RICE UNIVERSITY

Diversifying Selective Pressure on Influenza B Virus Hemagglutinin

By

Jun Shen

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Master of Science

APPROVED, THESIS COMMITTEE:

Jianpeng Ma, Professor, Chair
Department of Bioengineering, RICE
Biochemistry & Molecular Biology, BCM

Bruce R. Johnson, Distinguished Faculty Fellow
Department of Chemistry, RICE
Executive Director, Rice Quantum Institute

Oleg Igoshin, Assistant Professor
Department of Bioengineering
Rice Quantum Institute, RICE

HOUSTON, TEXAS
March 2009

ABSTRACT

Diversifying Selective Pressure on Influenza B Virus Hemagglutinin

by

Jun Shen

Influenza B virus hemagglutinin (HA) is a major surface glycoprotein with frequent amino-acid substitutions. However, the roles of antibody selection in the amino-acid substitutions of HA were still poorly understood. In order to gain insights into this important issue, an analysis was conducted on a total of 271 HA_1 sequences of influenza B virus strains isolated during 1940~2007. In this analysis, PAML (Phylogenetic Analysis by Maximum Likelihood) package was used to detect the existence of positive selection and to identify positively selected sites on HA_1. Strikingly, all the positively selected sites were located in the four major epitopes (120-loop, 150-loop, 160-loop and 190-helix) of HA identified in previous studies, thus supporting a predominant role of antibody selection in HA evolution. Of particular significance is the involvement of the 120-loop in positive selection, which may become increasingly important in future field isolates. Despite the absence of different subtypes, influenza B virus HA continued to evolve into new sublineages, within which the four major epitopes were targeted selectively in positive selection. Thus, any newly emerging strains need to be placed in the context of their evolutionary history in order to understand and predict their epidemic potential.

ACKNOWLEDGEMENTS

I would like to thank Dr. Ma's suggestion about the Master degree for application of H1b visa. An advanced degree will make the application of H1b much more smooth and affirmative.

I would like to thank my committee member Dr. Johnson and Dr. Igoshin, for providing me invaluable guidance and insight.

Thanks to Dr. Wang. All the calculation and analysis was based on her comprehensive and profound understanding of the structure and function of Influenza B. Thanks to Brian Kirk, Mingyang Lu, and other group members for their meaningful scientific discussion and friendship.

I would like to give my special thanks to my wife and my parents.

Contents

1 Introduction 1

2.1 Sequences preparation 5

2.2 Alignment of Sequence and tree drawing 6

2.4 Analysis of selective pressure and setting of PAML 8

3. Result 11

4 Discussion & Conclusion 29

Reference 34

CHAPTER 1

Introduction

Ever since the isolation of the first influenza B virus strain B/Lee/40 [Krystal et al., 1982], influenza B virus has remained a serious health problem, contributing to the seasonal "flu" epidemics each year. As a major glycoprotein on the surface of influenza B virus, HA undergoes constant amino-acid substitutions. The HA protein of current circulating influenza B virus strains belongs to one of the two major phylogenetic lineages: B/Victoria/2/87 (B/VI)-like and B/Yamagata /16/88 (B/YM)-like [Kanegae et al., 1990; Rota et al., 1990; Shaw et al., 2002].

Over the last 68 years, a large number of amino-acid substitutions on influenza B virus HA were observed in field isolates, in monoclonal-antibody escape mutants and in egg-adapted variants [Berton et al., 1984; Berton and Webster, 1985; Bootman and Robertson, 1988; Hovanec and Air, 1984; Krystal et al., 1982; Krystal et al., 1983; Lubeck et al., 1980; Rota et al., 1992; Rota et al., 1990; Verhoeyen et al., 1983; Webster and Berton, 1981]. However, it was unclear which of these substitutions the results of positive selection were, and what were the roles of antibody selection in the molecular evolution of influenza B virus HA. In this context, **positive selection** is defined as a significant excess of amino-acid altering substitutions over silent substitutions in nucleotide sequences, since, if completely random, only 24% of nucleotide substitutions would cause changes in the encoded amino acids [Air et al., 1990].

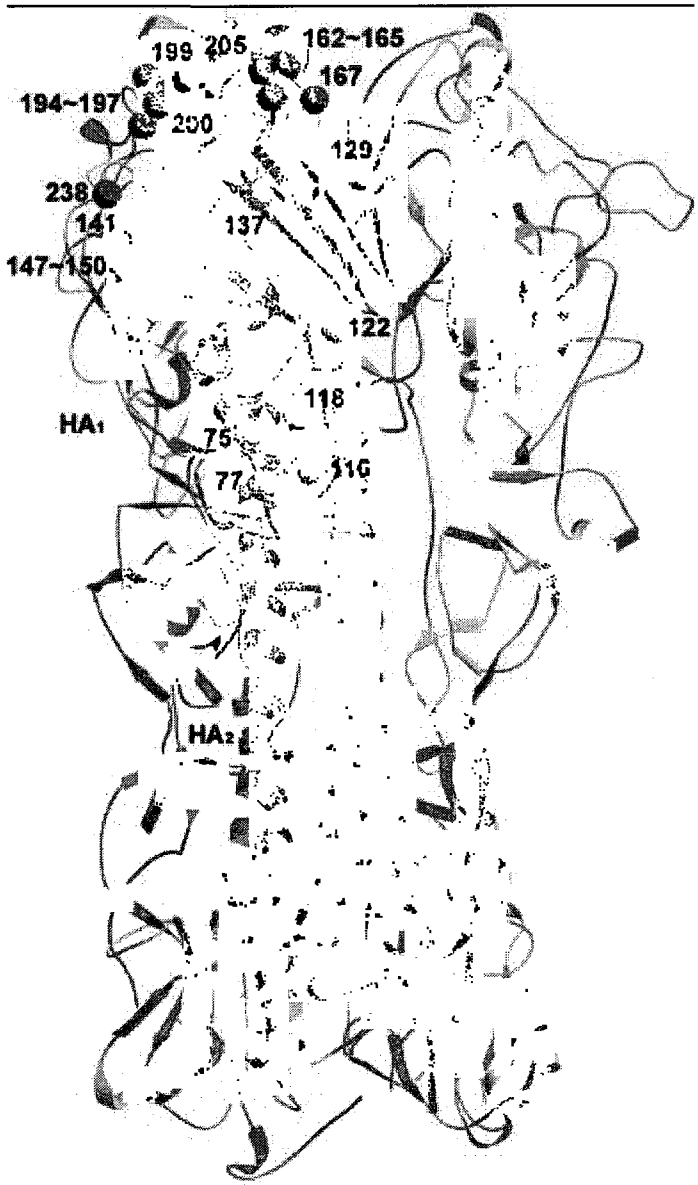

Fig 1.1 Major epitopes of influenza B virus HA. The trimeric HA is shown with one monomer highlighted in color: Pink for HA1 and yellow for HA2. Mutations in four regions, the 120-loop (cyan), 150-loop (green), 160-loop (blue), and 190-helix (red), have been found to cause antigenicity variation.

There was previously a sequence analysis on 49 HA_1 sequences of recent influenza B virus isolates, which identified HA_1 75, 197, and 199 (B/HongKong/8/73 HA numbering of 75, 194, and 196, respectively) to be under positive evolutionary selection

[Pechirra et al., 2005]. However, since it did not separate the B/YM-lineage and B/VI-lineage strains, this study might have failed to identify those amino-acid positions that were selected positively only in one but not the other lineage [Yang et al., 2000]. This was particularly problematic for the 150-loop and 160-loop (**Fig.1**), which had become specific for B/YM-like and B/VI-like strains, respectively [Nakagawa et al., 2001a; Nakagawa et al., 2001b; Nakagawa et al., 2003; Nakagawa et al., 2005; Wang et al., 2008]. Most recently, a larger-scale analysis that used 214 HA_1 sequences of influenza B virus strains has been published [Chen and Holmes, 2008]. Although it separated B/YM- and B/VI-lineage strains, the evolution of these lineages into distinct sublineages was not taken into account, which limited the accuracy of the positively selected sites derived therein [Chen and Holmes, 2008].

PAML is a package of programs that analyze DNA and protein sequences using maximum likelihood [Yang, 2007]. Using the program CODEML in PAML, the nonsynonymous (amino-acid altering substitutions)/synonymous (silent substitutions) rate ratio (ω) for each codon is calculated as an important indicator of selection pressure at the protein level: an $\omega > 1$ indicates positive selection [Yang, 2007; Yang et al., 2000]. Bayes Empirical Bayes analysis then calculates the posterior probability that each site belongs to a particular site class. Sites with high posterior probability of belonging to the site class of $\omega > 1$ are inferred to be under positive selection [Yang, 2007; Yang et al., 2000].

In order to gain insights into the amino-acid positions on influenza B virus HA that are truly under positive selective pressure, here a total of 271 HA_1 sequences of influenza B

virus strains isolated between 1940~2007 were analyzed. Based on the phylogenetic analysis, these HA_1 sequences were divided into three major groups: early strains (1940~1970), B/YM-like lineage (1972~2005) and B/VI-like lineage (1975~2007). The B/YM-lineage was further divided into four sublineages, and the B/VI lineage into two sublineages (**Fig.2**). These seven groups were analyzed by using CODEML in PAML version 4 [Yang, 2007; Yang et al., 2000]. The identified positively selected sites were located predominantly on the four major antigenic epitopes on HA_1: the 120-loop (HA_1 116~137), the 150-loop (HA_1 141~150), the 160-loop (HA_1 162~167), the 190-helix (HA_1 194~202), and their respective surrounding regions [Wang et al., 2008] (**Fig. 3**), suggesting the important roles of antibody selection in molecular evolution of influenza B virus HA.

CHAPTER 2

Materials and Methods

Phylogenetic analysis

2.1 Sequences preparation

This study focused on the first 340 amino acid residues of mature HA_1 (1~1020 nucleotides excluding those corresponding to the signal peptide). A total of 271 HA_1 sequences of influenza B virus strains isolated between 1940~2007 were used in the study. These sequences were selected to sample all the years in which influenza B viruses were active, and special cares were taken to avoid stains with high similarity and isolated in the same regions. All the sequences were obtained from the Influenza Sequence Database (Los Alamos National Laboratory, Los Alamos, NM, USA www.flu.lanl.gov) [Macken et al., 2001]. The name of the strain was composed of two part, strain type plus serial number and year, for example:

NewYork 2 – 90

NewYork: The location where the virus was collected.

2: Serial number, there may exit another strain NewYork3-90.

90: The virus strain was at 1990.

Fig 2.1 Meaning of the strain name.

All the adopted sequences are saved as fasta format, which are compatible with DNA star for further analysis.

2.2 Alignment of Sequence and tree drawing

The CLUSTAL W method [Thompson et al., 1994] with the MEGALIGN program of DNASTAR package (www.dnastar.com) was employed for sequencing alignment and phylogenetic analysis. Among all the available methods of sequence alignment, clustal W method provide the best sensitivity for the alignment of divergent protein sequences. Although most of the sequences used in this work share high similarity, the sequence shift in alignment is foremost step for calculation of substitution rate ratio. For Clustal W method, **PAM250** residue weight table was adopted.

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
C	12																				C
S	0	2																			S
T	-2	1	3																		T
P	-3	1	0	6																	P
A	-2	1	1	1	2																A
G	-3	1	0	-1	1	5															G
N	-4	1	0	-1	0	0	2														N
D	-5	0	0	-1	0	1	2	4													D
E	-5	0	0	-1	0	0	1	3	4												E
Q	-5	-1	-1	0	0	-1	1	2	2	4											Q
H	-3	-1	-1	0	-1	-2	2	1	1	3	6										H
R	-4	0	-1	0	-2	-3	0	-1	-1	1	2	6									R
K	-5	0	0	-1	-1	-2	1	0	0	1	0	3	5								K
M	-5	-2	-1	-2	-1	-3	-2	-3	-2	-1	-2	0	0	6							M
I	-2	-1	0	-2	-1	-3	-2	-2	-2	-2	-2	-2	-2	2	5						I
L	-6	-3	-2	-3	-2	-4	-3	-4	-3	-2	-2	-3	-3	4	2	6					L
V	-2	-1	0	-1	0	-1	-2	-2	-2	-2	-2	-2	-2	2	4	2	4				V
F	-4	-3	-3	-5	-4	-5	-4	-6	-5	-5	-2	-4	-5	0	1	2	-1	9			F
Y	0	-3	-3	-5	-3	-5	-2	-4	-4	-4	0	-4	-4	-2	-1	-1	-2	7	10		Y
W	-8	-2	-5	-6	-6	-7	-4	-7	-7	-5	-3	2	-3	-4	-5	-2	-6	0	0	17	W
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	

Table 2.1 Residue Weights of Clustal W method.

The alignment result of the early strains, one of the seven sublineages we studied, is listed as fig 2.2:

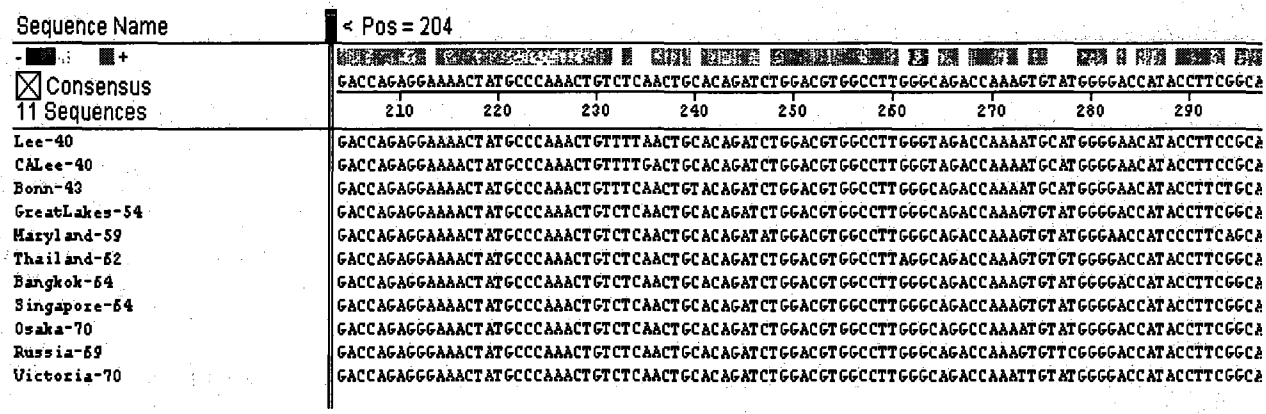

Fig. 2.2 Alignment of early strain.

In these position marked by red, all the strains possess same sequence, no variance. Our calculation focused on these non red areas, which are positions under evolution pressure, either synonymous (silent) or nonsynonymous(amino acid altering) substitution.

All the evolution tree pictures in this paper was also made by DNASTAR,

2.3 Analysis of selective pressure and setting of PAML

PAML is open source software for phylogenetic analysis of DNA or protein sequence. It analyses all the sequence through comparison of nested statistical models.

For this analysis, the CODEML program in PAML was used to calculate the codon-substitution models for heterogeneous selection pressure at amino-acid positions [Yang, 1997; Yang, 2007; Yang et al., 2000; Yang et al., 2005]. The models used in this study were M0, M1a, M2a, M7 and M8.

Model	NSsites	#p	Parameters
M0 (one ratio)	0	1	ω
M1a (neutral)	1	2	p_0 ($p_1 = 1 - p_0$), $\omega_0 < 1$, $\omega_1 = 1$
M2a (selection)	2	4	p_0, p_1 ($p_2 = 1 - p_0 - p_1$), $\omega_0 < 1$, $\omega_1 = 1$, $\omega_2 > 1$
M3 (discrete)	3	5	p_0, p_1 ($p_2 = 1 - p_0 - p_1$) $\omega_0, \omega_1, \omega_2$
M7 (beta)	7	2	p, q
M8 (beta&ω)	8	4	p_0 ($p_1 = 1 - p_0$), $p, q, \omega_s > 1$

Table 2.2. Parameters in the site models, #p is the number of free parameters in the ω distribution.

M1a (nearly neutral) and M7 (beta) were null models that did not support $\omega > 1$. In contrast, the alternative models M2a (positive selection) and M8 (beta and ω), compared to M1a and M7 respectively, each had an additional class that allowed $\omega > 1$. Likelihood ratio tests (LRT) comparing M2a versus M1a and M8 versus M7 provided test for the existence of positive selection. In LRT, twice the log likelihood difference, $2\Delta l = 2(l_1 - l_0)$, was compared with a χ^2 distribution to test whether the null model was to be rejected, where ℓ_0 and ℓ_1 were the log likelihood for the alternative

model and the null model, respectively. In addition, empirical Bayes analysis was employed to calculate the posterior probability that each site belonged to a particular site class. Sites with high posterior probability of belonging to the site class of $\omega > 1$ were inferred to be under positive selection. It was shown that Bayes Empirical Bayes, which assigned a prior to the model parameters [Deely and Lindley, 1981], worked well for both small and large datasets [Yang et al., 2005]. Since some of the subgroups used in this study were small, the results from Bayes Empirical Bayes analysis were used throughout this study. To account for the insertions and deletions in influenza B virus HA_1, the numbering of influenza B/HongKong/8/73 HA was used as a reference for all sequences [Wang et al., 2008].

The key control file for CODEML program, Codeml.ctl contains all important parameter for the calculation,

seqfile = FluB.NUC * *sequence data file name*

treefile = Flub..DND * *tree structure file name*

outfile = mlc * *main result file name*

noisy = 3 * *0,1,2,3,9: how much rubbish on the screen*

verbose = 0 * *1: detailed output, 0: concise output*

runmode = 0 * *0: user tree; 1: semi-automatic; 2: automatic*

* *3: StepwiseAddition; (4,5):PerturbationNNI; -2: pairwise*

CodonFreq = 2 * *0:1/61 each, 1:F1X4, 2:F3X4, 3:codon table*

clock = 0 * *0: no clock, unrooted tree, 1: clock, rooted tree*

aaDist = 0 * *0:equal, +:geometric; -:linear, {1-5:G1974,Miyata,c,p,v}*

model = 0

NSsites = 0 8 * *Mode0 amd Model 8 will be caculated in this run*

* *0:one w; 1:NearlyNeutral; 2:PositiveSelection; 3:discrete;*

* *4:freqs;5:gamma;6:2gamma;7:beta;*

* *8:beta&w;9:betaγ10:3normal*

icode = 0 * *0:standard genetic code; 1:mammalian mt; 2-10:see below*

Mgene = 0 * *0:rates, 1:separate; 2:pi, 3:kappa, 4:all*

fix_kappa = 0 * *1: kappa fixed, 0: kappa to be estimated*

kappa = .3 * *initial or fixed kappa*

fix_omega = 0 * *1: omega or omega_1 fixed, 0: estimate*

omega = 1.3 * *initial or fixed omega, for codons or codon-based AAs*

fix_alpha = 0 * 0: estimate gamma shape parameter; 1: fix it at alpha

alpha = 0.5 * initial or fixed alpha, 0:infinity (constant)

ncatG = 10 * *# of categories in the dG or AdG models of rates*

getSE = 0 * *0: don't want them, 1: want S.E.s of estimates*

RateAncestor = 0 * *(0,1,2): rates (alpha>0) or ancestral states (1 or 2)*

Small_Diff = .45e-6

cleandata = 1 * *remove sites with ambiguity data (1:yes, 0:no)*

fix_blength = 0 * *0: ignore, -1: random, 1: initial, 2: fixed*

Several pivotal input of CODEML worth to mention and discuss here:

Model: set equal to 0, that assume all the branches possess the same ratio ω.

Alpha: refers to the parameter α in gamma distribution for variable substitution rates across sites.

Clock: Without the clock (clock = 0), unrooted trees should be used, such as ((1,2),3,4) or (1,2,(3,4)).

Chapter 3

Results

Phylogenetic relationship of influenza B virus HA

According to phylogenetic analysis, the 271 HA_1 sequences were divided into three groups: early strains isolated between 1940~1970 (I), B/YM-like lineage since 1972 (II) and B/VI-like lineage since 1975 (III). The B/YM-lineage (II) was divided further into four (II-*i* ~ II-*iv*) sublineages (**Fig. 3.1**), among which (II-*i* ~ II-*iii*) sublineages had been described in a previous study [Nerome et al., 1998], whilst the (II-*iv*) sublineage was described here for the first time. The B/VI-lineage (III) was divided further into an earlier sublineage (III-*i*) and a more recent sublineage (III-*ii*) (**Fig. 3.1**). This large-scale phylogenetic analysis uncovered that the divergence of influenza B virus HA into B/YM- and B/VI-lineages can be dated back to early 1970s, which is much earlier than previously thought [Kanegae et al., 1990; Matsuzaki et al., 2004; McCullers et al., 2004; Rota et al., 1990] and agrees well with a just-published study [Chen et al., 2007].

Positive selection on influenza B virus HA

To detect positively selected sites in HA_1 sequence of influenza B virus strains between 1940~2007, the analysis using CODEML in PAML was performed individually on the seven subgroups (I, II-*i* ~ II-*iv*, and III-*i* ~ *ii*) (**Fig. 3.1**). In all but two cases, the LRT statistics ($2\Delta l$) for M2a versus M1a and M8 versus M7 were much larger than the critical value of $\chi^2_{1\%} = 6.63$ with degree of freedom (d.f.) set to 1 (**Table 3.1, 3.2**). Thus the LRT tests supported the existence of positive selection on influenza B virus HA. The sites with greater than 50% posterior probability to be under positive selective

pressure in models M2a and M8, obtained from Bayes Empirical Bayes analysis [Yang et al., 2005], were listed in **Table 3.3**. In general, M2a identified fewer sites under positive selection than M8 did. Nevertheless, the sites identified in M2a were those of the highest posterior probability in M8 (**Table 3.3**). In contrast, those identified only in M8 but not in M2a were generally of low posterior probability. To be more conservative, most of our discussion was focused on the sites that were identified in M8 model with greater than 95% posterior probability to be under positive selection. This cutoff limits the false-positive rate to 5~6% or lower [Yang et al., 2005]. It is important to emphasize that those of high posterior probability to be under positive selection were not necessarily those of the highest mutation rates. Different from influenza A virus HA [Bush et al., 1999; Yang et al., 2000], a much smaller number of sites on influenza B virus HA were subject to positive selection for antigenic drift, consistent with earlier studies [Air et al., 1990; Chen and Holmes, 2008].

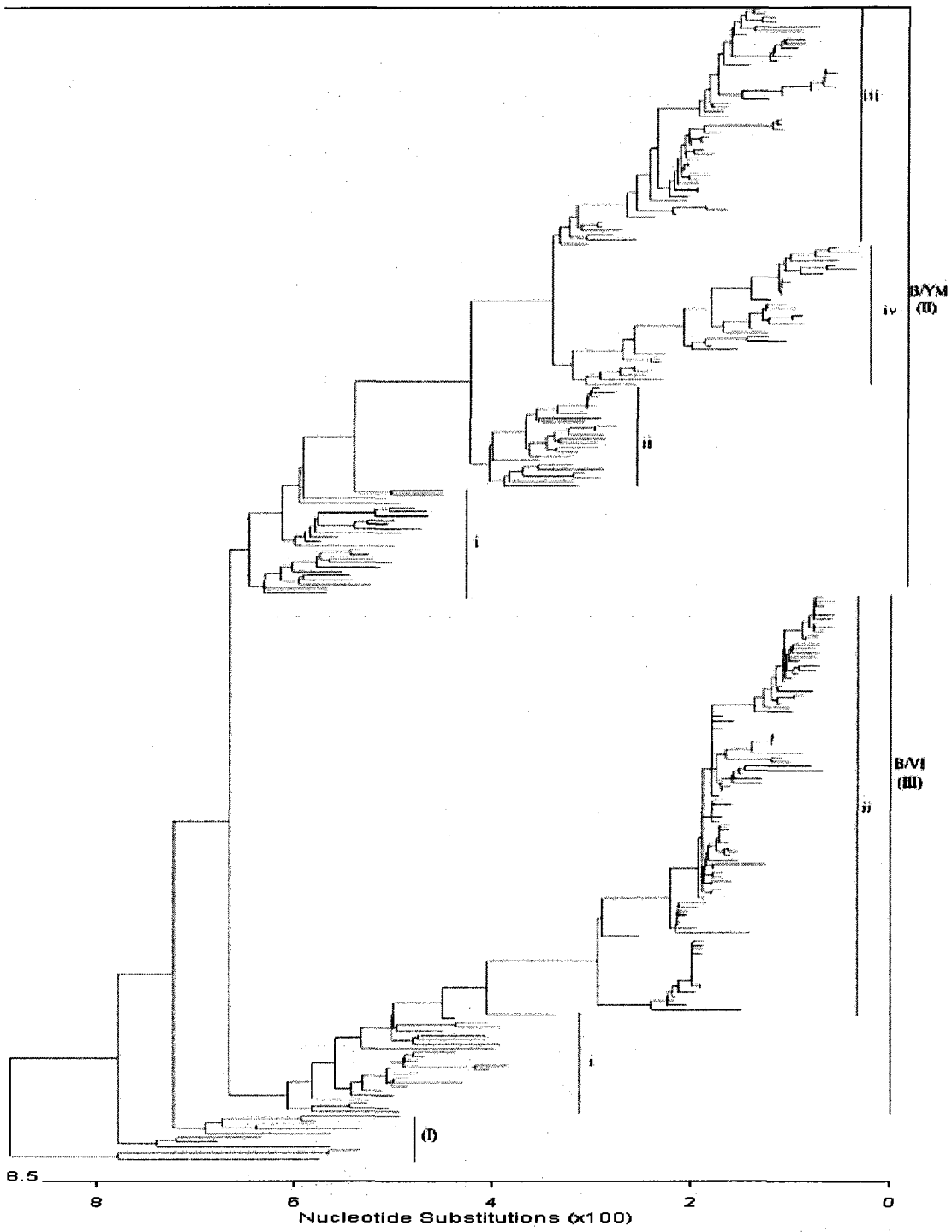

Fig. 3.1. Phylogenetic relationship of 271 HA1 sequences used in this study. For all sequences, the nucleotide sequences between 1 and 1,020, corresponding to residues HA1 1–340, were used. The phylogenetic tree was drawn using the program Megalign from DNASTAR package (www.dnastar.com).

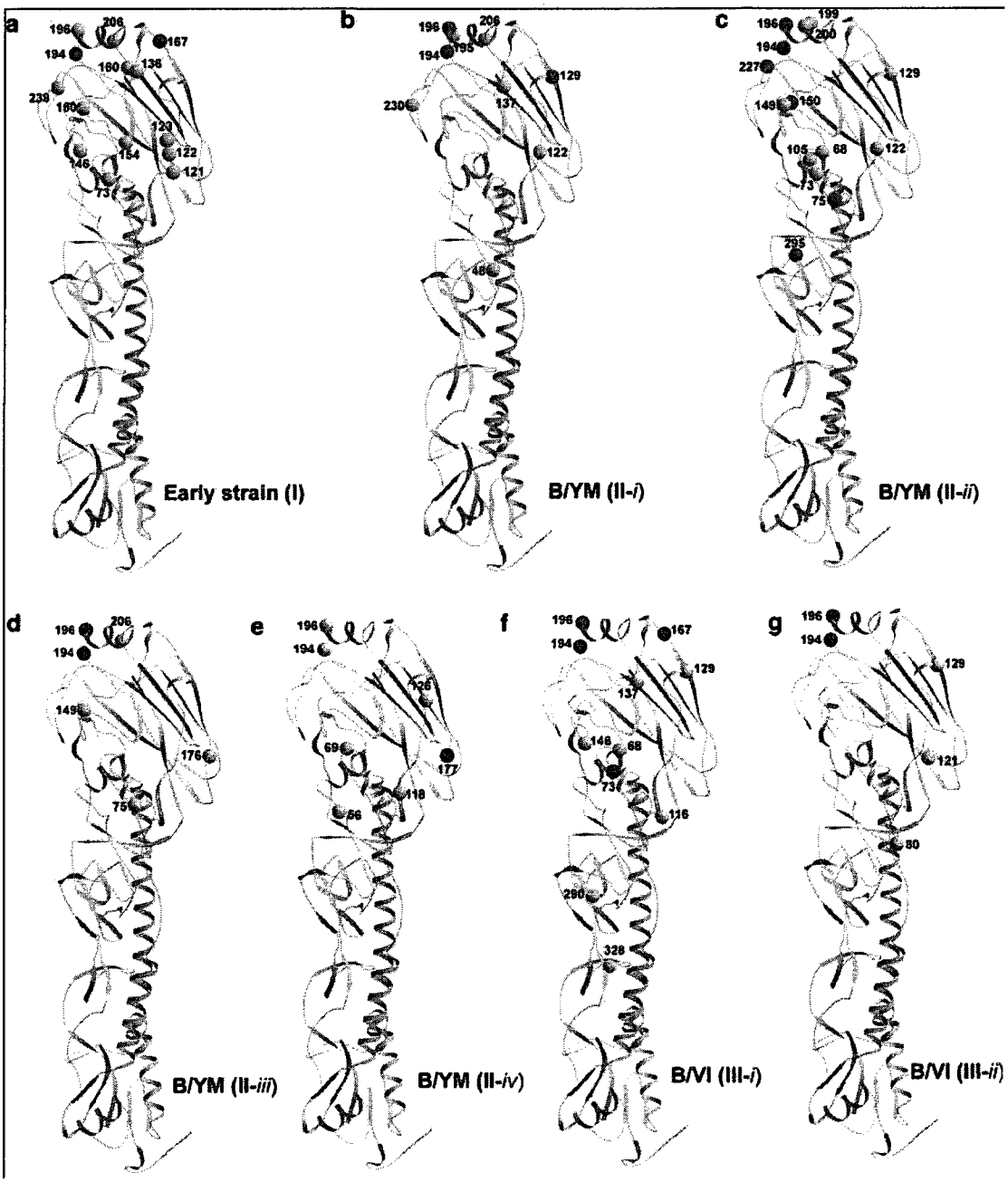

Fig 3.2 Sites with posterior probabilities of greater than50% to be under positive selection in the M8 models for the seven subgroups of influenza B virus HA, in the order of early strain (I) (a), B/YM-lineage (II-i–II-iv) (b–e) and B/VI-lineage (III-i and III-ii) (f, g). Each site is shown as a ball centered at its Ca atom in the structure (Protein Data Bank code 3BT6) [Wang et al., 2008]. Sites with greater than 95% posterior probability to be under positive selection are shown in dark color and the rest are in light color. The structure of one monomer of HA is in the same orientation as the monomer shown in color in Figure 1.1.

Table3.1 The values of log-likelihood (ℓ), d_N/d_S, and parameter estimates in the analysis of the HA_1 subunit of influenza B virus strains circulating between 1940~2007.

Model	ℓ	d_N/d_S	Parameters estimates
Early strain (I) 1940~1970 (11 strains)			
M0 (one-ratio)	-2343.24	0.271	ω=0.271
M1a (nearly neutral)	-2317.53	0.258	p_0=0.744 (p_1=0.256), ω_0=0.002 (ω_1=1)
M2a (positive selection)	-2314.85	0.297	p_0=0.744, p_1=0.251 (p_2=0.005), ω_0=0.005 (ω_1=1), ω_2=7.990
M7 (beta)	-2317.89	0.236	p=0.016, q=0.051
M8 (beta&ω)	-2315.01	0.287	p_0=0.994 (p_1=0.006), p=0.017, q=0.051, ω_s=7.428
B/YM-lineage (II-*i*) 1972~1984 (25 strains)			
M0 (one-ratio)	-2501.04	0.373	ω=0.373
M1a (nearly neutral)	-2461.96	0.262	p_0=0.755 (p_1=0.245), ω_0=0.022 (ω_1=1)
M2a (positive selection)	-2439.17	0.404	p_0=0.729, p_1=0.261 (p_2=0.011), ω_0=0.020 (ω_1=1), ω_2=11.904
M7 (beta)	-2462.35	0.300	p=0.005, q=0.012
M8 (beta&ω)	-2439.28	0.426	p_0=0.989 (p_1=0.011), p=0.017, q=0.042, ω_s=12.282
B/YM-lineage (II-*ii*) 1987~1996 (24 strains)			
M0 (one-ratio)	-2032.98	0.311	ω=0.311
M1a (nearly neutral)	-2002.40	0.188	p_0=0.812 (p_1=0.188), ω_0=0 (ω_1=1)
M2a (positive selection)	-1995.15	0.318	p_0=0.826, p_1=0.136 (p_2=0.038), ω_0=0 (ω_1=1), ω_2=4.779
M7 (beta)	-2002.45	0.200	p=0.005, q=0.020
M8 (beta&ω)	-1995.32	0.313	p_0=0.953 (p_1=0.047), p=0.012, q=0.076, ω_s=4.237
B/YM-lineage (II-*iii*) 1991~2002 (56 strains)			
M0 (one-ratio)	-2211.56	0.200	ω=0.200
M1a (nearly neutral)	-2196.38	0.168	p_0=0.901 (p_1=0.099), ω_0=0.077 (ω_1=1)
M2a (positive selection)	-2190.27	0.213	p_0=0.936, p_1=0.056 (p_2=0.009), ω_0=0.105 (ω_1=1), ω_2=6.802
M7 (beta)	-2197.92	0.184	p=0.098, q=0.437
M8 (beta&ω)	-2190.51	0.212	p_0=0.991 (p_1=0.009), p=0.431, q=2.312, ω_s=6.740
B/YM-lineage (II-*iv*) 1994~2005 (33 strains)			
M0 (one-ratio)	-2144.23	0.220	ω=0.220
M1a (nearly neutral)	-2127.98	0.211	p_0=0.789 (p_1=0.211), ω_0=0 (ω_1=1)
M2a (positive selection)	-2127.40	0.242	p_0=0.847, p_1=0.022 (p_2=0.131), ω_0=0.021 (ω_1=1), ω_2=1.537
M7 (beta)	-2128.03	0.200	p=0.005, q=0.020
M8 (beta&ω)	-2127.40	0.242	p_0=0.868 (p_1=0.132), p=0.052, q=1.023, ω_s=1.560
B/VI-lineage (III-*i*) 1975~1993 (24 strains)			
M0 (one-ratio)	-2530.27	0.336	ω=0.336
M1a (nearly neutral)	-2492.37	0.251	p_0=0.749 (p_1=0.251), ω_0=0 (ω_1=1)
M2a (positive selection)	-2477.70	0.351	p_0=0.736, p_1=0.254 (p_2=0.010), ω_0=0.005 (ω_1=1), ω_2=9.700
M7 (beta)	-2492.81	0.222	p=0.009, q=0.029
M8 (beta&ω)	-2477.89	0.354	p_0=0.983 (p_1=0.017), p=0.016, q=0.050, ω_s=7.267
B/VI-lineage (III-*ii*) 1996~2007 (98 strains)			
M0 (one-ratio)	-2924.64	0.299	ω=0.299
M1a (nearly neutral)	-2899.56	0.266	p_0=0.805 (p_1=0.195), ω_0=0.088 (ω_1=1)
M2a (positive selection)	-2887.74	0.320	p_0=0.796, p_1=0.199 (p_2=0.005), ω_0=0.094 (ω_1=1), ω_2=9.871
M7 (beta)	-2899.97	0.266	p=0.156, q=0.430
M8 (beta&ω)	-2887.00	0.309	p_0=0.994 (p_1=0.006), p=0.244, q=0.698, ω_s=8.534

Table 3.2 Likelihood ratio tests (LRT) between M2a versus M1a and M8 versus M7 for the seven subgroups of HA_1 subunit of influenza B virus strains circulating between 1940~2007

LRT	$2\Delta\ell^*$
Early strain (I) 1940~1970 (11 strains)	
M2a – M1a	**5.36**
M8 – M7	**5.76**
B/YM-lineage (II-*i*) 1972~1984 (25 strains)	
M2a – M1a	**44.44**
M8 – M7	**46.14**
B/YM-lineage (II-*ii*) 1987~1996 (24 strains)	
M2a – M1a	**14.50**
M8 – M7	**14.26**
B/YM-lineage (II-*iii*) 1991~2002 (56 strains)	
M2a – M1a	**12.22**
M8 – M7	**14.82**
B/YM-lineage (II-*iv*) 1994~2005 (33 strains)	
M2a – M1a	**1.16**
M8 – M7	**1.26**
B/VI-lineage (III-*i*) 1975~1993 (24 strains)	
M2a – M1a	**29.34**
M8 – M7	**29.84**
B/VI-lineage (III-*ii*) 1996~2007 (98 strains)	
M2a – M1a	**23.64**
M8 – M7	**25.94**

* In LRT tests, the values of $2\Delta\ell$ were compared with the critical values of χ^2 distribution (9.21 and 5.99 for $\chi^2_{1\%}$ and $\chi^2_{5\%}$, respectively, with d.f.=2) [Yang, 1997; Yang, 2007; Yang et al., 2000; Yang et al., 2005]. Significantly larger values of $2\Delta\ell$ over those of χ^2 distributions led to the rejection of the null models M1a and M7.

Table 3.3 Sites with higher than 50% posterior probabilities of being under positive selective pressure for the HA_1 subunit of influenza B virus strains circulating between 1940~2007

Model	Positively selected sites[1]
	Early strain (I) 1940~1970 (11 strains)
M2a (positive selectio	73, 150, 167*, 194**, 196, 238
M8 (beta&ω)	73*, 121, 122, 123, 136, 146, 150*, 154*, 160, 167***, 194***, 196**, 206, 238*
	B/YM-lineage (II-*i*) 1972~1984 (25 strains)
M2a (positive selectio	122, 129***, 194***, 196***, 206
M8 (beta&ω)	48, 122**, 129***, 137, 194***, 195, 196***, 206**, 230
	B/YM-lineage (II-*ii*) 1987~1996 (24 strains)
M2a (positive selectio	68*, 75**, 122, 129, 150**, 194***, 196***, 200, 227**, 295**
M8 (beta&ω)	68**, 73, 75***, 105, 122*, 129**, 149, 150***, 194***, 196***, 199, 200, 227***, 295***
	B/YM-lineage (II-*iii*) 1991~2002 (56 strains)
M2a (positive selectio	176, 194***, 196***
M8 (beta&ω)	75, 149, 176*, 194***, 196***, 206
	B/YM-lineage (II-*iv*) 1994~2005 (33 strains)
M2a (positive selectio	69*, 177**, 194, 196
M8 (beta&ω)	56, 69**, 118, 126, 177***, 194**, 196**
	B/VI-lineage (III-*i*) 1975~1993 (24 strains)
M2a (positive selectio	73*, 116, 167**, 194***, 196***, 290, 328
M8 (beta&ω)	68, 73***, 116**, 129, 137, 146, 167***, 194***, 196***, 290**, 328**
	B/VI-lineage (III-*ii*) 1996~2007 (98 strains)
M2a (positive selectio	194***, 196***
M8 (beta&ω)	80, 121*, 129, 194***, 196***

[1]Positively selected sites from Bayes Empirical Bayes analysis [Yang et al., 2005].
*Posterior probability of positive selective pressure is between 75~84%.
**Posterior probability of positive selective pressure is between 85~94%.
***Posterior probability of positive selective pressure is higher than 95%.

Early strain (I) (1940~1970).

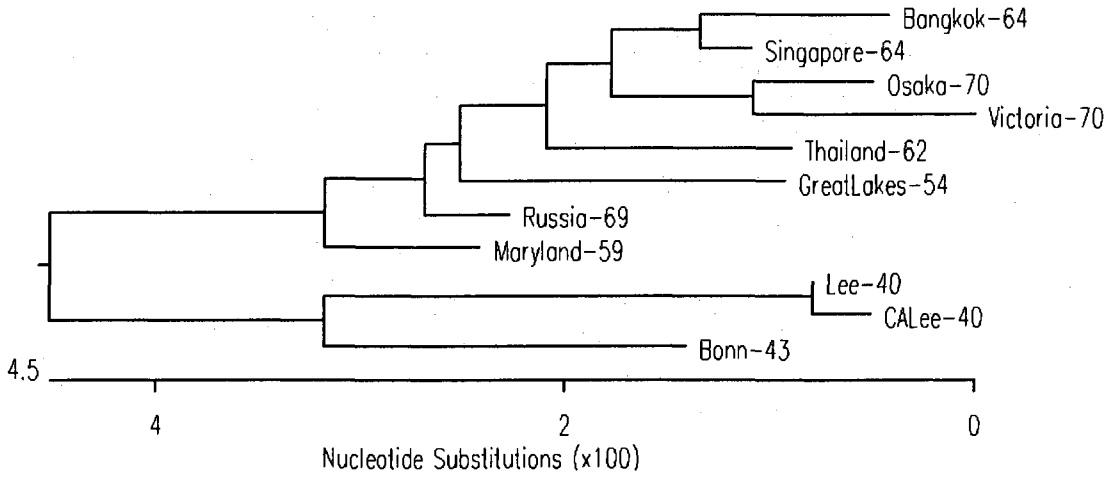

Fig. 3.3. Early strain lineage (I).

Among the 271 HA_1 sequences analyzed in this study, a total of 11 sequences over a time span of 31 years belong to this group (**Fig. 3.1, Fig. 3.3**). To limit the uncertainties related to the relatively small number of samples in this group, the results from Bayes Empirical Bayes analysis were used throughout this study [Anisimova et al., 2002; Yang et al., 2005]. In LRT tests, the values of $2\Delta l$ were 5.36 for M2a versus M1a, and 5.76 for M8 versus M7 (**Table 3.2**). These values were larger than the critical value of $\chi^2_{5\%}$ = 3.84, but smaller than $\chi^2_{1\%}$ = 6.63 with d.f. = 1 [Yang, 1997; Yang, 2007; Yang et al., 2000; Yang et al., 2005]. The M2a model suggested ~0.5% sites to be under positive selection with ω_2=7.990 (**Table 3.1**). Similarly, the M8 model suggested ~0.6% sites to be under positive selection with ω_s=7.428. The M2a model identified a total of six sites to be under positive selective pressure (>50% posterior probability) (**Table 3.3**). The M8 model identified 14 sites of being under positive selective pressure (>50% posterior probability) (**Fig. 3.3**). Among them, two sites were of greater than 95% posterior

probability to be under positive selection: HA_1 167 (95%) on the 160-loop and 194 (99%) on the 190-helix.

B/YM-like lineage (II).

A total of 138 HA_1 sequences in this analysis belong to B/YM-like lineage. It was further divided into four sublineages, II-*i* (25 sequences), II-*ii* (24 sequences), II-*iii* (56 sequences) and II-*iv* (33 sequences) (**Fig. 3.1**).

Early strain sublineage (II-*i*) (1972~1984).

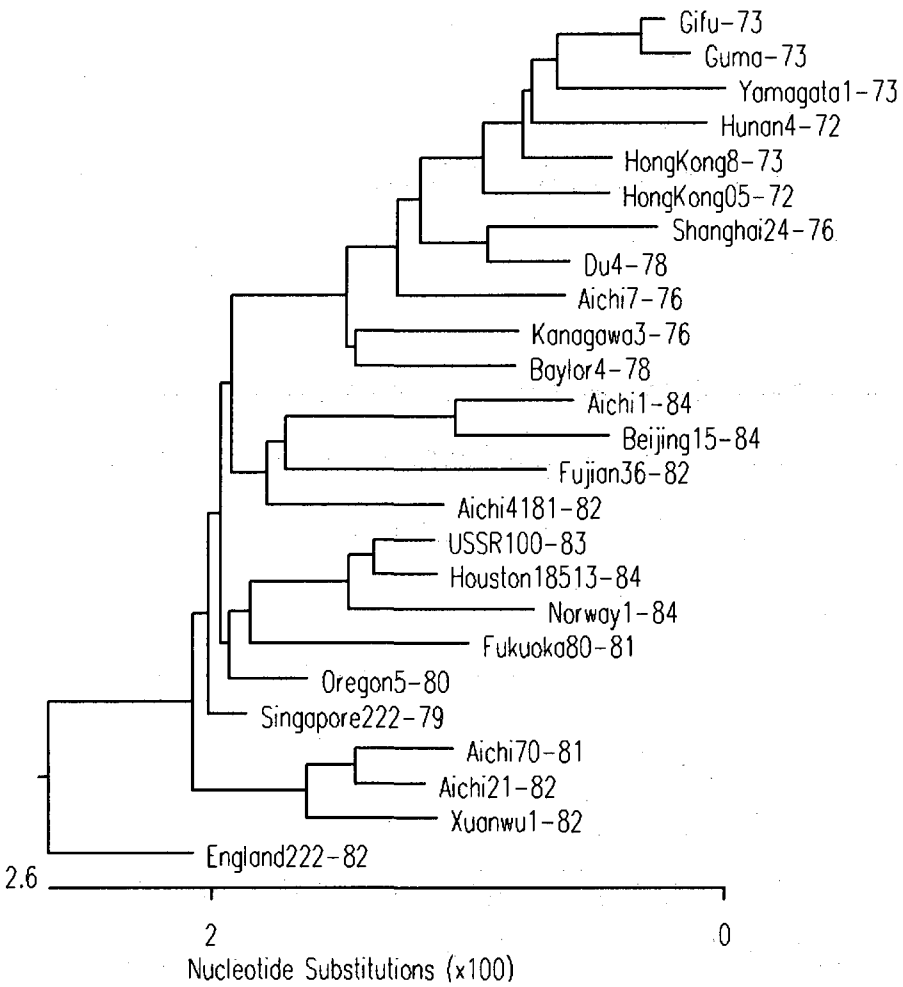

Fig. 3.4 B/YM-lineage II-*i*(1972~1984)

These early strains of B/YM-lineage spanned a period of 13 years (Fig. 3.4). The values of M2a versus M1a and M8 versus M7 were much greater than = 6.63 with d.f. = 1 (Table 3.1, 3.2), resulting in the rejection of the null models M1a and M7. Both M2a and M8 models suggested ~1.1% sites to be under strong positive selection with large values (Table 3.1). The M2a model identified a total of five sites to be under positive selection (>50% posterior probability) (Table 3.3), three of which were of greater than 95% posterior probability: HA1 129 (97%) on the 120-loop, 194 (100%) and 196 (100%) on the 190-helix. These three sites were again with >95% posterior probability in the M8 model: HA1 129 (99%), 194 (100%) and 196 (100%) (Fig. 3.4).

Sublineage (II-*ii*) (1987~1996).

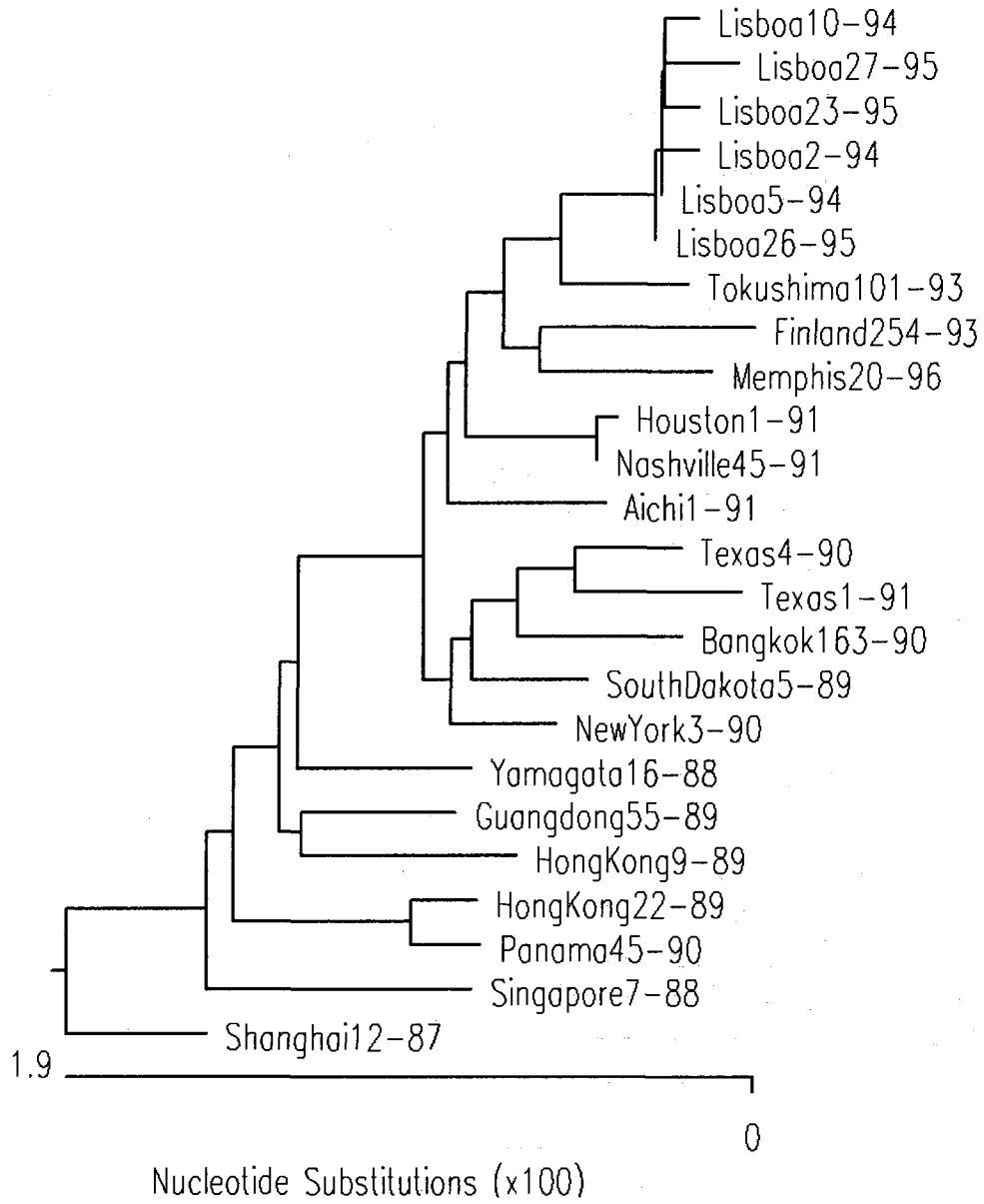

Fig. 3.5 B/YM-lineage II-*ii*

The B/YM-lineage strains in this group covered a 10-year period (**Fig. 3.5**). In LRT tests, the $2\Delta l$ values of M2a versus M1a and M8 versus M7 provided strong support for the existence of positive selection (**Table 3.1, 3.2**). Both M2a and M8 models suggested ~4% sites to be under positive selection with $\omega \approx 4$ (**Table 3.1**). The

M2a model identified two sites with higher than 95% posterior probability of being positively selected (**Table 3.3**): HA_1 194 (99%) and 196 (97%) on the 190-helix. The M8 model identified a total of six sites with greater than 95% posterior probability of being positively selected (**Table 3.3**): HA_1 75 (97%) and 295 (98%) on the 120-loop, 150 (96%) on the 150-loop, 194 (100%), 196 (99%) and 227 (97%) on the 190-helix (**Fig. 3.5**). It is noteworthy that HA_1 150 on the 150-loop was inferred to be under positive selection with very high confidence, in excellent agreement with previous conclusions that the 150-loop is an important epitope for B/YM-lineage [Nakagawa et al., 2001a; Nakagawa et al., 2003].

Sublineage (II-*iii*) (1991~2002). This sublineage of B/YM-like strains covered a 12-year period (**Fig. 3.6**). The LRT tests led to the rejection of the null models M1a and M7 (**Table 3.1, 3.2**). Both M2a and M8 models suggested ~0.9% sites to be under positive selection with $\omega \approx 7$ (**Table 3.1, 3.2**). The M2a model revealed three sites of being under positive selection, including HA_1 194 (97%) and 196 (99%) on the 190-helix (**Table 3.3**). These two sites were of 99% and 100% posterior probability of positive selection in the M8 model (**Table 3.3 and Fig. 3.6**).

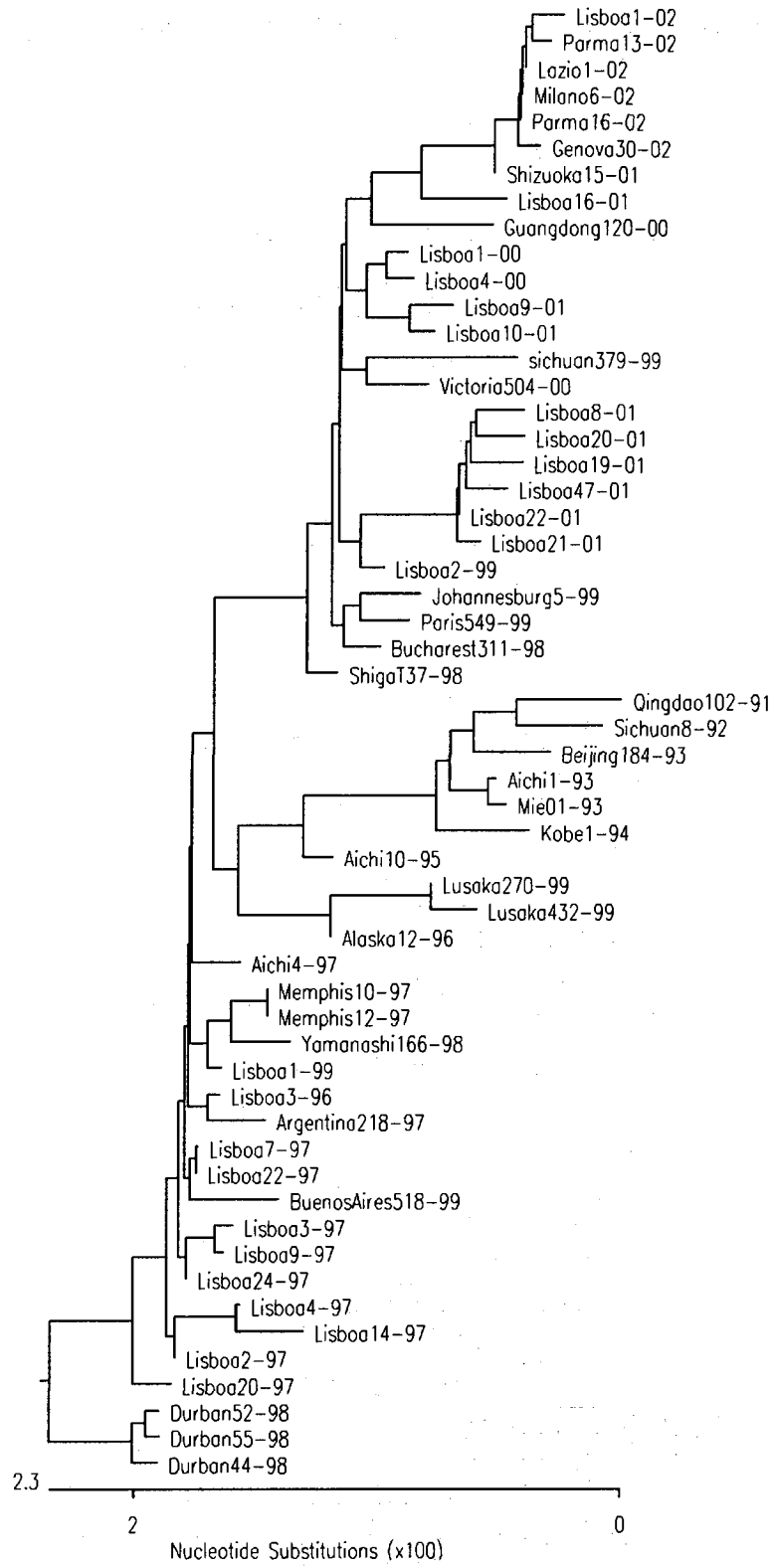

Fig. 3.6 B/YM-lineage II-*iii*.

Sublineage (II-*iv*) (1994~2005).

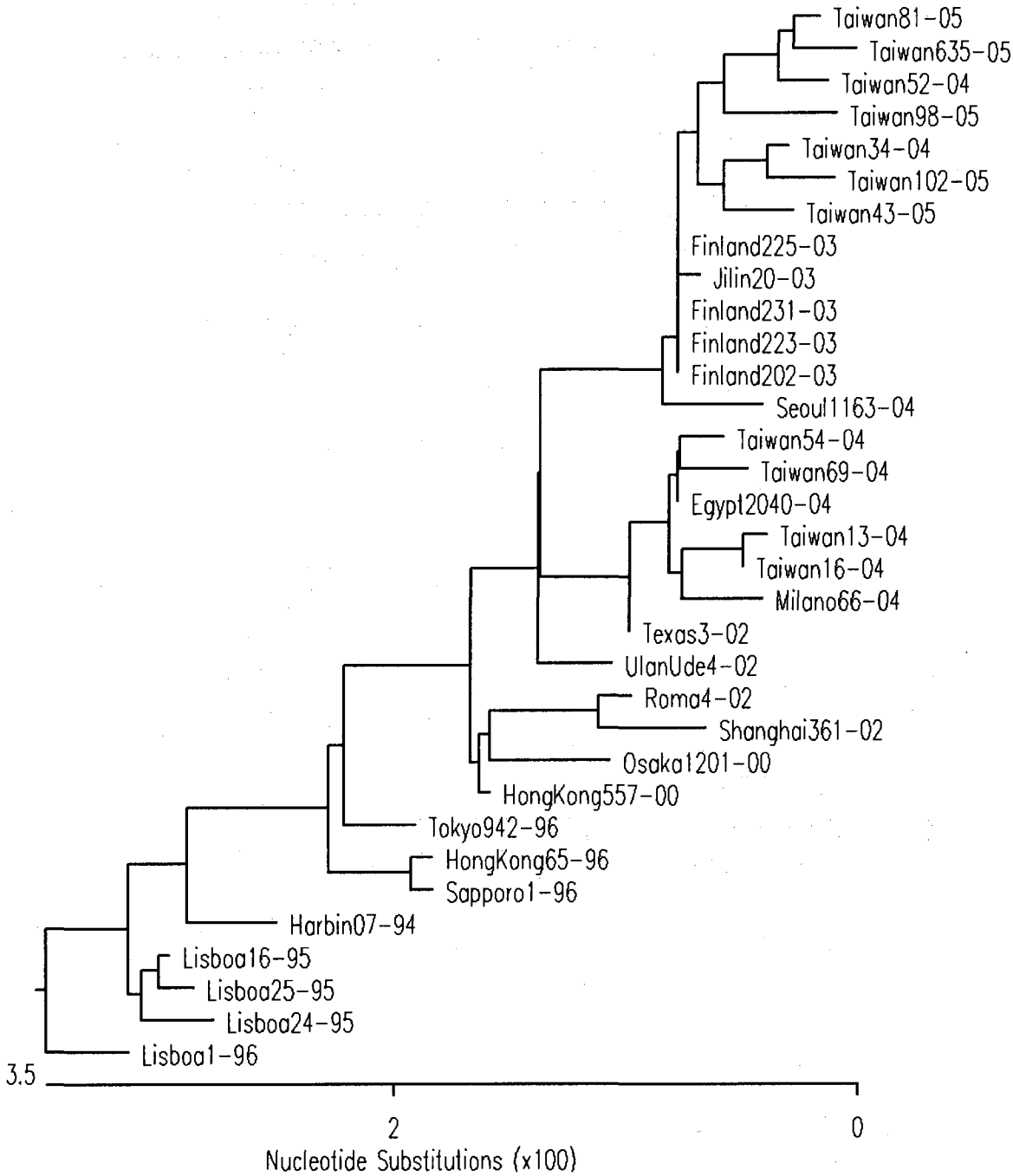

Fig. 3.7 B/YM-lineage II-*iv*.

This sublineage of B/YM-like strains contained some of the most recently circulating

strains of B/YM-lineage (**Fig. 3.7**). In sharp contrast to all other sublineages of B/YM-like strains and to all B/VI-like strains, the LRT statistics were $2\Delta l$ =1.16 and 1.26 for M2a versus M1a and M8 versus M7, respectively (**Table 3.2**), suggesting a low confidence for the existence of positive selection. In Bayes Empirical Bayes analysis, both M2a and M8 models suggested a relatively large percentage of sites (~13%) to be under very weak positive selection with ω_2=1.537 and ω_s=1.560, respectively (**Table 3.1**). The M2a model identified a total of four positively selected sites with >50% posterior probability (**Table 3.3**). In the M8 model, a total of seven sites were identified, with only one site, HA_1 177 (98%) on the 120-loop, with > 95% posterior probability (**Table 3.3 and Fig. 3.7**). This sublineage was the only group in which HA_1 194 and 196 are of lower than 95% probability to be under positive selection.

B/VI-like lineage (III). A total of 122 HA_1 sequences of influenza B virus strains belong to this lineage. They were grouped into two sublineages, early strains (III-*i*) containing 24 sequences and more recent strains (III-*ii*) containing 98 sequences (**Fig. 3.1**).

Early strain sublineage (III-*i*) (1975~1993). These early strains of B/VI-lineage spanned a time period of 19 years and exhibited significant sequence differences from the recent circulating B/VI-like strains (III-*ii*) (**Fig.2, Fig. 3.8**). The LRT statistics supported the rejection of the null models (M1a and M7) and strongly supporting the presence of positive selection (**Table 3.1, 3.2**). The M2a model suggested 1.0% sites to be under positive selection with ω_2=9.700 (**Table 3.1**). Similarly, the M8 model suggested 1.7% sites to be under positive selection with ω_s=7.267. The

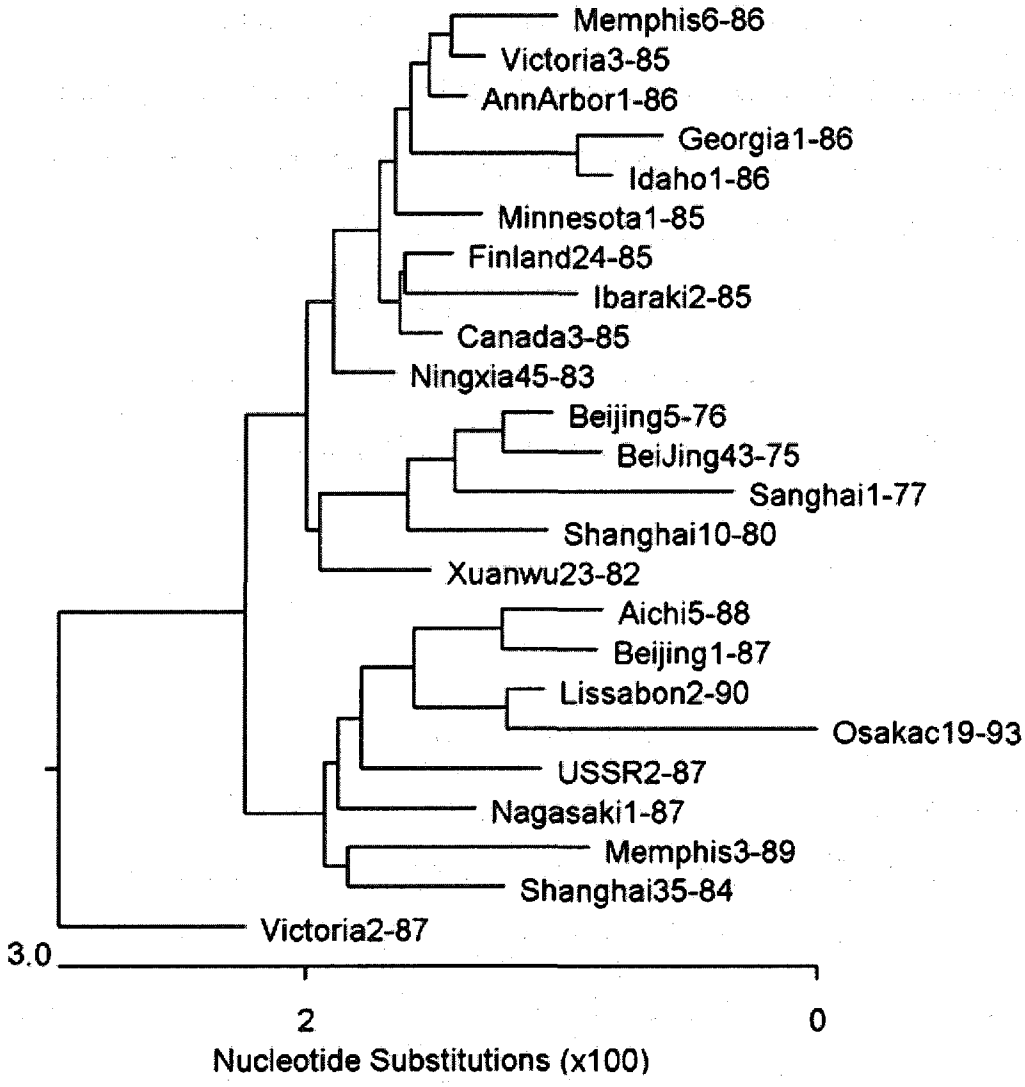

Fig. 3.8 B/VI-lineage III-*i*.

M2a model identified seven sites to be under positive selective pressure (**Table 3.3**), with HA_1 194 (100%) and 196 (100%) on the 190-helix of higher than 95% posterior probability. The M8 model revealed a total of 11 positively selected sites, among which four sites were of greater than 95% posterior probability. They were HA_1 73 (95%) (120-loop), 167 (97%) (160-loop), 194 (100%) and 196 (100%) (190-helix) (**Table 3.3,**

Fig. 3.8). It is important to note that among these four sites with the highest posterior probability, one site, HA_1 167, is on the 160-loop, while none is located on the 150-loop. In sharp contrast, B/YM-like (II-*ii*) sublineage, which circulated in an overlapping time period and contained the same number of sequences, had HA_1 150 on the 150-loop to be under positive selection. These observations further supported earlier conclusions that the 160-loop epitope is specific for the B/VI-lineage strains [Nakagawa et al., 2001b; Nakagawa et al., 2005] while the 150-loop epitope is specific for the B/YM-lineage strains [Nakagawa et al., 2001a; Nakagawa et al., 2003].

Recent strain sublineage (III-*ii*) (1996~2007).

The more recent isolates of B/VI-lineage strains remained to be a single group over the time period of 12 years (**Fig. 3.9**). The LRT statistics supported strongly the presence of positive selection (**Table 3.1, 3.2**). The M2a model suggested ~0.5% sites to be under positive selection with ω_2=9.871 (**Table 3.1**). Similarly, the M8 model suggested ~0.6% sites to be under positive selection with ω_s=8.534. HA_1 194 had a posterior probability of 95% and 99% to be under positive selection in M2a and M8 models, respectively, while HA_1 196 has a 100% posterior probability in both M2a and M8 models (**Table 3.3 and Fig. 3.9**). Compared to the earlier B/VI-like (III-*i*) sublineage, one noticeable difference is that the 160-loop was no longer under positive selection in these recent strains (III-*ii*). Rather, positive selection was focused on the 190-helix.

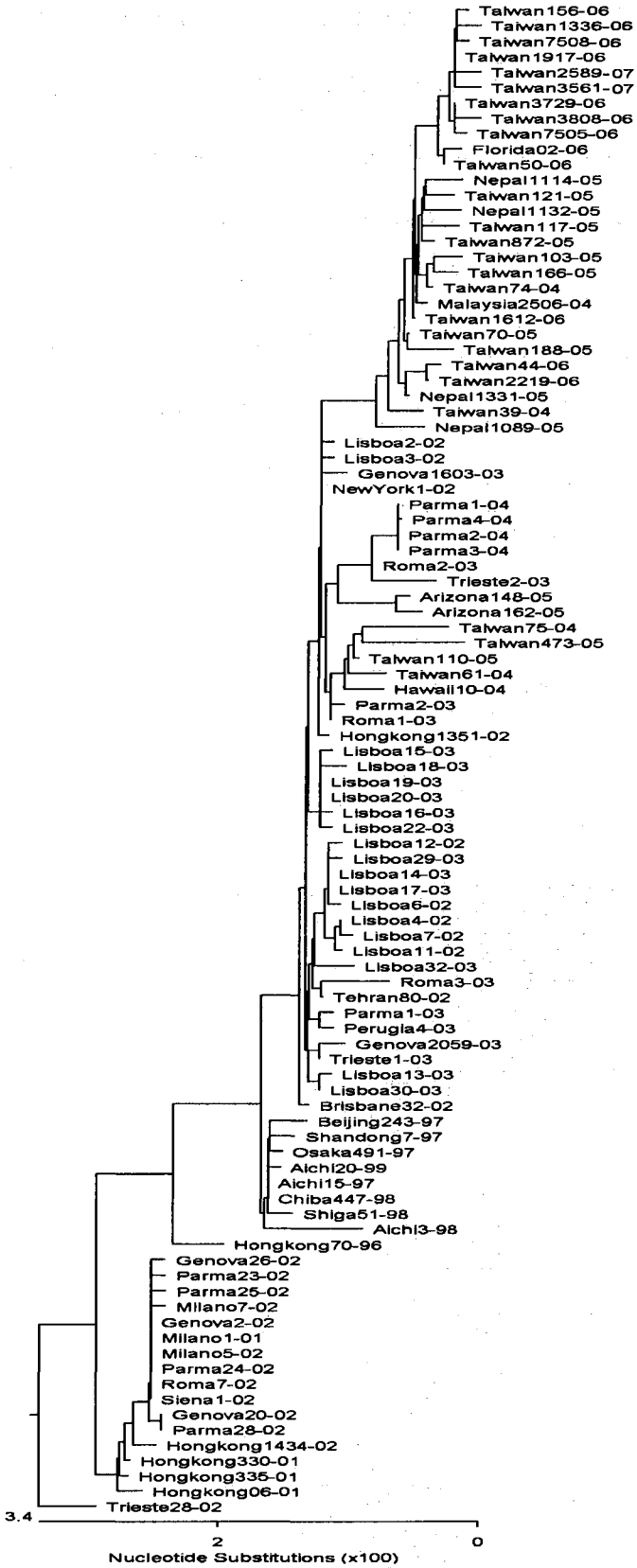

Fig. 3.9

B/VI-lineage III-ii.

Chapter 4

Discussion and Conclusions

Roles of antibody selection in the evolution of influenza B virus HA

In previous studies, four major antigenic epitopes of influenza B virus HA, the 120-loop, the 150-loop, the 160-loop, and the 190-helix, were identified on the membrane-distal domain of HA_1 [Wang et al., 2008] (**Fig.1.1**). Strikingly, in this study, all the identified positively selected sites in the seven subgroups were located on these four major antigenic epitopes, supporting the important roles of antibody selection in the molecular evolution of influenza B virus HA.

The 150-loop is an important epitope on HA. Antigenic properties were altered for influenza B virus with mutations on this loop in laboratory-selected escape mutants [Berton et al., 1984; Berton and Webster, 1985; Nakagawa et al., 2003; Webster and Berton, 1981], field isolates [Abed et al., 2003; Nakagawa et al., 2001a] and egg-adapted variants [Lugovtsev et al., 2005; Lugovtsev et al., 2007; Oxford et al., 1991; Oxford et al., 1990]. In more recent influenza B virus isolates, the 150-loop region appeared to be the neutralizing epitope specific for B/YM-like strains [Nakagawa et al., 2001a; Nakagawa et al., 2003]. Consistent with that finding, HA_1 150 was under positive selection with 96% posterior probability in B/YM-like (II-*ii*) sublineage.

The 160-loop is the only region in influenza B virus HA where insertions, deletions and single amino-acid substitutions were detected in field isolates [McCullers et al., 1999; Nakagawa et al., 2005; Nerome et al., 1998] and mAb-escape mutants [Berton et al., 1984; Berton and Webster, 1985; Nakagawa et al., 2001a; Nakagawa et al., 2001b; Webster and Berton, 1981], as an effective way for influenza B virus to survive a long

period of time without antigenic shifts as observed in influenza A virus [Nerome et al., 1998]. In recent isolates, the 160-loop became specific for B/VI-like lineage [Nakagawa et al., 2001b; Nakagawa et al., 2005]. In agreement with this observation, HA_1 167 on the 160-loop was selected positively in early strains (I) and in B/VI-like (III-*i*) sublineage (**Table 3.3, fig.3.2**), with 95% and 96% posterior probability, respectively.

The 190-helix, which forms part of the receptor-binding site (RBS) of influenza B virus HA, is inarguably one of the most important epitopes. The hot spot is at HA_1 194~196, a potential glycosylation site. Similar to influenza A virus HA [Caton et al., 1982; Daniels et al., 1983; Schulze, 1997; Skehel et al., 1984; Skehel and Wiley, 2000], influenza B virus HA also utilized the addition or removal of glycosylation as a mechanism for antigenic drift [Berton et al., 1984; Berton and Webster, 1985; Gambaryan et al., 1999; Ikonen et al., 2005; Muyanga et al., 2001; Nakagawa et al., 2000; Nakagawa et al., 2004; Oxford et al., 1991; Oxford et al., 1990; Robertson et al., 1990; Robertson et al., 1985; Saito et al., 2004; Schild et al., 1983; Wang et al., 2008]. In this current analysis, HA_1 194 and 196 were constantly identified to be under positive selective pressure, with greater than 99% probability in 11 out of 14 cumulative cases (combining both sites in seven groups), and over 85% in three other cases. HA_1 227 in sublineage (II-*ii*) was another positively selected site on 190-helix with high posterior probability (97%) (**Table 3.3, Fig.3.2**).

Perhaps one of the most important observations from this study is positive selection of the 120-loop region. The 120-loop epitope was defined as HA_1 116~137 and its surrounding regions [Wang et al., 2008]. In this context of this article, we refer to all sites not adjoining the 150-loop, 160-loop or the 190-helix epitopes as the 120-loop

region due to spatial proximity (**Fig.1.1**). Although the 120-loop region appeared to be one of the most frequently mutated regions in field isolates [Verhoeyen et al., 1983], its role in antigenicity of influenza B virus HA was not recognized until most recently [Lugovtsev et al., 2007; Nakagawa et al., 2006; Wang et al., 2008]. One possibility for such a delay in recognition is that the 120-loop is proximal to the viral envelope membrane, making the access by antibodies more difficult, as observed for influenza A virus HA [Barbey-Martin et al., 2002; Bizebard et al., 2001; Bizebard et al., 1995; Fleury et al., 1999; Gigant et al., 2000; Knossow et al., 2002]. Thus, it is very important that this current study provided strong evidence for positive selection of the 120-loop region, further supporting its significance in antigenicity of influenza B virus HA.

Trends of positive selection on influenza B virus HA

The early strains (I) seemed to have rather even distribution of positive selective pressure on all four major epitopes, although the positive selection on the 160-loop and 190-helix appeared to be stronger and/or more prevailing (**Table 3.3, Fig. 3.3**). In contrast, the early strains of B/YM-lineage and B/VI-lineage, sublineages (II-*ii*) and (III-*i*) respectively, were sharply divided. HA_1 150 on the 150-loop in B/YM-like (II-*ii*) sublineage, and HA_1 167 on the 160-loop in B/VI-like (III-*i*) sublineage, were inferred to be under positive selection with high posterior probability (**Table 3.3**). These observations agreed very well with earlier studies in which the 150-loop and 160-loop were found to be specific epitopes for the B/YM- and B/VI-lineages, respectively [Nakagawa et al., 2001a; Nakagawa et al., 2001b; Nakagawa et al., 2003; Nakagawa et al., 2005]. However, despite large sequence differences, the recent B/YM-like (II-*iii*)

sublineage and B/VI-like (III-*ii*) sublineage converged at focusing on the 190-helix for antigenic drift (**Table 3.3 and Fig. 3.2**). Most strikingly, in the newest B/YM-like (II-*iv*) sublineage, a large number of sites were found to be under rather weak positive selection, and the only positively selected site identified with high confidence was HA_1 177 on the 120-loop. The new trends of positive selection among these most recent strains, in conjunction with results from other studies [Lugovtsev et al., 2007; Nakagawa et al., 2006], stress the increasingly important role of the 120-loop in antigenicity of influenza B virus HA.

Concluding remarks

This study reports a large-scale systematic analysis of diversifying positive selective pressure on HA of distinct lineages/sublineages of influenza B virus isolated in the past 68 years. The highlights of the results from this study are:

a). The number of positively selected sites in influenza B virus HA were much fewer than those of influenza A virus HA [Air et al., 1990];

b). Although it does not have subtypes as influenza A virus HA, influenza B virus HA did and continue to diverge into different sublineages. This was particularly true for B/YM-lineage, as exemplified by the newly emerging B/YM-like (II-*iv*) sublineage that had not been previously described.

c). The study revealed the predominant roles of antibody selection in the molecular evolution of influenza B virus HA.

d). Despite the differences among different lineages/sublineages, HA_1 194 and 196 were constantly under positive selective pressure in all but one cases.

e). The 120-loop was an important epitope under constant positive selection. It may play an increasingly important role in antigenicity in future field isolates, as evidenced in the most recent B/YM-like (II-*iv*) sublineage (**Table 3.3**).

f). Each lineage/sublineage utilized their respective favorite sites in positive selection. Thus, for any newly emerging strains of influenza B virus, it is important to put them in the context of their evolutionary history in order to understand and appreciate their full epidemic potential.

References

Abed Y, Coulthart MB, Li Y, Boivin G. 2003. Evolution of surface and nonstructural-1 genes of influenza B viruses isolated in the Province of Quebec, Canada, during the 1998-2001 period. Virus Genes 27(2):125-135.

Air GM, Gibbs AJ, Laver WG, Webster RG. 1990. Evolutionary changes in influenza B are not primarily governed by antibody selection. Proc Natl Acad Sci U S A 87(10):3884-3888.

Anisimova M, Bielawski JP, Yang Z. 2002. Accuracy and power of bayes prediction of amino acid sites under positive selection. Mol Biol Evol 19(6):950-958.

Barbey-Martin C, Gigant B, Bizebard T, Calder LJ, Wharton SA, Skehel JJ, Knossow M. 2002. An antibody that prevents the hemagglutinin low pH fusogenic transition. Virology 294(1):70-74.

Berton MT, Naeve CW, Webster RG. 1984. Antigenic structure of the influenza B virus hemagglutinin: nucleotide sequence analysis of antigenic variants selected with monoclonal antibodies. J Virol 52(3):919-927.

Berton MT, Webster RG. 1985. The antigenic structure of the influenza B virus hemagglutinin: operational and topological mapping with monoclonal antibodies. Virology 143(2):583-594.

Bizebard T, Barbey-Martin C, Fleury D, Gigant B, Barrere B, Skehel JJ, Knossow M. 2001. Structural studies on viral escape from antibody neutralization. Curr Top Microbiol Immunol 260:55-64.

Bizebard T, Gigant B, Rigolet P, Rasmussen B, Diat O, Bosecke P, Wharton SA, Skehel JJ, Knossow M. 1995. Structure of influenza virus haemagglutinin complexed with a

neutralizing antibody. Nature 376(6535):92-94.

Bootman JS, Robertson JS. 1988. Sequence analysis of the hemagglutinin of B/Ann Arbor/1/86, an epidemiologically significant variant of influenza B virus. Virology 166(1):271-274.

Bush RM, Fitch WM, Bender CA, Cox NJ. 1999. Positive selection on the H3 hemagglutinin gene of human influenza virus A. Mol Biol Evol 16(11):1457-1465.

Caton AJ, Brownlee GG, Yewdell JW, Gerhard W. 1982. The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). Cell 31(2 Pt 1):417-427.

Chen JM, Guo YJ, Wu KY, Guo JF, Wang M, Dong J, Zhang Y, Li Z, Shu YL. 2007. Exploration of the emergence of the Victoria lineage of influenza B virus. Arch Virol 152(2):415-422.

Chen R, Holmes EC. 2008. The Evolutionary Dynamics of Human Influenza B Virus. J Mol Evol.

Daniels RS, Douglas AR, Skehel JJ, Wiley DC. 1983. Analyses of the antigenicity of influenza haemagglutinin at the pH optimum for virus-mediated membrane fusion. J Gen Virol 64 (Pt 8):1657-1662.

Deely JJ, Lindley DV. 1981. Bayes empirical bayes. J Am Stat Assoc 76:833-841.

Fleury D, Barrere B, Bizebard T, Daniels RS, Skehel JJ, Knossow M. 1999. A complex of influenza hemagglutinin with a neutralizing antibody that binds outside the virus receptor binding site. Nat Struct Biol 6(6):530-534.

Gambaryan AS, Robertson JS, Matrosovich MN. 1999. Effects of egg-adaptation on the receptor-binding properties of human influenza A and B viruses. Virology

258(2):232-239.

Gigant B, Barbey-Martin C, Bizebard T, Fleury D, Daniels R, Skehel JJ, Knossow M. 2000. A neutralizing antibody Fab-influenza haemagglutinin complex with an unprecedented 2:1 stoichiometry: characterization and crystallization. Acta Crystallogr D Biol Crystallogr 56 (Pt 8):1067-1069.

Hovanec DL, Air GM. 1984. Antigenic structure of the hemagglutinin of influenza virus B/Hong Kong/8/73 as determined from gene sequence analysis of variants selected with monoclonal antibodies. Virology 139(2):384-392.

Ikonen N, Pyhala R, Axelin T, Kleemola M, Korpela H. 2005. Reappearance of influenza B/Victoria/2/87-lineage viruses: epidemic activity, genetic diversity and vaccination efficacy in the Finnish Defence Forces. Epidemiol Infect 133(2):263-271.

Kanegae Y, Sugita S, Endo A, Ishida M, Senya S, Osako K, Nerome K, Oya A. 1990. Evolutionary pattern of the hemagglutinin gene of influenza B viruses isolated in Japan: cocirculating lineages in the same epidemic season. J Virol 64(6):2860-2865.

Knossow M, Gaudier M, Douglas A, Barrere B, Bizebard T, Barbey C, Gigant B, Skehel JJ. 2002. Mechanism of neutralization of influenza virus infectivity by antibodies. Virology 302(2):294-298.

Krystal M, Elliott RM, Benz EW, Jr., Young JF, Palese P. 1982. Evolution of influenza A and B viruses: conservation of structural features in the hemagglutinin genes. Proc Natl Acad Sci U S A 79(15):4800-4804.

Krystal M, Young JF, Palese P, Wilson IA, Skehel JJ, Wiley DC. 1983. Sequential

mutations in hemagglutinins of influenza B virus isolates: definition of antigenic domains. Proc Natl Acad Sci U S A 80(14):4527-4531.

Lubeck MD, Schulman JL, Palese P. 1980. Antigenic variants of influenza viruses: marked differences in the frequencies of variants selected with different monoclonal antibodies. Virology 102(2):458-462.

Lugovtsev VY, Vodeiko GM, Levandowski RA. 2005. Mutational pattern of influenza B viruses adapted to high growth replication in embryonated eggs. Virus Res 109(2):149-157.

Lugovtsev VY, Vodeiko GM, Strupczewski CM, Ye Z, Levandowski RA. 2007. Generation of the influenza B viruses with improved growth phenotype by substitution of specific amino acids of hemagglutinin. Virology 365(2):315-323.

Macken C, Lu H, Goodman J, Boykin L. 2001. The value of a database in surveillance and Vaccine selection. In: Osterhaus A, Cox N, Hampson AW, editors. Options for the control of the influenza IV. Amsterdam: Elsevier Sciences. p 103-106.

Matsuzaki Y, Sugawara K, Takashita E, Muraki Y, Hongo S, Katsushima N, Mizuta K, Nishimura H. 2004. Genetic diversity of influenza B virus: the frequent reassortment and cocirculation of the genetically distinct reassortant viruses in a community. J Med Virol 74(1):132-140.

McCullers JA, Saito T, Iverson AR. 2004. Multiple genotypes of influenza B virus circulated between 1979 and 2003. J Virol 78(23):12817-12828.

McCullers JA, Wang GC, He S, Webster RG. 1999. Reassortment and insertion-deletion are strategies for the evolution of influenza B viruses in nature. J Virol 73(9):7343-7348.

Muyanga J, Matsuzaki Y, Sugawara K, Kimura K, Mizuta K, Ndumba I, Muraki Y, Tsuchiya E, Hongo S, Kasolo FC, Numazaki Y, Nakamura K. 2001. Antigenic and genetic analyses of influenza B viruses isolated in Lusaka, Zambia in 1999. Arch Virol 146(9):1667-1679.

Nakagawa N, Kubota R, Maeda A, Nakagawa T, Okuno Y. 2000. Heterogeneity of influenza B virus strains in one epidemic season differentiated by monoclonal antibodies and nucleotide sequences. J Clin Microbiol 38(9):3467-3469.

Nakagawa N, Kubota R, Maeda A, Okuno Y. 2004. Influenza B virus victoria group with a new glycosylation site was epidemic in Japan in the 2002-2003 season. J Clin Microbiol 42(7):3295-3297.

Nakagawa N, Kubota R, Morikawa S, Nakagawa T, Baba K, Okuno Y. 2001a. Characterization of new epidemic strains of influenza B virus by using neutralizing monoclonal antibodies. J Med Virol 65(4):745-750.

Nakagawa N, Kubota R, Nakagawa T, Okuno Y. 2001b. Antigenic variants with amino acid deletions clarify a neutralizing epitope specific for influenza B virus Victoria group strains. J Gen Virol 82(Pt 9):2169-2172.

Nakagawa N, Kubota R, Nakagawa T, Okuno Y. 2003. Neutralizing epitopes specific for influenza B virus Yamagata group strains are in the 'loop'. J Gen Virol 84(Pt 4):769-773.

Nakagawa N, Kubota R, Okuno Y. 2005. Variation of the conserved neutralizing epitope in influenza B virus victoria group isolates in Japan. J Clin Microbiol 43(8):4212-4214.

Nakagawa N, Suzuoki J, Kubota R, Kobatake S, Okuno Y. 2006. Discovery of the

neutralizing epitope common to influenza B virus victoria group isolates in Japan. J Clin Microbiol 44(4):1564-1566.

Nerome R, Hiromoto Y, Sugita S, Tanabe N, Ishida M, Matsumoto M, Lindstrom SE, Takahashi T, Nerome K. 1998. Evolutionary characteristics of influenza B virus since its first isolation in 1940: dynamic circulation of deletion and insertion mechanism. Arch Virol 143(8):1569-1583.

Oxford JS, Newman R, Corcoran T, Bootman J, Major D, Yates P, Robertson J, Schild GC. 1991. Direct isolation in eggs of influenza A (H1N1) and B viruses with haemagglutinins of different antigenic and amino acid composition. J Gen Virol 72 (Pt 1):185-189.

Oxford JS, Schild GC, Corcoran T, Newman R, Major D, Robertson J, Bootman J, Higgins P, al-Nakib W, Tyrrell DA. 1990. A host-cell-selected variant of influenza B virus with a single nucleotide substitution in HA affecting a potential glycosylation site was attenuated in virulence for volunteers. Arch Virol 110(1-2):37-46.

Pechirra P, Nunes B, Coelho A, Ribeiro C, Goncalves P, Pedro S, Castro LC, Rebelo-de-Andrade H. 2005. Molecular characterization of the HA gene of influenza type B viruses. J Med Virol 77(4):541-549.

Robertson JS, Bootman JS, Nicolson C, Major D, Robertson EW, Wood JM. 1990. The hemagglutinin of influenza B virus present in clinical material is a single species identical to that of mammalian cell-grown virus. Virology 179(1):35-40.

Robertson JS, Naeve CW, Webster RG, Bootman JS, Newman R, Schild GC. 1985. Alterations in the hemagglutinin associated with adaptation of influenza B virus to

growth in eggs. Virology 143(1):166-174.

Rota PA, Hemphill ML, Whistler T, Regnery HL, Kendal AP. 1992. Antigenic and genetic characterization of the haemagglutinins of recent cocirculating strains of influenza B virus. J Gen Virol 73 (Pt 10):2737-2742.

Rota PA, Wallis TR, Harmon MW, Rota JS, Kendal AP, Nerome K. 1990. Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983. Virology 175(1):59-68.

Saito T, Nakaya Y, Suzuki T, Ito R, Saito H, Takao S, Sahara K, Odagiri T, Murata T, Usui T, Suzuki Y, Tashiro M. 2004. Antigenic alteration of influenza B virus associated with loss of a glycosylation site due to host-cell adaptation. J Med Virol 74(2):336-343.

Schild GC, Oxford JS, de Jong JC, Webster RG. 1983. Evidence for host-cell selection of influenza virus antigenic variants. Nature 303(5919):706-709.

Schulze IT. 1997. Effects of glycosylation on the properties and functions of influenza virus hemagglutinin. J Infect Dis 176 Suppl 1:S24-28.

Shaw MW, Xu X, Li Y, Normand S, Ueki RT, Kunimoto GY, Hall H, Klimov A, Cox NJ, Subbarao K. 2002. Reappearance and global spread of variants of influenza B/Victoria/2/87 lineage viruses in the 2000-2001 and 2001-2002 seasons. Virology 303(1):1-8.

Skehel JJ, Stevens DJ, Daniels RS, Douglas AR, Knossow M, Wilson IA, Wiley DC. 1984. A carbohydrate side chain on hemagglutinins of Hong Kong influenza viruses inhibits recognition by a monoclonal antibody. Proc Natl Acad Sci U S A 81(6):1779-1783.

Skehel JJ, Wiley DC. 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem 69:531-569.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673-4680.

Verhoeyen M, Van Rompuy L, Jou WM, Huylebroeck D, Fiers W. 1983. Complete nucleotide sequence of the influenza B/Singapore/222/79 virus hemagglutinin gene and comparison with the B/Lee/40 hemagglutinin. Nucleic Acids Res 11(14):4703-4712.

Wang Q, Cheng F, Lu M, Tian X, Ma J. 2008. Crystal Structure of Unliganded Influenza B Virus Hemagglutinin. J Virol 82:3011-3020.

Webster RG, Berton MT. 1981. Analysis of antigenic drift in the haemagglutinin molecule of influenza B virus with monoclonal antibodies. J Gen Virol 54(Pt 2):243-251.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13(5):555-556.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24(8):1586-1591.

Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155(1):431-449.

Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22(4):1107-1118.