SUPPLEMENTARY INFORMATION

The Genomic and Epidemiological Dynamics of Human Influenza A Virus

Supplementary Information

Andrew Rambaut[1], Oliver G. Pybus[2], Martha I. Nelson[3], Cecile Viboud[4], Jeffery K. Taubenberger[5], Edward C. Holmes[3,4]

[1]Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Edinburgh EH9 3JT, Scotland, UK.

[2]Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, U.K.

[4]Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802. USA.

[4]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892. USA.

[5]Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892. USA.

Corresponding Authors:

Dr. Edward C. Holmes, Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802. USA.

Tel: +1 814 863 4689; Fax 1 814 865 9131; E-mail: ech15@psu.edu

Dr. Andrew Rambaut, Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Edinburgh EH9 3JT, Scotland, UK.

Tel: +44 131 650 8624; E-mail: a.rambaut@ed.ac.uk

Supplementary Tables, Figures and Legends

Fig. S1. Plots of genetic diversity, g, and $\frac{d^2(g)}{dt^2}$ against time (in calendar years), for each hemisphere and for each subtype of influenza A virus (and for HA and NA genomic segments). g is shown as a blue line and is measured on the left hand y-axis. A numerical approximation of $\frac{d^2(g)}{dt^2}$ is shown as a black line and is measured on the right hand y-axis. For each influenza season, the value $p = \min\left[\frac{d^2(g)}{dt^2}\right]$ is a measure of the relative flatness/peakedness of g for each epidemic. A sharply defined epidemic peak produces a large negative *p*. If g changes at a constant rate or is flat then *p* is zero.

Figure S1 (a) New York H3N2 HA segment

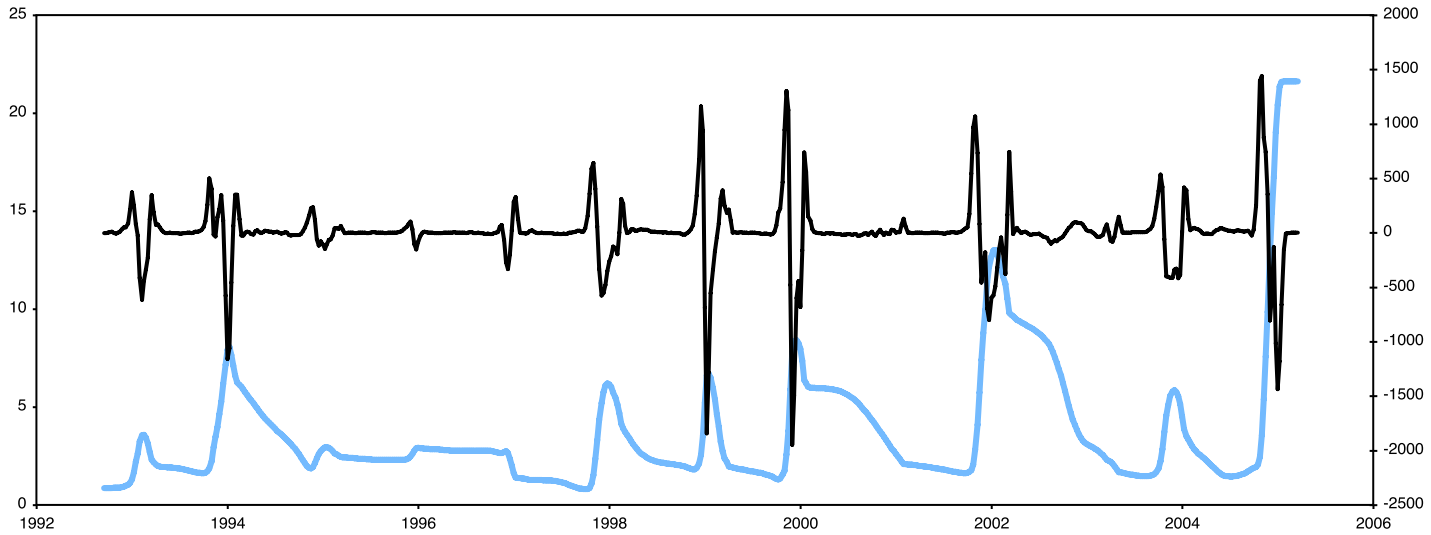

Figure S1 (b) New York H1N1 HA segment

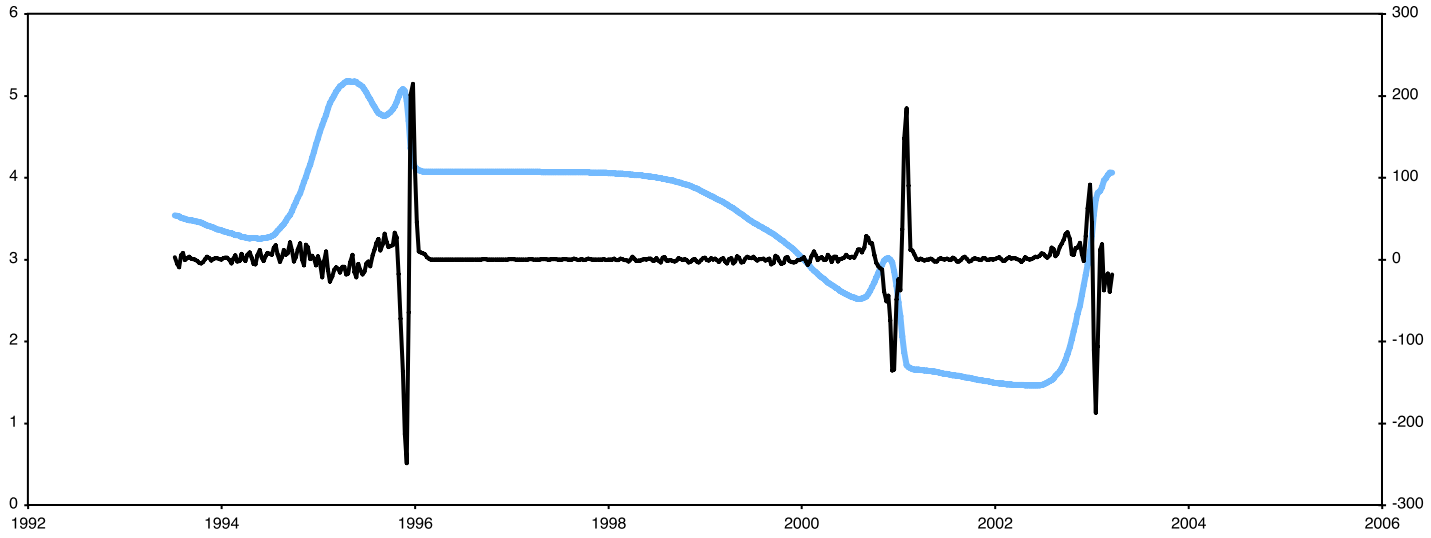

Figure S1 (c) New York H3N2 NA segment

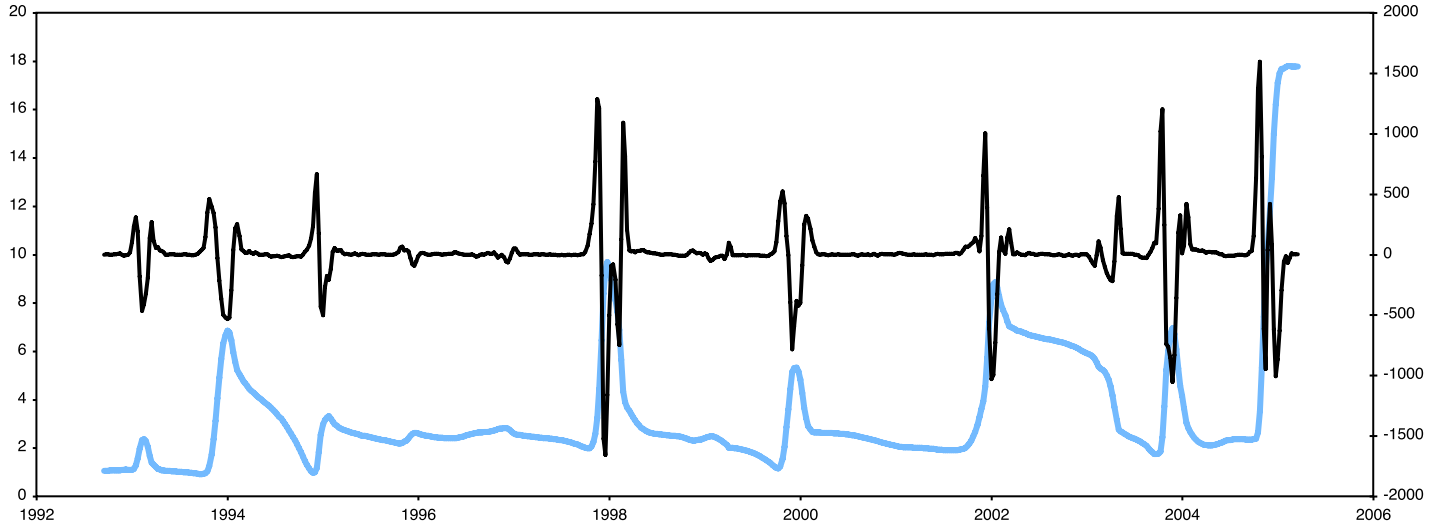

Figure S1 (d) New York H1N1 NA segment

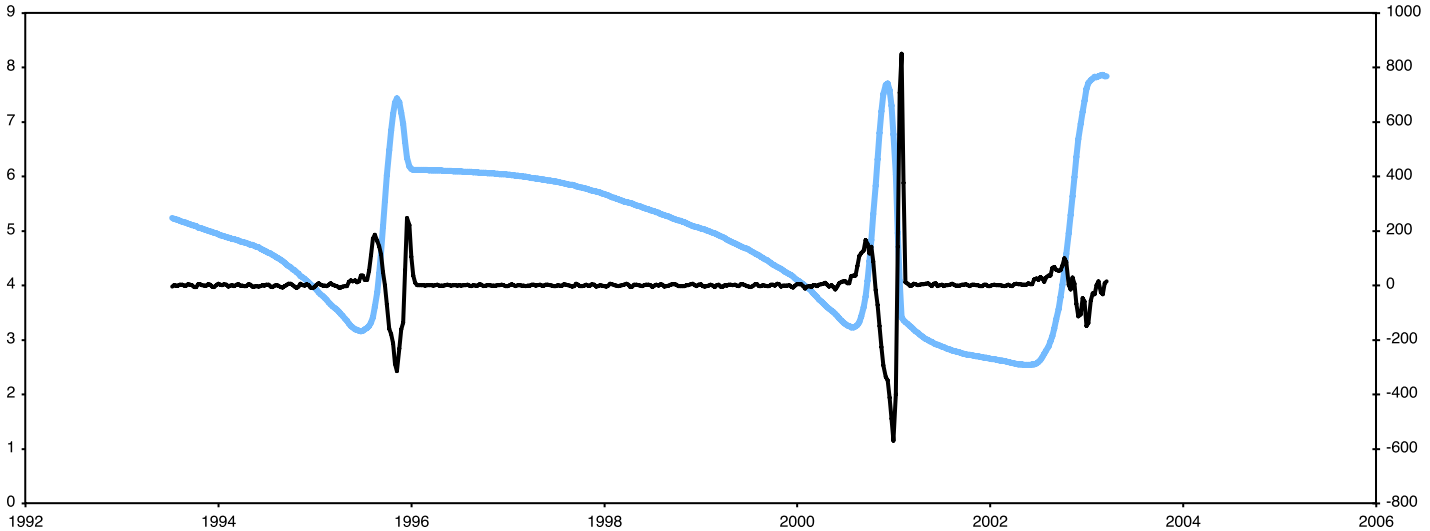

Figure S1 (e) New Zealand H3N2 HA segment

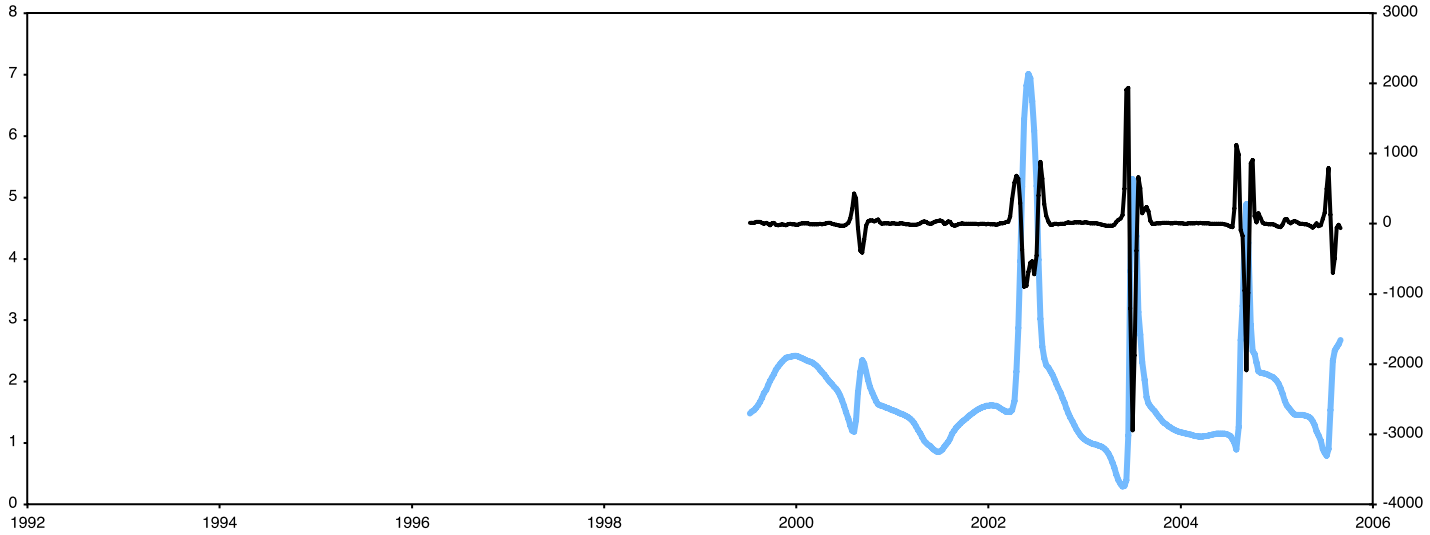

Figure S1 (f) New Zealand H1N1 HA segment

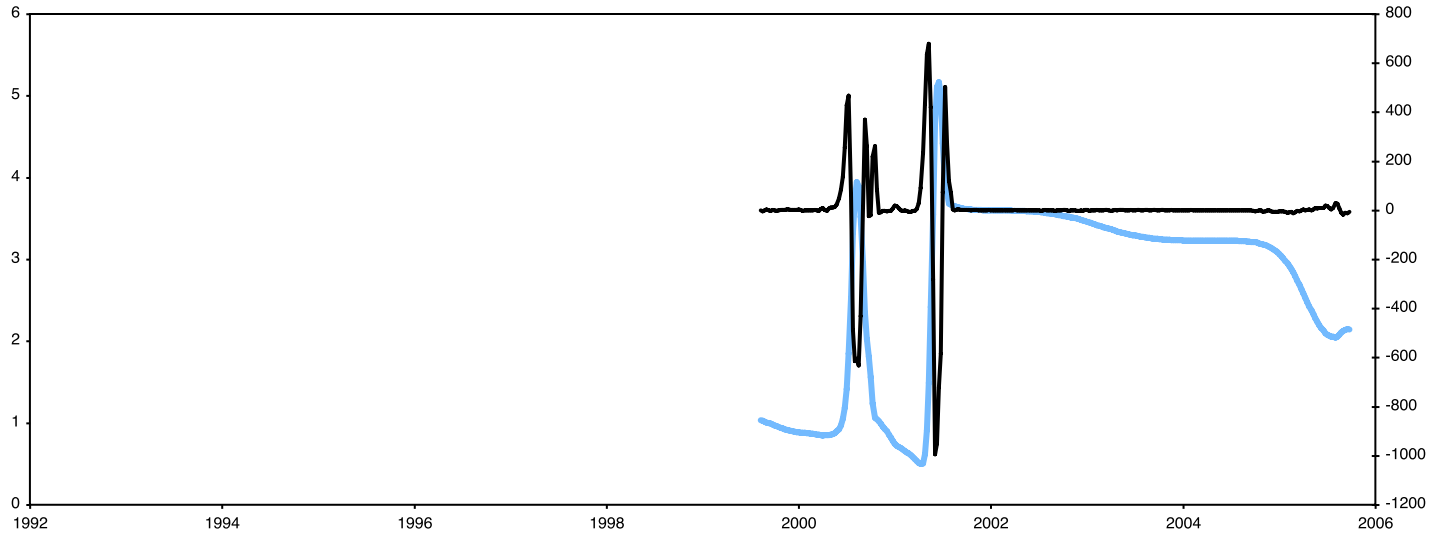

Figure S1 (g) New Zealand H3N2 NA segment

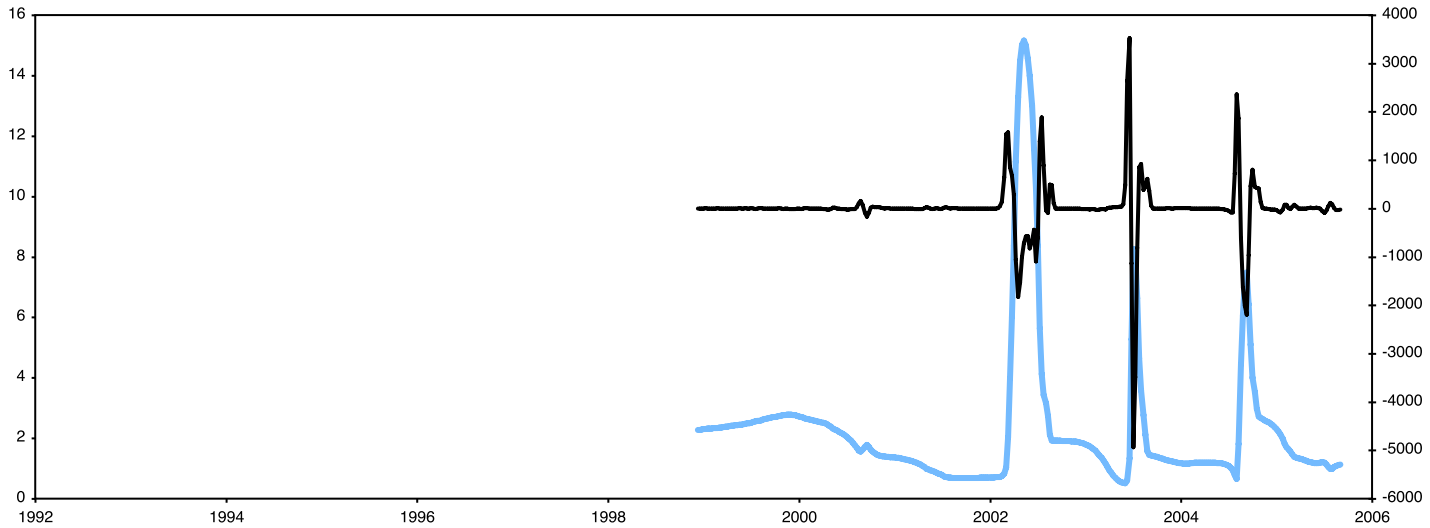

Figure S1 (h) New Zealand H1N1 NA segment

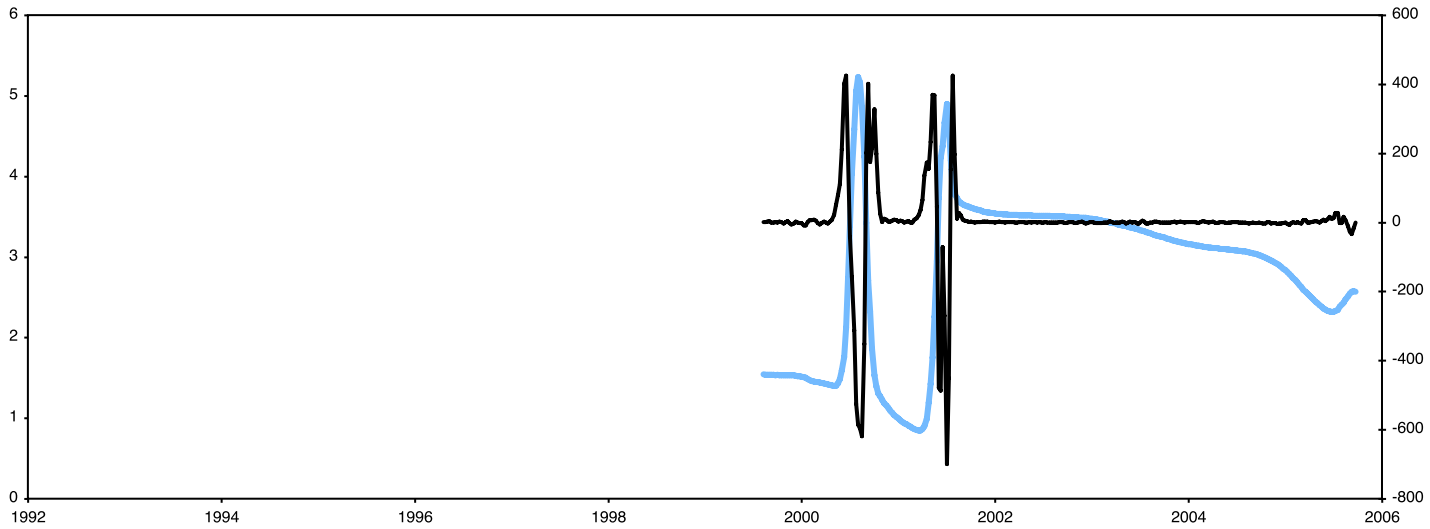

Fig. S2. The values of p were calculated for each season (see legend to Figure S1 for details). A scatterplot (not shown) indicates that the p values for A/H1N1 and A/H3N2 seasons are inversely related. A Wilcoxon Sign Rank test of the p values is strongly significant (prob < 0.002). The p values for each subtype were then ranked, and the ranks also found to be negatively correlated (this figure). Therefore, there is an inverse relationship between the estimated 'peakedness' of the A/H1N1 and A/H3N2 epidemics in each season.

Figure S2 Rank peakedness of H1N1 against H3N2

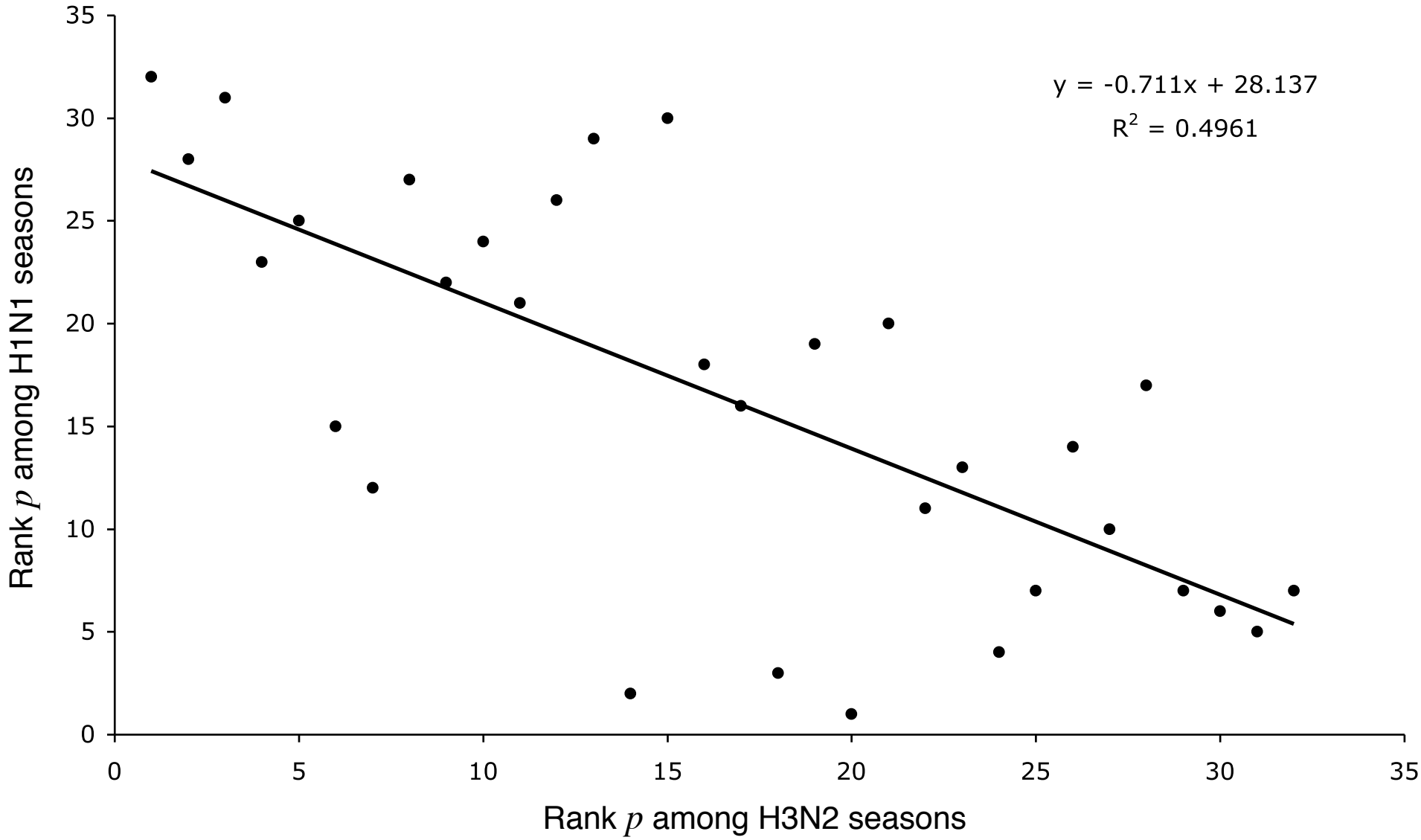

Fig. S3. Phylogenetic trees for each genome segment (a-h) of H3N2 human influenza A virus from New York State, USA. Nexus tree files for each segment are also provided. The trees presented are the Maximum Clade Credibility (MCC) trees summarized from the 10,000 trees sampled using the MCMC method available in BEAST. The MCC tree is the sampled tree with the highest product of the individual clade posterior probabilities. The node ages are shown as the mean age of that node across the posterior sample.

Figure S3 (a) PB2 segment

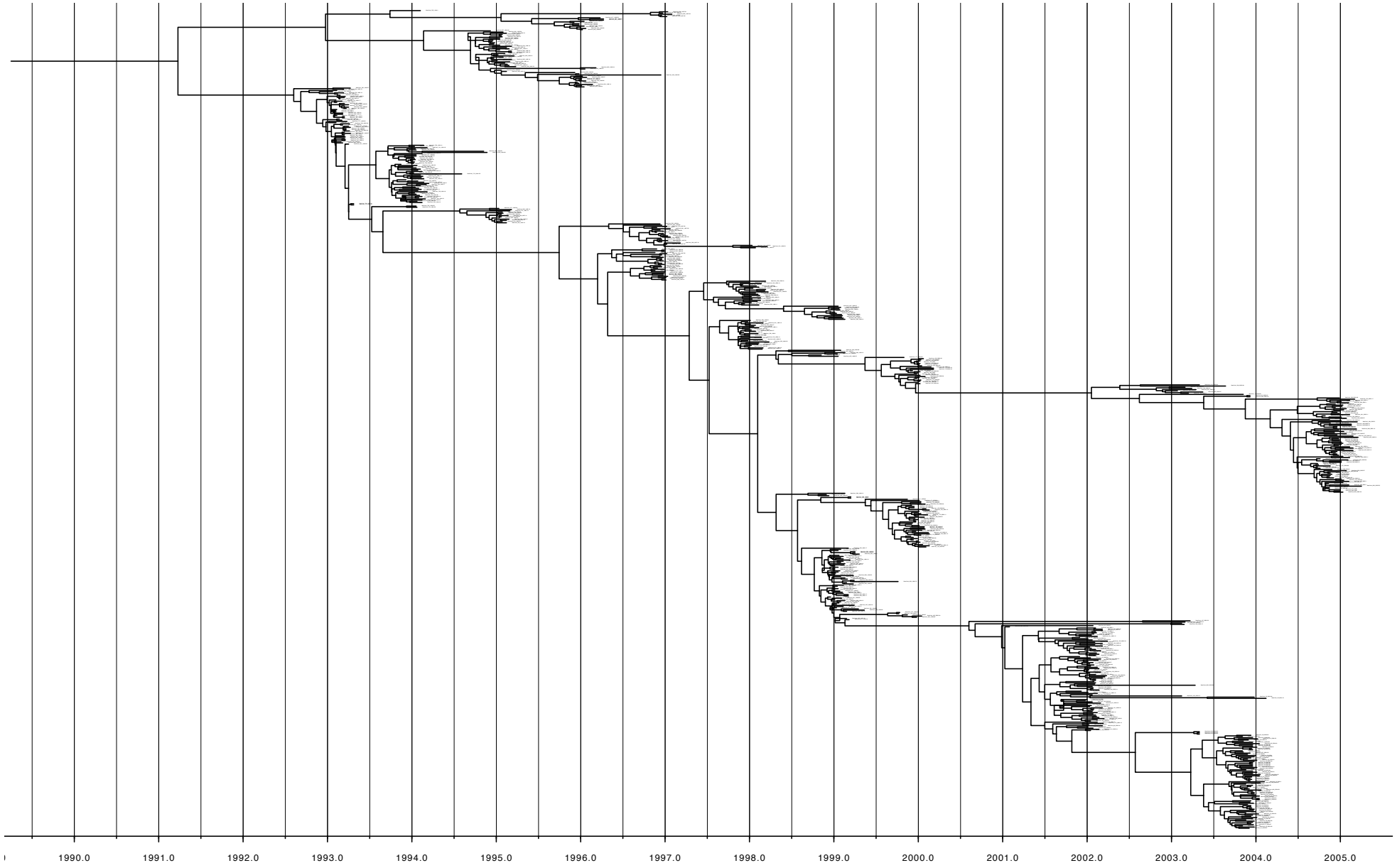

Figure S3 (b) PB1 segment

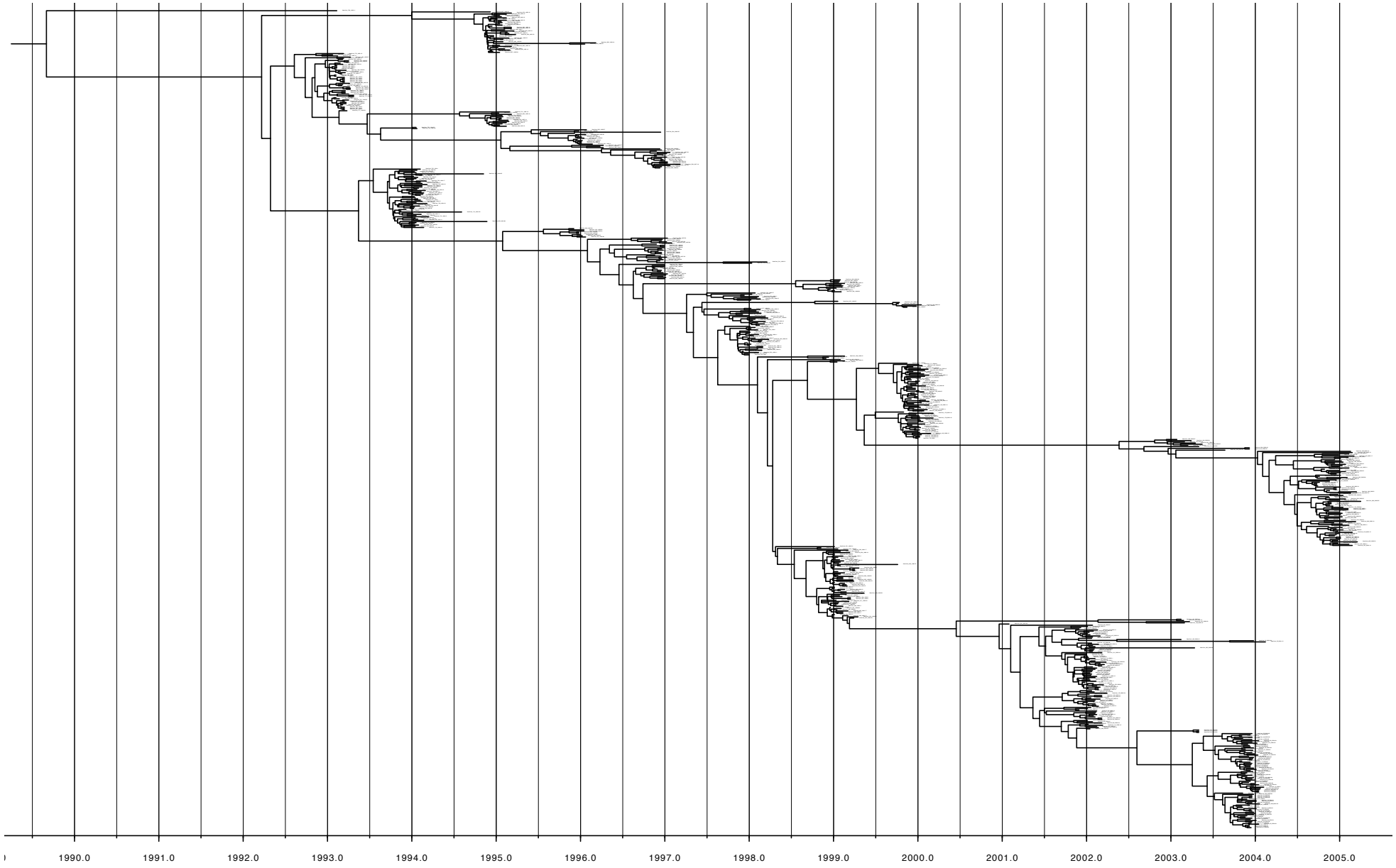

Figure S3 (c) PA segment

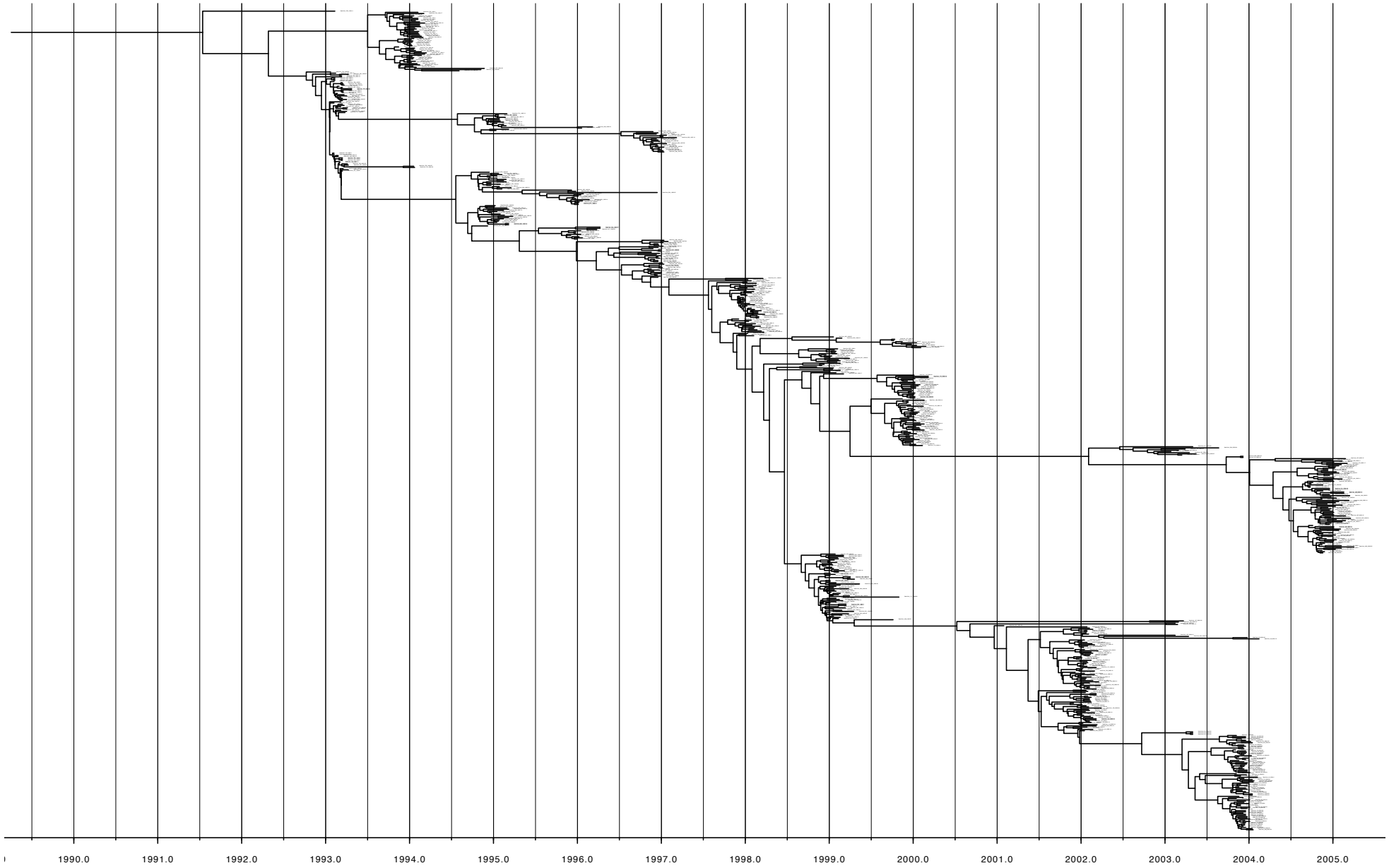

Figure S3 (d) NP segment

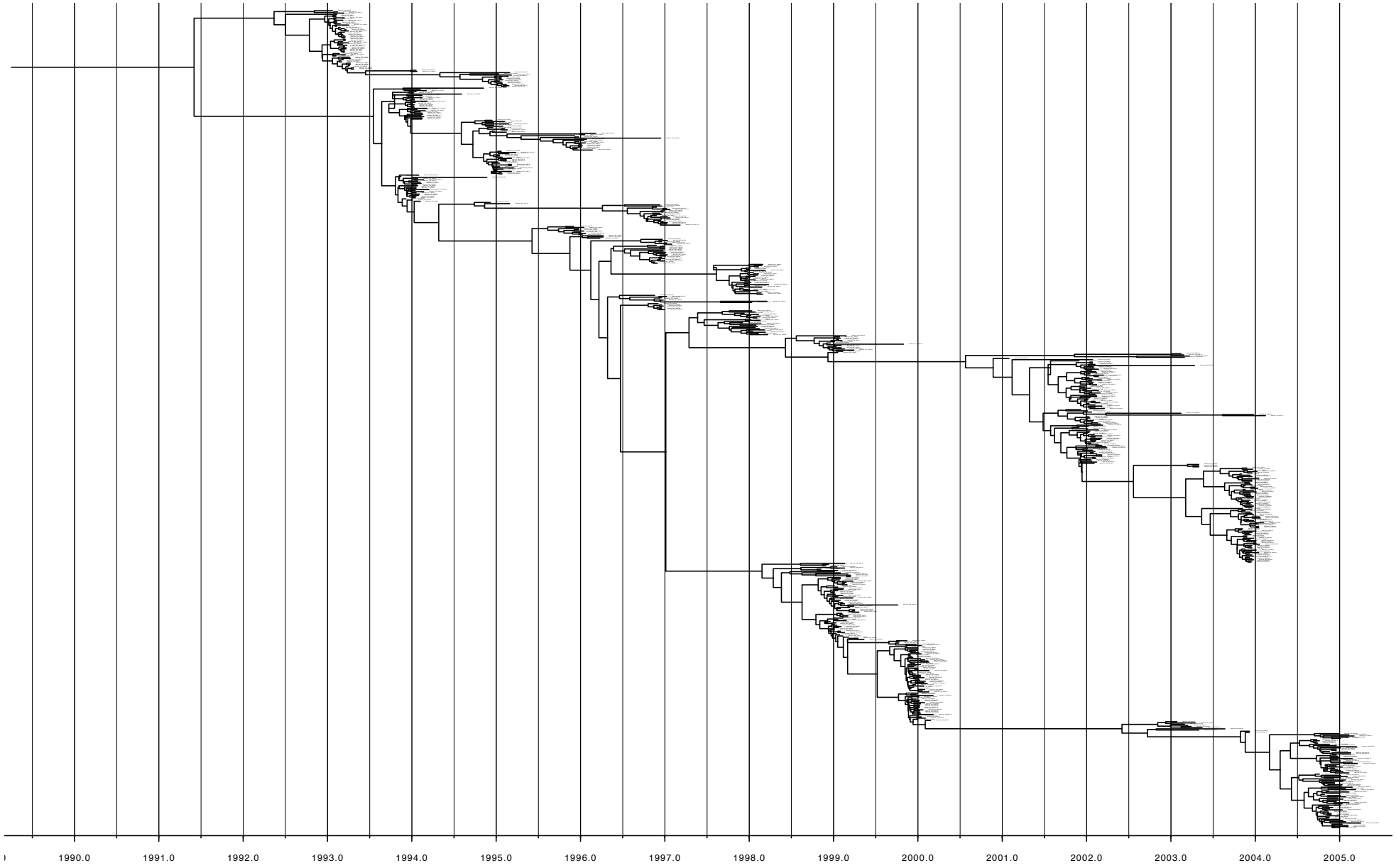

Figure S3 (e) HA segment

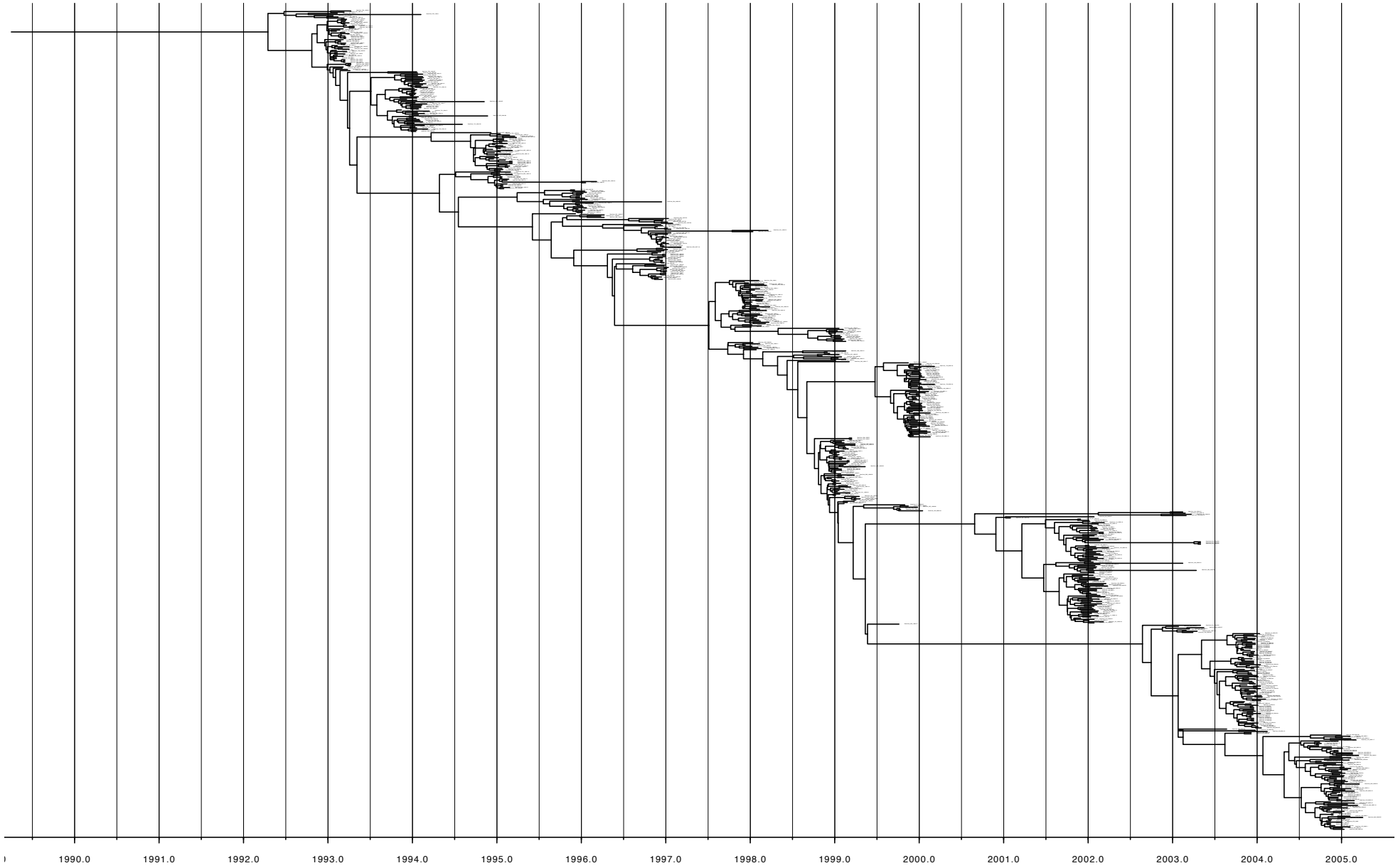

Figure S3 (f) NA segment

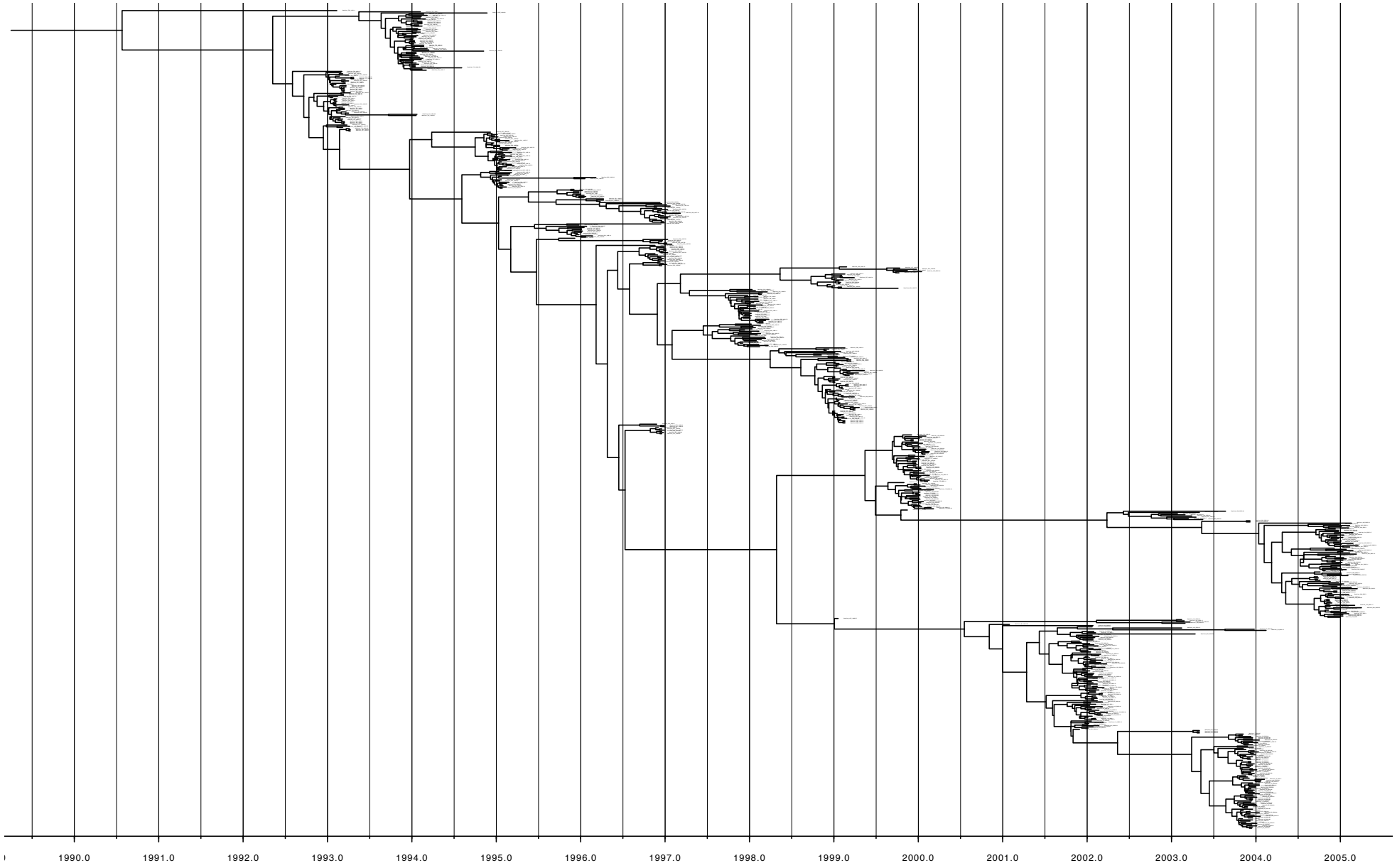

Figure S3 (g) M1/2 segment

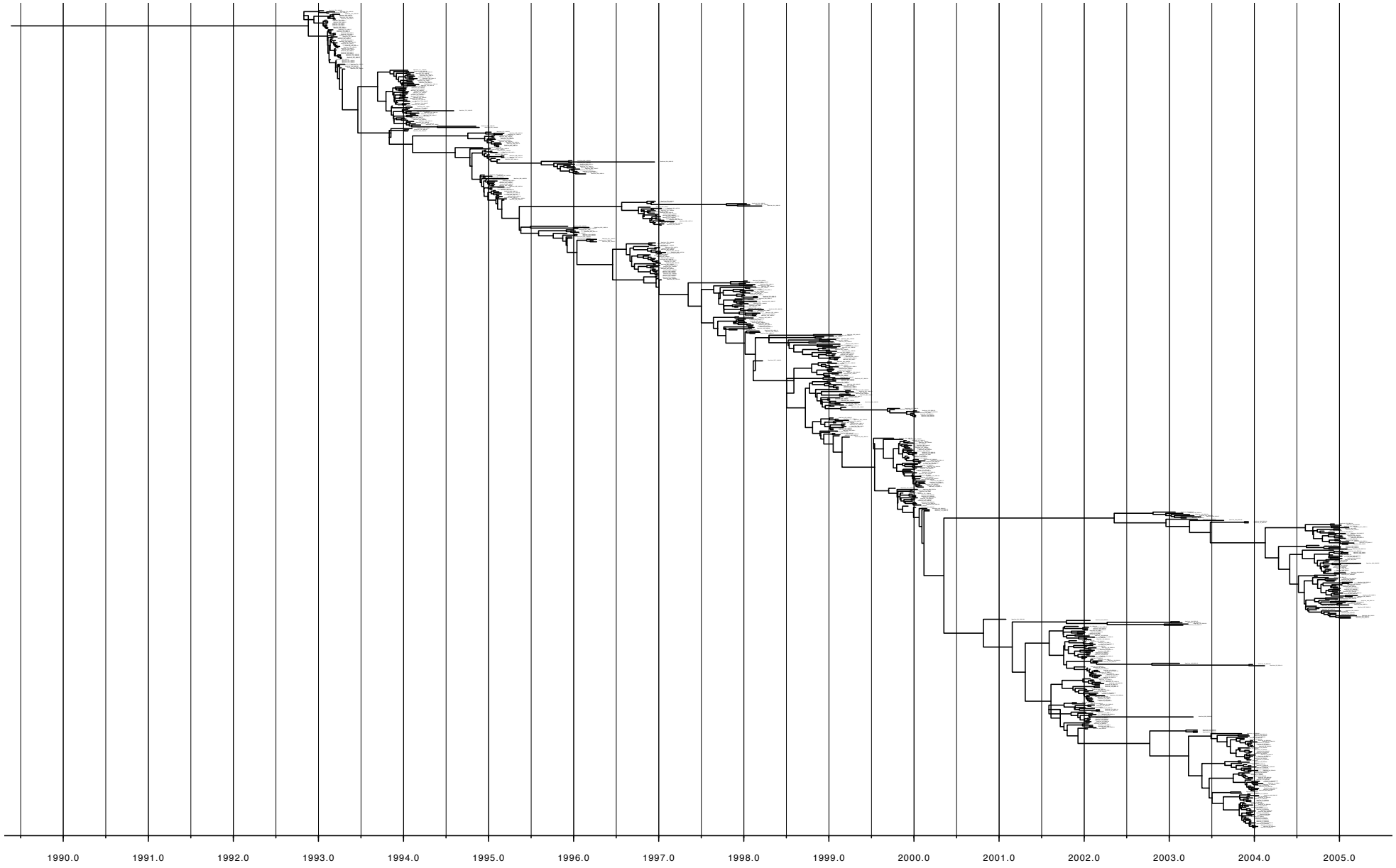

Figure S3 (h) NS1/2 segment

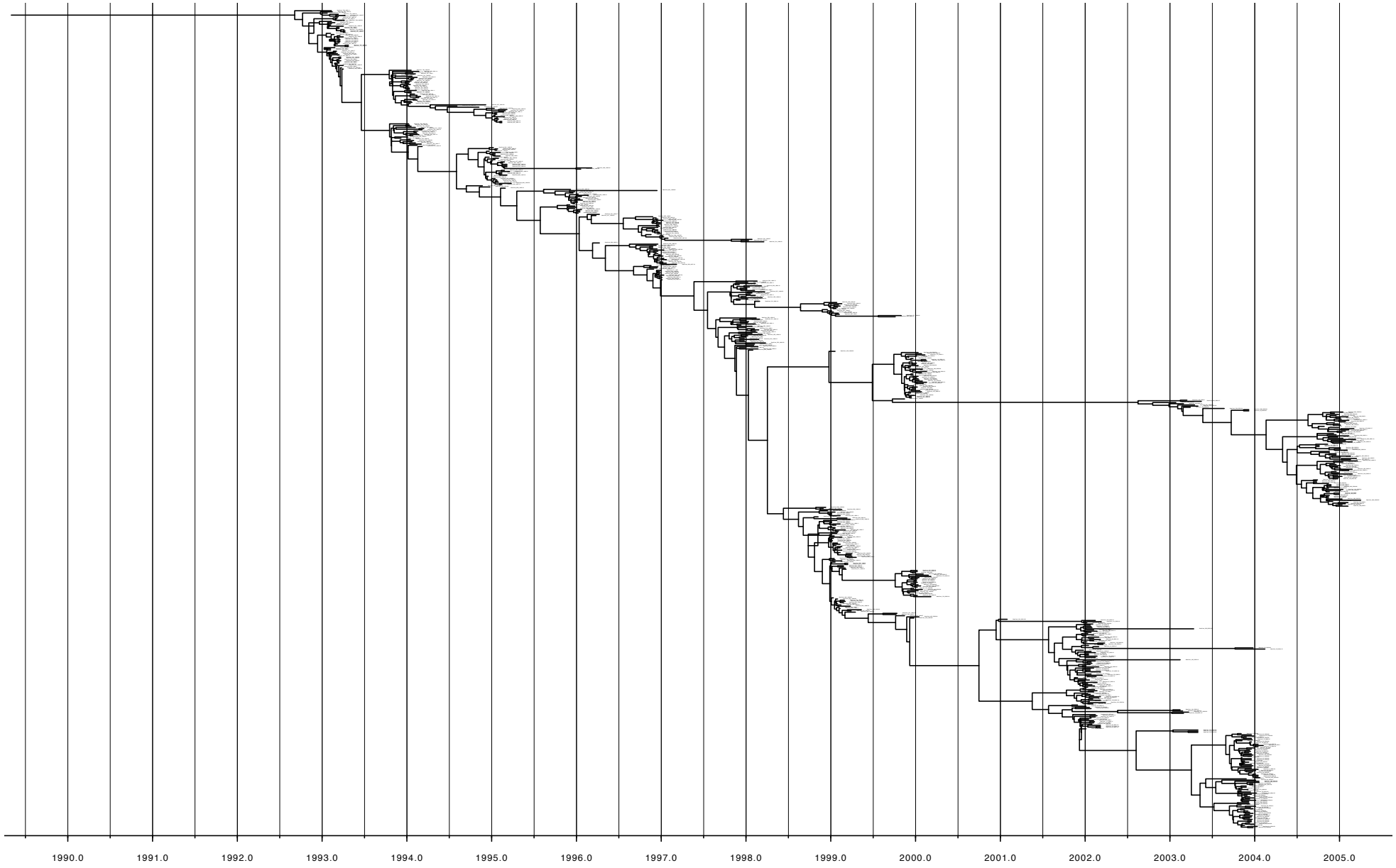

Fig. S4. Bayesian skyline plots each genome segment (a-h) of H3N2 human influenza A virus from New York State, USA. A measure of relative genetic diversity is shown on the y-axis (see Methods for a more complete description) with the 95% highest posterior density (HPD) intervals shown in blue. The thin black line is a generalized skyline plot reconstructed for the respective Maximum Clade Credibility tree given in Figure S3. (i) The plots for all genome segments of A/H3N2 from New York State, USA overlaid for comparison.

Figure S4 (a) PB2 segment

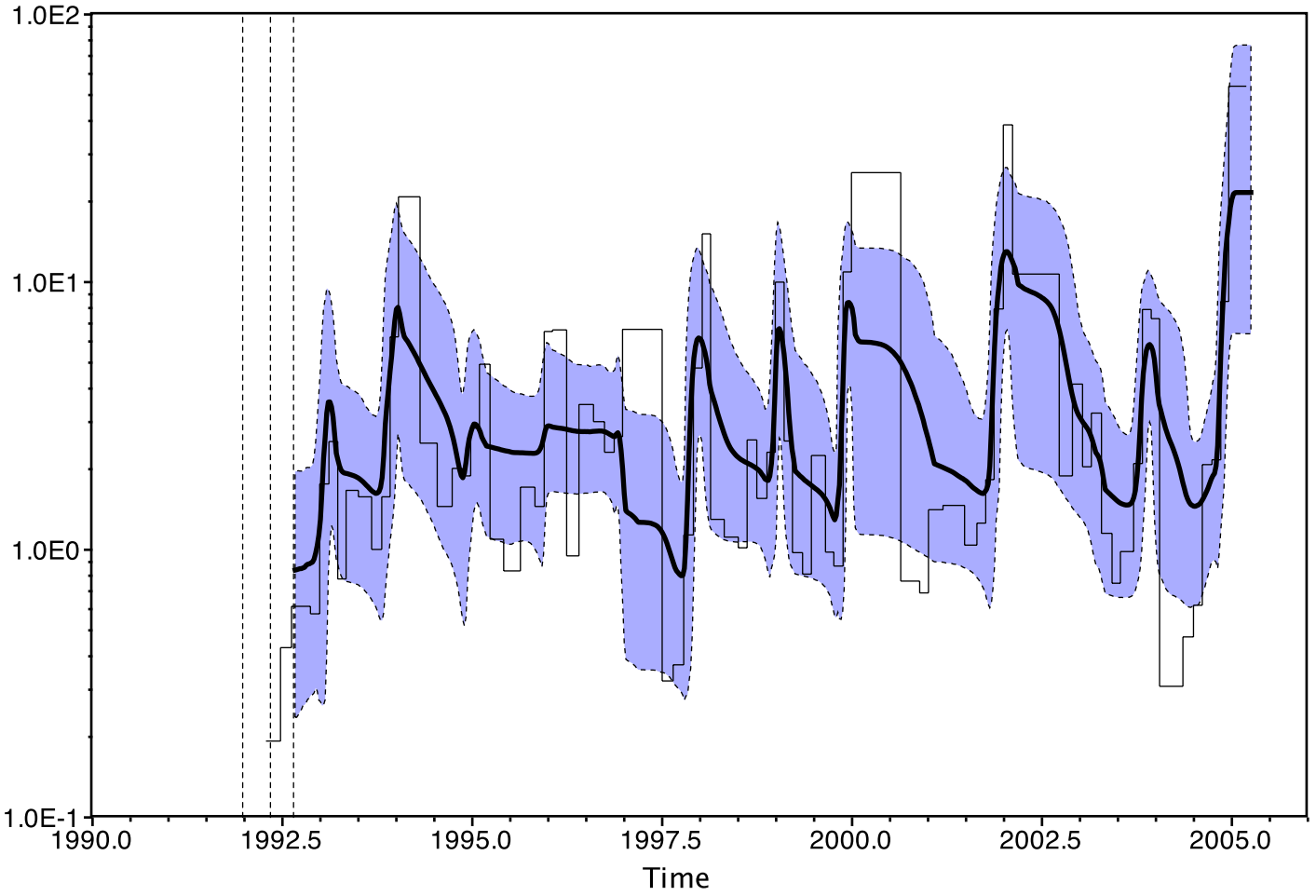

Figure S4 (b) PB1 segment

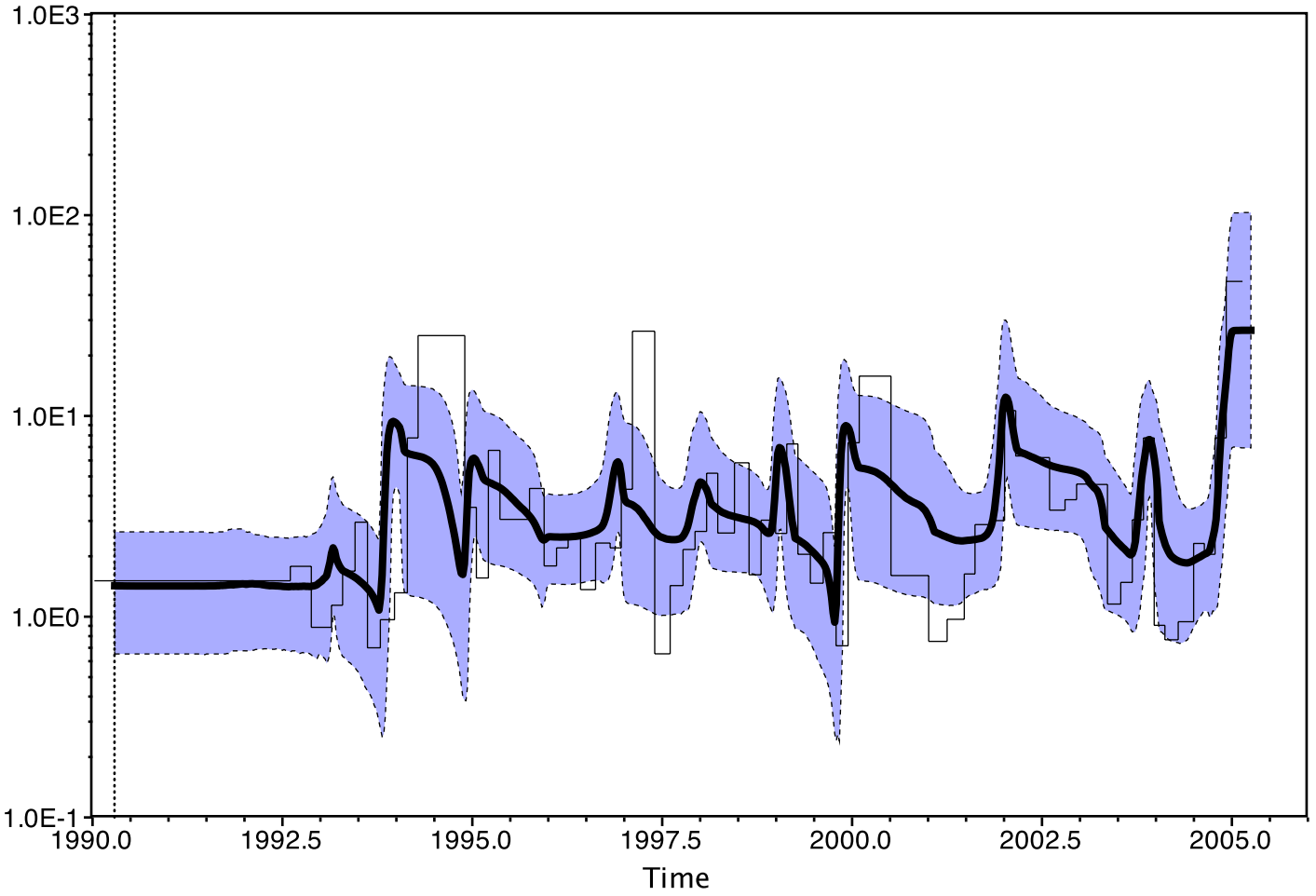

Figure S4 (c) PA segment

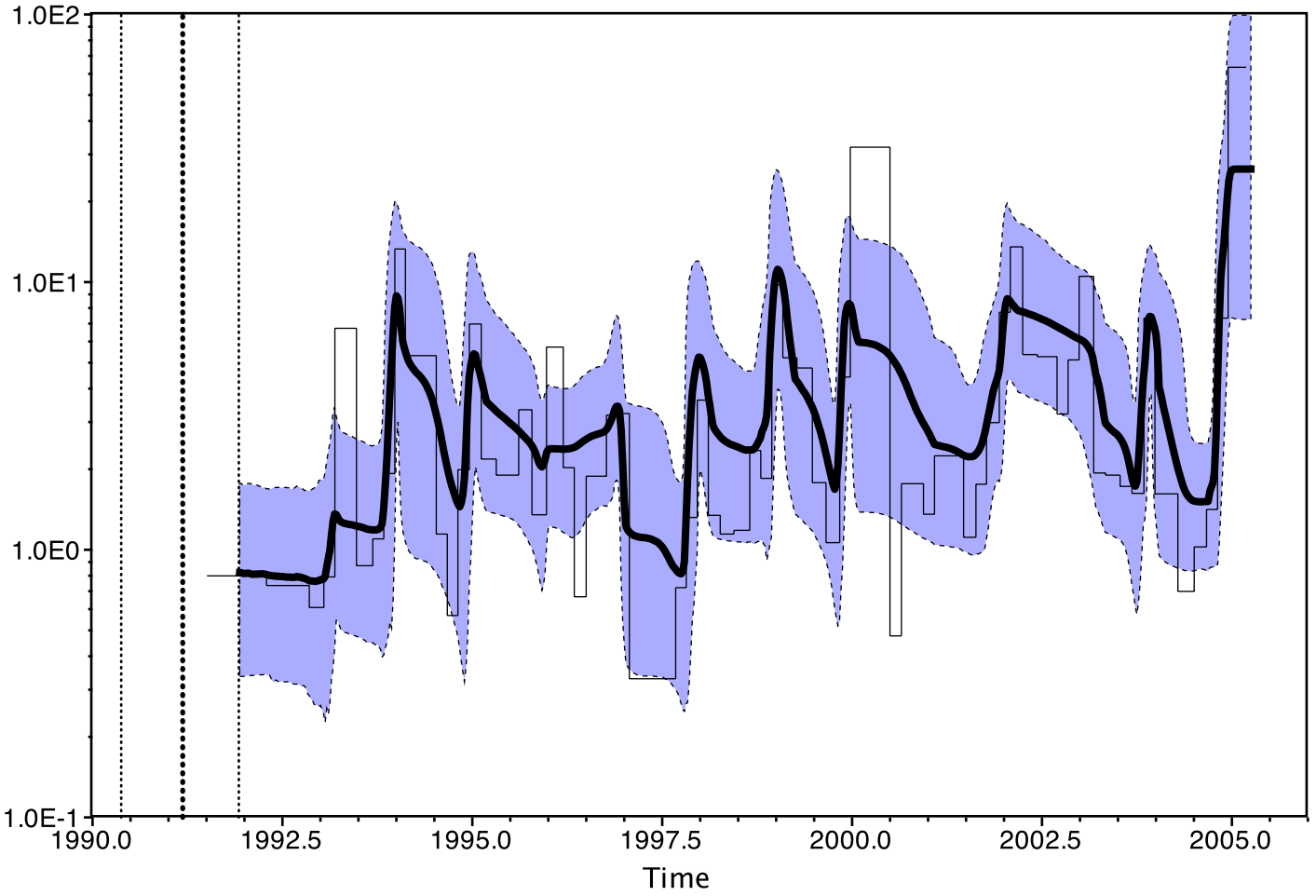

Figure S4 (d) NP segment

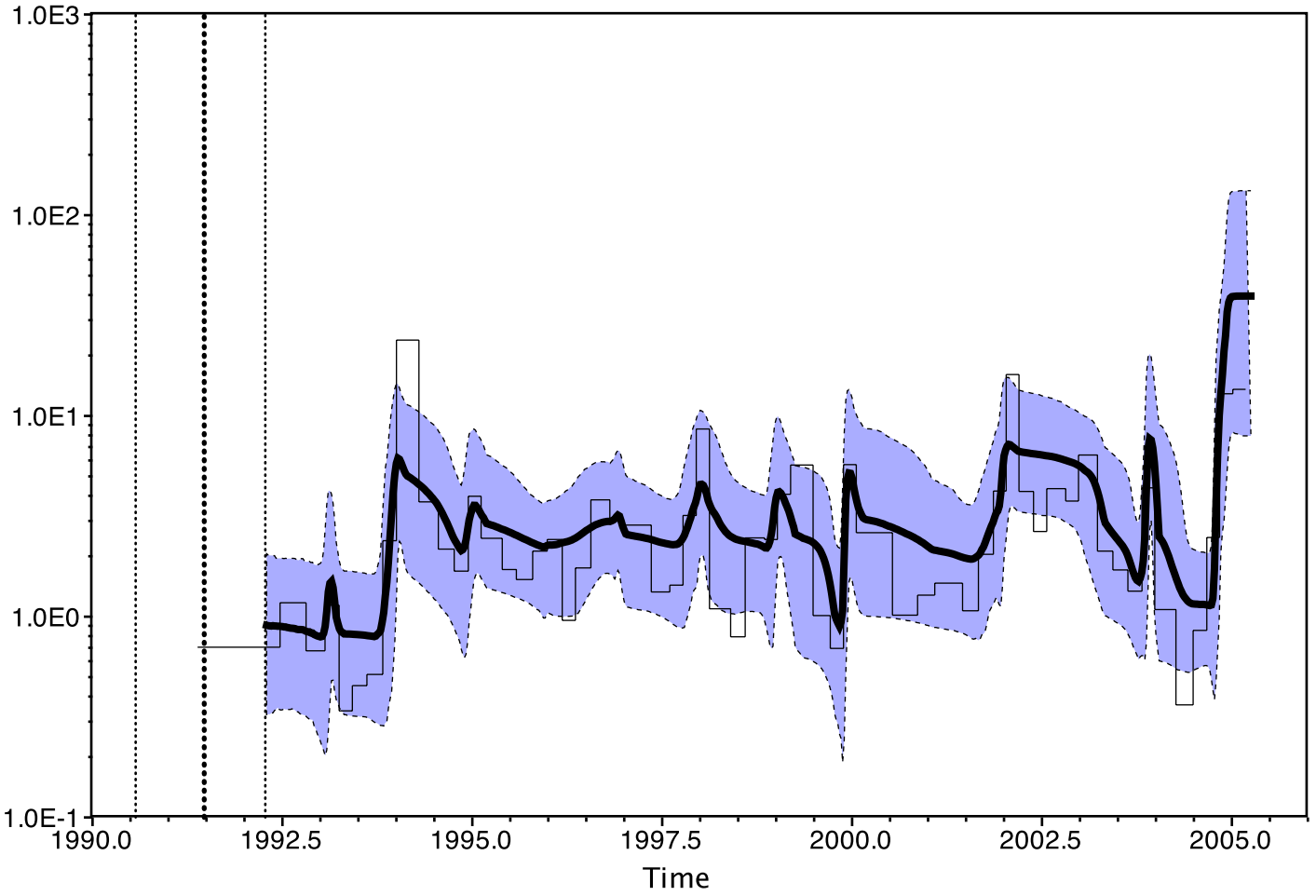

Figure S4 (e) HA segment

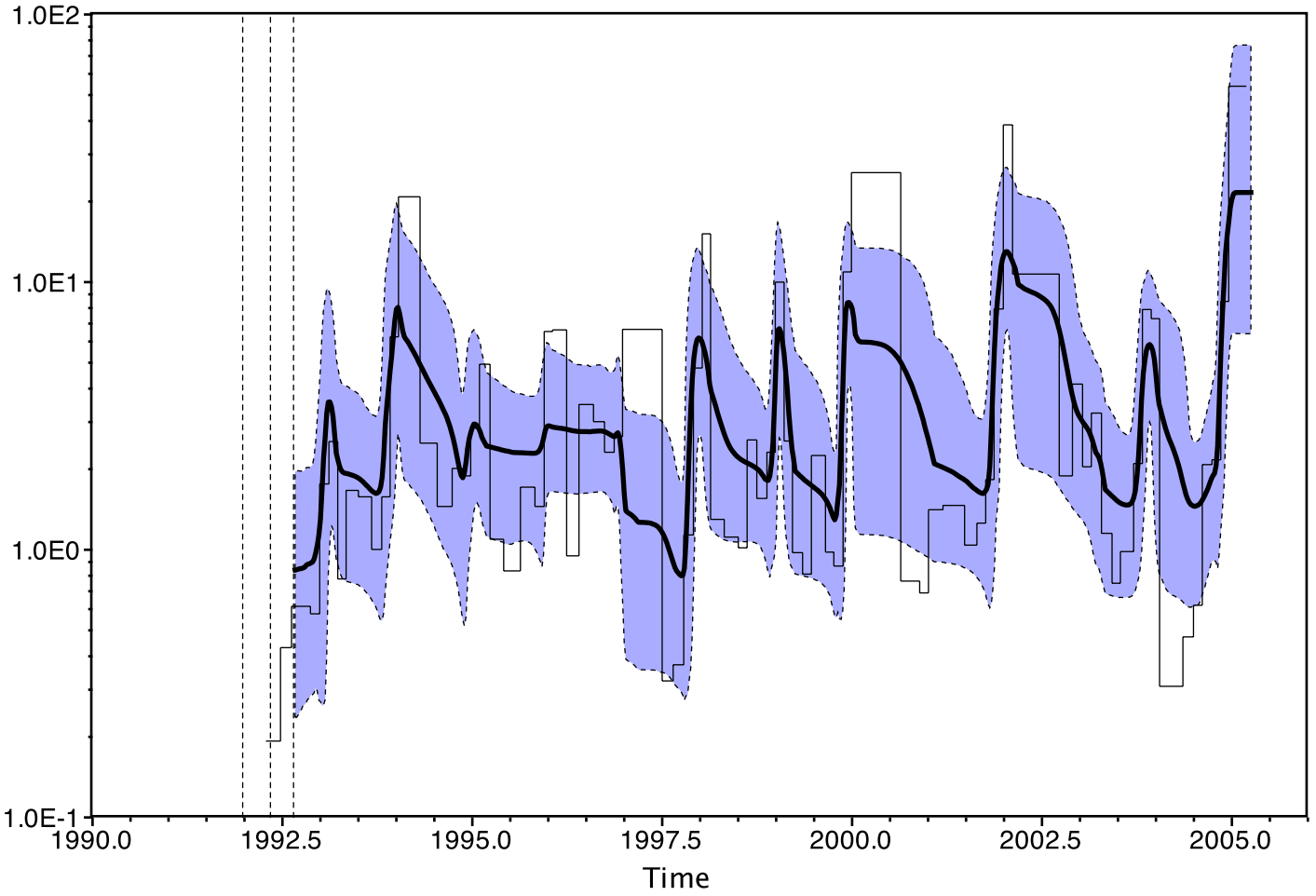

Figure S4 (f) NA segment

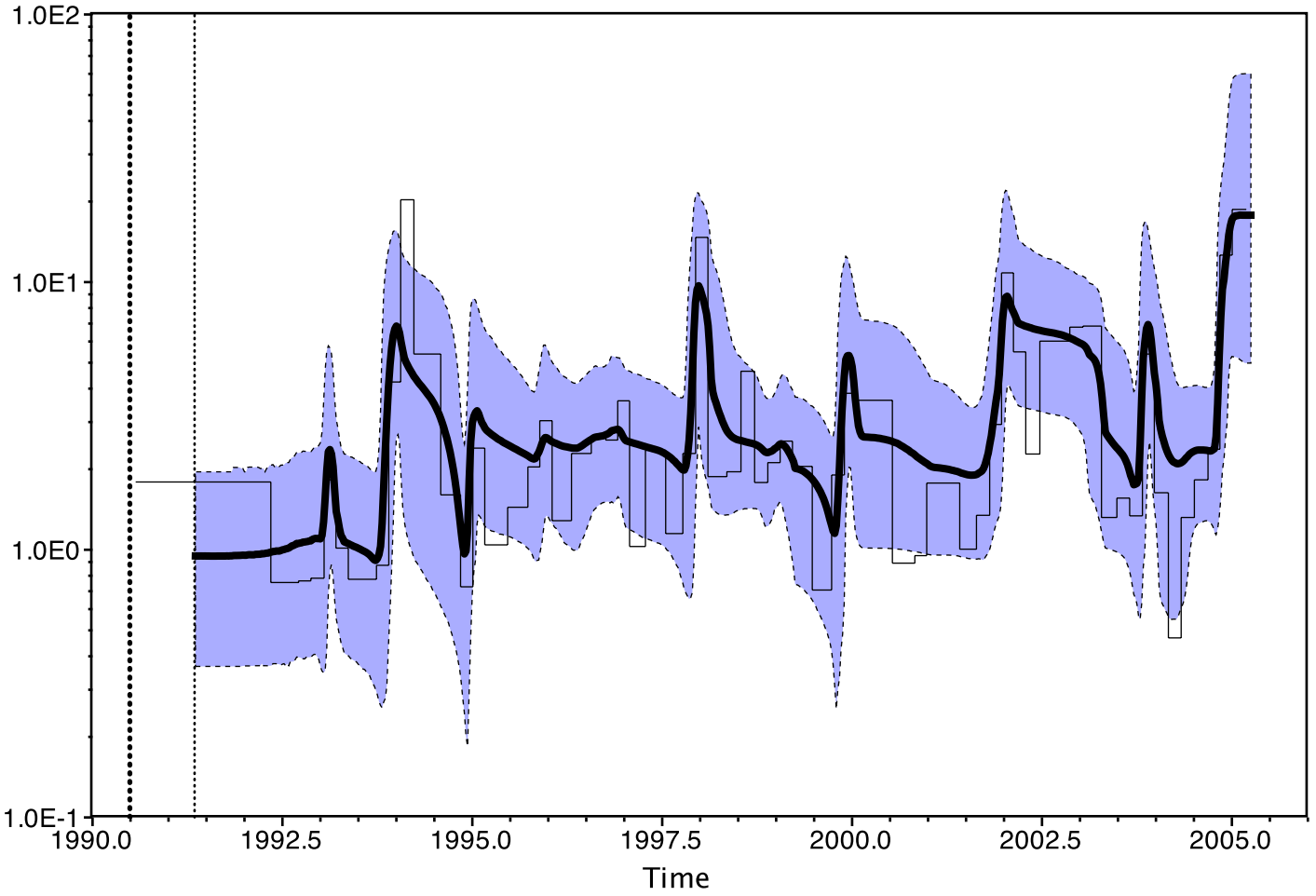

Figure S4 (g) M1/2 segment

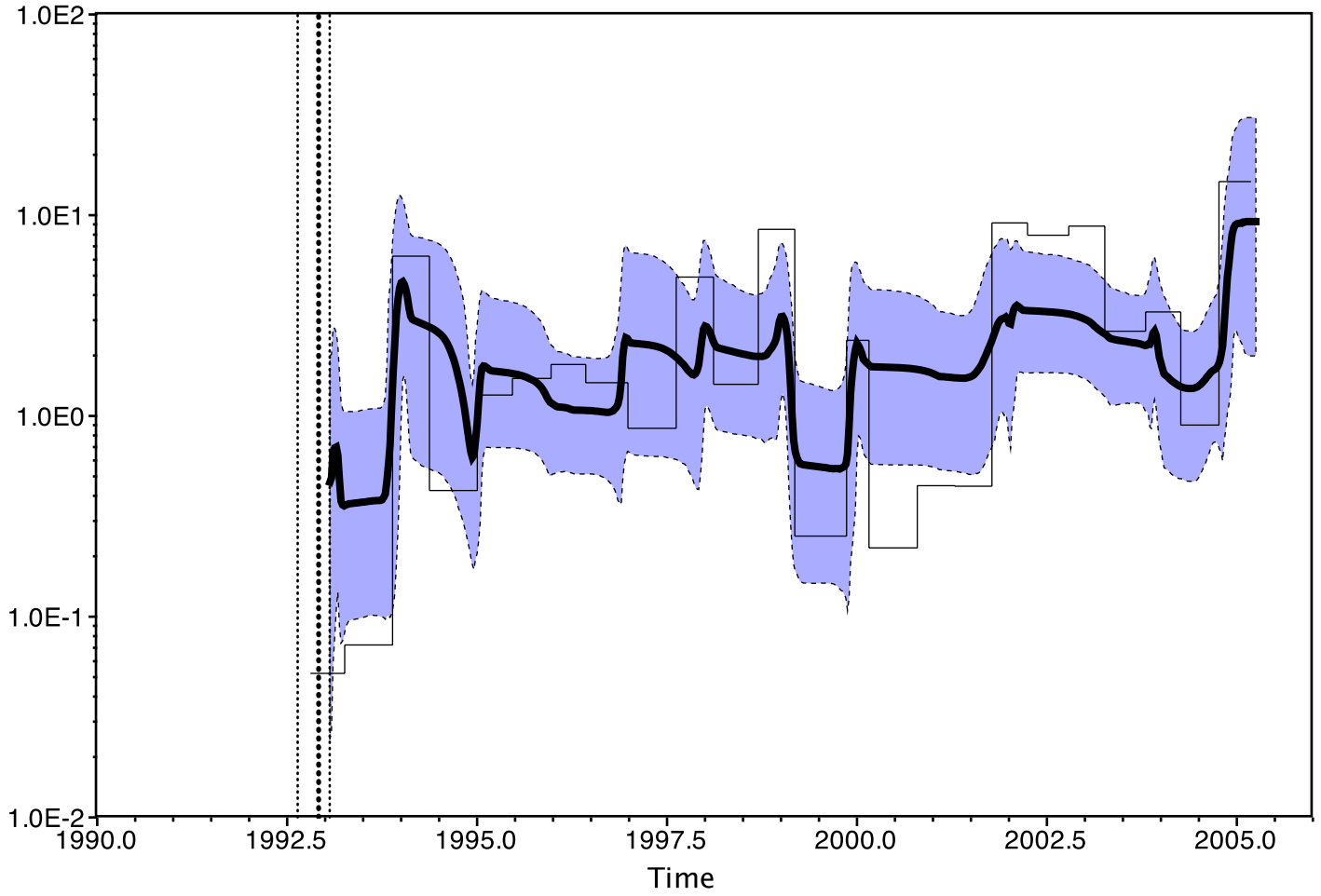

Figure S4 (h) NS1/2 segment

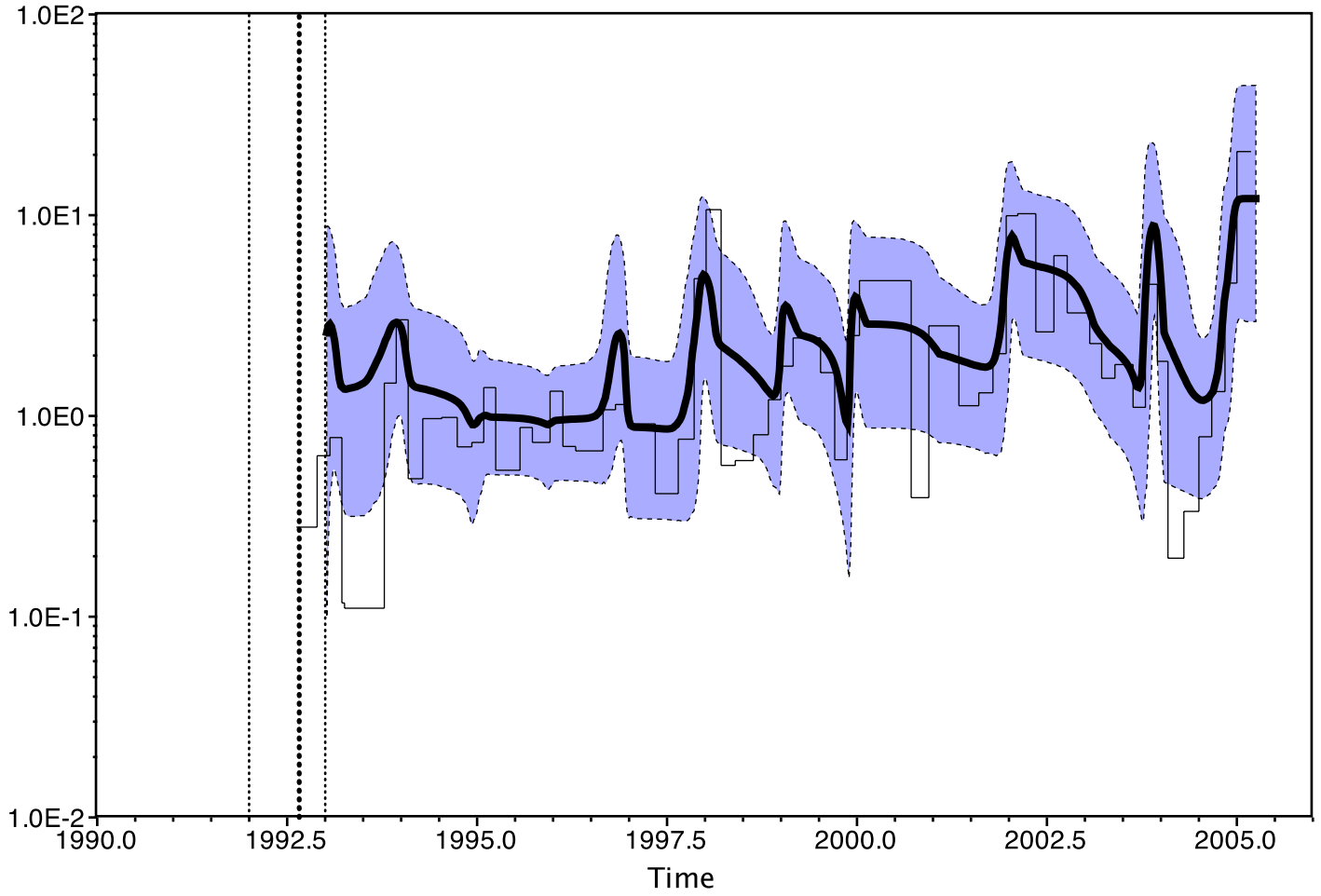

Figure S4 (i) New York H3N2 all gene segments

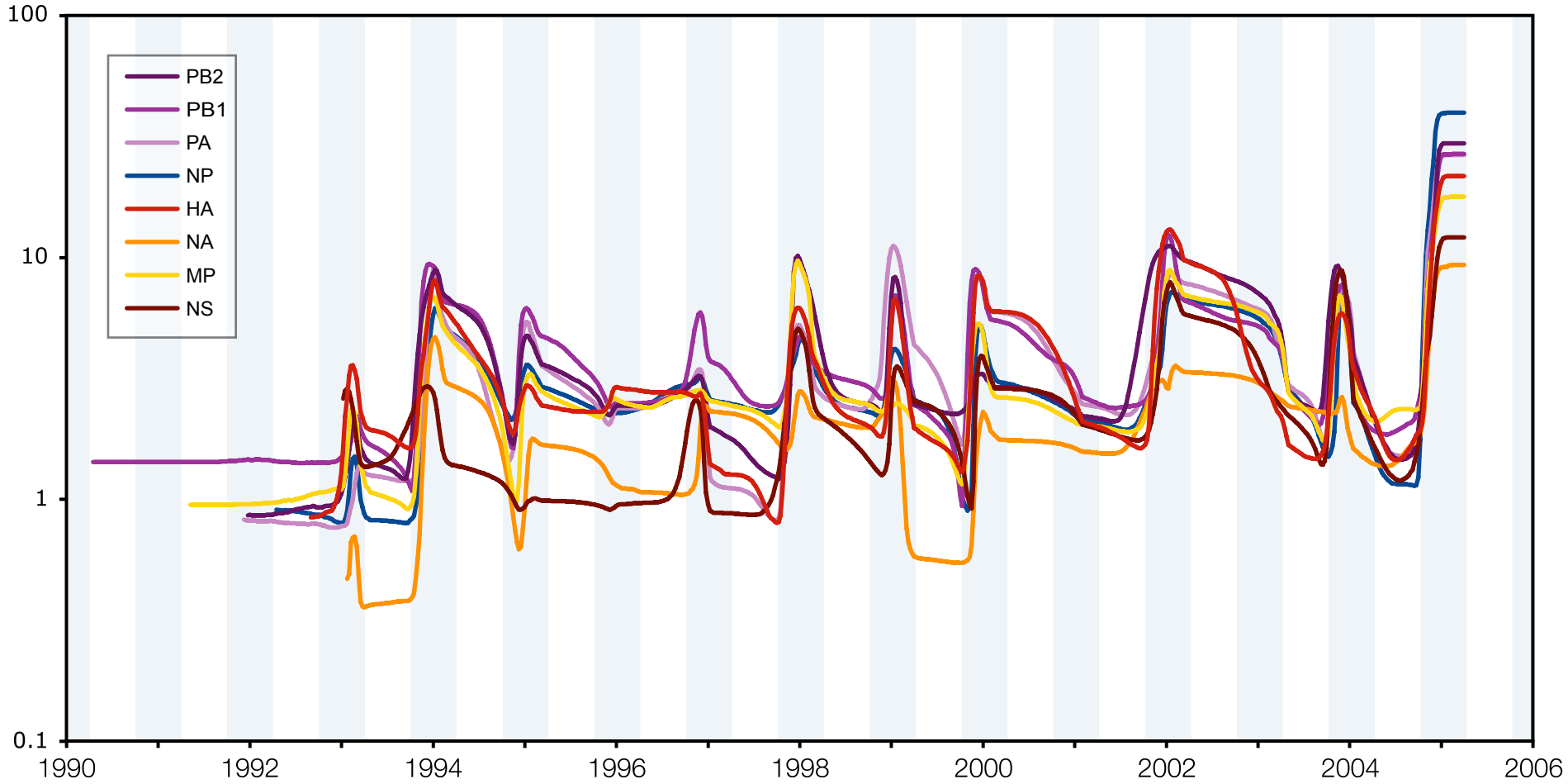

Fig. S5. Bayesian skyline plots of (a) the HA segment, (b) the NA segment of H3N2 human influenza A virus from New Zealand. See legend for S4, above, for details.

Figure S5 (a) H3N2, New Zealand, HA segment

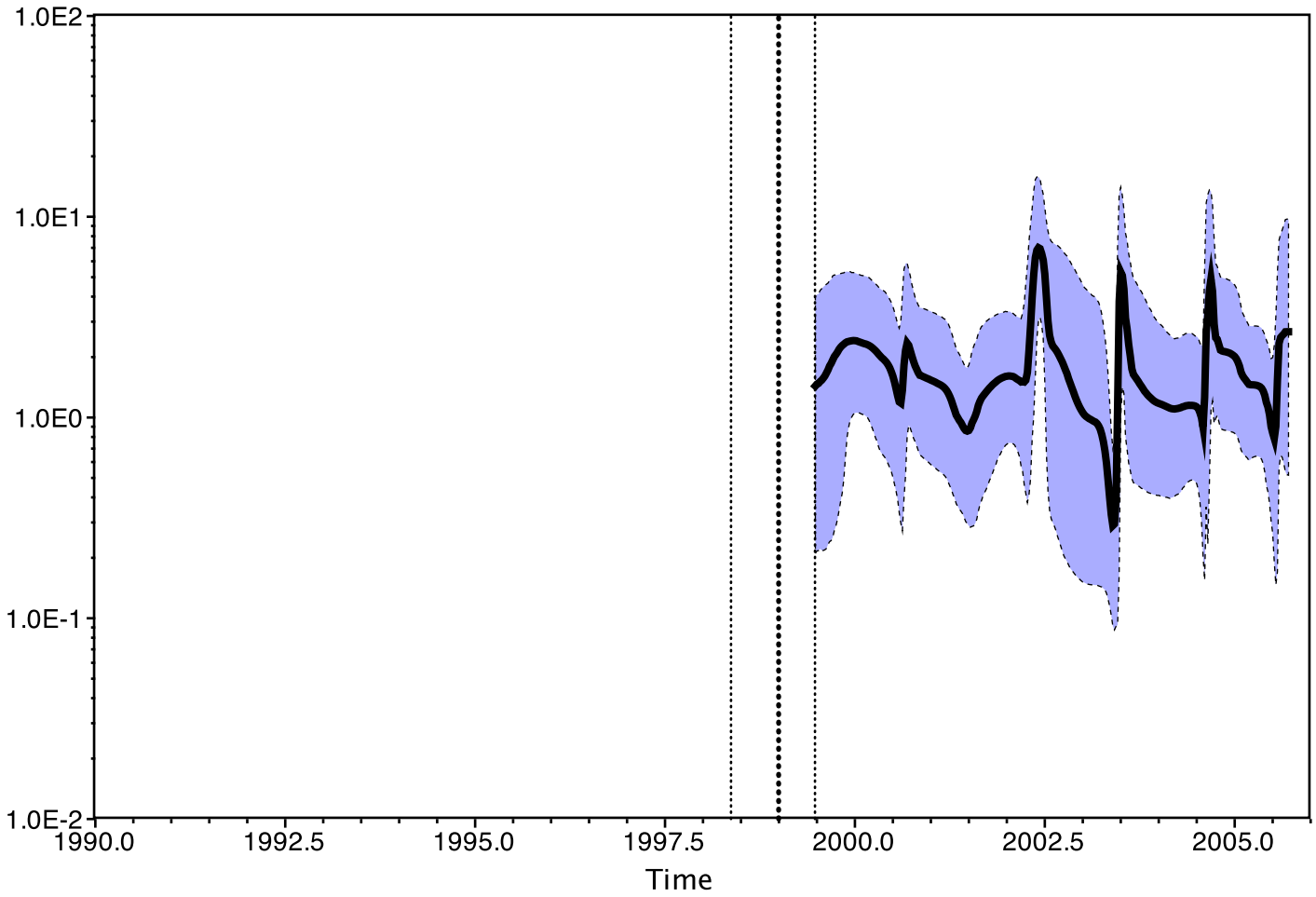

Figure S5 (b) H3N2, New Zealand, NA segment

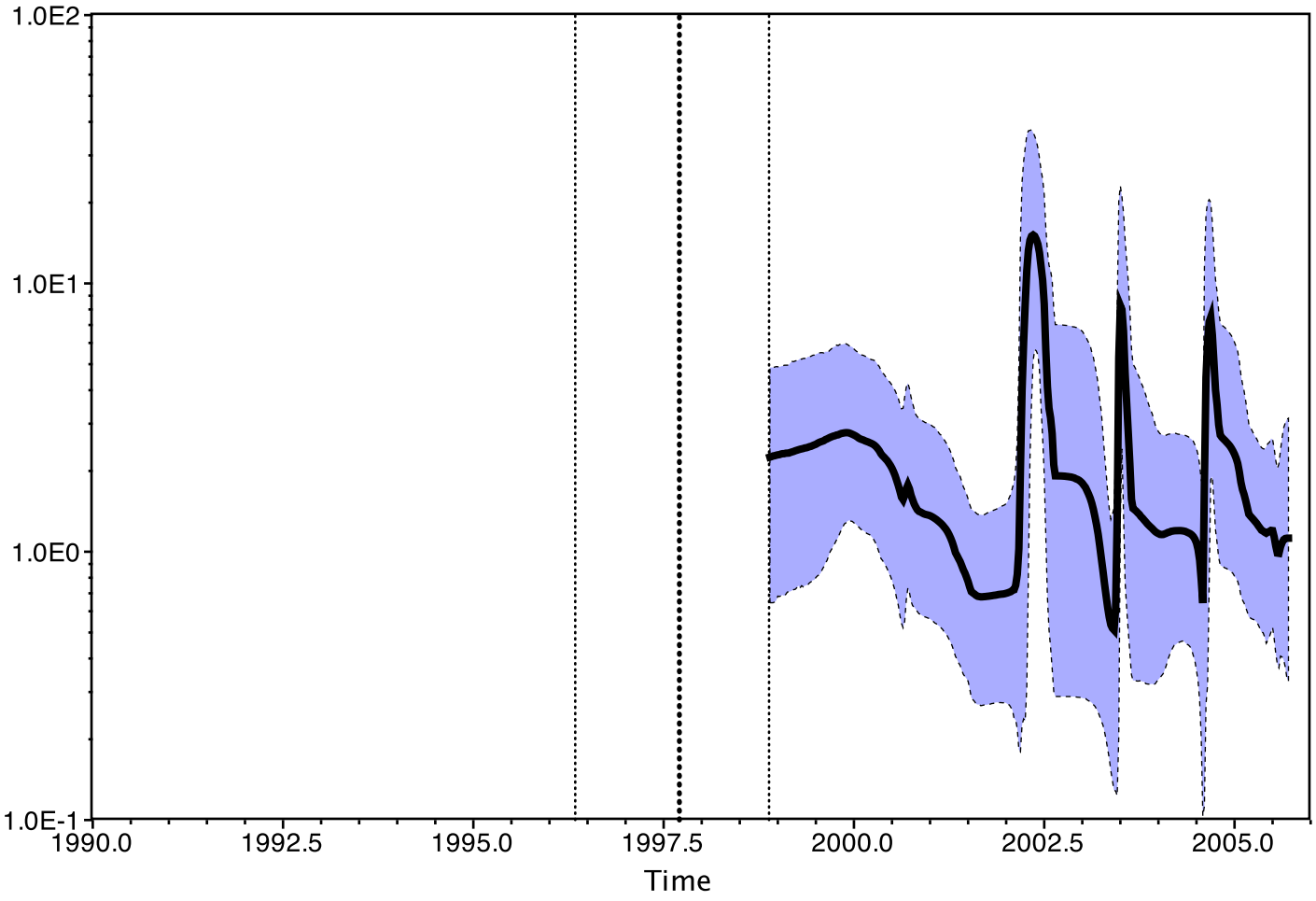

Fig. S6. Bayesian skyline plots of (a) the HA segment, (b) the NA segment of H1N1 human influenza A virus from New York State, USA. See legend for S4, above, for details.

Figure S6 (a) H1N1, New York, HA segment

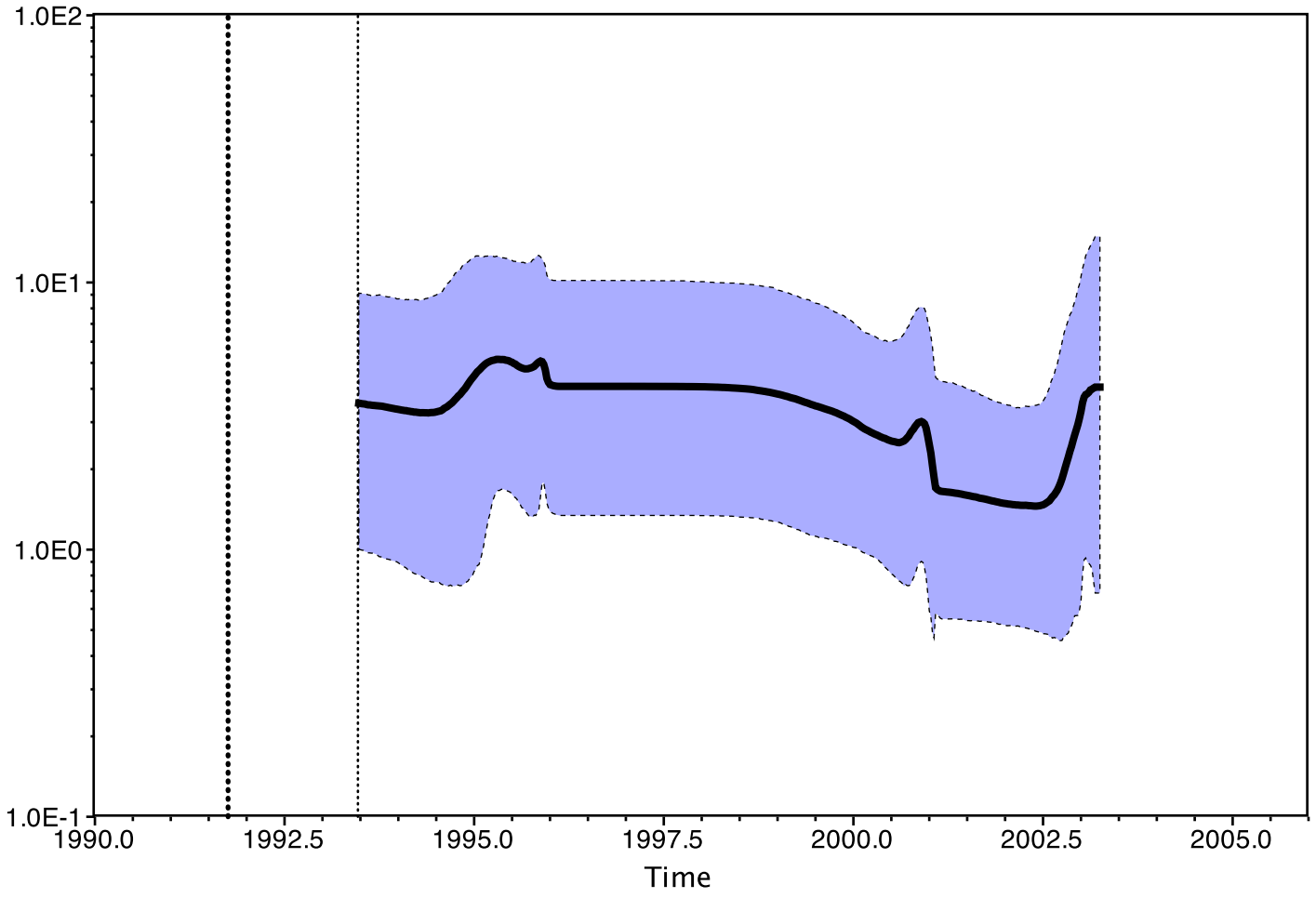

Figure S6 (b) H1N1 New York, NA segment

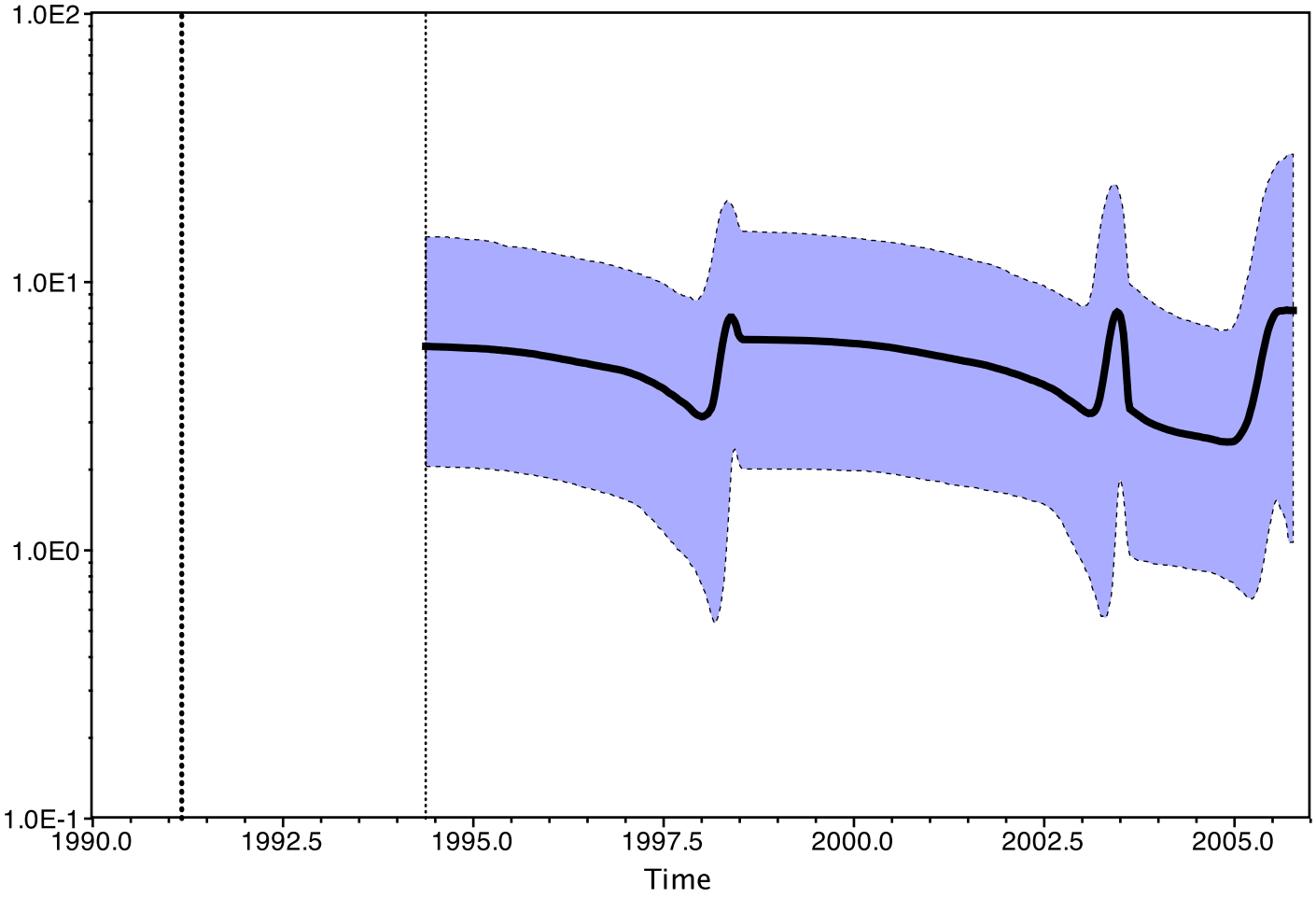

Fig. S7. Bayesian skyline plots of (a) the HA segment, (b) the NA segment of H1N1 human influenza A virus from New Zealand. See legend for S4, above, for details.

Figure S7 (a) H1N1, New Zealand, HA segment

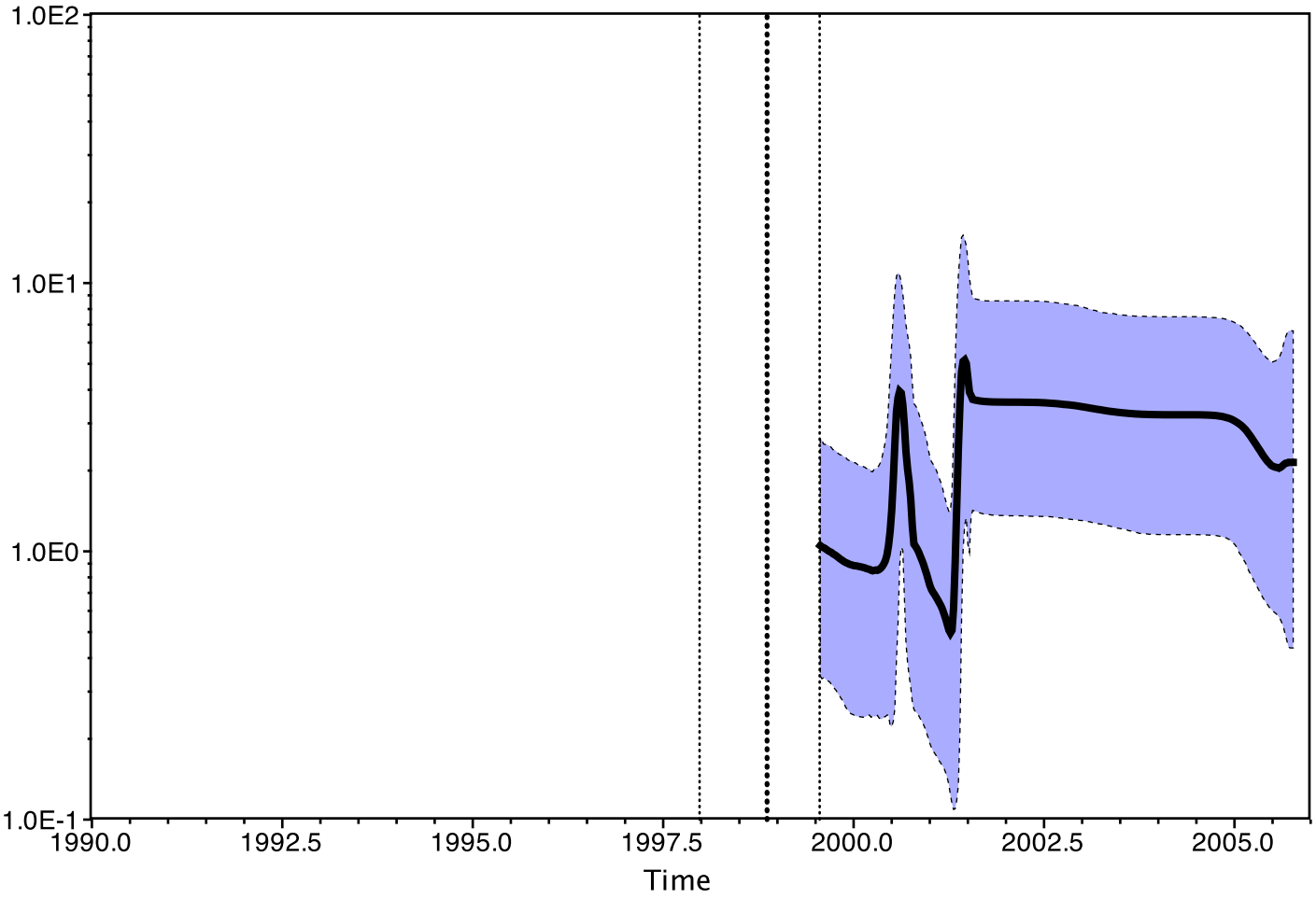

Figure S7 (b) H1N1, New Zealand, NA segment

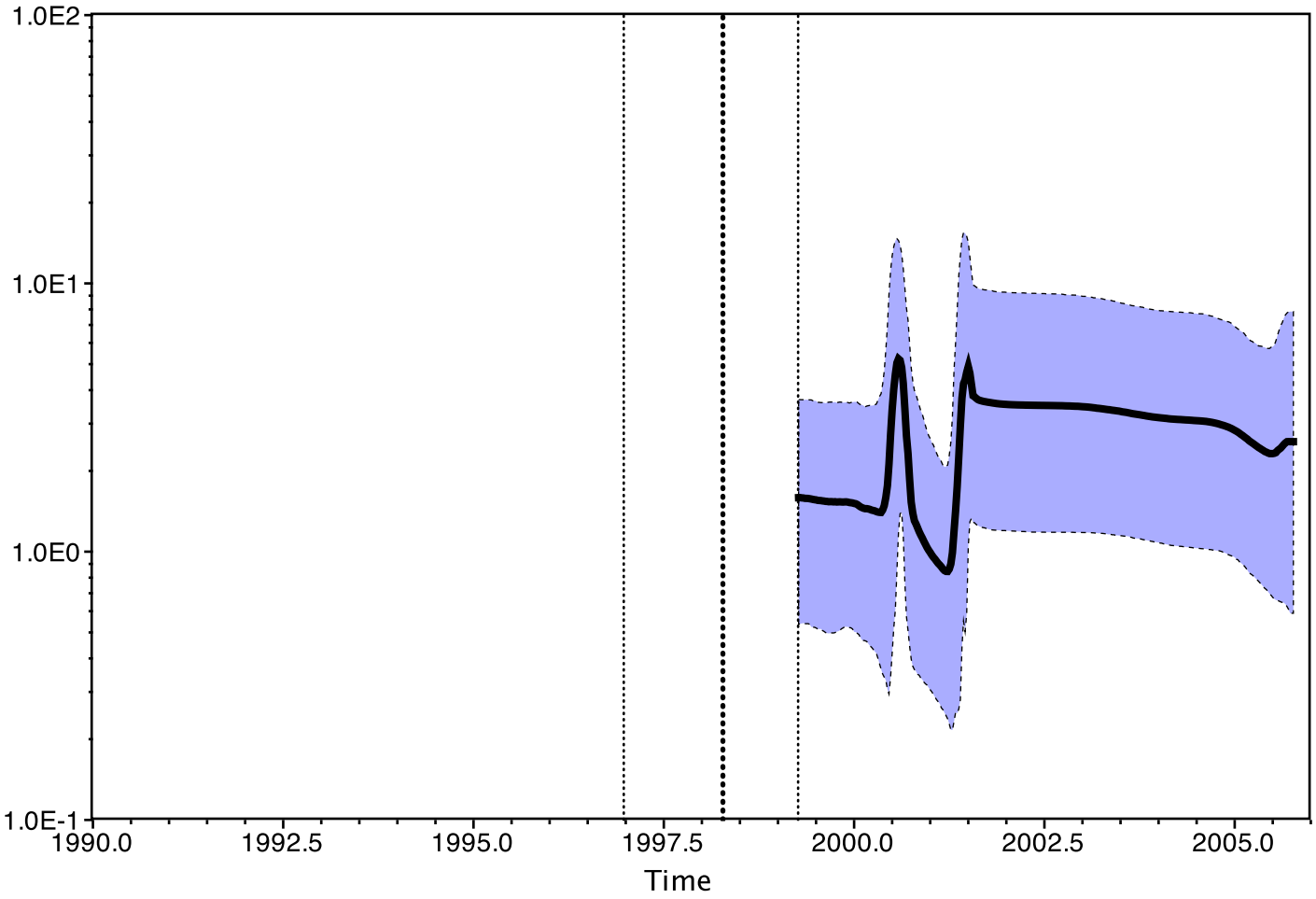

Fig. S8. Simulations of a range of demographic models using an identical sampling of viruses as for the A/H3N2 New York State data set. In each case, the red line describes the simulated model. The thin black lines present generalized skyline plots for five independent trees simulated under this model. The thick black line is a Bayesian Skyline estimated using BEAST. Each of the five simulated trees for each demographic model was analyzed and the runs then combined. The demographic models are as follows: (a) constant population size; N(t) = 10, (b) exponential growth; N(t) = 10*[exp{-0.25*t}], (c) logistic growth; N(t) = 10*[(1+c) / (1+{c*exp[0.65*t]})] where c=0.0025, (d) long period sinusoidal curve; N(t) = 5*[1 + {0.95*sin(t)}], (e) short period (annual) sinusoidal curve; N(t) = 5*[1 + {0.95*sin(2*pi*t)}]. There are a few points to note here. First, the earliest sampling point is early January 1993, and once the population has declined prior to this point, there is no information about any earlier peaks (this is especially clear in part (e)). Second, there are very few samples from the winters of 1996 and 2001 as these were A/H1N1 dominated seasons; this is reflected in a flat reconstruction and large HPD uncertainty for these seasons. Finally, it is clear from part (a) that the heterogeneous sampling in time of the virus does not in itself produce the fluctuating dynamics we observe for the real data.

Figure S8 (a) Simulation: Constant population size

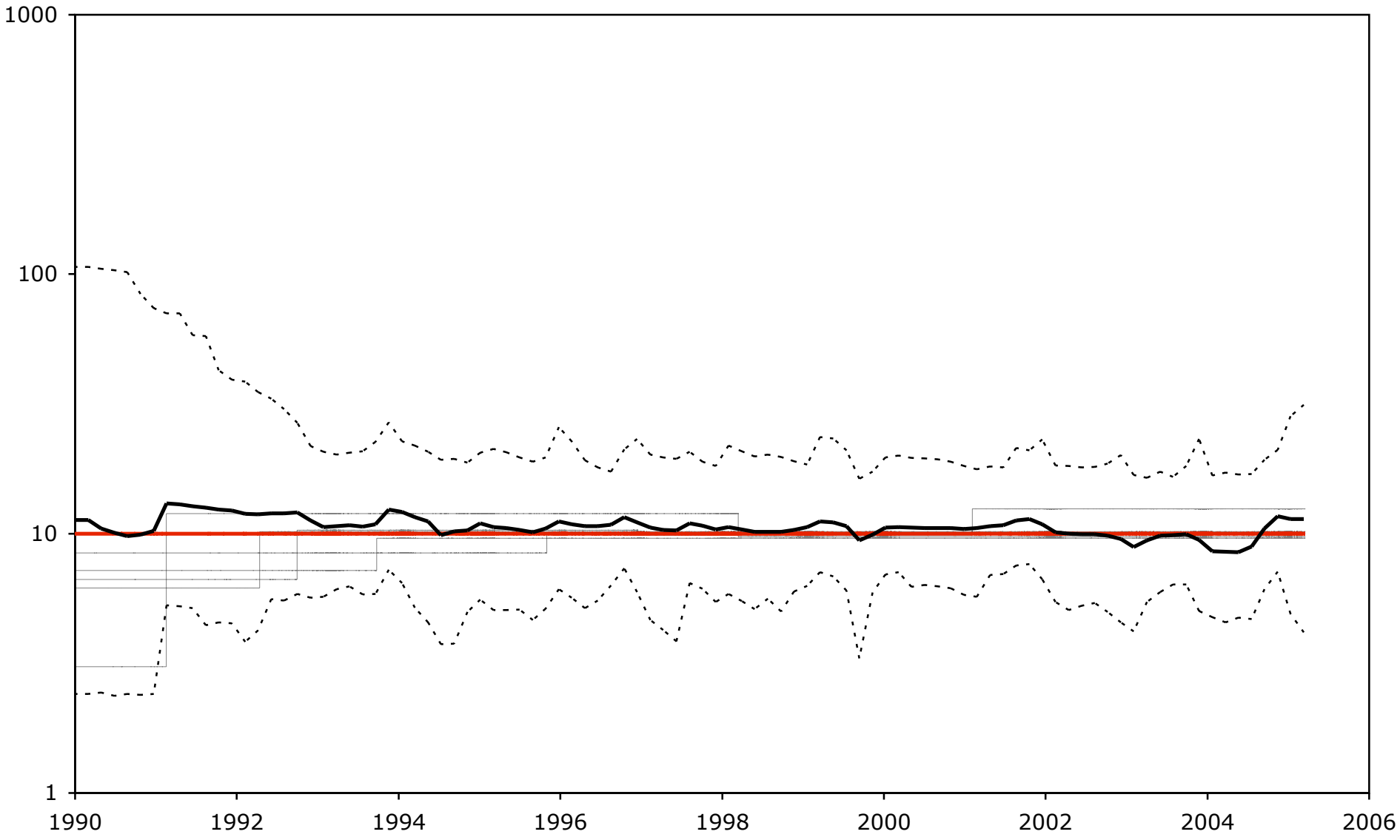

Figure S8 (b) Simulation: Exponential growth

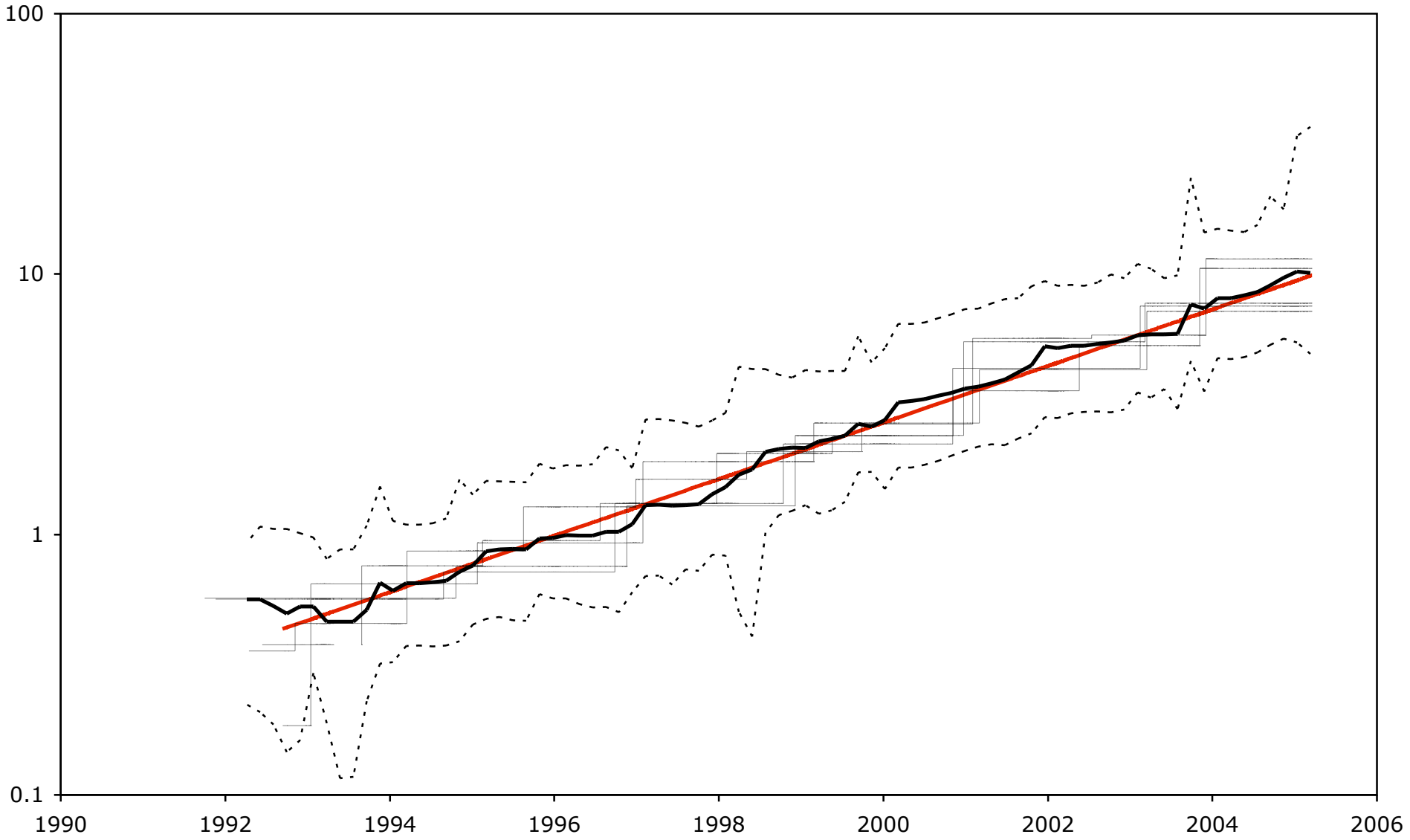

Figure S8 (c) Simulation: Logistic growth

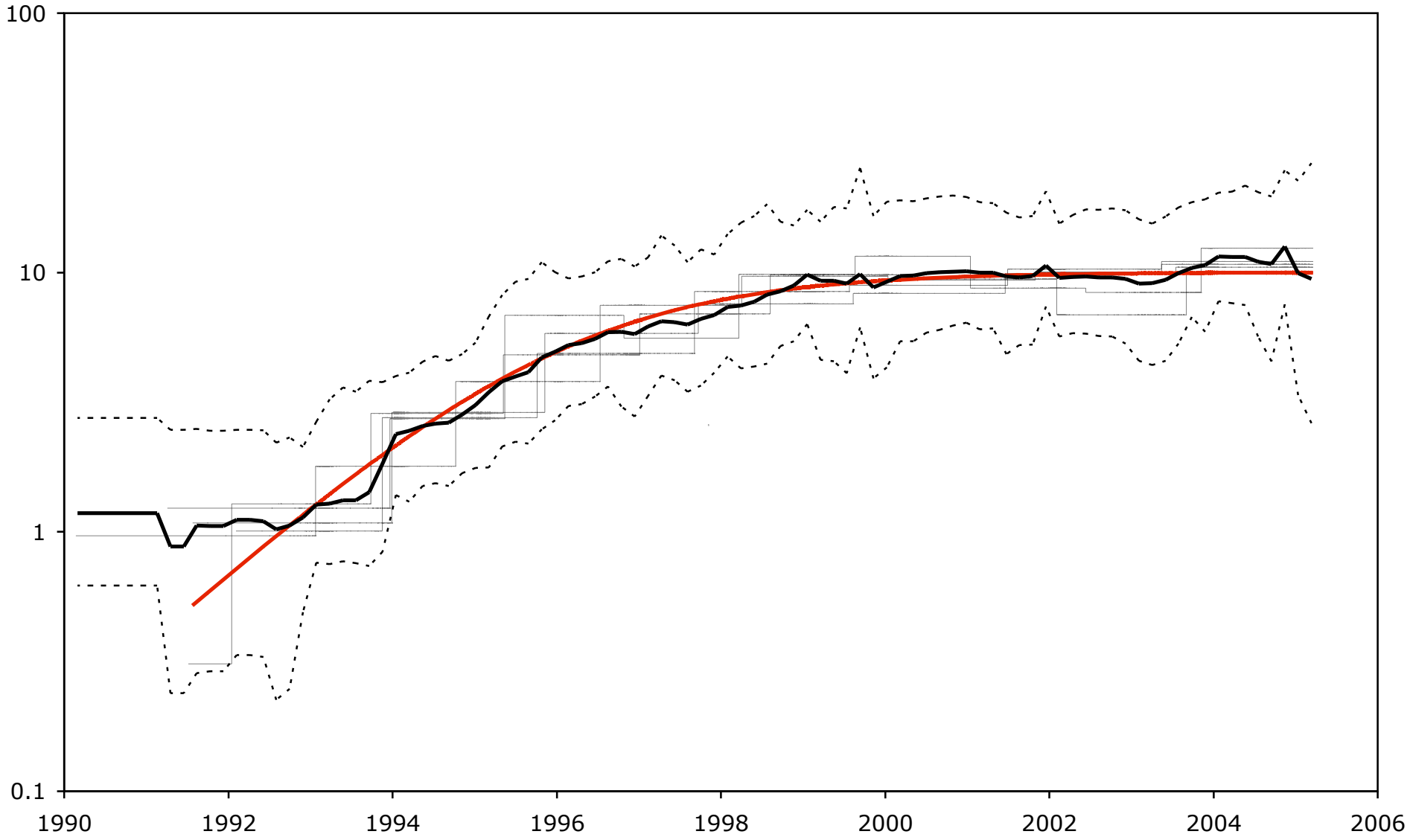

Figure S8 (d) Simulation: Long period sinusoidal

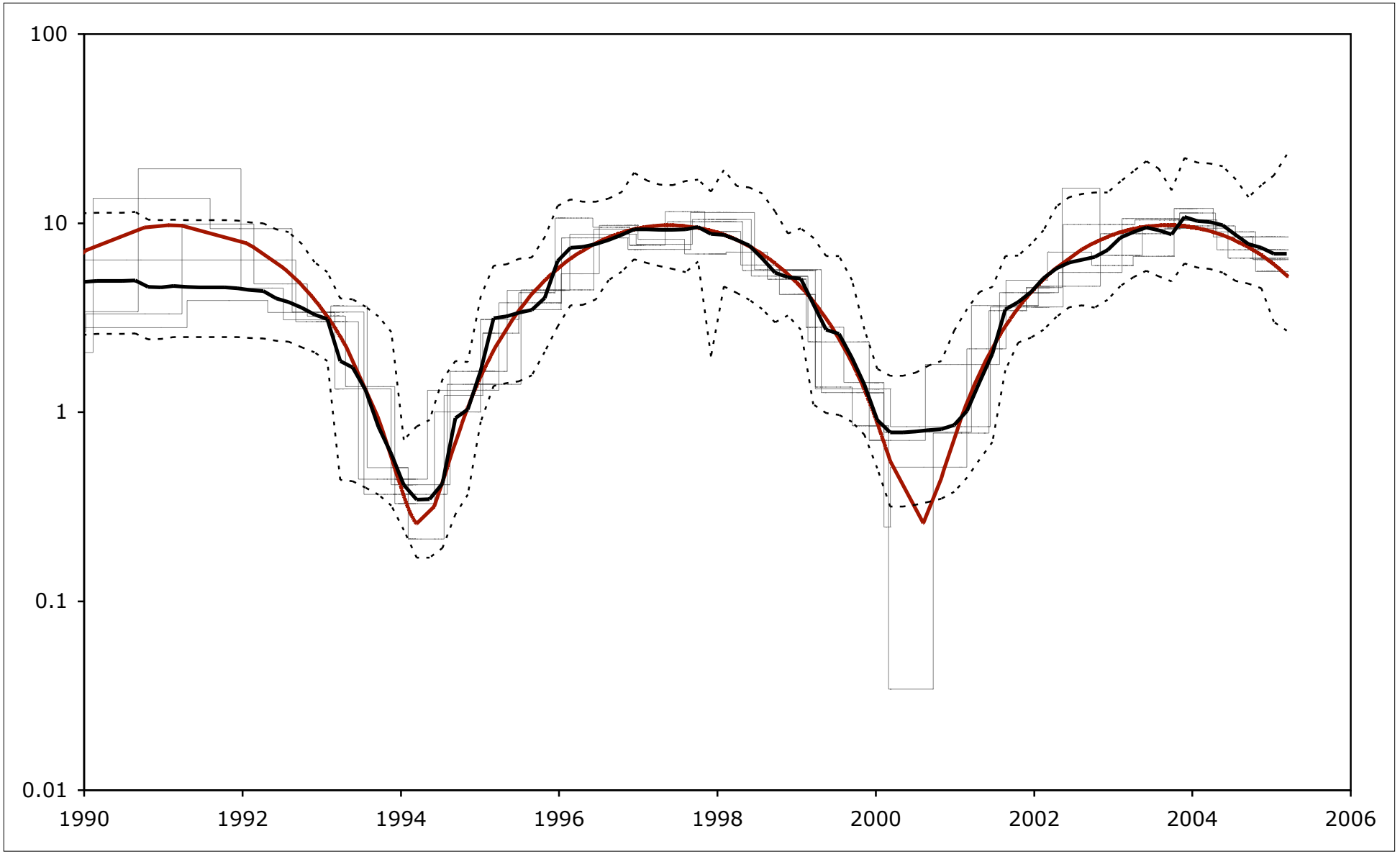

Figure S8 (e) Simulation: Short period sinusoidal

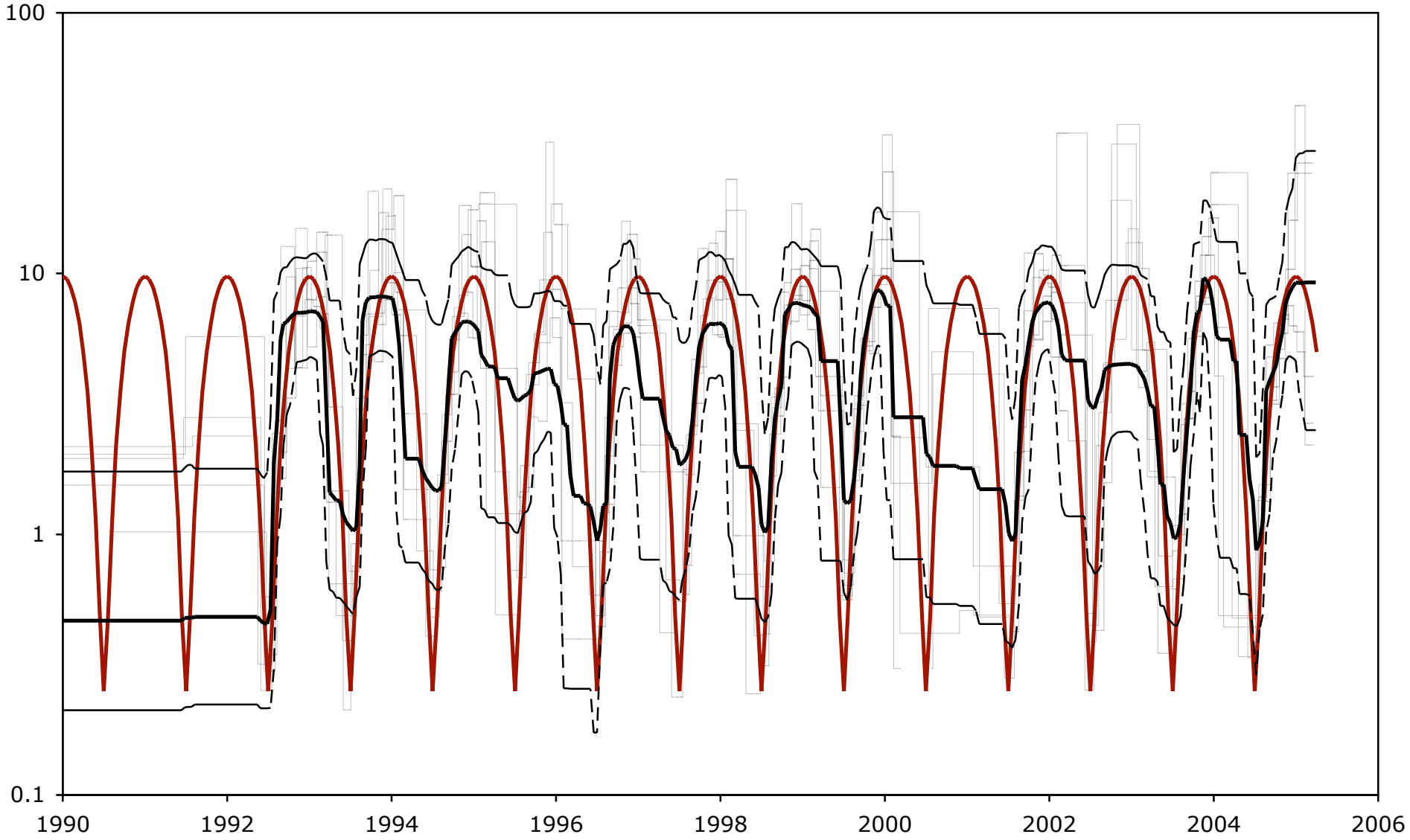

Fig. S9. (a) Rates of nucleotide substitution of the eight genome segments of H3N2 human influenza A virus sampled from New York State, USA. (b) Evolutionary rates of the HA and NA segments of H3N2 and H1N1 influenza A virus sampled from New York State, USA and New Zealand. Evolutionary rates are measured as the total number of nucleotide substitutions per site, per year x 10^{-3} (x-axis) and the relative substitution rate at 1^{st} and 2^{nd} positions compared to that at 3^{rd} codon positions (y-axis).

Figure S9 (a) New York H3N2; all segments

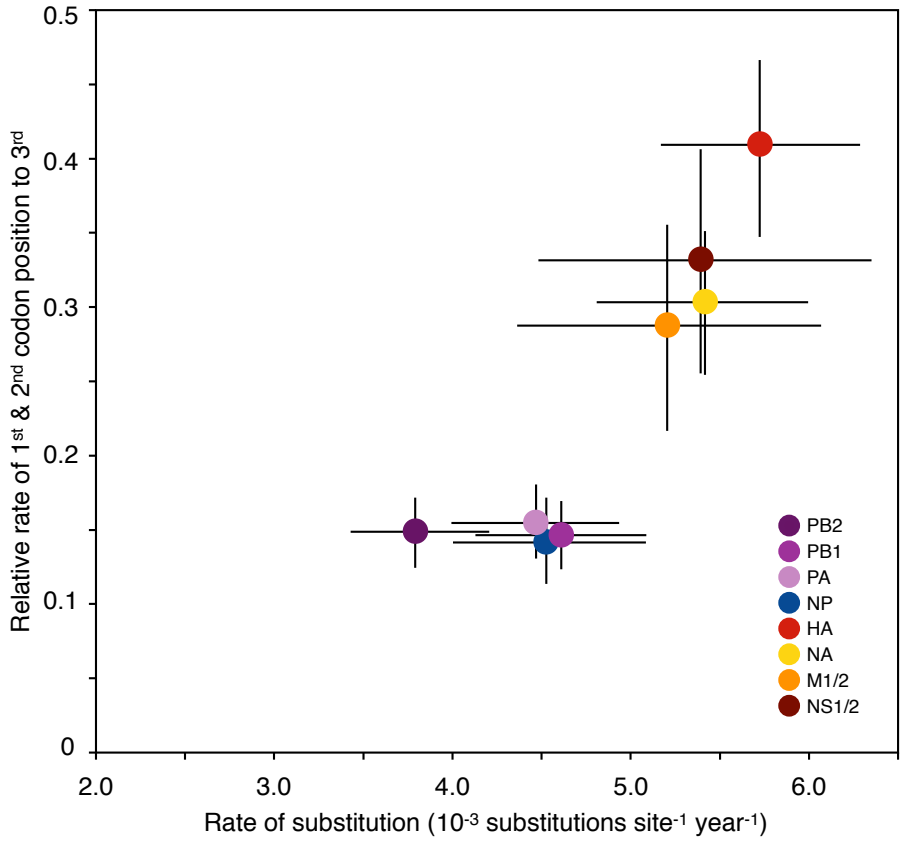

Figure S9 (b) H3N2, H1N1; HA and NA segments

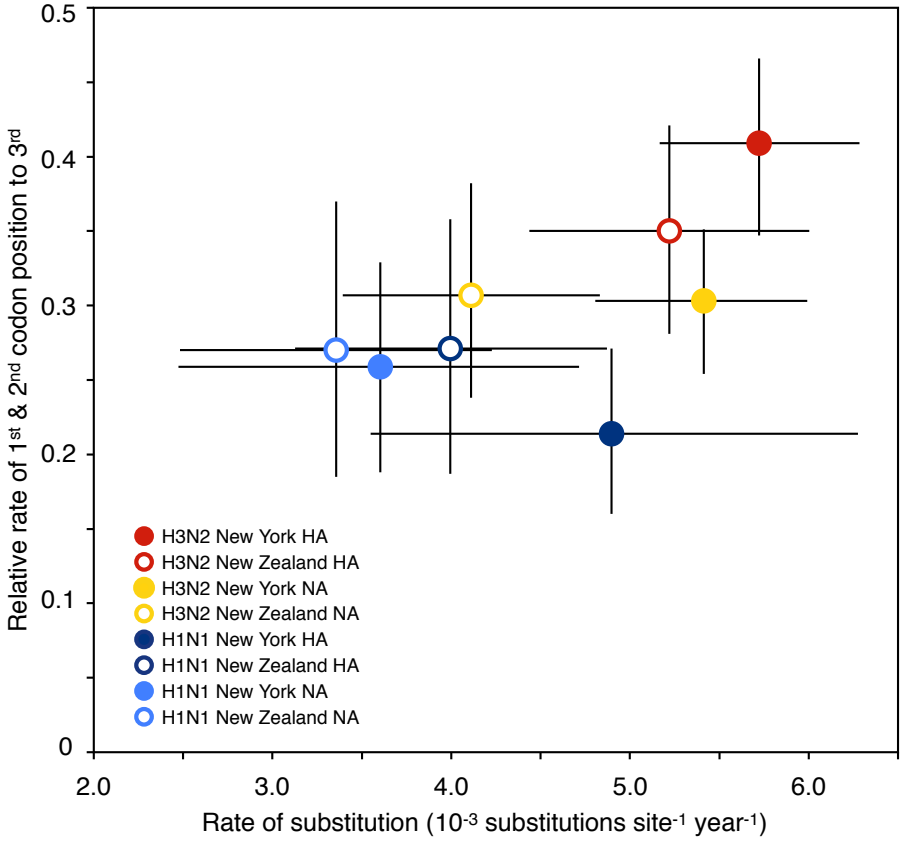

Fig. S10. A comparison of the Bayesian skyline plots for the 407 New Zealand A/H3N2 sequences (blue line and solid area; mean and 95% HPD interval) to those reconstructed with the addition of a smaller data set from Australia (123 sequences) sampled over the same seasons (red line and solid area; mean and 95% HPD interval).

Figure S10 (a) New Zealand & Australia, H3N2; HA segment

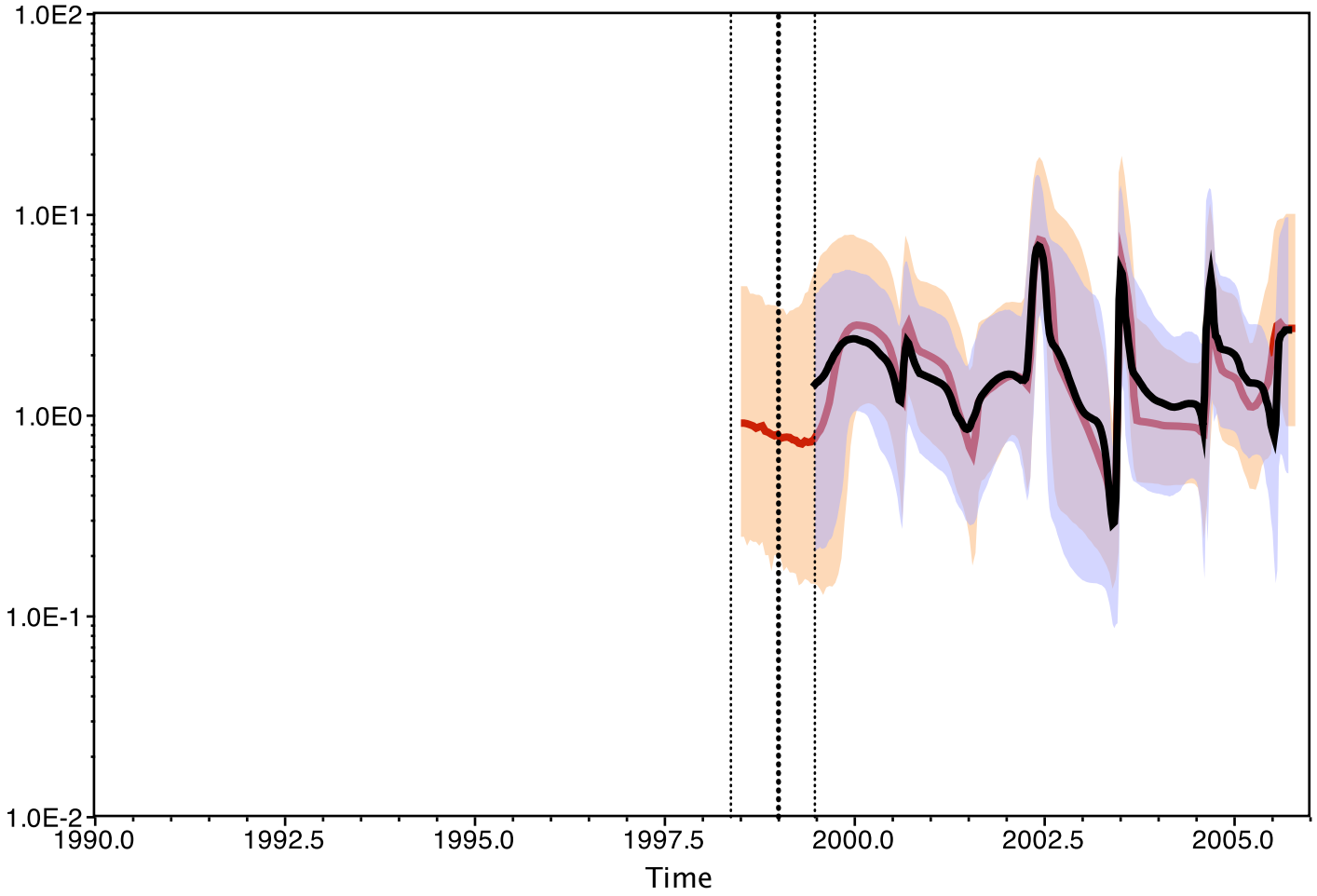

Figure S10 (b) New Zealand & Australia, H3N2; NA segment

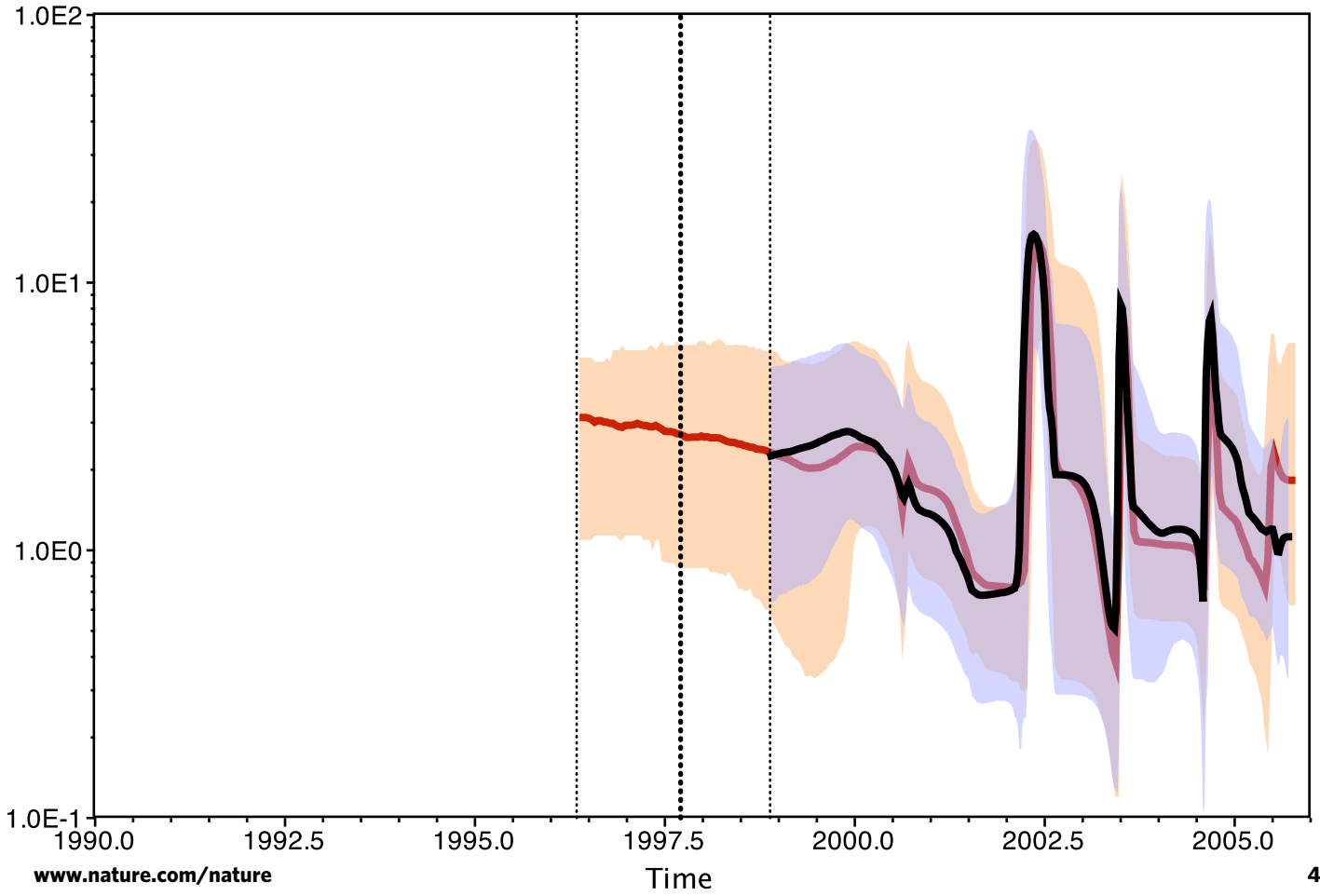

Fig. S11. Phylogenetic analysis of the HA segment of A/H3N2 (114 sequences, 1701 nt) and A/H1N1 (189 sequences, 1698 nt) viruses sampled from the 2006-2007 season in the northern hemisphere (with identical sequences removed). The trees were inferred using the maximum likelihood procedure available in the PAUP* package[1] and utilizing the best-fit GTR+G_4+I model of nucleotide substitution as determined by MODELTEST[2]. Both trees were mid-point rooted and all horizontal branches are drawn to a scale of substitutions per site. Branches are colour-coded as follows: New York State, red; other North America, black; Europe, blue. Both phylogenies show the clear mixing of isolates across the northern hemisphere.

[1]Swofford, D. L. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates. Sunderland, Mass (2003).

[2]Posada, D. & Crandall K. A., Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818 (1998).

Figure S11 (a) 2006-2007 season, H3N2; HA segment

Figure S11 (b) 2006-2007 season, H1N1; HA segment

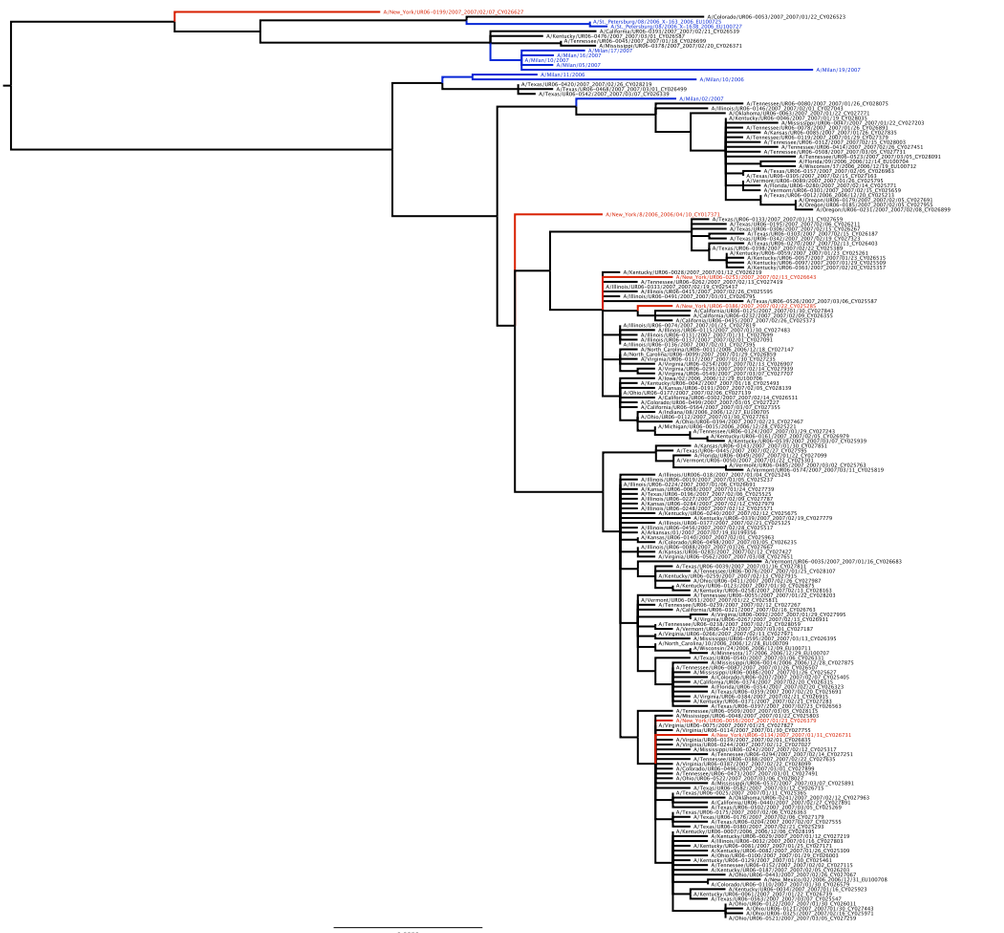

Fig. S12. The posterior density of the correlation coefficient of the TMRCAs for each season calculated between pairs of trees sampled from the MCMC. The right hand most curve in each case gives the distribution of this measure between pairs of trees sampled for the given genomic segment. The other curves in each panel give the distributions of the measure between the given segment and the others (given by the colour key). These distances are same as those summarized in the main text Figure 4 but taking each genomic segment independently will not be prone to any distortion of the spread by the dimension reduction technique used in Figure 4.

Figure S12 (a) New York H3N2 HA segment

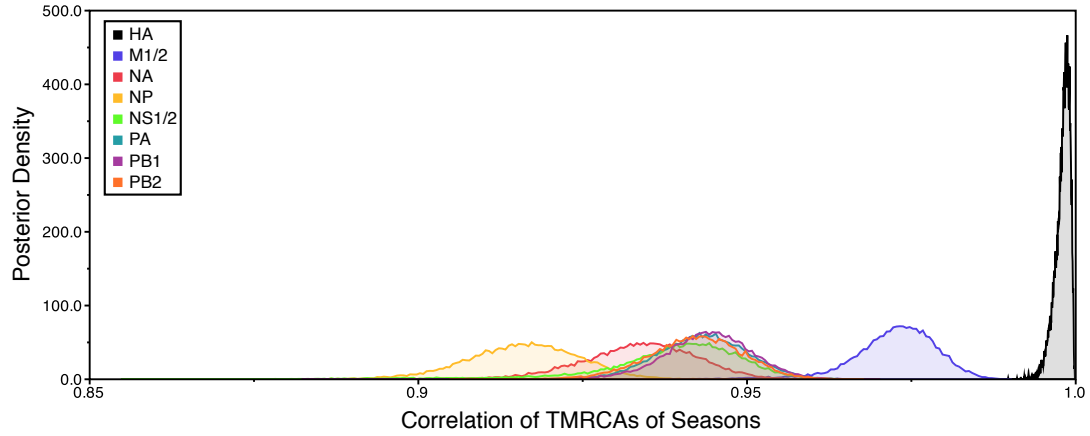

Figure S12 (b) New York H3N2 M1/2 segment

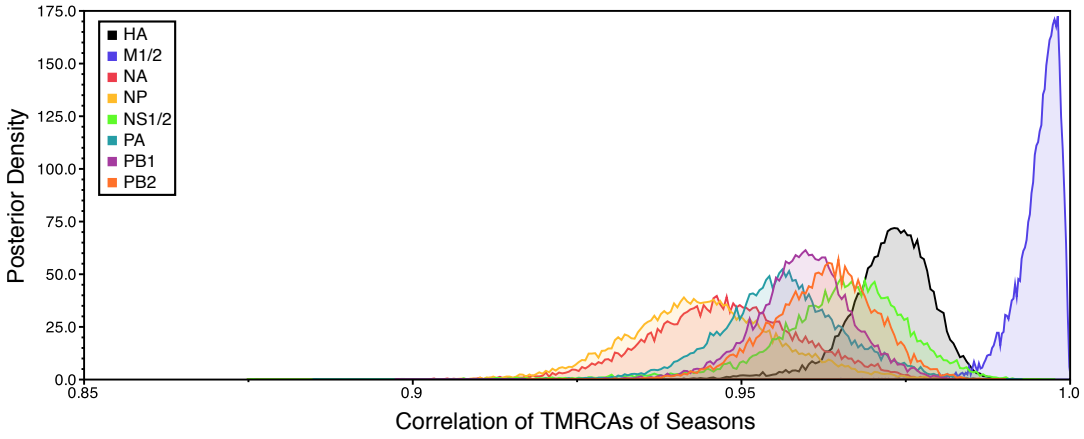

Figure S12 (c) New York H3N2 NA segment

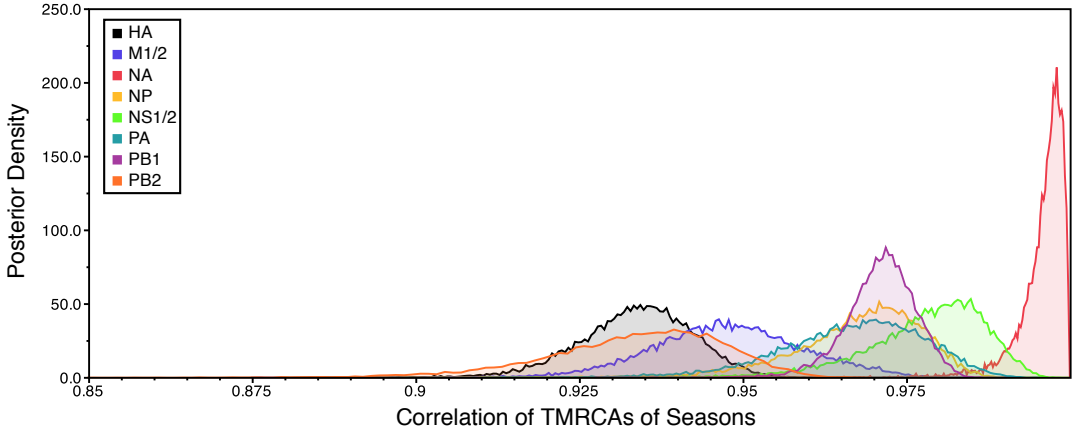

Figure S12 (d) New York H3N2 NP segment

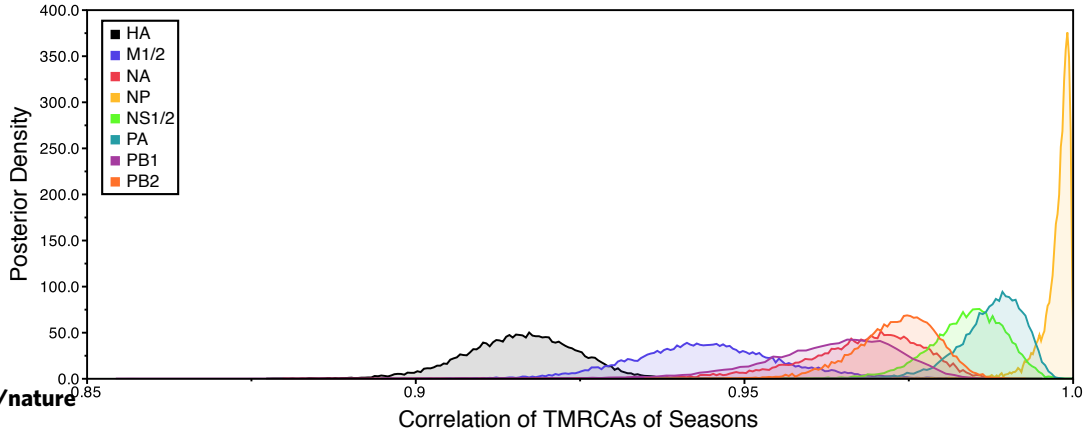

Figure S12 (e) New York H3N2 NS1/2 segment

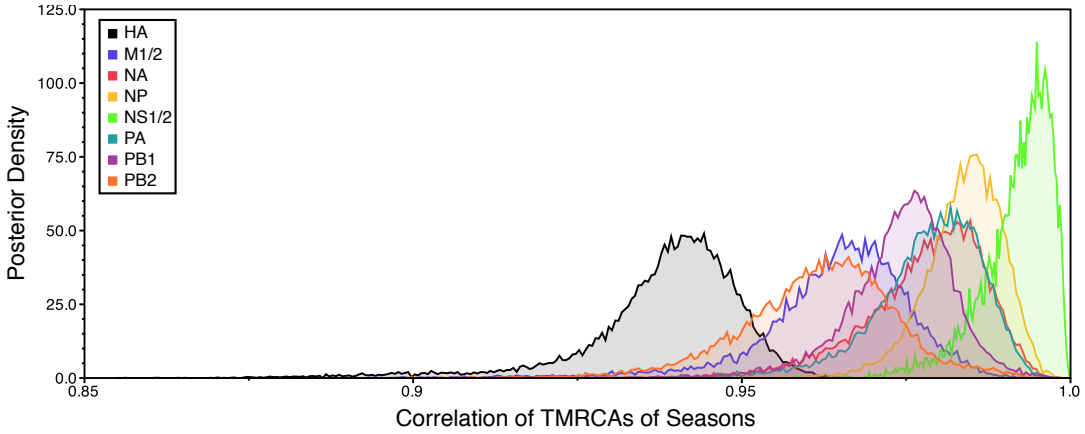

Figure S12 (f) New York H3N2 PA segment

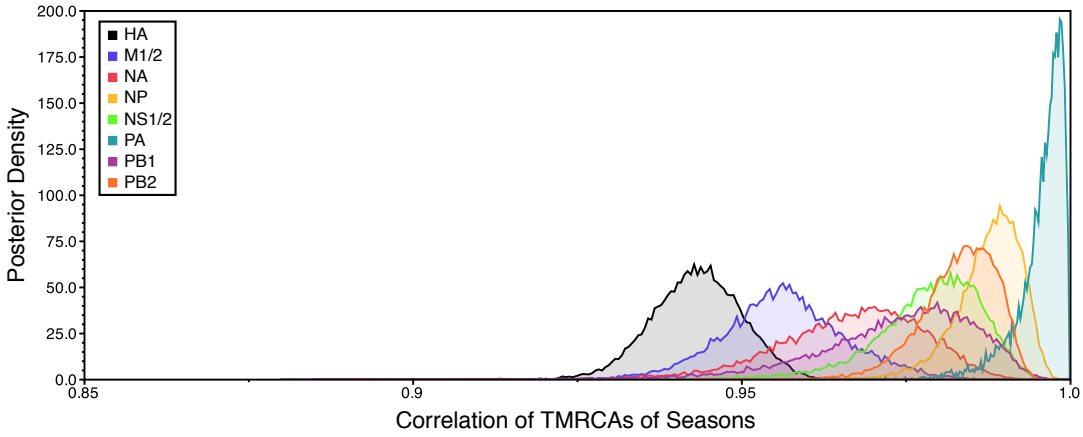

Figure S12 (g) New York H3N2 PB1 segment

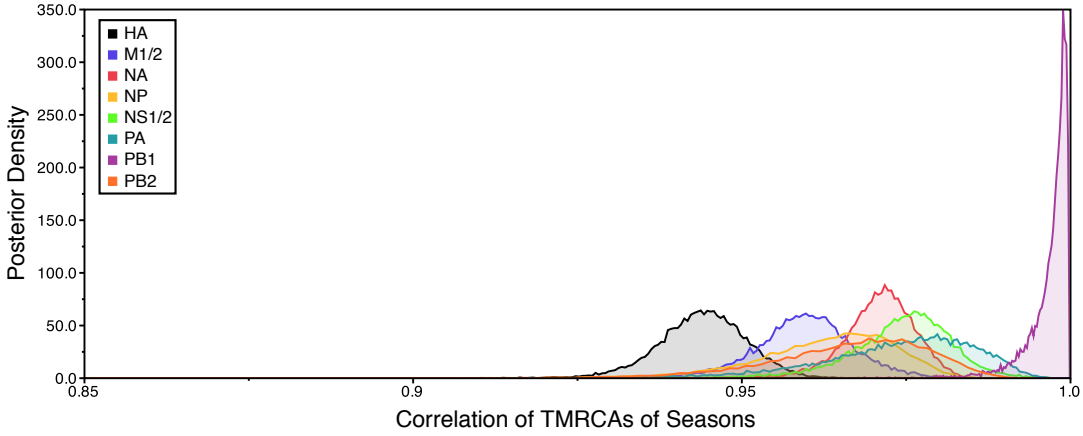

Figure S12 (h) New York H3N2 PB2 segment

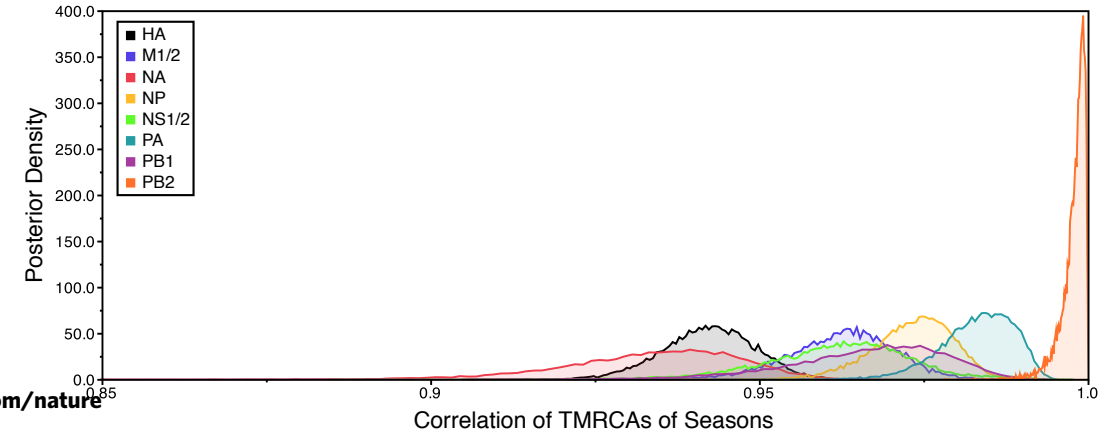

Table S1. Tables providing the posterior probabilities that, for each genomic segment (a – h), any specific season (rows) has a TMRCA that is older than that of each of the seasons that preceded it (columns). Probabilities that lie between 0.025 and 0.975 are highlighted in bold. In these cases, the TMRCAs of the genomic segment for the two respective seasons are not significantly different.

(a) PB2 genomic segment

	2003/ 2004	2002/ 2003	2001/ 2002	1999/ 2000	1998/ 1999	1997/ 1998	1996/ 1997	1995/ 1996	1994/ 1995	1993/ 1994	1992/ 1993
2004/2005	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003/2004		**0.50**	1.00	**0.49**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2002/2003			1.00	**0.50**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001/2002				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1999/2000					0.00	0.00	0.00	0.00	0.00	0.00	0.00
1998/1999						0.00	0.00	0.00	0.00	0.00	0.00
1997/1998							0.00	0.00	0.00	0.00	0.00
1996/1997								1.00	**0.50**	**0.49**	1.00
1995/1996									0.00	0.00	**0.17**
1994/1995										**0.50**	1.00
1993/1994											1.00

(b) PB1 genomic segment

	2003/ 2004	2002/ 2003	2001/ 2002	1999/ 2000	1998/ 1999	1997/ 1998	1996/ 1997	1995/ 1996	1994/ 1995	1993/ 1994	1992/ 1993
2004/2005	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003/2004		**0.50**	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2002/2003			1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001/2002				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1999/2000					0.00	0.00	0.00	0.00	0.00	0.00	0.00
1998/1999						0.00	0.00	0.00	0.00	0.00	0.00
1997/1998							0.00	0.00	0.00	0.00	0.00
1996/1997								**0.27**	**0.27**	**0.50**	0.00
1995/1996									**0.50**	**0.73**	0.00
1994/1995										**0.73**	0.00
1993/1994											0.00

(c) PA genomic segment

	2003/2004	2002/2003	2001/2002	1999/2000	1998/1999	1997/1998	1996/1997	1995/1996	1994/1995	1993/1994	1992/1993
2004/2005	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003/2004		**0.50**	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2002/2003			1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001/2002				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1999/2000					**0.40**	0.00	0.00	0.00	0.00	0.00	0.00
1998/1999						0.00	0.00	0.00	0.00	0.00	0.00
1997/1998							0.00	0.00	0.00	0.00	0.00
1996/1997								**0.50**	0.00	0.00	0.00
1995/1996									0.00	0.00	0.00
1994/1995										**0.50**	**0.91**
1993/1994											**0.91**

(d) HA genomic segment

	2003/2004	2002/2003	2001/2002	1999/2000	1998/1999	1997/1998	1996/1997	1995/1996	1994/1995	1993/1994	1992/1993
2004/2005	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003/2004		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2002/2003			1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001/2002				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1999/2000					0.00	0.00	0.00	0.00	0.00	0.00	0.00
1998/1999						0.00	0.00	0.00	0.00	0.00	0.00
1997/1998							0.00	0.00	0.00	0.00	0.00
1996/1997								**0.27**	0.00	0.00	0.00
1995/1996									0.00	0.00	0.00
1994/1995										0.00	0.00
1993/1994											**0.37**

(e) NP genomic segment

	2003/2004	2002/2003	2001/2002	1999/2000	1998/1999	1997/1998	1996/1997	1995/1996	1994/1995	1993/1994	1992/1993
2004/2005	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003/2004		**0.50**	1.00	**0.50**	**0.49**	0.00	0.00	0.00	0.00	0.00	0.00
2002/2003			1.00	**0.50**	**0.50**	0.00	0.00	0.00	0.00	0.00	0.00
2001/2002				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1999/2000					**0.50**	0.00	0.00	0.00	0.00	0.00	0.00
1998/1999						0.00	0.00	0.00	0.00	0.00	0.00
1997/1998							0.00	0.00	0.00	0.00	0.00
1996/1997								**0.50**	0.00	0.00	0.00
1995/1996									0.00	0.00	0.00
1994/1995										**0.50**	0.98
1993/1994											0.98

(f) NA genomic segment

	2003/2004	2002/2003	2001/2002	1999/2000	1998/1999	1997/1998	1996/1997	1995/1996	1994/1995	1993/1994	1992/1993
2004/2005	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003/2004		**0.50**	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2002/2003			1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001/2002				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1999/2000					**0.42**	**0.57**	0.00	0.00	0.00	0.00	0.00
1998/1999						**0.64**	0.00	0.00	0.00	0.00	0.00
1997/1998							0.00	0.00	0.00	0.00	0.00
1996/1997								0.00	0.00	0.00	0.00
1995/1996									0.00	0.00	0.00
1994/1995										**0.49**	0.00
1993/1994											0.00

(g) M1/2 genomic segment

	2003/2004	2002/2003	2001/2002	1999/2000	1998/1999	1997/1998	1996/1997	1995/1996	1994/1995	1993/1994	1992/1993
2004/2005	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003/2004		**0.50**	1.00	**0.03**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2002/2003			1.00	**0.03**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001/2002				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1999/2000					0.00	0.00	0.00	0.00	0.00	0.00	0.00
1998/1999						0.00	0.00	0.00	0.00	0.00	0.00
1997/1998							**0.39**	**0.39**	0.00	0.00	0.00
1996/1997								0.48	0.00	0.00	0.00
1995/1996									0.00	0.00	0.00
1994/1995										**0.50**	0.01
1993/1994											0.01

(h) NS genomic segment

	2003/2004	2002/2003	2001/2002	1999/2000	1998/1999	1997/1998	1996/1997	1995/1996	1994/1995	1993/1994	1992/1993
2004/2005	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003/2004		**0.51**	1.00	**0.04**	**0.04**	0.00	0.00	0.00	0.00	0.00	0.00
2002/2003			1.00	**0.04**	**0.03**	0.00	0.00	0.00	0.00	0.00	0.00
2001/2002				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1999/2000					**0.39**	0.00	0.00	0.00	0.00	0.00	0.00
1998/1999						0.00	0.00	0.00	0.00	0.00	0.00
1997/1998							0.00	0.00	0.00	0.00	0.00
1996/1997								**0.06**	0.02	0.00	0.00
1995/1996									**0.14**	0.02	0.01
1994/1995										**0.06**	0.02
1993/1994											**0.28**