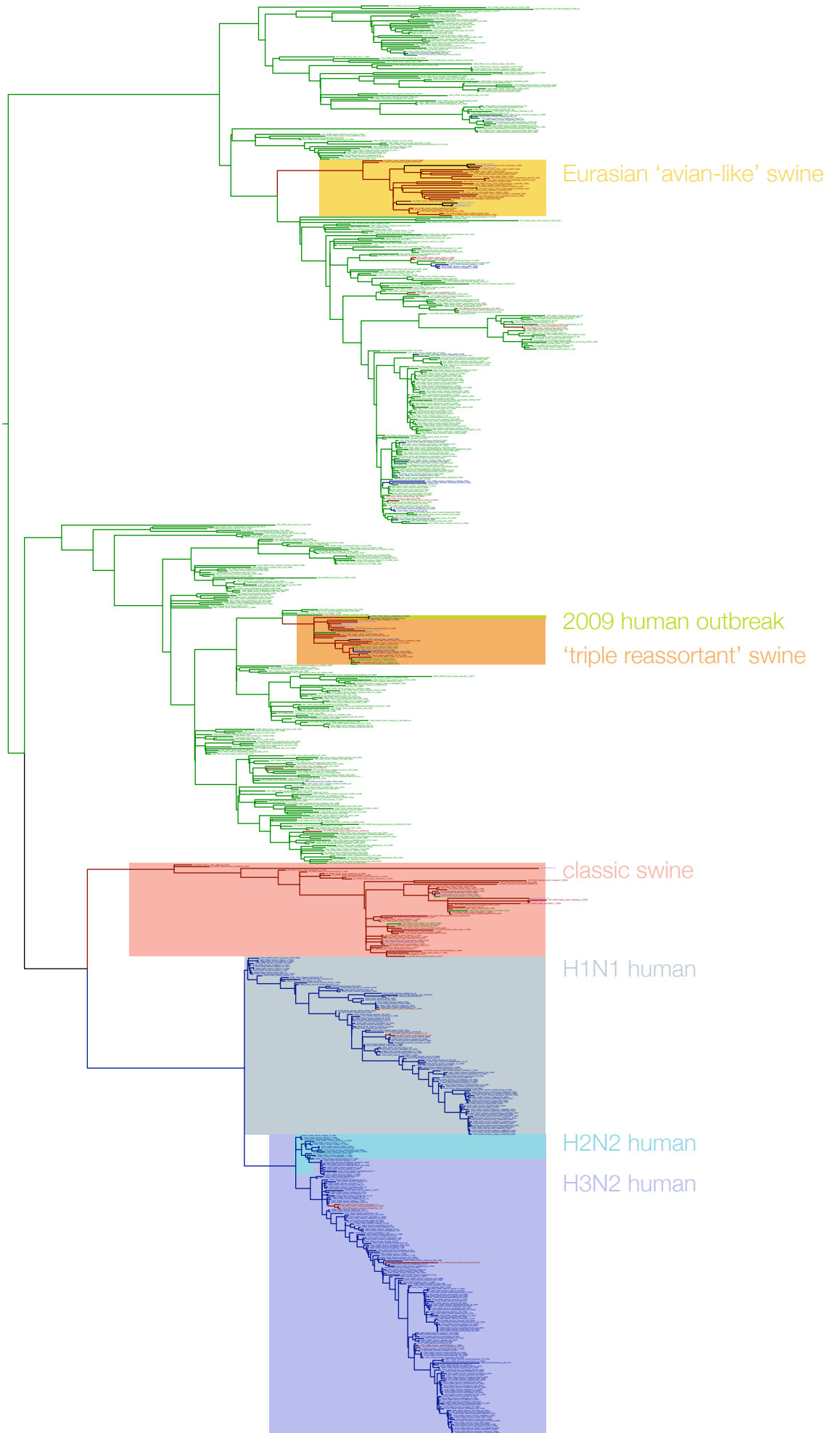
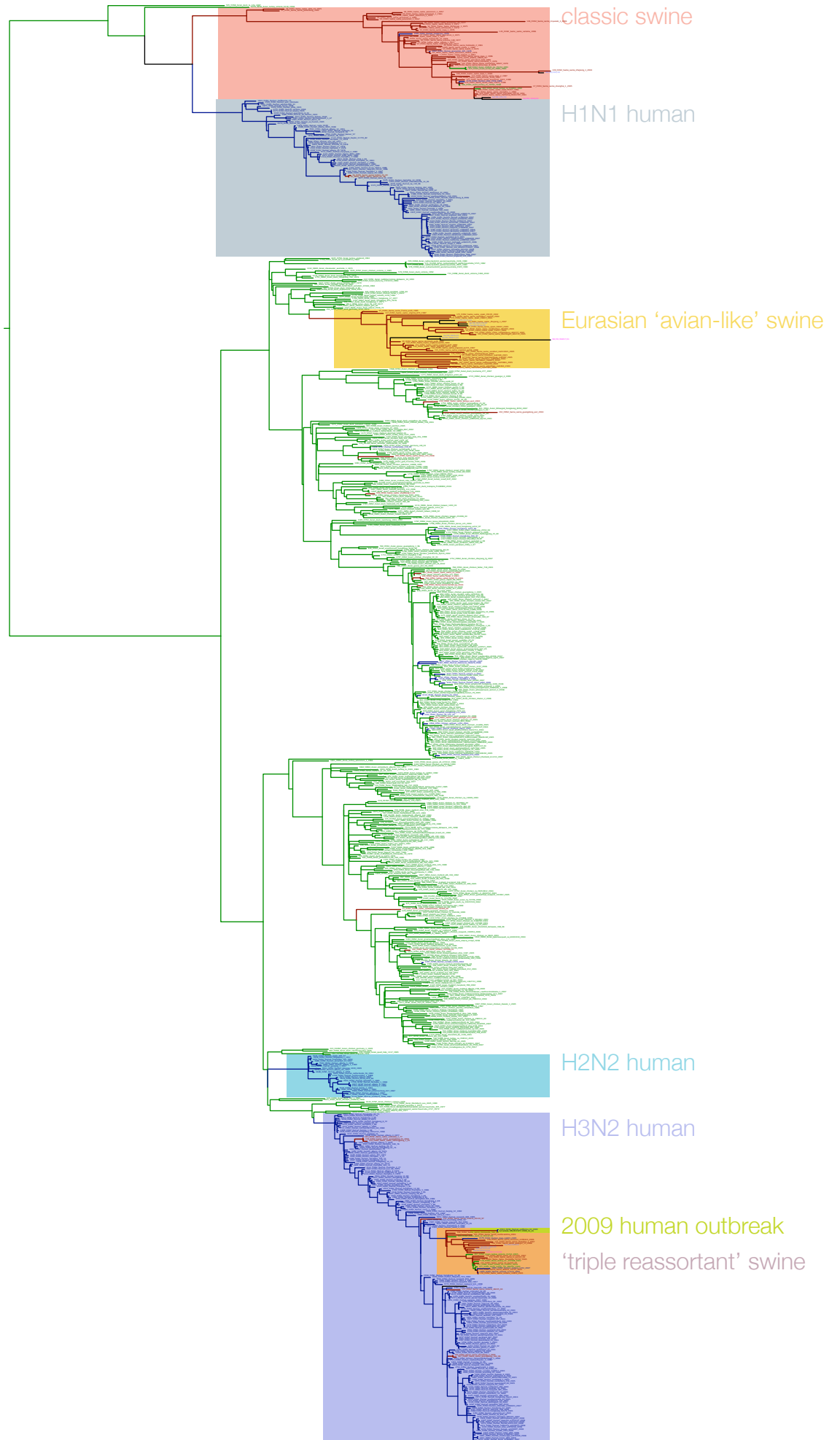


Supplementary Figure 1. Phylogenetic relationships of each gene segment (PB2, PB1, PA, HA, NP, NA, M & NS) of swine influenza A viruses indicating genetic components of the swine-origin influenza A (H1N1) virus. Clade labels indicate major swine lineages. Human viruses are coloured blue, swine viruses in red and avian viruses in green.



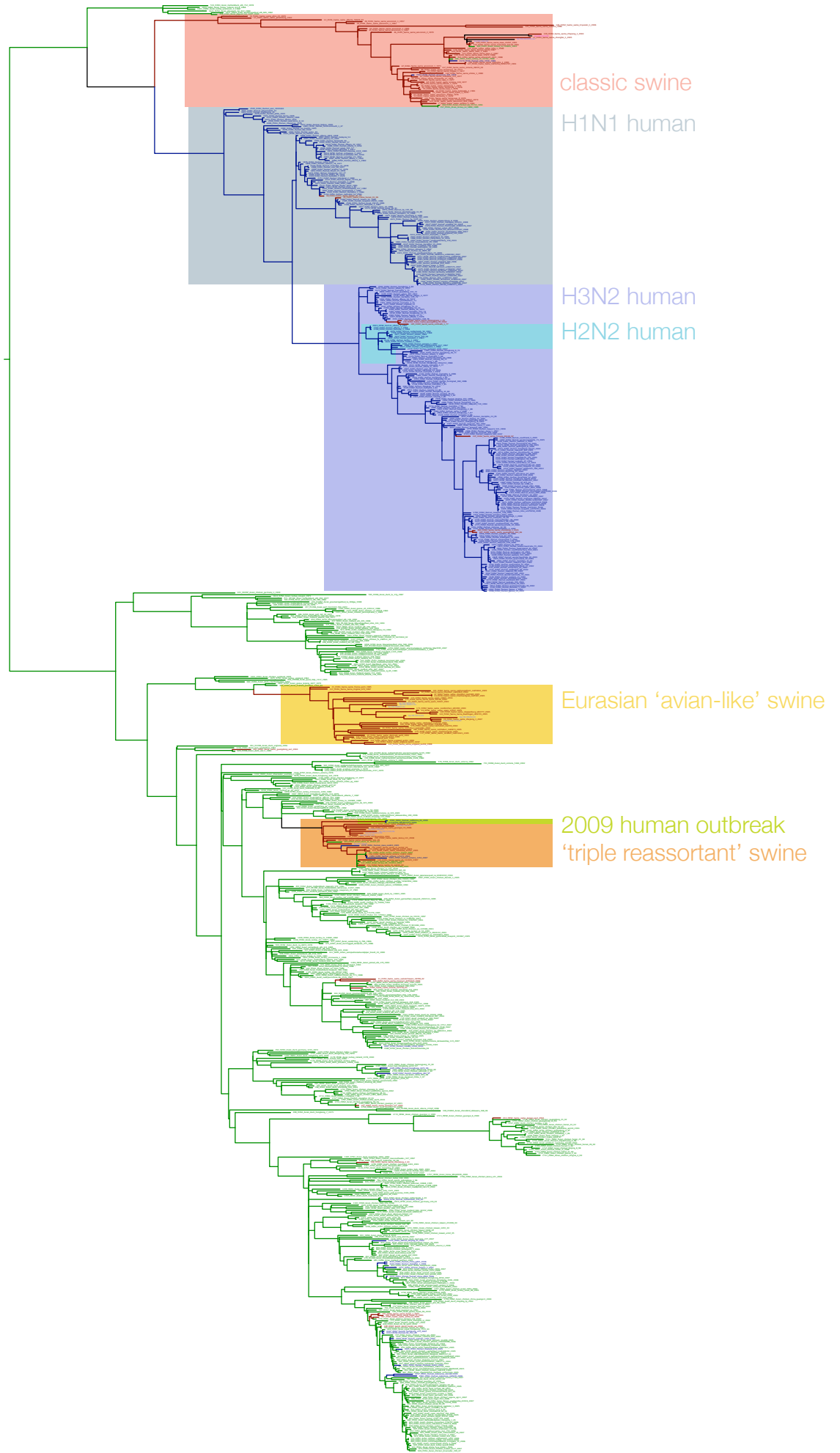
0.03

PB1



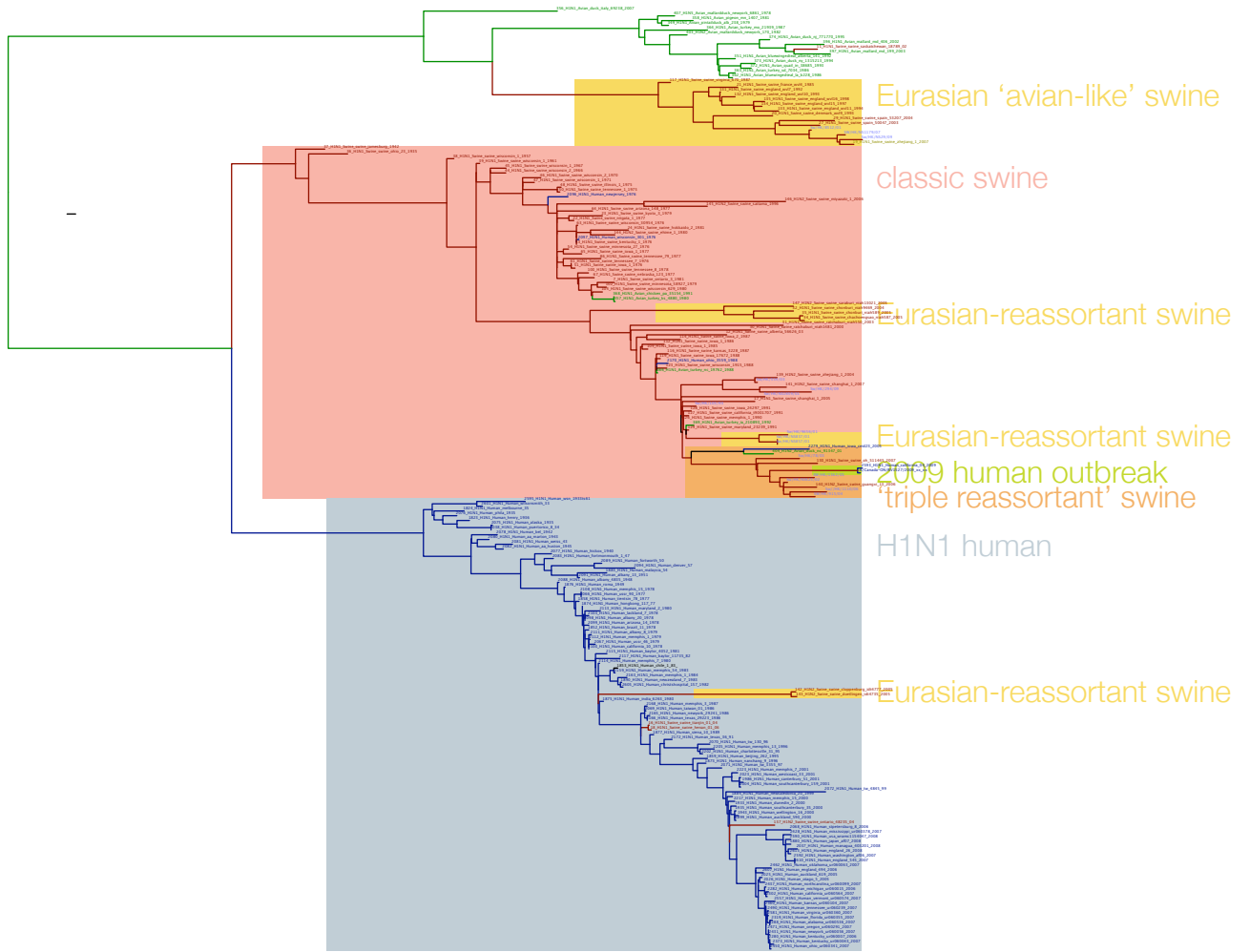
0.05

PA



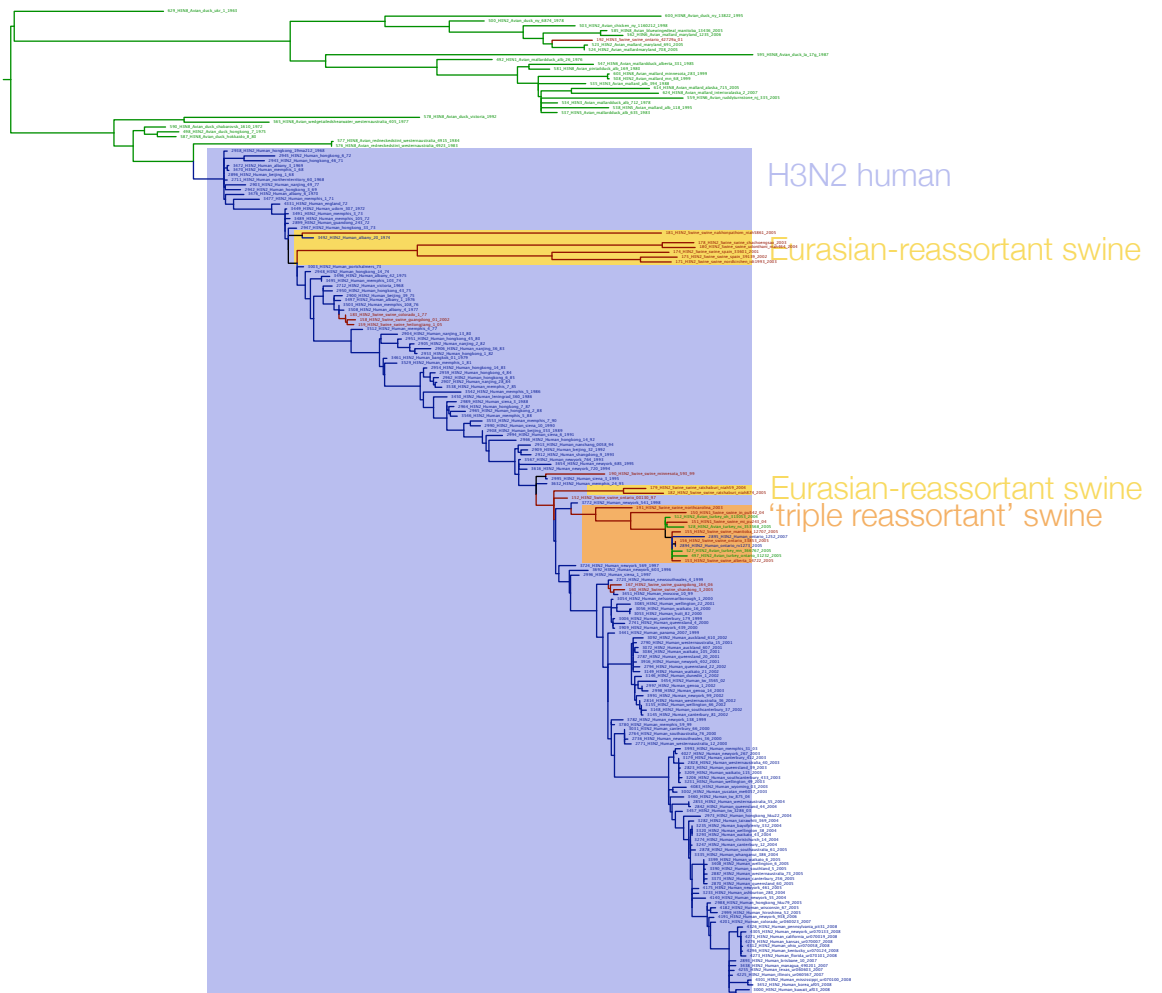
0.05

HA H1



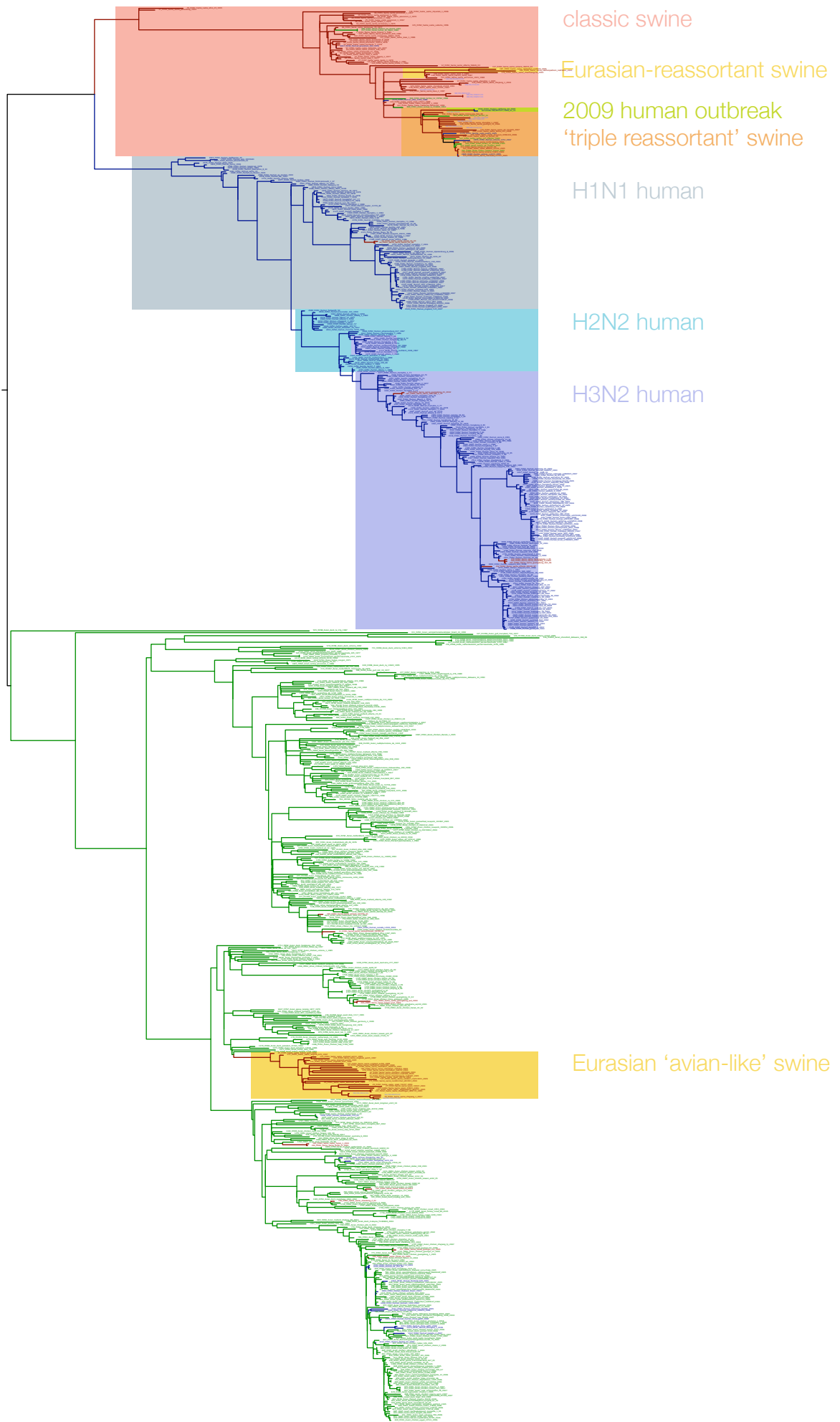
0.07

H3



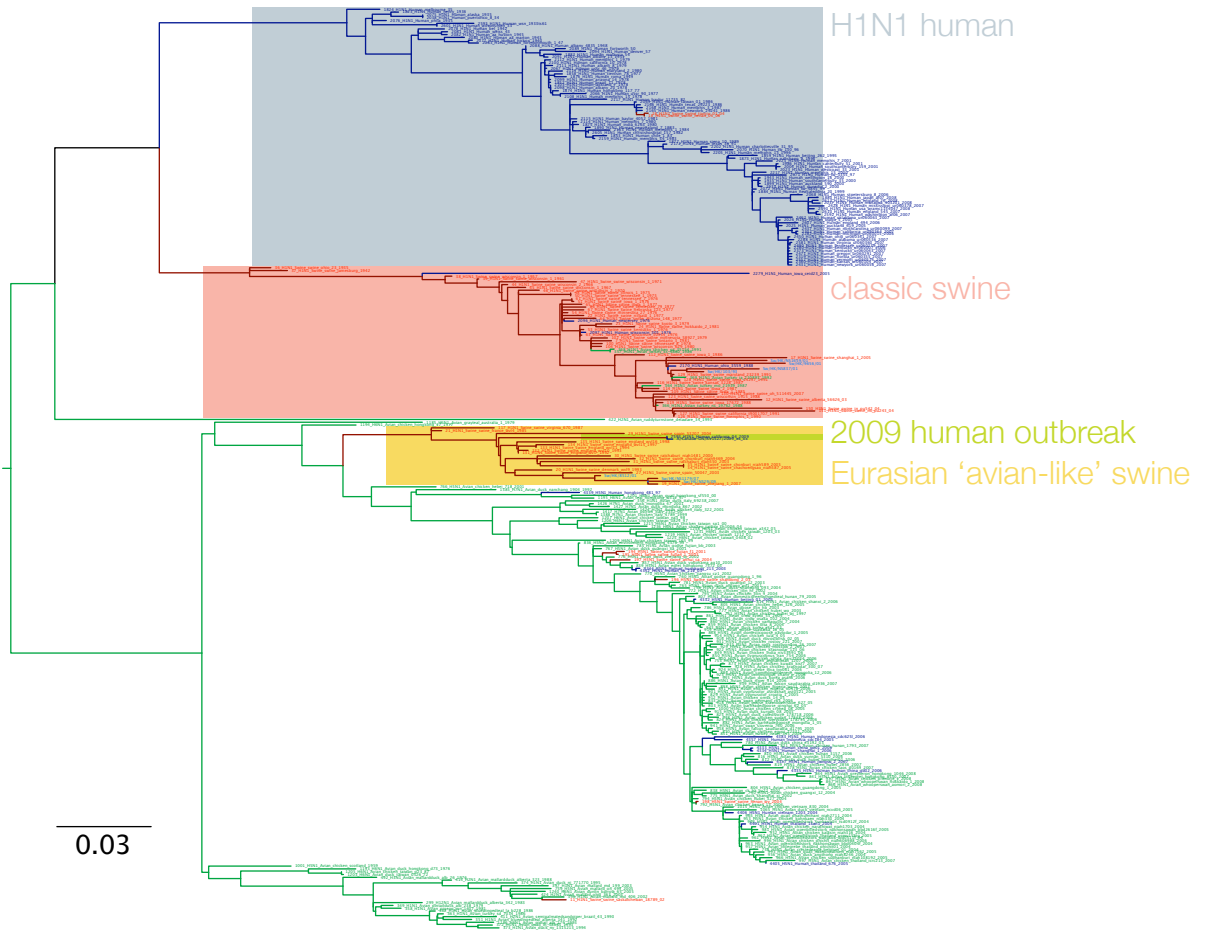
0.05

NP

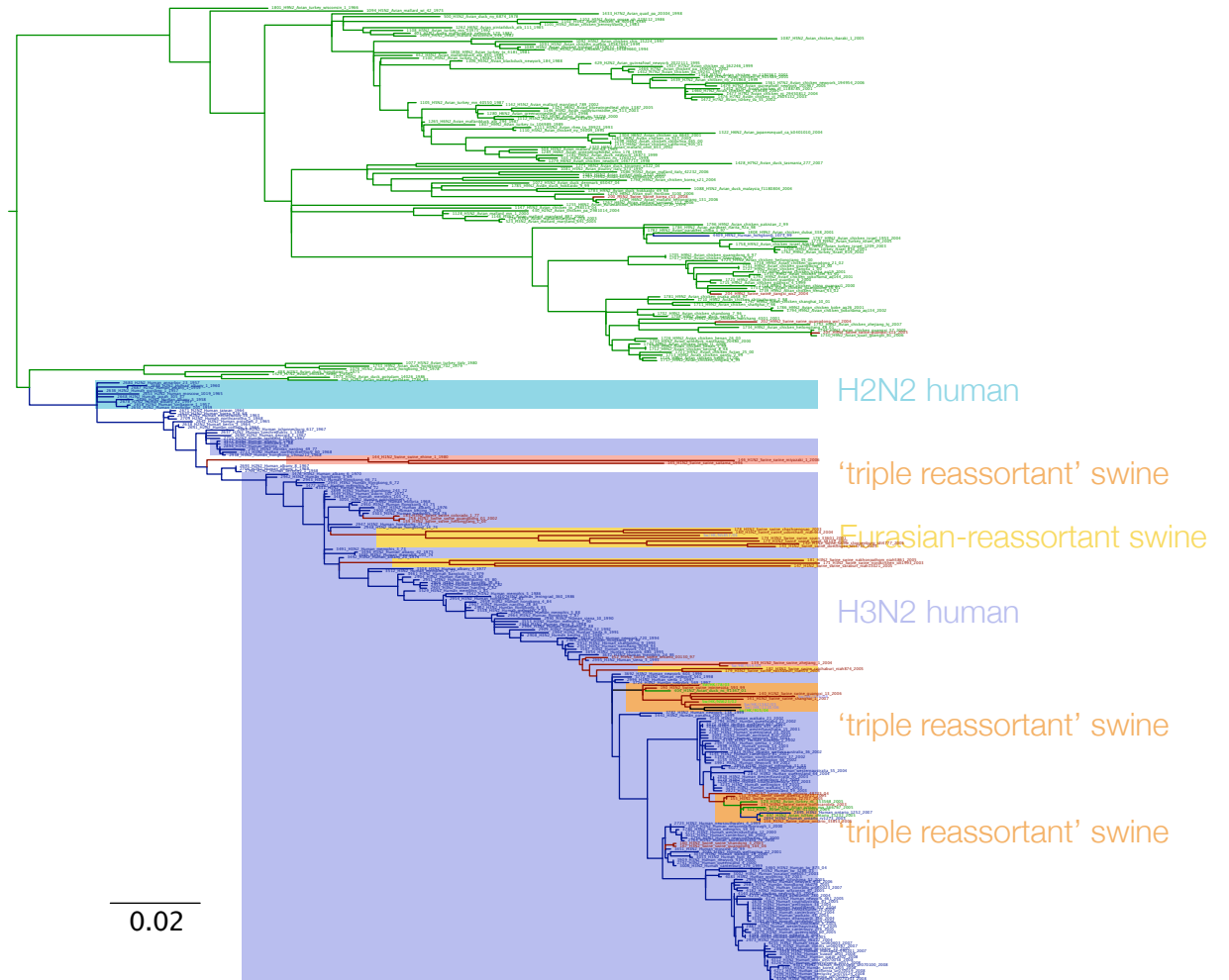


0.05

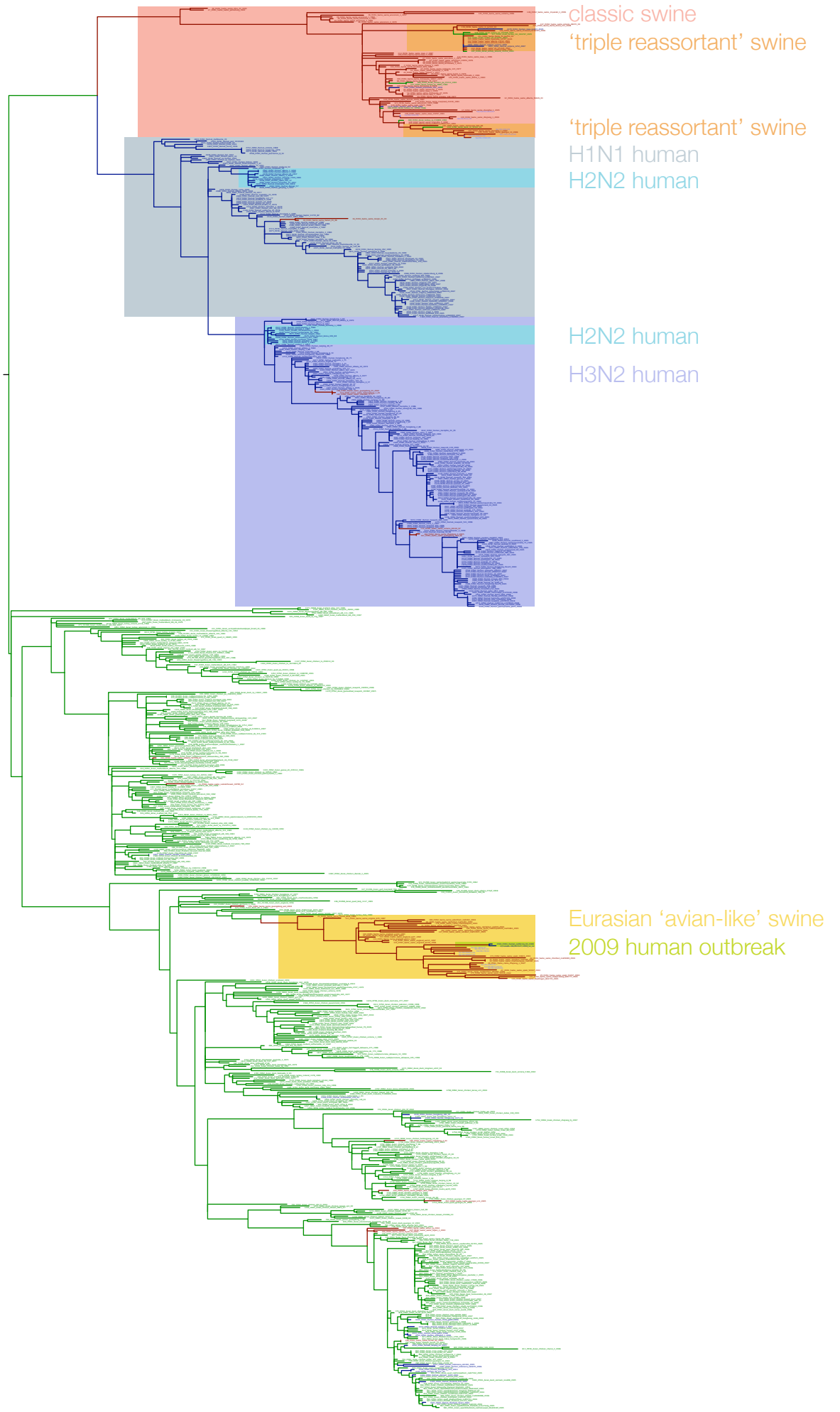
NA N1



N2

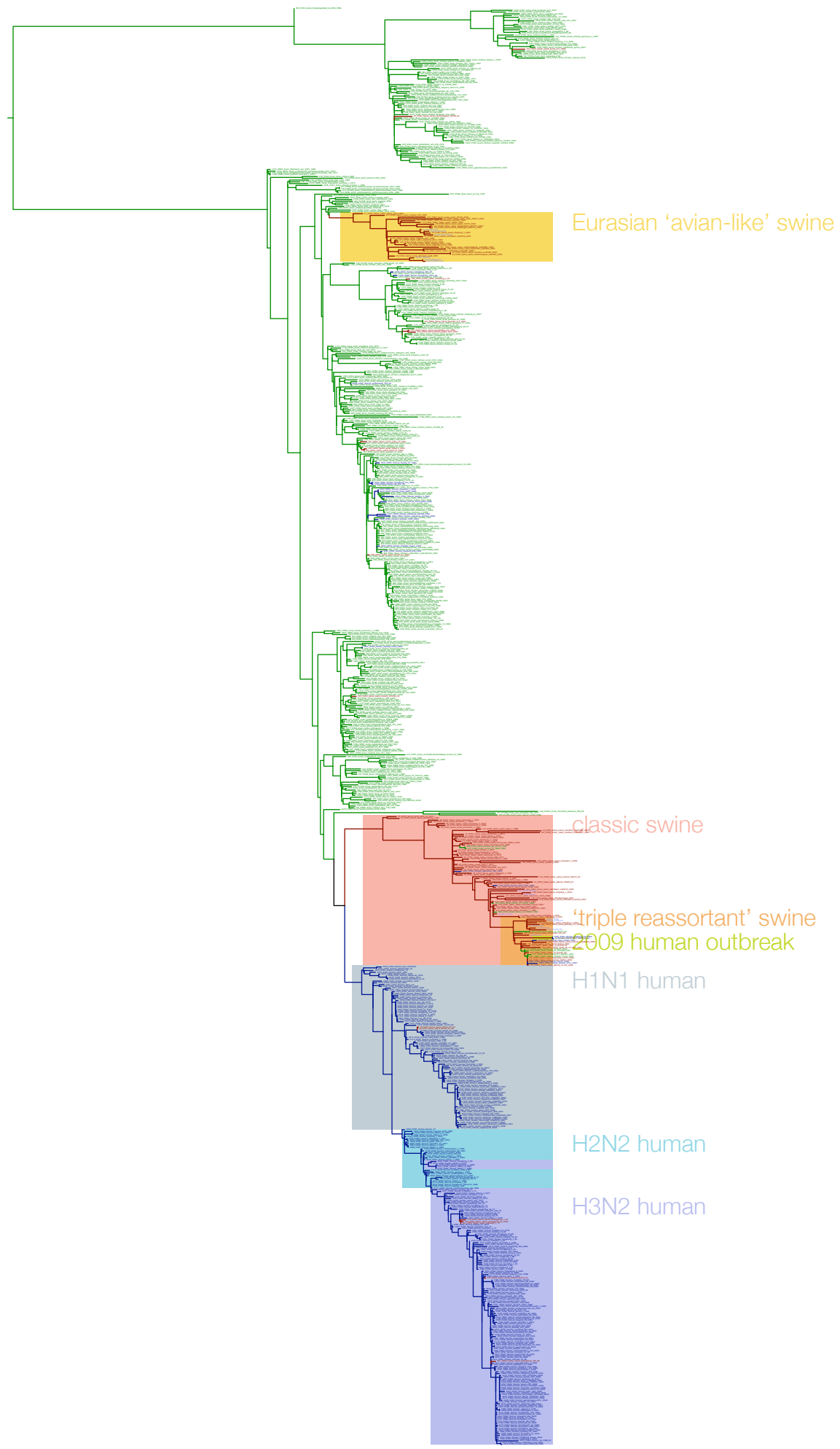


M



0.03

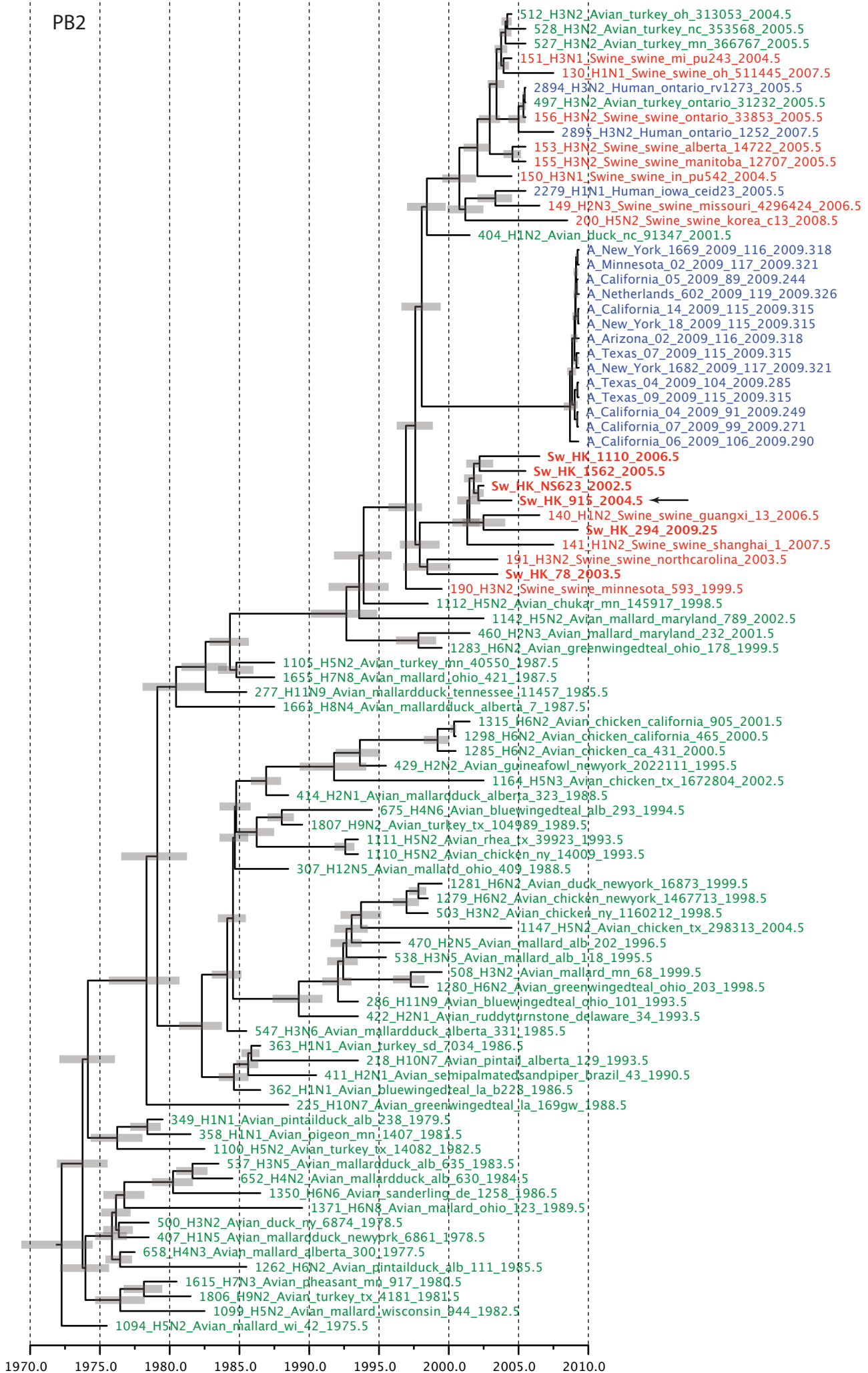
NS

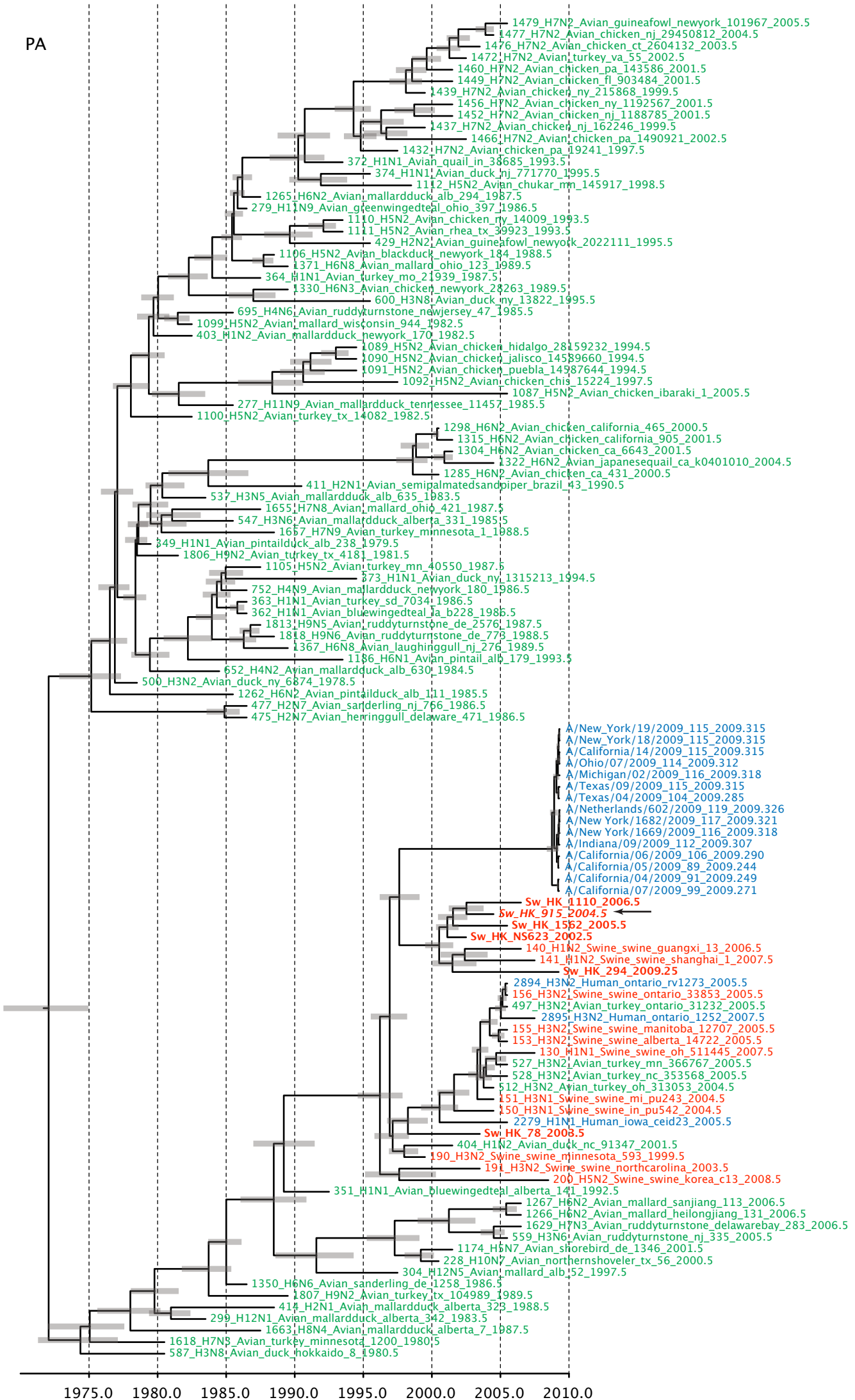


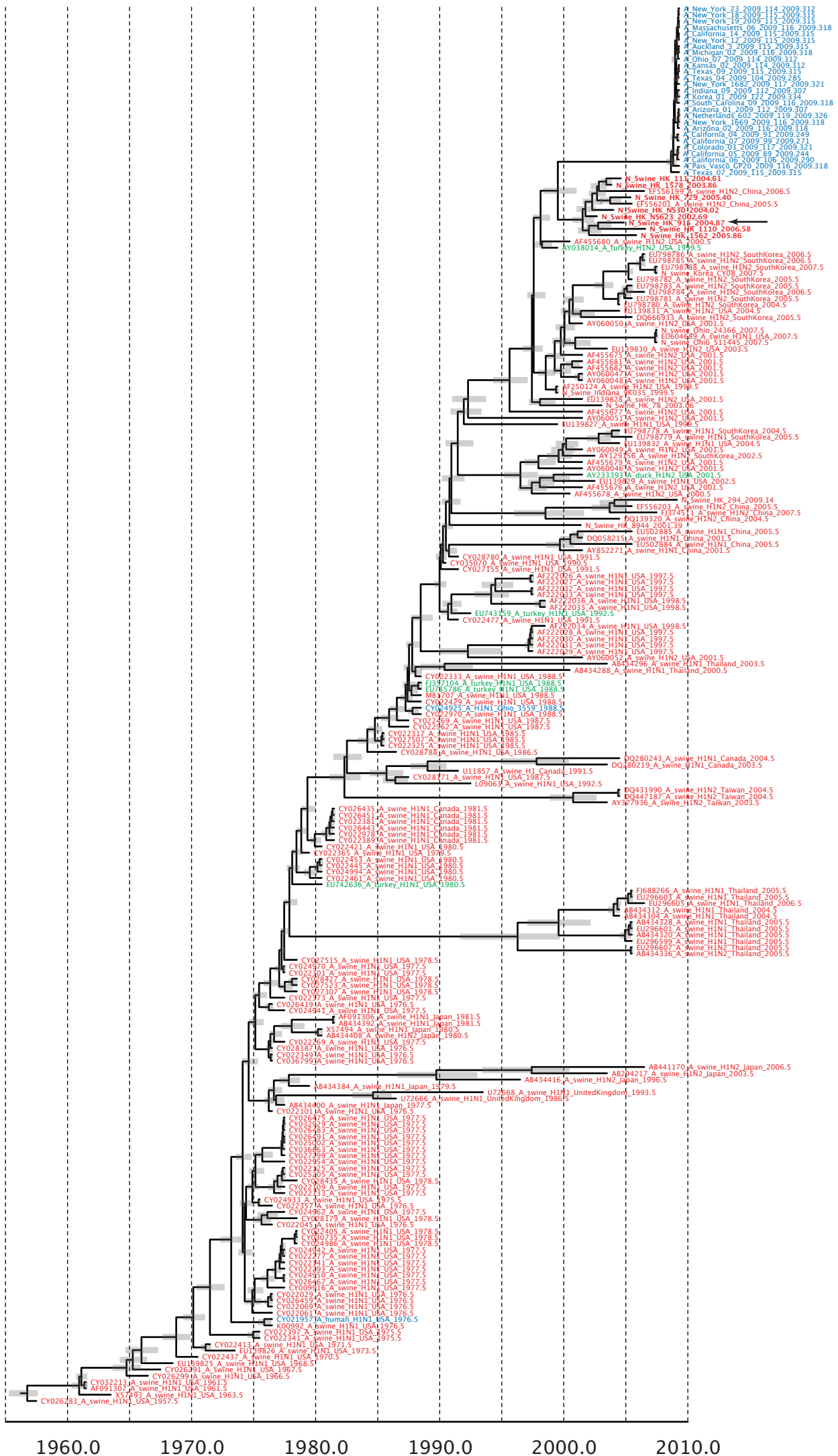
0.1

Supplementary Figure 2. Phylogenetic relationships scaled to time for each gene segment (PB2, PB1, PA, HA, NP, NA, M & NS) of the swine-origin influenza A (H1N1) virus as represented in Figure 2 of the main text but with full virus names and GenBank accession numbers. Internal nodes are reconstructed common ancestors with 95% credible intervals on their date given by the grey bars. Human viruses are coloured blue, swine viruses in red and avian viruses in green. Newly described Hong Kong sequences have bold labels. Sw/HK/915/04 (H1N2) is highlighted with an arrow when present (see main text). Timeline is in years.

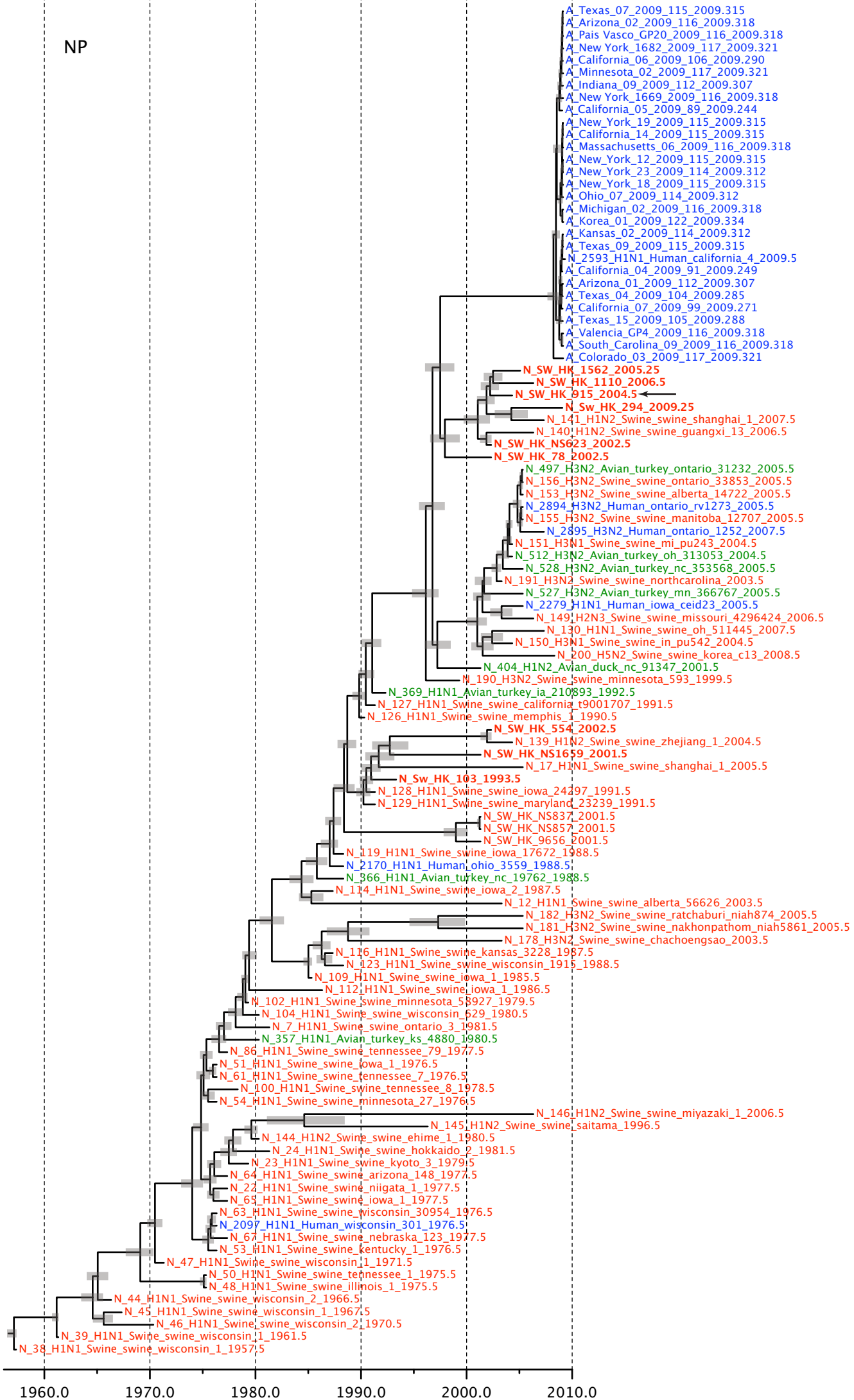
PB2





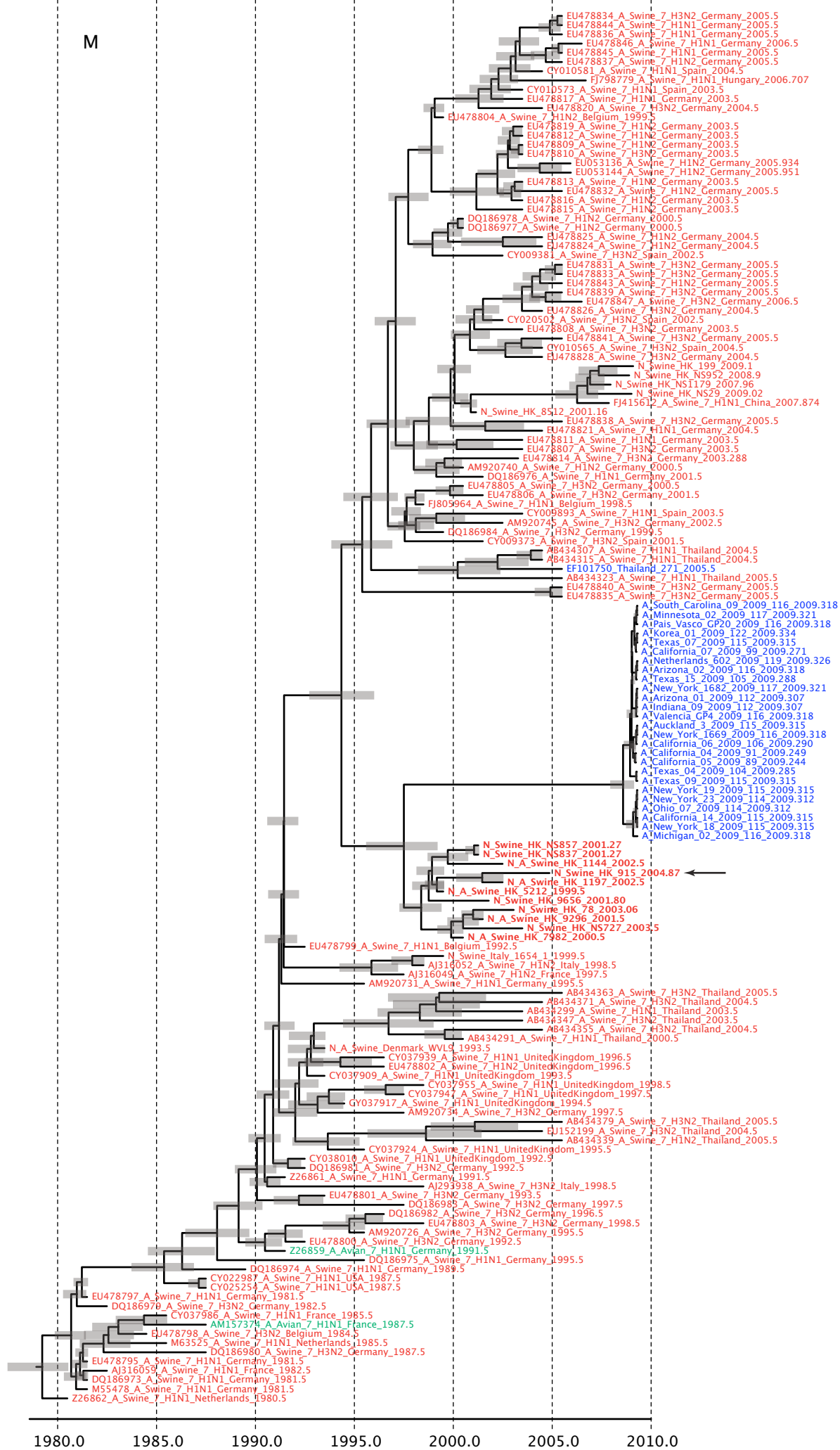


NP

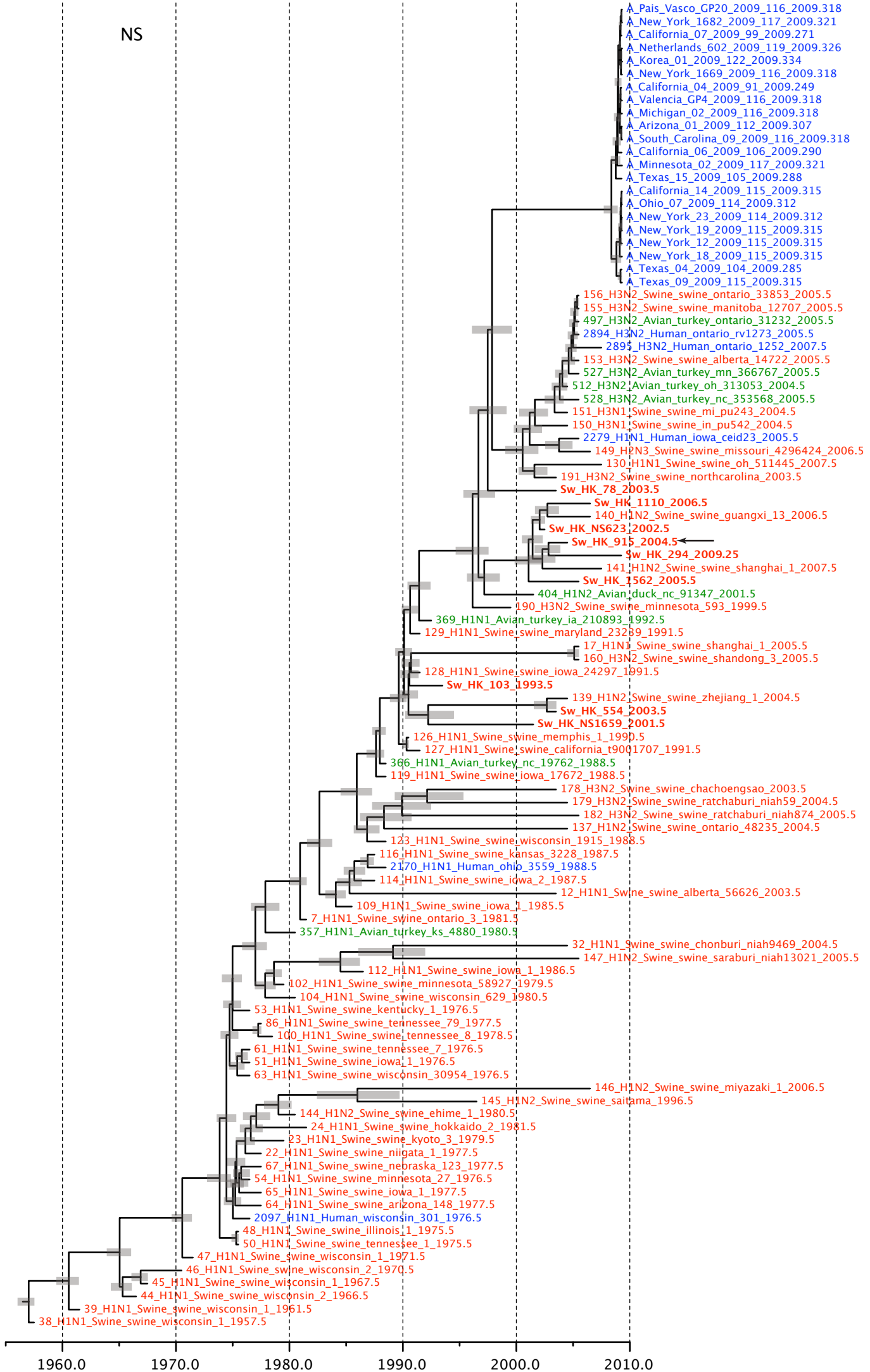




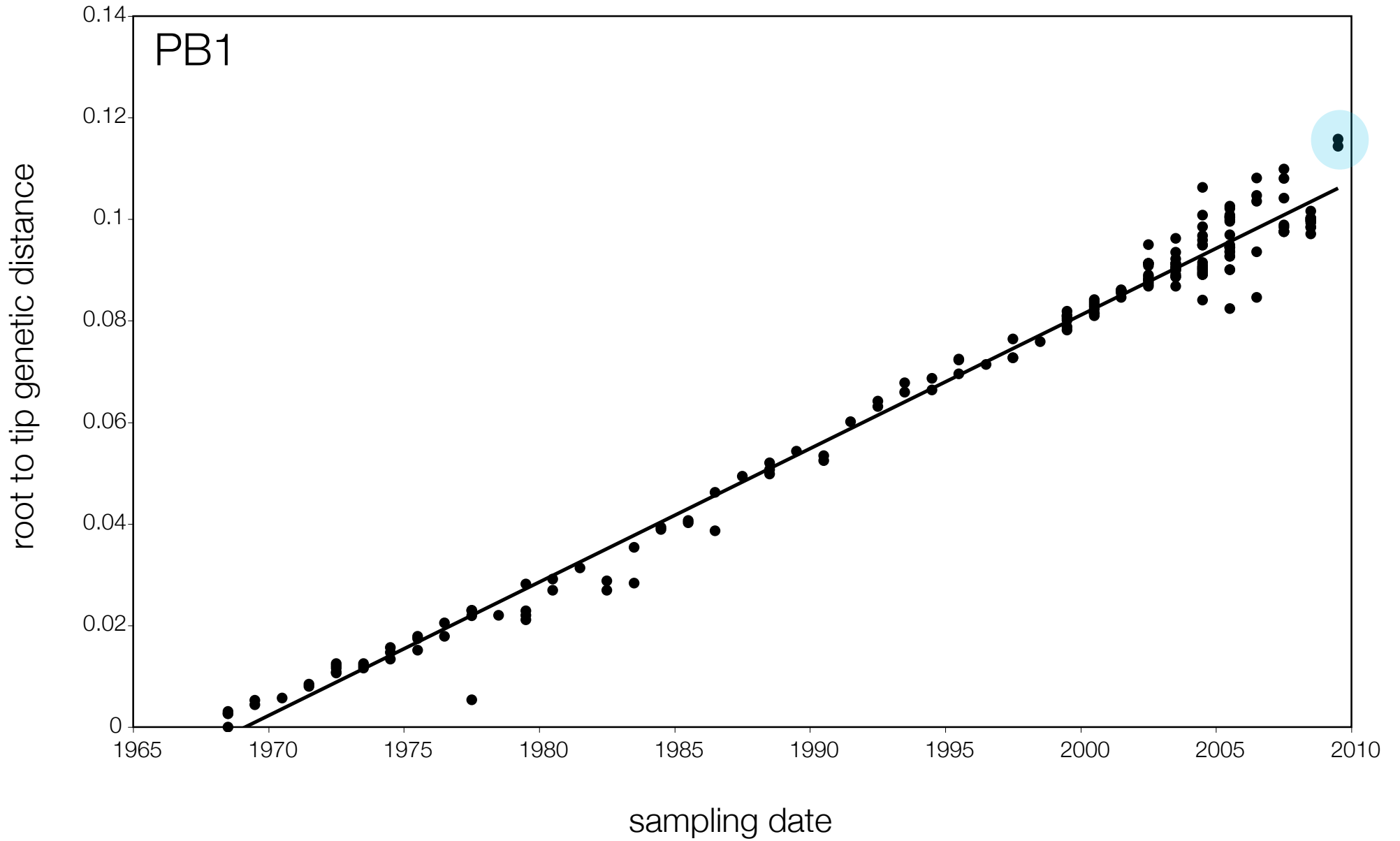
M

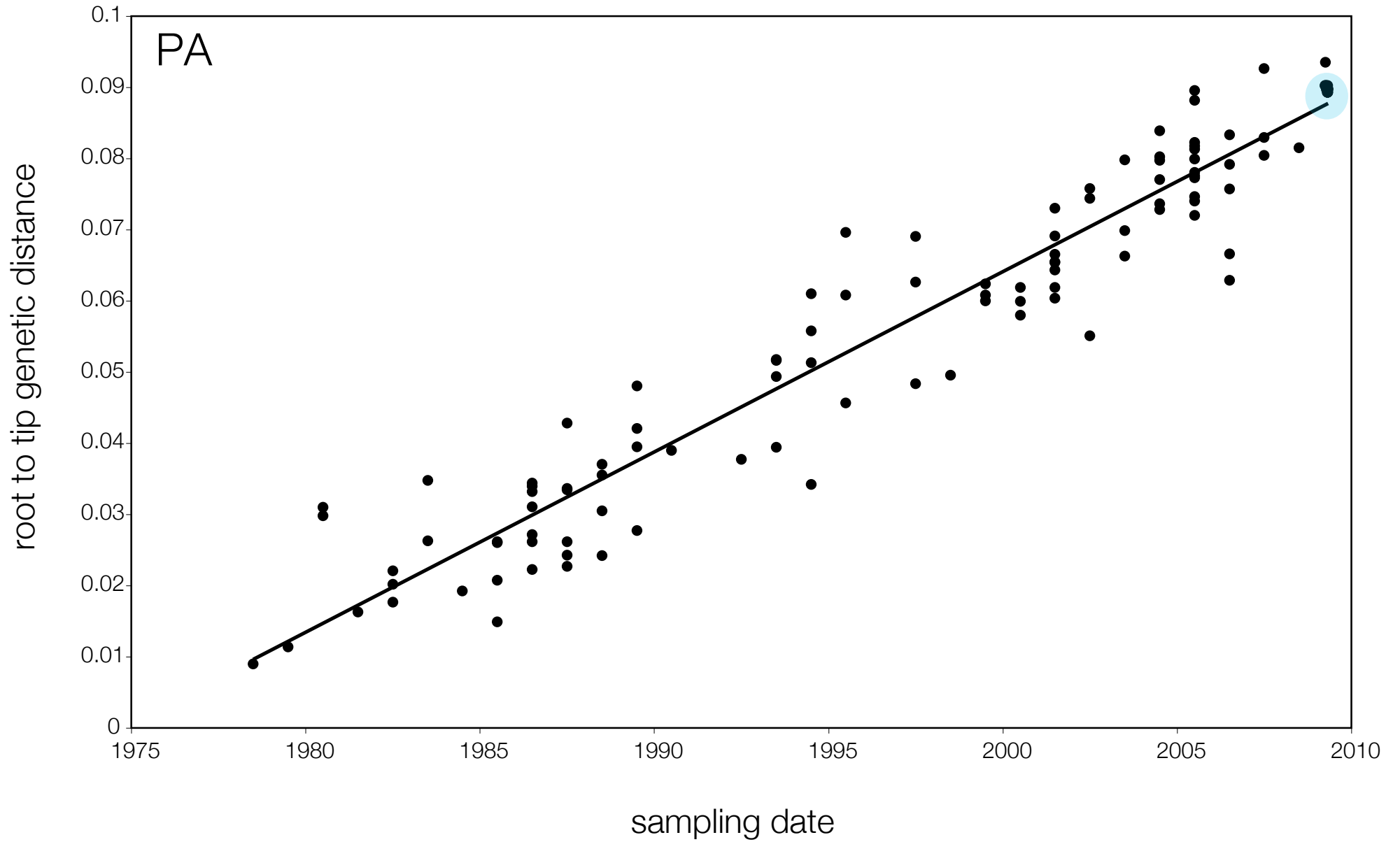


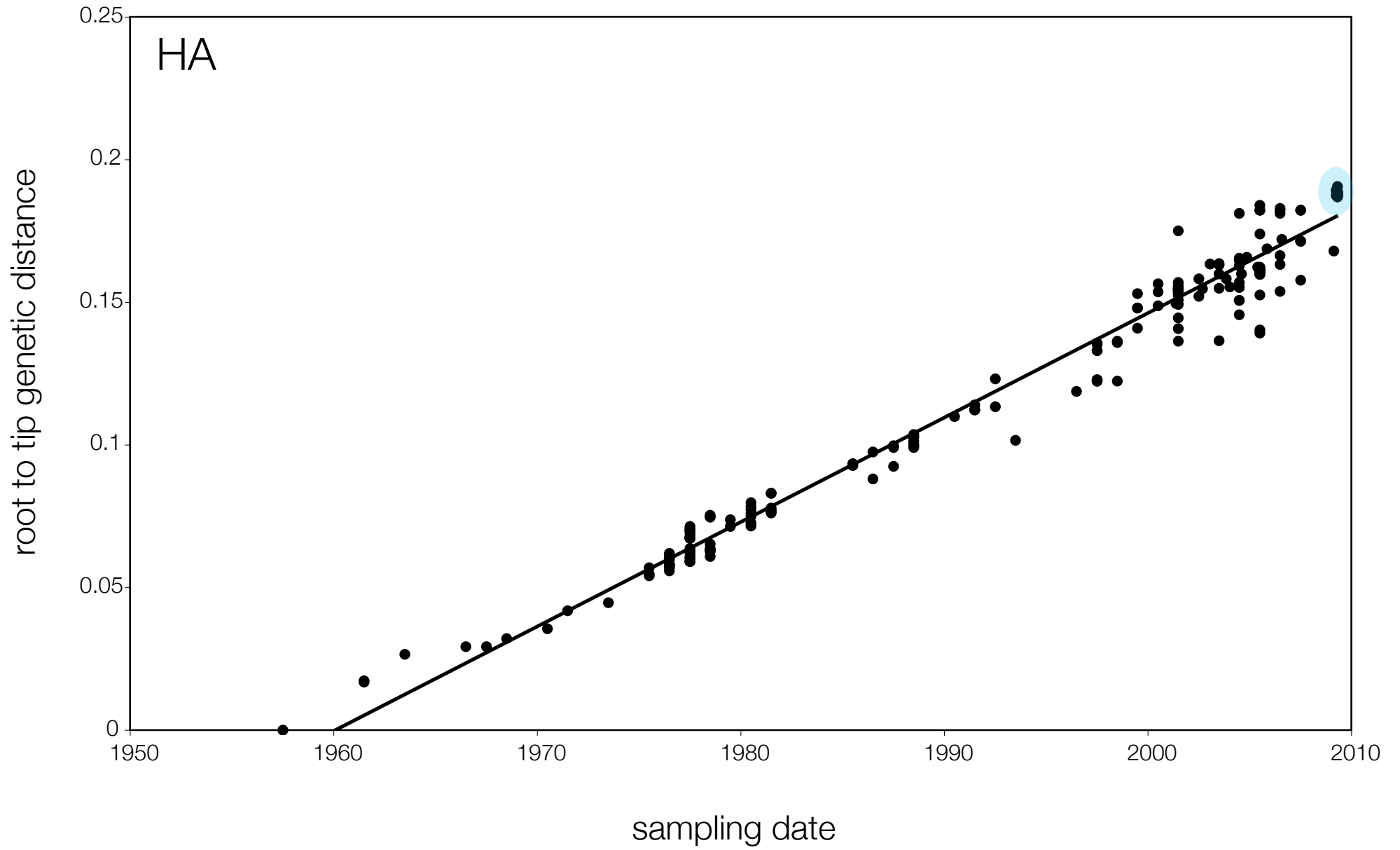
NS

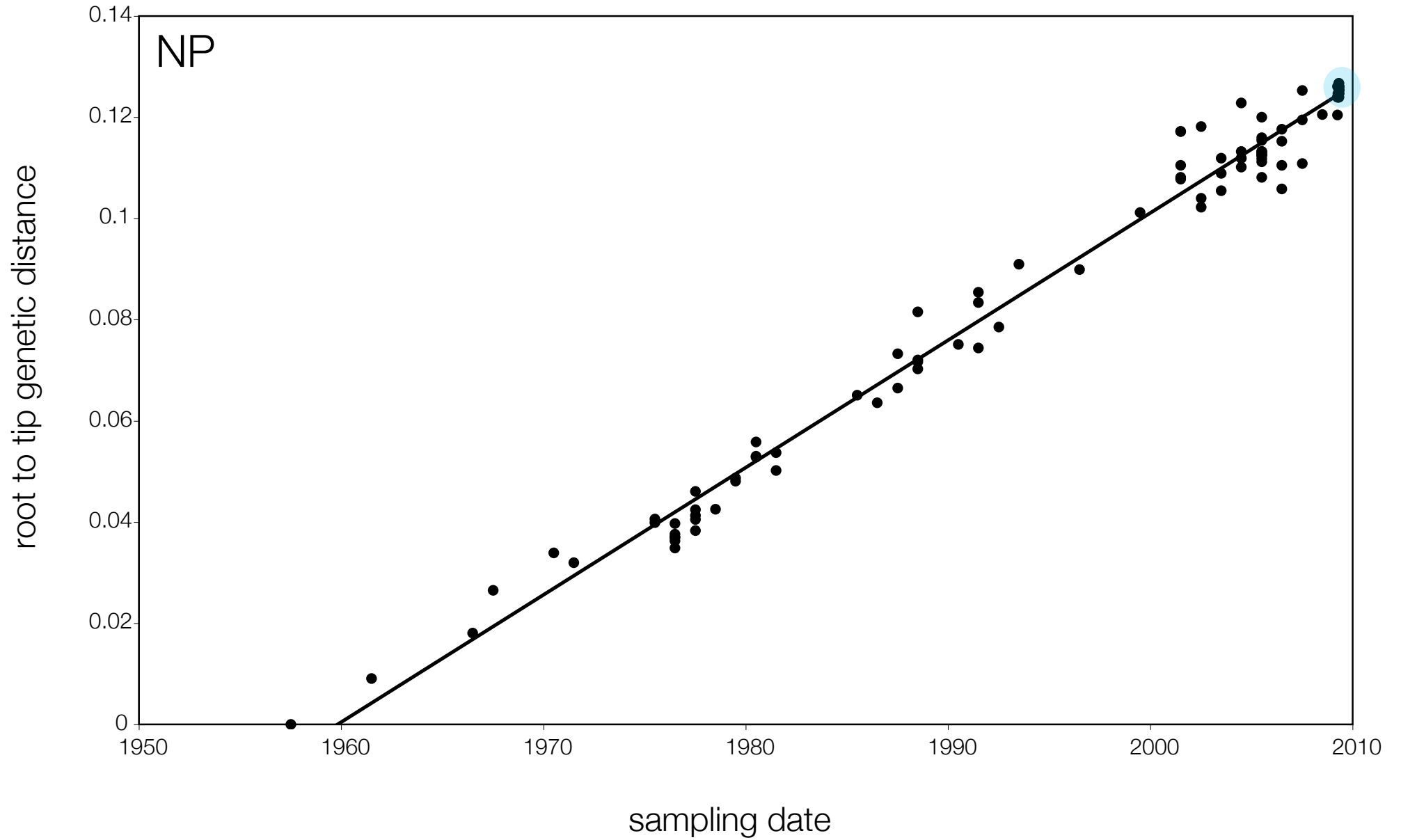


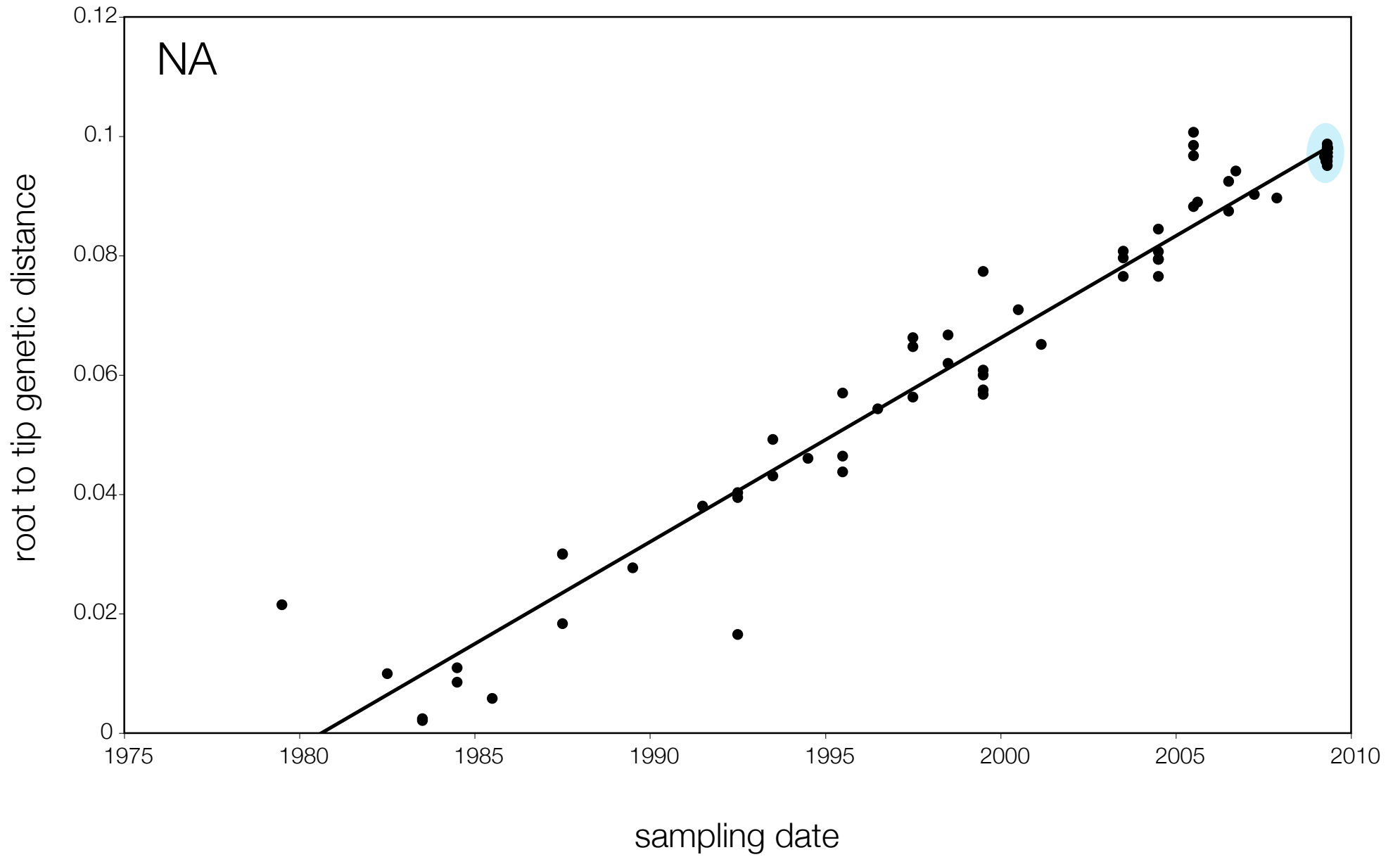
Supplementary Figure 3. For each gene segment (PB2, PB1, PA, HA, NP, NA, M & NS), we plot the isolation date of each influenza sequence against the genetic distance from that sequence to the root of the phylogeny. The linear regression gradient is therefore an estimate of the rate of sequence evolution and the x-intercept is an estimate of the TMRCA of the whole phylogeny. Phylogenies were estimated using neighbour-joining, with rooting chosen to maximise the regression fit. The chosen root was typically very close to the earliest sampled sequence. Residual analysis was performed to identify and remove significant outliers, which most likely result from isolation data annotation errors in the sequence database. For each gene, the degree of scatter about the linear regression reflects evolutionary rate heterogeneity among lineages, such that a "strict clock" corresponds to all the points falling exactly on the regression line. The 2009 outbreak sequences (highlighted in light blue) are entirely typically of the long term trends in divergence, hence there is no evidence that the branch leading to the outbreak has evolved unusually rapidly or slowly. For further discussion of this methodology, see Drummond AJ, Pybus OG, Rambaut A. 2003. Inference of viral evolutionary rates from molecular sequences. *Advances in Parasitology* **54**:331-358.

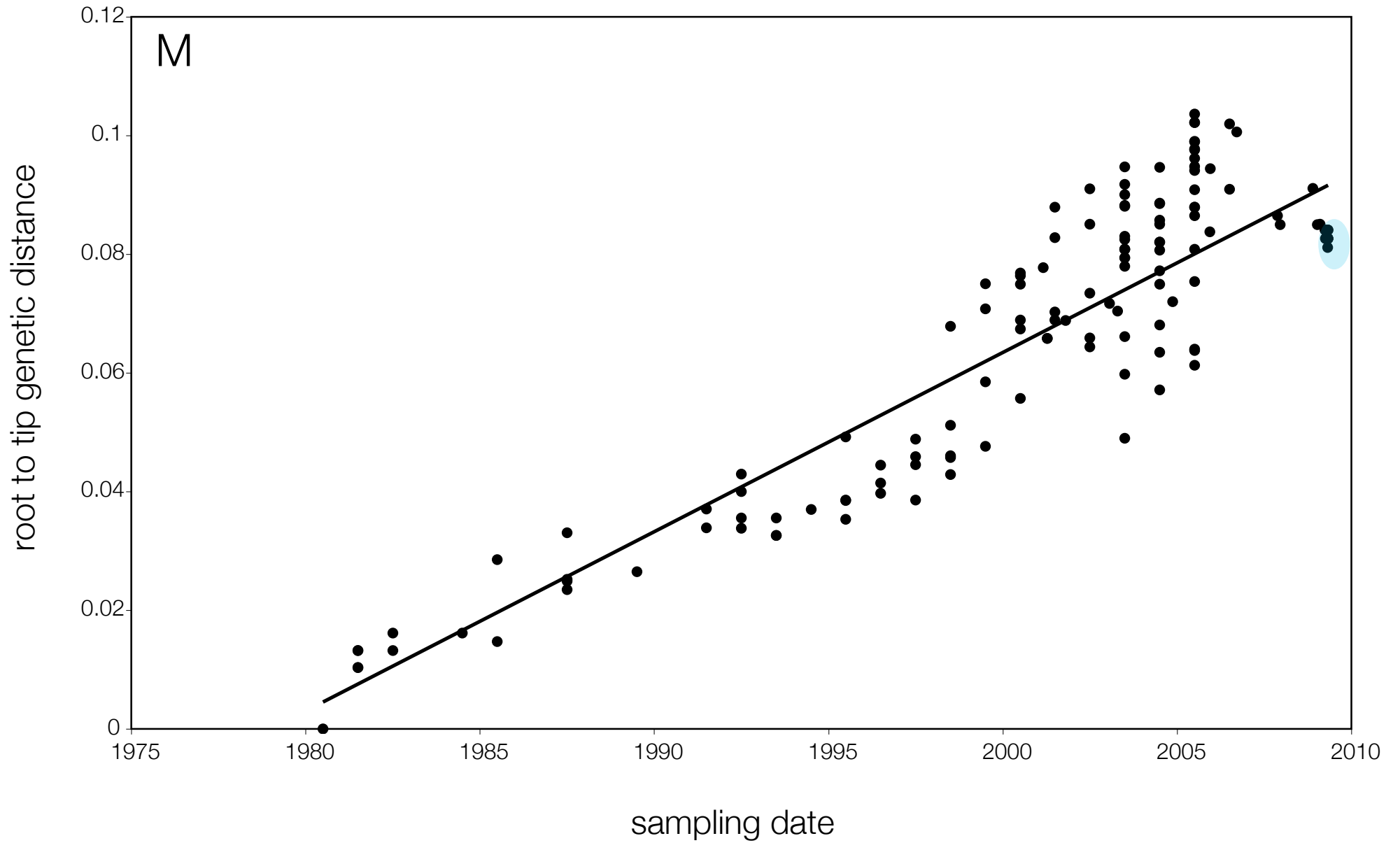












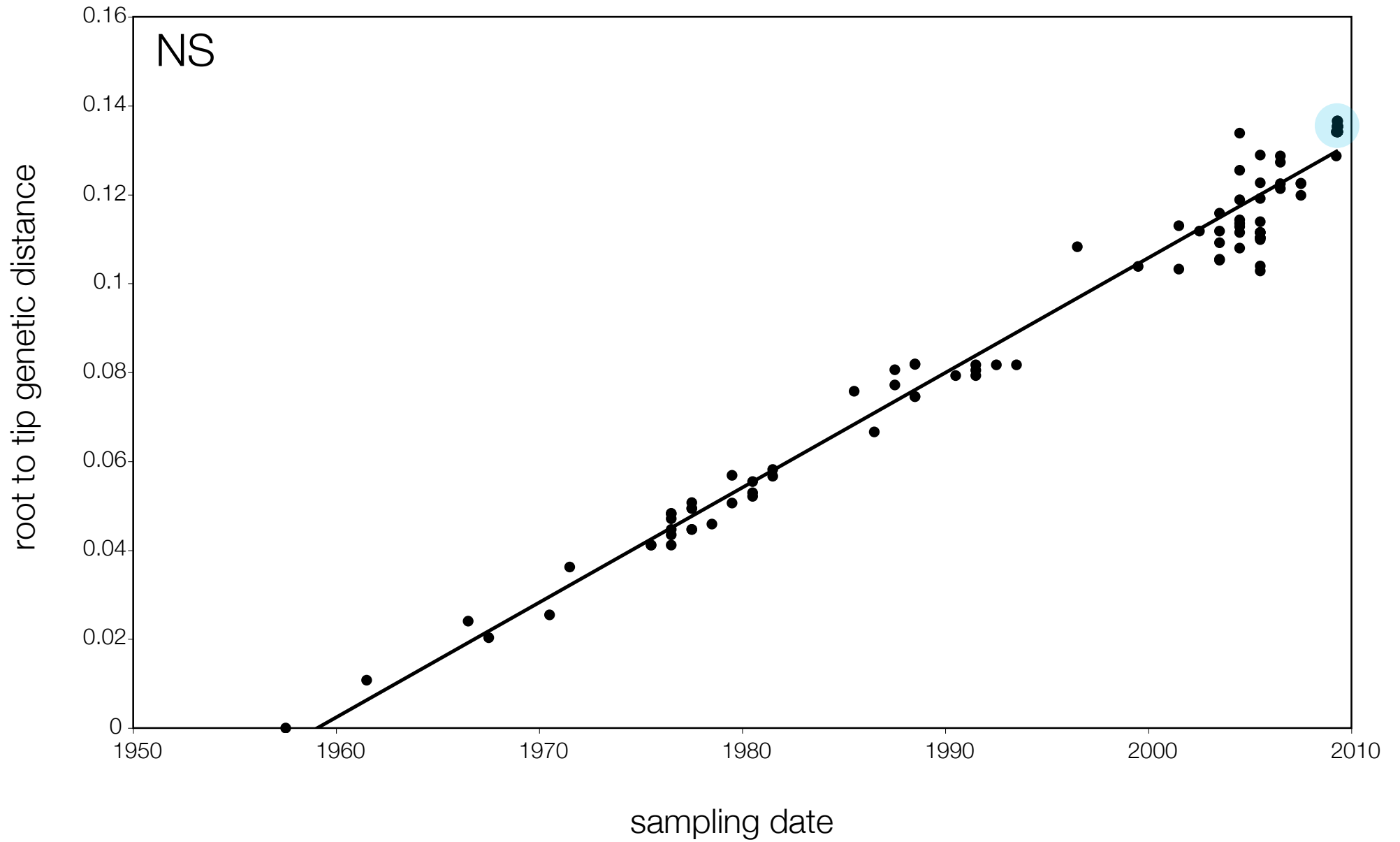


Table S1 SLAC Results

gene	outbreak	reference	estimated	mean TMRCA	
	dN/dS ^a	dN/dS	relative rate ^b	mean TMRCA	(adjusted)
HA	0.32	0.21	1.22	28-Aug-2008	11-Oct-2008
NA	0.26	0.18	1.16	8-Aug-2008	12-Sep-2008
MP	0.19	0.05	1.43	3-Aug-2008	26-Oct-2008
NP	0.18	0.05	1.39	27-Mar-2008	19-Jul-2008
NS	2.15	0.23	3.85	21-May-2008	6-Feb-2009
PA	0.11	0.06	1.15	7-Oct-2008	6-Nov-2008
PB1	0.212	0.06	1.45	24-Oct-2008	22-Dec-2008
PB2	0.15	0.11	1.11	9-Sep-2008	4-Oct-2008

^aCalculated using SLAC (S. Kosakovsky Pond, available at <http://datamonkey.org>)

^bsubstitution rate in outbreak clade relative to non-outbreak sequences based on excess non-synonymous mutations inferred from the higher dN/dS ratio.

Table S2 SNAP Results

gene	outbreak	reference	estimated	mean TMRCA	
	dN/dS ^a	dN/dS	relative rate ^b	mean TMRCA	(adjusted)
HA	0.24	0.09	1.41	28-Aug-2008	9-Nov-2008
NA	0.32	0.1	1.59	8-Aug-2008	17-Nov-2008
MP	0.3	0.03	1.82	3-Aug-2008	5-Dec-2008
NP	0.14	0.02	1.39	27-Mar-2008	19-Jul-2008
NS	0.31	0.13	1.43	21-May-2008	5-Sep-2008
PA	0.18	0.05	1.43	7-Oct-2008	10-Dec-2008
PB1	0.16	0.03	1.45	24-Oct-2008	22-Dec-2008
PB2	0.13	0.07	1.15	9-Sep-2008	10-Oct-2008

^aCalculated using SNAP (B. Korber, available at <http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html>)

^bsubstitution rate in outbreak clade relative to non-outbreak sequences based on excess non-synonymous mutations inferred from the higher dN/dS ratio.

Table S3 Hong Kong swine genetic origins

Lineage of segments from Sw/HK sequences based on
Phylogenetic analysis

All Seqs	PB2	PB1	PA	HA (H1)	NP	NA	M	NS
Swine/HK/103/93	C	C	C	C	C	N1	C	C
Sw/HK/NS1659/01	C	C	C	C	C	N1	C	C
Sw/HK/8512/01	C	E	E	E	E	N1	E	E
Sw/HK/NS837/01	E	E	E	E	C	N1	E	E
Sw/HK/9656/01	E	E	E	E	C	N1	E	E
Sw/HK/NS1179/07	E	E	E	E	E	N1	E	E
Sw/HK/NS29/09	E	E	E	E	E	N1	E	E
Sw/HK/554/03	C	C	C	C	C	N2	C	C
Sw/HK/NS857/01	E	E	E	E	C	N2	E	E
Sw/HK/NS623/02	T	T	T	T	T	N2	T	T
Sw/HK/78/03	T	H	T	T	T	N2	E	T
Sw/HK/1110/06	T	T	T	T	T	N2	T	T
Sw/HK/915/04	T	T	T	T	T	N2	E	T
Sw/HK/1562/05	T	T	T	T	T	N2	T	T
Sw/HK/294/09	T	T	T	C	T	N2	T	T
A/California/04/2009	T	T	T	T	T	N1	E	T

Classical Swine	C
Eurasian Swine	E
Triple Reassortant	T
Human	H

Table S4 S-OIV sequences available on NCBI Influenza virus database at time of analysis.

	Days ¹	PB2 ²	PB1	PA	HA	NP	NA	MP	NS
A/Arizona/01/2009	112				GQ117067	GQ117063	GQ117064	GQ117066	GQ117065
A/Arizona/02/2009	116	GQ117076	GQ117075		GQ117079	GQ117074	GQ117077	GQ117078	
A/Auckland/3/2009	115						FJ973552	FJ973553	
A/California/04/2009	91	FJ966079	FJ966080	FJ966081	FJ966082	FJ966083	FJ966084	FJ966085	FJ966086
A/California/05/2009	89	FJ966955	FJ966958	FJ966957	FJ966952	FJ966953	FJ966956	FJ966954	
A/California/06/2009	106	FJ966963	FJ966965	FJ966964	FJ966960	FJ966961	FJ971075	FJ966962	FJ971074
A/California/07/2009	99	FJ984387	FJ969531	FJ969529	FJ981613		FJ984386	FJ966975	
A/California/14/2009	115	GQ117035	GQ117034	GQ117037	GQ117040	GQ117033	GQ117036	GQ117039	GQ117038
A/Colorado/03/2009	117				GQ117119	GQ117117	GQ117118		
A/Indiana/09/2009	112		GQ117093	GQ117095	GQ117097	GQ117092	GQ117094	GQ117096	
A/Kansas/02/2009	114				GQ117059	GQ117057	GQ117058		
A/Korea/01/2009	122				GQ131023	GQ131024	GQ132185	GQ131025	GQ131026
A/Massachusetts/06/2009	116				GQ117043	GQ117041	GQ117042		
A/Michigan/02/2009	116			GQ117109	GQ117112	GQ117107	GQ117108	GQ117111	GQ117110
A/Minnesota/02/2009	117	GQ117070	GQ117069			GQ117068	GQ117071	GQ117073	GQ117072
A/Netherlands/602/2009	119				CY039527		CY039528		
A/New York/1669/2009	116	CY039900	CY039899	CY039898	CY039893	CY039896	CY039895	CY039894	CY039897
A/New York/1682/2009	117	CY039908	CY039907	CY039906	CY039901	CY039904	CY039903	CY039902	CY039905
A/New York/12/2009	115				FJ984337	FJ984336	FJ984335		FJ984334
A/New York/18/2009	115	FJ984351	FJ984353	FJ984354	FJ984355	FJ984352	FJ984350	FJ984348	FJ984349
A/New York/19/2009	115		FJ984392	FJ984393	FJ984394	FJ984391	FJ984390	FJ984388	FJ984389
A/New York/23/2009	114				FJ984364	FJ984363	FJ984362		FJ984361
A/Ohio/07/2009	114			FJ984401	GQ117100	GQ117098	GQ117099		
A/Pais Vasco/GP20/2009	116				FJ985763	FJ985761	FJ985764	FJ985760	FJ985762
A/South Carolina/09/2009	116				GQ117056	GQ117052	GQ117053	GQ117055	GQ117054
A/Texas/04/2009	104	FJ969525	FJ981615	FJ981618	FJ981614	FJ981617	FJ981620	FJ966980	
A/Texas/07/2009	115	GQ117089	GQ117088		GQ117091	GQ117087		GQ117090	
A/Texas/09/2009	115	GQ117027	GQ117026	GQ117029	GQ117032	GQ117025	GQ117028	GQ117031	GQ117030
A/Texas/15/2009	105		GQ122094		GQ122097	GQ122092	GQ122096	GQ122095	GQ122093
A/Valencia/GP4/2009	116				FJ985758	FJ985756	FJ985759	FJ985755	FJ985757

1. The sampling date as number of days since 1-Jan-2009.

2. Accession numbers for each available genomic segment

Supplementary notes:**Adaptation and purifying selection**

We suggest in the main text that the increased presence of amino-acid polymorphisms may be present in the outbreak sequences relative to the swine influenza reference sequences due to the rapid growth in number of infections and the small window of time over which they were sampled. The fairly uniform distribution of non-synonymous changes across the outbreak genomes (data not shown) and the relatively consistent dN/dS ratios across genes (Supplementary Tables S1 & S2) support this view. We predict that many of these mutations will be mildly deleterious and, given time, will be removed by purifying selection along an emergent ‘trunk’ lineage, and that the estimated evolutionary rate and gene-specific dN/dS pattern will approach that observed in the swine influenza data sets sampled over a time period of decades (see below). An interesting alternative is that the elevated rate in the outbreak sequences (and higher dN/dS ratio) is due to a burst of adaptive evolution in a new host, rather than a relaxation of purifying selection and that many of the ‘extra’ non-synonymous mutations will become fixed in the population. If the current outbreak persists in the human population, it will be possible to distinguish these hypotheses.

Detailed molecular characterization

The presence of residue Asn 31 in the M2 protein invariably confers resistance to the adamantanes, a group of antiviral drugs used for treatment of human influenza¹.

Sequence analysis revealed that Asn 31 was present in all human isolates of the current outbreak, this mutation was also present in swine H1N1 and H1N2 viruses that were most closely related in the M gene. The widespread occurrence of the Asn 31 mutation in closely related viruses indicate that it may have descended from the

swine lineage of amantadine-resistant M2 genes rather than independently acquired. It should be noted that majority of the seasonal H1N1 and H3N2 viruses are resistant to amantadine, except Brisbane-like H1N1 viruses². No mutations were observed in the NA gene that confers Oseltamivir resistance.

The molecular determinants of interspecies transmission of the A/California/04/2009-like virus to humans are unclear. In these viruses the amino acid residue at the receptor binding pocket of HA1— position Gln 226 and Gly 228 retain configurations (2,3-NeuAcGal linkages) predicted to have affinity for mammalian cell-surface receptors³⁻⁴. Amino acid residues relevant to receptor binding were identical to those of classical swine H1N1 and North American H1N1 and recent H1N1 viruses in both the 130-loop and 190-helix. However, the amino acids at positions 133 and 135 are different from the current seasonal vaccine strain A/Brisbane/59/2007(H1N1) used in both the northern and southern hemispheres, indicating that these viruses may show less cross-reaction to the current vaccine strain⁵⁻⁶.

Mutations Glu627Lys and Asp701Asn of the PB2 gene that are thought to be associated with adaptation to mammals and increased virulence of influenza viruses in mice were not present⁷⁻¹². The C terminal of the NS1 gene is truncated in the A/California/04/2009-like viruses and four residues of the PDZ ligand domain are not present¹¹.

Supplementary Methods

Viral RNA was directly extracted from infected allantoic fluid or cell culture using QIAamp viral RNA minikit (Qiagen, Inc., Valencia, Calif.). cDNA were synthesized by reverse transcription reaction and gene amplification by PCR were performed

using specific primers for each gene segments. PCR products were purified with the QIAquick PCR purification kit (Qiagen Inc.) and sequenced by synthetic oligonucleotides. Reactions were performed using Big Dye-Terminator v3.1 Cycle Sequencing Reaction Kit on an ABI PRISM 3700 DNA Analyzer (Applied Biosystems) following the manufacturer's instructions. All sequences were assembled and edited with Lasergene version 6.1 (DNASTAR, Madison, WI). Full genome sequences of these viruses are available for download at GISAID under the accession numbers (EPI177540 to EPI77658 and EPI177947).

References:

1. 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).
2. Barr, I. G., Deng, Y. M., Iannello, P., Hurt, A. C. & Komadina, N. Adamantane resistance in influenza A (H1) viruses increased in 2007 in South East Asia but decreased in Australia and some other countries. *Antiviral Research* **80**, 200–205 (2008).
3. 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).
4. Rogers, G. N. *et al.* Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* **304**, 76–78 (1983).

5. Winter, G., Fields, S. & Brownlee G. G. Nucleotide sequence of the haemagglutinin gene of a human influenza virus H1 subtype. *Nature* **292**, 72–75.
6. Canton, A., Brownlee, G. G., Yewdell, J. W. & Gerhard, W. The antigenic structure of the influenza virus A/PR/*34 hemagglutinin (H1 subtype). *Cell* **31**, 417–27 (1982).
7. Hatta, M., Gai, P., Halfmann, P. & Kawaoka, Y. Molecular basis of high virulence of Hong Kong H5N1 influenza A viruses. *Science* **7**, 1840–1842 (2001)
8. Gabriel, G. I., Herwig, A. & Klenk, H.–D. Interaction of polymerase subunit PB2 and NP with importin α 1 is a determinant of host range of influenza A virus. *PLoS Pathog.* **4**, e11 (2008).
9. Le, Q. M., Sakai-Tagawa, Y., Ozawa, M., Ito, M. & Kawaoka, Y. Selection of H5N1 Influenza Virus PB2 during Replication in Humans. *J. Virol.* **83**, 5278–5281 (2009).
10. Steel, J., Lowen, A. C., Mubareka, S. & Palese, P. Transmission of Influenza Virus in a Mammalian Host Is Increased by PB2 Amino Acids 627K or 627E/701N. *PLoS Pathog.* **5**, e1000252 (2009).
11. Obenauer, J. C. *et al.* Large-Scale Sequence Analysis of Avian Influenza Isolates. *Science* **311**, 1576–1580 (2006).
12. Jackson D, Hossain MJ, Hickman D, Perez DR, Lamb RA (2008) A new influenza virus virulence determinant: The NS1 protein four C-terminal residues modulate pathogenicity. *Proc. Natl. Acad. Sci. USA* **105**, 4381–4386 (2008).