gene and mutations at known antigenic sites indicating evidence of compensatory mutation.

The PB1-F2 protein, which is encoded by an overlapping (+1) open reading frame of the PB1 gene, has been implicated with virulence in mammals³⁵. The majority of avian isolates express a full length PB1-F2 protein (87–91 amino acid). The H1N1/1918 pandemic virus expresses a full-length PB1-F2 protein, however is truncated in human H1N1 viruses isolated since 1948–50. Human H3N2 viruses maintain the full-length avian-derived PB1-F2 protein.

The PB1-F2 protein of the swine isolates (Table S5) are truncated at various sites, except for viruses with a PB1 gene derived directly from avian or human H3N2 clades that have a full length PB1-F2 protein. Truncation at position 12 is seen in all EA viruses isolated in our study, however full-length PB1-F2 is maintained in a majority of all other EA viruses. PB1-F2 truncation at positions 12, 26 and 80 are sporadically observed in European EA viruses. H1N1/2009 and TRIG viruses isolated from swine are also truncated at positions 12 and 58 of the PB1-F2 proteins, respectively. However, the majority of North American TRIG viruses maintain a full length (91 amino acid) PB1-F2 protein. This phenotype continues to be isolated in North American swine (e.g. Sw/Oklahoma/020736-1/2008) with truncation at positions 26, 58 and 60 only rarely detected.

While the full-length PB1-F2 continues to be isolated in European and North American swine, these viruses were predominantly isolated during disease outbreak investigations. In contrast, PB1-F2 truncation was observed in almost all viruses isolated from systematic surveillance of apparently healthy pigs in our study. Taken together, these findings correspond to the observed increased virulence associated with full-length PB1-F2. Furthermore, it is also possible that viruses with various PB1-F2 truncations may be circulating at a higher-rate than detected in North American and European pig populations, however asymptotically, highlighting the importance of

influenza surveillance to capture the full genetic diversity of SwIV. Alternatively, it has been argued that the PB1-F2 genetic variation is under weak or no selective constraint³⁶.

Other amino acid mutations in Sw/Hong Kong/72/2007-like EA viruses. In addition to the mutations in the antigenic determinant domain of the HA gene of Sw/Hong Kong/72/2007-like viruses, a number of amino acid substitutions were observed in the remaining genes and those that are unique when compared to EA viruses are listed: PB2: His127Asn, Val495Ala, Gln591Lys; PB1: Ser154Asn, Ile182Lys, Phe254Thr, Asp375Tyr; PA: Val379Met; NP: Asn319Lys; NA: Phe339Ser, Arg382Gly.

3.4 Molecular evolution

The establishment of EA lineage viruses in swine provides an opportunity to investigate the molecular mechanisms of adaptation of an avian influenza virus to a mammalian host. We compared the ratio of non-synonymous to synonymous substitution (d_N/d_S) of EA virus HA genes isolated in Europe shortly after inter-species transmission of an avian H1N1 virus to swine (1980–2000) with those isolated later in swine from Europe (2001–2009) (Table S6). The d_N/d_S ratio was modestly but significantly higher ($p < 0.01$) for EA SwIV isolated earlier ($d_N/d_S = 0.24$; 95% CIs = 0.22–0.27) than for later EA viruses ($d_N/d_S = 0.17$; 95% CIs = 0.14–0.20), consistent with the hypothesis that host-specific selection pressure resulted in an increase in viral adaptation after the introduction of EA viruses to swine.

Overall, SwIV of all the lineages we studied were under relatively strong purifying selection (Table S6). Eight amino acid residues across the PB1, PA, N1 and NS2 proteins of CS viruses were inferred to be under positive selection ($p < 0.1$), while only one residue (position 578, $d_N/d_S = 2.3$, $p < 0.1$) in the PB1 protein of the EA viruses was positively selected.

3.5. Seroprevalence in humans

Lack of cross-neutralizing antibody to the H1N1/2009 virus allowed this virus to cause a pandemic. It is therefore relevant to define the herd-immunity in human populations to swine virus lineages identified in this study. Seroprevalence to a recent seasonal influenza (H1N1) virus, the pandemic H1N1 virus and representatives of the different swine influenza viruses was determined by microneutralization tests in a panel of human sera stratified by age collected prior to the first wave of the 2009 H1N1 pandemic in Hong Kong. As expected, seroprevalence to recent seasonal influenza H1N1 virus was highest in children and that for the pandemic H1N1 virus was low in all age groups, with the exception of those born prior to 1944 (Table S7). It is notable that sero-prevalence to a number of these swine viruses was low in persons born after 1977 (Table S7, S8), especially so for EA virus Sw/HK/NS29/2009. There appears to be higher sero-prevalence to swine influenza viruses in those born between 1968-1976 (i.e. those who may have been infected with the re-emergence H1N1 virus in 1977) and also in those born pre-1944. Overall, these results indicate that some of the swine haemagglutinins may pose pandemic threat if acquired through reassortment with a virus adapted for efficient transmissibility in humans, e.g. the pandemic H1N1/2009 virus that has been repeatedly detected in swine in many parts of the world.

4. Supplementary References:

1. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
2. Smith, D. J. *et al.* Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**, 371–376 (2004).
3. Cai, Z., Zhang, T. & Wan X. F. A computational framework for influenza antigenic cartography. *PLoS Comput. Biol.* **6**, e1000949 (2010).

4. Humphrey, W., Dalke, A., & Schulten, K. VMD – visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
5. Stone, J. An efficient library for parallel ray tracing and animation. Thesis, University of Missouri-Rolla (1998).
6. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. P. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
7. Rambaut, A. *et al.* The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453**, 615–619 (2008).
8. Russell, C. A. *et al.* The global circulation of seasonal influenza A (H3N2) viruses. *Science* **320**, 340-346 (2008).
9. Peiris, J. S. M. *et al.* Cocirculation of avian H9N2 and contemporary "human" H3N2 influenza A viruses in pigs in southeastern China: potential for genetic reassortment? *J. Virol.* **75**, 9679–9686 (2001).
10. Vijaykrishna, D. *et al.* Reassortment of pandemic H1N1 viruses in swine. *Science* **328**, 1529 (2010).
11. Zhou, N. N. *et al.* Genetic reassortment of avian, swine, and human influenza A viruses in American pigs. *J. Virol.* **73**, 8851–8856 (1999).
12. Webby, R. J. *et al.* Evolution of swine H3N2 influenza viruses in the United States. *J. Virol.* **74**, 8243–8251 (2000).
13. Shope R. E. Swine Influenza. *J. Exp. Med.* **54**, 349–359 (1931).
14. Smith G. J. D. *et al.* Dating the emergence of pandemic influenza A viruses. *Proc. Natl. Acad. Sci. USA* **106**, 11709–11712 (2009).

15. Pensaert, M., Ottis, K., Vanderputte, J., Kaplan, M. M. & Buchmann, P. A. Evidence for the natural transmission of influenza A virus from wild ducks to swine and its potential for man. *Bull. World Health Organ.* **59**, 75–78 (1981).
16. Scholtissek, C., Hinshaw, V. S. & Olsen, C. W. Influenza in pigs and their role as the intermediate host. In: Nicholson KG, Webster RG, Hay AJ (eds.), *Textbook of Influenza*, Blackwell Scientific, Oxford, 137–145 (1998).
17. Castrucci, M. R. *et al.* Genetic reassortment between avian and human influenza A viruses in Italian pigs. *Virology* **193**, 503–506 (1993).
18. Van Reeth, K. *et al.* Seroprevalence of H1N1, H3N2 and H1N2 influenza viruses in pigs in seven European countries in 2002–2003. *Influenza and Other Respiratory Viruses* **2**, 99–105 (2008).
19. Bálint, A. *et al.* The first Swedish H1N2 swine influenza virus isolate represents an uncommon reassortant. *Virology J.* **6**, 180 (2009).
20. Kundin, W. D. Hong Kong A2 influenza virus infection among swine during a human epidemic in Taiwan. *Nature* **228**, 957–958 (1970).
21. Shortridge, K. F., Webster, R. G., Butterfield, W. K. & Campbell, C. H. Persistence of Hong Kong influenza virus variants in pigs. *Science* **4297**, 1454–1455 (1977).
22. Karasin, A. I. *et al.* Genetic characterization of H1N2 influenza A viruses isolated from pigs throughout the United States. *J. Clin. Microbiol.* **40**, 1073–1079 (2002).
23. Shinde, V. *et al.* Triple-reassortant swine influenza A (H1) in humans in the United States, 2005–2009. *N. Engl. J. Med.* **360**, 2616–2625 (2009).
24. Smith, G. J. D. *et al.* Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **459**, 1122–5 (2009).
25. Yu, H. *et al.* Genetic evolution of swine influenza A (H3N2) viruses in China from 1970–2006. *J. Virol.* **46**, 1067–1075 (2008).

26. Bi, Y. *et al.*, Novel swine influenza virus reassortants in pigs, China. *Emerg. Infect. Dis.* **16**, 1162–1164 (2010).
27. Ma, W. *et al.* Viral reassortment and transmission after coinfection of pigs with classical H1N1 and triple reassortant H3N2 swine influenza viruses. *J Gen. Virol.* **91**, 2314–2321 (2010).
28. Rogers, G. N. *et al.* Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* **304**, 76–78 (1983).
29. Wiley, D. C., Wilson, I. A. & Skehel, J. J. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature* **289**, 373–378 (1981).
30. World Health Organization. Preliminary review of D222G amino acid substitution in the haemagglutinin of pandemic influenza A(H1N1) 2009 viruses. *Wkly Epidemiol Rec.* **85**, 21-22 (2010).
31. Scholtissek, C., Quack, G., Klenk, H. D. & Webster, R. G. How to overcome resistance of influenza A viruses against adamantane derivatives. *Antiviral Res.* **37**, 83–95 (1998).
32. Yamada, S. *et al.* Biological and Structural Characterization of a host-adapting amino acid in influenza virus. *PLoS Pathog.* **6**, e1001034 (2010).
33. Obenauer, J. C. *et al.* Large-scale sequence analysis of avian influenza isolates. *Science* **311**, 1576–1580 (2006).
34. Taubenberger, J. K., Reid, A. H., Lourens, R. M., Wang, R., Jin, G. and Fanning, T. G. (2005) Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889–893 (2005).
35. Chen, W. *et al.* A novel influenza A virus mitochondrial protein that induces cell death. *Nat. Med.* **7**, 1306–1312 (2001).

36. Holmes, E. C., Lipman, D. J., Zamarin, D. & Yewdell, J. W. Comment on “Large-scale sequence analysis of avian influenza isolates”. *Science* **313**, 1573b (2006).