SAGE-Hindawi Access to Research
International Journal of Evolutionary Biology
Volume 2011, Article ID 143498, 11 pages
doi:10.4061/2011/143498

Research Article

Computational Analysis Suggests That Lyssavirus Glycoprotein Gene Plays a Minor Role in Viral Adaptation

Kevin Tang[1] and Xianfu Wu[2]

[1] *BCFB, DSR, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA*
[2] *Rabies, PRB, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA*

Correspondence should be addressed to Xianfu Wu, xwu@cdc.gov

Received 15 October 2010; Revised 15 December 2010; Accepted 3 January 2011

Academic Editor: Hiromi Nishida

Copyright © 2011 K. Tang and X. Wu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Lyssavirus glycoprotein (G) is a membrane protein responsible for virus entry and protective immune responses. To explore possible roles of the glycoprotein in host shift or adaptation of *Lyssavirus*, we retrieved 53 full-length glycoprotein gene sequences from NCBI GenBank. The sequences were from different host isolates over a period of 70 years in 21 countries. Computational analyses detected 1 recombinant (AY987478, a dog isolate of CHAND03, genotype 1 in India) with incongruent phylogenetic support. No recombination was detected when AY98748 was excluded in the analyses. We applied different selection models to identify selection pressure on the glycoprotein gene. One codon at amino acid residual 483 was found to be under weak positive selection with marginal probability of 95% by using the maximum likelihood method. We found no significant evidence of positive selection on any site of the glycoprotein gene when the putative recombinant AY987478 was excluded. The computational analyses suggest that the G gene has been under purifying selection and that the evolution of the G gene may not play a significant role in Lyssavirus adaptation.

1. Introduction

Positive selection and recombination are important mechanisms in microbial pathogen adaption to new hosts, resistance to antibiotics, and evasion of immune responses [1]. RNA viruses have high mutation rates due to lack of both proofreading and postreplicative repair activities associated with RNA replicases and reverse transcriptases [2], which benefits RNA viruses in adapting to the changing environment. Recombination is a general phenomenon in evolution and plays a significant role in viral fitness [3, 4]. Rabies virus is a single-stranded negative RNA virus belonging to the order *Mononegavirales*, family *Rhabdoviridae*, genus *Lyssavirus*, which causes rabies in all warm-blooded mammals. Host shift and spillover events are frequently reported in rabies [5–9]. The nucleotide substitution rate of lyssaviruses is estimated to be around 10^{-4} per site per year [7]. The RNA-dependent RNA-polymerase (RdRp or L) together with phosphoprotein (P), functions as the transcriptase and replicase complex. The glycoprotein (G) is the only outer membrane protein responsible for virus entry and inducing protective immune responses [10, 11]. The role of the G gene in rabies spillover, host shift, and adaptation has not been analyzed thoroughly. The information could help understand viral pathogenesis and develop a vaccine for a broad spectrum of lyssavirus infections.

Here, we used newly developed computational algorithms as well as traditional methods to investigate potential recombination events and selection pressures in the G gene of Lyssaviruses. The dataset for the study was comprised of 53 full-length glycoprotein gene sequences isolated from different hosts in 21 countries over a period of 70 years. We hypothesized that if different hosts with rabies infections over decades did not lead to positive selection or recombination events in the G gene, the gene does not play a significant role in lyssavirus adaptation.

2. Methods

2.1. Dataset. We choose a dataset that covers lyssavirus isolates spatially and geographically over a long period of time in various animal hosts. Fifty-three full-length G sequences

from 21 countries isolated over a period of 70 years were retrieved from NCBI GenBank. The sequences were aligned using fast statistical alignment (FSA, [12]). Briefly, FSA is a probabilistic multiple-sequence alignment algorithm, which uses a "distance-based" approach to aligning homologous protein, RNA, or DNA sequences. It produces superior alignments of homologous sequences that are subject to very different evolutionary constraints. The nucleotide (nt) sequence alignment of the lyssavirus G genes was corrected manually by visual inspection using the amino acid sequence alignment. Gaps were removed if they existed in majority of the sequences.

2.2. Phylogenetic Analyses. A phylogenetic tree was reconstructed by using the neighbor joining algorithm in the MEGA 4 package [13]. The maximum composite likelihood model was used as well as the pairwise deletion option for gaps. The statistical significance of the phylogeny was measured by bootstrap with 1,000 replicates.

2.3. Recombination Detection. We first applied PHI [14], NSS [15], and Max χ^2 [16] tests (implemented in PhiPack [14]) with 1,000 permutations to detect recombination. Sequences involved in the recombination and breakpoints were determined by using 3SEQ [17] and GARD implemented in the Datamonkey web interface [18, 19]. The recombination was further verified by bootscanning and phylogenetic incongruence analysis. Bootscanning was performed using SimPlot software version 3.5.1 [20]. The parameters for bootscanning were window size, 200 bp; step, 10 bp; Gap-Strip, on; bootstrap replicate, 1000; distance model, Kimura (2-parameter); tree algorithm, neighbor-joining.

2.4. Selection Analyses. To test positive selection on sites of the G gene in Lyssaviruses, the Codeml program in PAML software package version 4.4 was employed [21]. Codeml implements the maximum likelihood method to test if positive selection has taken place at sites within a gene. This method uses different codon substitution models to estimate the number of nonsynonymous (dN) and synonymous substitutions (dS) per site among codons, since different amino acids in a protein could be under different selective pressures, thus creating a different ω (dN/dS) ratio. The models in our dataset analyses were M0 (one-ratio), M1 (nearly neutral), M2 (positive selection), M7 (β distribution), and M8 ($\beta + \omega > 1$) [22]. The M0 model estimates overall ω for the data. The M1 model estimates codon site proportion p_0 with $\omega_0 < 1$ and proportion p_1 ($p_1 = 1 - p_0$) with $\omega_1 = 1$. The M2 model allows an additional class of positively selected sites with proportion p_2 ($p_2 = 1 - p_1 - p_0$) with ω_2 estimated from the data. The M7 model specifies that ω follows a beta distribution and the value of ω is allowed to change between 0 and 1. Parameters p and q of the beta distribution are estimated from the data in the M7 model. In the M8 model, a proportion of sites p_0 has a ω in the beta distribution and the proportion p_1 sites are assumed to be positively selected. Two sets of comparisons (M2 versus M1, M8 versus M7) were made to test the hypothesis of selection. Within the comparison, the likelihood ratio test statistic used to determine the level of significance was calculated as twice the difference of the likelihood scores ($2\Delta l$) estimated by each model. The significance was determined under χ^2 distribution. The degrees of freedom for the M1 versus M2 and M7 versus M8 tests are 2 [22]. If M8 or M2 is significantly favored and it contains codons with $\omega > 1$, positive selection is significantly evident. Posterior probabilities of the inferred positively selected sites were estimated by the Bayes empirical Bayes (BEB) approach [23].

We also applied single-likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL), and random-effects likelihood (REL) [18] to indentify selection pressure on individual codons of the G gene in lyssaviruses.

3. Results

3.1. Recombination Analyses. Our dataset covered lyssaviruses isolated over a period of 70 years from 21 countries (Table 1), including the new and old continents. The hosts included bats, cows, dogs, foxes, humans, raccoons, sheep, and skunks.

The PHI and Max χ^2 tests suggested significant evidence of recombination in the G gene. By 1000 permutations, the P-values of PHI and Max χ^2 test were .006 and 0, respectively. However, no significant evidence ($P = .796$) of recombination was detected by using the NSS test.

By using 3SEQ, 6 long recombinant sequences (>100 bp) were detected: AF233275, AY237121, AY987478, DQ074978, DQ849071, and L04523 (Table 2). Two breakpoints were identified in all recombinants. The first breakpoint was at nucleotide position between 400 and 800. The second breakpoint was around nucleotide position of 1080. However, the two breakpoints for DQ074978 and L04523 were at the very beginning and around nucleotide position of 109, respectively.

The analysis by using GARD also suggested evidence of recombination with significant topological incongruence at the 2 breakpoints (Table 3). The first breakpoint was at nucleotide position of 441 and the second was at nucleotide position of 1089. The significance value for the 2 breakpoints was 0.01. The left hand side (LHS) and the right hand side (RHS) P-values for the 2 breakpoints were .0004.

We analyzed the recombination events by using Boot-Scanning as implemented in SimPlot. Sequence AY987478 was used as a query sequence in all four cases (Figures 1(a)–1(d)). The analysis confirmed the recombination event in the G gene of lyssavirus. The high bootstrap values support clustering sequence AY987478 with AF325489 (Figures 1(a) and 1(b)) and with AY237121 (Figures 1(c) and 1(d)) at positions from 1 to around 440 and at positions from around 1130 to the end of the sequences. The bootstrap values are also high for clustering AY987478 with AF23375 (Figures 1(a) and 1(c)) and DQ074978 (Figures 1(b) and 1(d)) at positions from around 540 to 1000. The switches of the high bootstrap values at nucleotide positions from around 440 to 540 and from 1000 to 1130 indicate two possible breakpoints for the recombination.

Table 1: Sequences of glycoprotein gene used in this study.

Accession no.	Country	Host	Year of isolation	Strain/isolate	Genotype	References
AB115921	Indonesia	Dog	2001	SN01-23	GT1	Unpublished
AF233275	India	Sheep		PV11	GT1	Unpublished
AF298141	USA	Bat	1979	USA7-BT	GT1	Badrane et al. [24]
AF298142	Poland	Bat	1985	EBL1POL	GT5	Badrane et al. [24]
AF298143	France	Bat	1989	EBL1FRA	GT5	Badrane et al. [24]
AF298144	Finland	Bat	1986	EBL2FIN	GT6	Badrane et al. [24]
AF298145	Holland	Bat	1986	EBL2HOL	GT6	Badrane et al. [24]
AF298146	S. Africa	Bat	1970	DuvSAF1	GT4	Badrane et al. [24]
AF298147	S. Africa	Bat	1981	DuvSAF2	GT4	Badrane et al. [24]
AF325487	Malaysia	Human	1985	MAL1-HM	GT1	Badrane and Tordo [7]
AF325489	Nepal	Dog	1989	NEP1-DG	GT1	Badrane and Tordo [7]
AF325490	French	Bovine	1985	GUY1-BV	GT1	Badrane and Tordo [7]
AF325491	Brazil	Bovine	1986	BRA1-BV	GT1	Badrane and Tordo [7]
AF325492	Mexico	Bat	1987	MEX2-VP	GT1	Badrane and Tordo [7]
AF325494	USA	Bat	1981	USA8-BT	GT1	Badrane and Tordo [7]
AF325495	USA	Bat	1982	USA9-BT	GT1	Badrane and Tordo [7]
AF401285	Thailand			8743THA	GT1	Unpublished
AF426297	Australia	Bat	1997	ABLSF12NB	GT7	Guyatt et al. [25]
AF426298	Australia	Bat	1997	ABLSF11KW	GT7	Guyatt et al. [25]
AJ871962	China	Vaccine		PM	GT1	Unpublished
AY009098	China	Human	1986	CNX8601	GT1	Tang et al. [26]
AY009099	China	Human	1986	CNX8511	GT1	Tang et al. [26]
AY009100	China	Dog (Vaccine)	1955	CTN	GT1	Tang et al. [26]
AY237121	India	Dog		RVD	GT1	Unpublished
AY257980	Thailand	Human		HM65	GT1	Hemachudha et al. [27]
AY257982	Thailand	Human		HM88	GT1	Hemachudha et al. [27]
AY257983	Thailand	Human		HM208	GT1	Hemachudha et al. [27]
AY987478	India	Dog	1999	CHAND03	GT1	Unpublished
D14873	Japan	Vaccine		RC-HL	GT1	Unpublished
D16330	Japan	Vaccine		RC-HL	GT1	Ito et al. [28]
DQ074978	India	Dog			GT1	Agrawal et al. [29]
DQ076097	S. Korea	Bovine		SKRBV0404HC	GT1	Hyun et al. [30]
DQ076099	S. Korea	Dog		SKRRD9903YG	GT1	Hyun et al. [30]
DQ767897	China	Vaccine		CTN-35	GT1	Unpublished
DQ849071	China	Dog	1994	GX4	GT1	Meng et al. [31]
DQ849072	China	Dog	1992	CQ92	GT1	Meng et al. [31]
L04522	China	Vaccine (Dog)	1931	3aG	GT1	Bai et al. [32]
L04523	China	Vaccine (dog)	1993	CGX89-1	GT1	Bai et al. [32]
L40426				CVS	GT1	Yelverton et al. [33]
M81058	Algeria	Dog		ALG1-DG	GT1	Benmansour et al. [34]
M81059	Algeria	Human			GT1	Benmansour et al. [34]
M81060	Algeria	Human			GT1	Benmansour et al. [34]
U03765	Canada	Vulpes		8480FX	GT1	Nadin-Davis et al. [35]
U03766	Arctic Circle	Dog	1992	Arctic A1-1090DG	GT1	Nadin-Davis et al. [35]
U03767	Canada	Dog	1993	Hudson Bay-4055DG	GT1	Nadin-Davis et al. [35]
U11736	Canada	Arctic Fox		91RABN1035	GT1	Nadin-Davis et al. [36]
U11755	Canada	Skunk		91RABN1578	GT1	Nadin-Davis et al. [36]
U27214	USA	Raccoon		NY 516	GT1	Nadin-Davis et al. [37]
U27215	USA	Raccoon		NY 771	GT1	Nadin-Davis et al. [37]
U27216	USA	Raccoon		FLA 125	GT1	Nadin-Davis et al. [37]
U27217	USA	Raccoon		PA R89	GT1	Nadin-Davis et al. [37]
U52946	USA	Bat	1994	SHBRV	GT1	Morimoto et al. [38]
X69122	India	Vaccine		Flury	GT1	Unpublished

Table 2: Recombination detection in glycoprotein gene of lyssavirus by using 3SEQ.

P	Q	C	*P*-value	Dunn Sidak	Breakpoints	
M81058	AY987478	AF233275	0	$2.08E-08$	432–440, 1080–1089	456–496, 1080–1089
M81060	AY987478	AF233275	$1E-12$	$1.31E-07$	432–440, 1080–1089	456–496, 1080–1089
AY987478	M81059	AY237121	0	$2.81E-11$	441–455, 1077–1079	
AY987478	M81058	AY237121	0	$2.13E-13$	441–455, 1077–1079	
AY987478	M81060	AY237121	0	$1.13E-13$	441–455, 1077–1079	
AY987478	AF233275	AY237121	$1.3E-10$	$1.88E-05$	432–455, 1068–1089	465–518, 1068–1089
AY987478	L04522	AY237121	$1.1E-08$	$1.48E-03$	627–638, 1077–1089	663–666, 1077–1089
AY987478	AF325489	AY237121	0	$2.71E-15$	700-701, 1077–1097	
AY987478	U11755	AY237121	$3.2E-10$	$4.42E-05$	717–719, 1077–1082	729–734, 1077–1082
AY987478	U11736.2	AY237121	$3.3E-09$	$4.61E-04$	717–719, 1077–1082	729–734, 1077–1082
AY987478	DQ849071	AY237121	$6.1E-11$	$8.61E-06$	736-737, 1077–1079	
AY987478	DQ076097	AY237121	$1.2E-10$	$1.69E-05$	630–638, 1077–1089	699–701, 1077–1089
AY987478	DQ076099	AY237121	$9E-12$	$1.31E-06$	700-701, 1077–1089	714–719, 1077–1089
AY987478	L04523	AY237121	$2.2E-09$	$3.04E-04$	736-737, 1077–1079	
AY987478	X69122	AY237121	$4E-12$	$6.00E-07$	666–669, 1032–1049	666–669, 1077–1089
AY987478	AY009098	AY237121	$4E-12$	$4.99E-07$	693–701, 1077–1079	705–711, 1077–1079
AY987478	AY009099	AY237121	$4E-12$	$4.99E-07$	693–701, 1077–1079	705–711, 1077–1079
AY987478	DQ849072	AY237121	$2.1E-11$	$3.02E-06$	693–701, 1077–1079	705–711, 1077–1079
AY987478	AJ871962	AY237121	$1E-12$	$7.29E-08$	750–794, 1077–1089	
AY987478	AF325487	AY237121	0	$1.36E-08$	780–794, 1077–1079	
AY987478	L40426	AY237121	$4.8E-11$	$6.71E-06$	750–794, 1077–1089	
AY987478	AF401285	AY237121	0	$9.99E-10$	780–795, 1077–1079	
AY987478	AY257983	AY237121	$2.3E-11$	$3.27E-06$	780–795, 1077–1079	
AY987478	AY257980	AY237121	0	$5.70E-09$	750–761, 1077–1079	780–795, 1077–1079
AY987478	AY257982	AY237121	$5.9E-11$	$8.33E-06$	780–795, 1032–1043	780–795, 1077–1079
AY987478	DQ767897	AY237121	$1E-07$	$1.46E-02$	759–767, 972–974	
AY987478	U52946	AY237121	$5.3E-08$	$7.43E-03$	741–748, 900-901	741–748, 918–938
AY987478	U03766	AY237121	$2.4E-07$	$3.27E-02$	717–719, 876–889	717–719, 894–914
AY987478	U03765	AY237121	$2.6E-07$	$3.55E-02$	717–719, 876–889	717–719, 894–914
AY237121	AF233275	AY987478	0	$1.38E-39$	432–452, 1077–1089	
AY237121	DQ074978	AY987478	0	$1.16E-38$	432–452, 1077–1089	
AY237121	L04522	AY987478	0	$7.15E-25$	627–647, 1065–1089	
AY237121	DQ076097	AY987478	0	$1.47E-08$	630–647, 1056–1058	630–647, 1065–1089
AY237121	U03767	AY987478	$1E-11$	$1.42E-06$	630–638, 1041–1058	630–638, 1065–1079
AY237121	AJ871962	AY987478	0	$6.36E-20$	642–647, 1041–1058	642–647, 1065–1089
AY237121	X69122	AY987478	0	$3.16E-26$	642–647, 1041–1049	654–659, 1041–1049
AY237121	L40426	AY987478	0	$1.38E-17$	642–647, 1041–1058	642–647, 1065–1089
AY237121	M81058	AY987478	0	$9.60E-21$	441–452, 1065–1079	618–710, 1065–1079
AY237121	M81060	AY987478	0	$6.49E-23$	441–452, 1065–1079	618–710, 1065–1079
AY237121	D14873	AY987478	0	$6.82E-17$	685–701, 1065–1085	705–710, 1065–1085
AY237121	D16330	AY987478	0	$5.23E-17$	685–701, 1065–1085	705–710, 1065–1085
AY237121	AY257980	AY987478	$5.3E-08$	$7.49E-03$	705–710, 1041–1046	
AY237121	DQ849071	AY987478	$1.9E-09$	$2.68E-04$	708–710, 1041–1046	
AY237121	L04523	AY987478	$1.3E-07$	$1.85E-02$	708–710, 1041–1046	
AY237121	DQ076099	AY987478	0	$1.07E-08$	634–647, 1056–1058	634–647, 1065–1089
AY237121	U11755	AY987478	$1E-12$	$1.60E-07$	630–647, 1056–1058	630–647, 1065–1082
AY237121	U11736.2	AY987478	0	$2.29E-08$	630–647, 1056–1058	630–647, 1065–1082
AY237121	DQ767897	AY987478	$1.7E-09$	$2.37E-04$	708–710, 1017–1022	708–710, 1041–1046
AY237121	AY009098	AY987478	$1.8E-08$	$2.58E-03$	705–710, 1041–1046	736-737, 1041–1046
AY237121	AY009099	AY987478	$1.8E-08$	$2.58E-03$	705–710, 1041–1046	736-737, 1041–1046

Table 2: Continued.

P	Q	C	*P*-value	Dunn Sidak	Breakpoints	
AY237121	AF325487	AY987478	$7.5E-10$	$1.06E-04$	705–710, 1041–1046	732–734, 1041–1046
AY237121	U03766	AY987478	$1.2E-09$	$1.75E-04$	630–638, 1041–1058	630–638, 1065–1079
AY237121	U03765	AY987478	$2.3E-10$	$3.26E-05$	630–638, 1041–1058	630–638, 1065–1079
AY237121	M81059	AY987478	0	$1.27E-18$	441–452, 993–998	441–452, 1017–1034
AY237121	AF325490	AY987478	$9.2E-09$	$1.30E-03$	705–710, 993–995	705–710, 1017–1019
AY237121	AF325491	AY987478	$7E-12$	$9.93E-07$	705–710, 993–995	
AY237121	AF325492	AY987478	$9.6E-09$	$1.34E-03$	700-701, 993–995	705–710, 993–995
AY237121	DQ849072	AY987478	$1.3E-07$	$1.83E-02$	705–710, 924–935	705–710, 945–950
AY237121	AY009100	AY987478	$3.4E-07$	$4.62E-02$	708–710, 885–887	708–710, 924–938
AY237121	AF401285	AY987478	$6.4E-09$	$8.95E-04$	736-737, 883–887	
AY237121	AY257983	AY987478	$9.9E-08$	$1.38E-02$	732–734, 883–887	732–734, 1041–1046
M81059	AY987478	DQ074978	0	$9.05E-10$	519–522, 1080–1089	
M81058	AY987478	DQ074978	0	$4.12E-09$	519–522, 1080–1089	
M81060	AY987478	DQ074978	0	$2.80E-08$	519–522, 1080–1089	
AY009100	M81059	DQ849071	$2E-08$	$2.80E-03$	0–3, 108–119	
AY009100	M81058	DQ849071	$3E-08$	$4.20E-03$	0–3, 108–119	
AY009100	M81060	DQ849071	$1.2E-07$	$1.70E-02$	0–3, 108–119	
AY009100	AJ871962	DQ849071	$2.2E-07$	$3.02E-02$	0–3, 108–110	
AY009100	M81059	L04523	$9.6E-09$	$1.34E-03$	0–3, 108–119	
AY009100	M81058	L04523	$1.5E-08$	$2.04E-03$	0–3, 108–119	
AY009100	M81060	L04523	$6.7E-08$	$9.37E-03$	0–3, 108–119	0–3, 139–161
AY009100	AJ871962	L04523	$1.3E-07$	$1.88E-02$	0–3, 108–110	

Note: P and Q are putative parent sequences, and C is the putative child sequence in the recombination.

Table 3: KH tests verify the significance of breakpoints estimated by GARD analysis.

Breakpoint	LHS *P*-value	RHS *P*-value	Significance
441	.00040	.00040	0.01
1089	.00040	.00040	0.01

Since recombination with 2 breakpoints was predicted by 3SEQ, GARD, and Bootscanning, we constructed phylogenetic trees by using sequences from the beginning to the first breakpoint and the sequences from the second breakpoint to the end (Figure 2(a)) and a phylogenetic tree with sequences between the two breakpoints (Figure 2(b)). The reconstructed trees presented conflicting topological positions of the putative recombinant AY987478. The putative recombinant was clustered with AY237121 and AF325489 in Figure 2(a), but clustered with DQ074978 and AF233275 in Figure 2(b). All other 5 putative recombinants did not present phylogenetic incongruence. The same result was also verified by GARD (data not shown). When AY987478 was excluded from the dataset, the *P*-values of Phi, Max χ^2, and NSS were .121, .209, and .791, respectively, suggesting no evidence of recombination. The GARD analysis did not indicate evidence of recombination either.

3.2. Selection Pressure Analyses. The selection pressure analysis with the glycoprotein gene by using PAML is presented in Table 4. The likelihood ratio test statistic ($2\Delta l$) estimated by M2 and M1 was 0. The corresponding *P* value was .99, which is not significant to reject the nearly null hypothesis of neutral selection in M1. In the comparison between the null neutral site model (M7) and the selection model (M8), the $2\Delta l$ was 18.18 and the corresponding *P*-value was .0001, indicating that the positive selection model was significantly favored over the null neutral site model. Posterior probabilities of the inferred positively selected sites estimated by the BEB approach were shown in Table 5. Four amino acid sites at 466, 483, 486, and 490 were identified to be under positive selection. But only the site at position 483 had a marginal significance support with posterior probability of 95% and weak positive selection pressure with ω of 1.466. The corresponding posterior probabilities for sites at 466, 486 and, 490 were 68%, 56%, and 82%, respectively.

To test the effect of recombination on positive selection analysis, we excluded the putative recombinant AY987478 from the dataset. Similar results were observed, and the BEB posterior probability supports for amino acid sites under positive selection were nonsignificant (Table 5). When all six putative recombinants were excluded in our analysis, no evidence was found to support positive selection either in M1 or M7 (data not shown). In all cases, the ω in M0 was either 0.07 or 0.08. Overall, 87% of the sites in the G gene had a very low ω value of 0.05 in M2 and M7, indicating strong selective constraints on those sites.

To study the effect of viral passages and possible genetic bottlenecks on the results, we repeated the analysis with a dataset excluding six vaccine sequences and the sequence

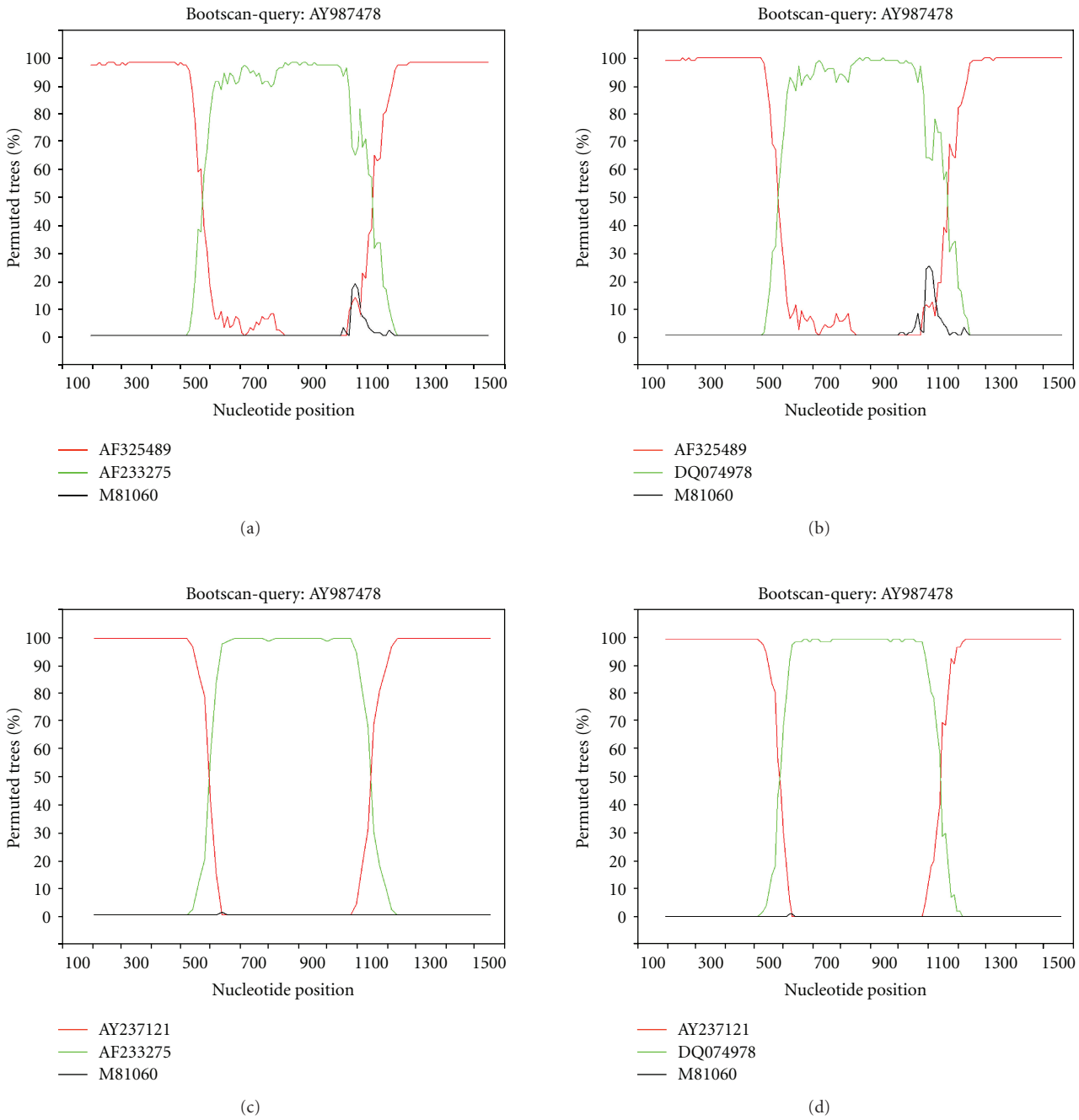

Figure 1: Bootscanning analysis of recombination in glycoprotein gene of lyssavirus by using the SimPlot program with a window size of 200 nucleotides and a step size of 10 nucleotides.

AF233275 (PV11) from cell culture of lyssaviruses under intensive cell culture. We found no significant evidence for positive selection pressure on any site of the G gene.

Analyses using SLAC, REL, and FEL found no evidence of any amino acid in the G gene under positive selection, instead most of the amino acids were found to be under negative selection (Table 6). One site at position 416 was under marginal positive selection by FEL with P-value of .0999, narrowly passing the significance level of 0.1. However, this result was not supported by SLAC and REL.

4. Discussion

Lyssaviruses can infect all warm-blooded mammals, and spillover events and host shift have been well documented [5–9]. The molecular mechanism of rabies infection and transmission is still not completely understood, and the phenomenon usually leads to the connection with rabies virus G protein, since G is the only membrane protein responsible for virus entry both in vitro and in vivo. Therefore, it is a reasonable assumption that rabies virus

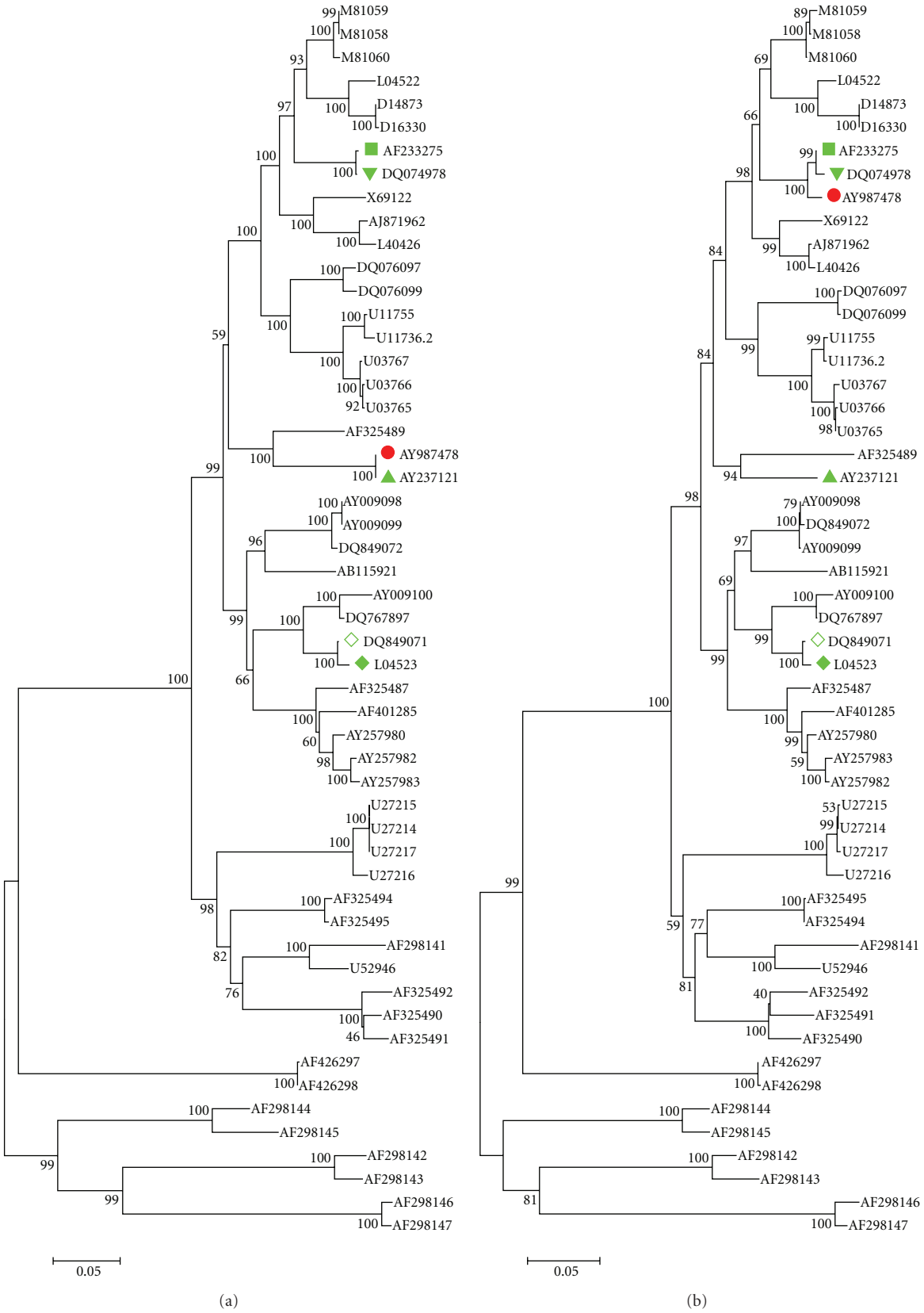

Figure 2: (a) NJ phylogenetic tree of 53 glycoprotein gene sequences with regions concatenated from position of 1 to 441 and position of 1090 to 1572. Bootstrap values of 1000 replicates are shown above the branches. The red marker represents the putative recombinant. (b) NJ phylogenetic tree of 53 glycoprotein gene sequences with region from position of 441 to 1089. Bootstrap values of 1000 replicates are shown above the branches. The red marker represents the putative recombinant.

Table 4: Parameter estimates, dN/dS ratio, likelihood score, and test statistics under models of variable ω ratios among sites for the glycoprotein gene in lyssavirus.

	Parameter estimates	dN/dS	Likelihood scores (l)	Model comparison ($2\Delta l$, d.f., P)	Positive selection
M0: one ratio	$\omega = 0.08$	0.08	−24586.10		None
M1: Nearly neutral	$\omega_0 = 0.05$, $\omega_1 = 1$, ($p_0 = 0.87$, $p_1 = 0.13$)	0.17	−24010.40		Not allowed
M2: Positive selection	$\omega_0 = 0.05$, $\omega 1 = 1$, $\omega_2 = 1$, ($p_0 = 0.87$, $p_1 = 0.06$, $p_2 = 0.07$)	0.17	−24010.40	M2 versus M1:0, d.f. = 2, $P = .99$	None
M7: β, Neutral	$p = 0.26$, $q = 2.11$	0.10	−23443.16		Not allowed
M8: $\beta + \omega > 1$, Selection	$p_0 = 0.98$, $p = 0.28$, $q = 2.92$, ($p_1 = 0.02$), $\omega = 1.0$	0.10	−23434.07	M7 versus M8: 18.18, d.f. = 2, $P = .0001$	See Table 6

Table 5: Positive selection sites in the glycoprotein gene predicted by using Bayes empirical analysis under different PAML models.

Codon		Amino acid		Posterior probability		Post mean ± S.E.	
Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
466	466	A	A	0.68	0.72	1.27 ± 0.35	1.29 ± 0.34
483	483	V	V	0.95	0.84	1.46 ± 0.16	1.39 ± 0.26
486	486	T	T	0.56	0.53	1.19 ± 0.36	1.16 ± 0.36
490	490	Q	Q	0.82	0.80	1.38 ± 0.27	1.36 ± 0.29

Dataset I: The whole 53 nucleotide sequences. Dataset II: AY987478 was excluded.

adaptation is due to the G gene. Positive selection is an important evolutionary force that drives adaptation. It is not surprising that evolutionary scientists first applied selection analysis to the G gene of lyssaviruses [39]. One notable difference between the previous investigations and our study was the dataset. Previous dataset with 55 complete G gene sequences were from isolates of natural rabies infections, excluding passages and vaccine strains. Our dataset included street 53 rabies isolates and vaccine strains collected over a period of 70 years from 21 countries. The neutrality tests on the G in lyssavirus indicated that the protein was under negative selection. Analysis of heterogeneous selective pressures on the amino acid sites across the gene found no evidence for positive selection on any site when the putative recombinant AY987478 was excluded. Instead, most of the sites were under strong negative selection, which was consistent with previous investigations using only street rabies isolates [39, 40]. The only weak positive selection identified by our analyses was at amino acid residue 483 (not in the ectodomain). No positive selection has been detected in the main epitope II or III, the site of virus escape identified by monoclonal antibody binding selections in vitro. It is possible that the results were confounded by the sequences from isolates under intensive cell culture. Repeated passages of an RNA virus resulted in loss of fitness due to Muller's ratchet [41]. Serial virus passages severely reduce population size when a small set of founder population is reintroduced into an identical unpopulated environment, which may lead to the stochastic loss of certain genotypes, especially the rare genotypes [42, 43]. However, exclusion of sequences of passaged lyssaviruses from the dataset in this study did not affect the readout of the analyses. It appears that rabies spillover, host shift (happened naturally), virus escape by monoclonal antibody selection, and vaccine strains (under various in vitro and in vivo conditions) is not the result of positive selection in the G gene.

Recombination is another important evolutionary driving force in adaptation, and it is a mechanism that prevents the accumulation of deleterious substitutions [44]. It allows the acquisition of multiple genetic changes in a single step and can combine genetic information to produce advantageous genotypes. It may be important for incremental host adaptation after switching to new host has occurred [45]. Recombination in rabies viruses had been proposed, but it was not thoroughly inspected [46, 47]. Our study suggested one recombinant event. The recombinant sequence AY987478 was from a dog isolate (CHAND03, genotype 1) and the possible parental sequences were isolated from dogs and sheep from the same geographic area (India and Nepal). However, the putative recombinant AY987478 could be an artifact from sequencing or sample contamination. Generation of recombinants in the course of reverse transcription of RNA and subsequent PCR is a well-known phenomenon [48–50]. From the bootscanning analysis in this research, the 3 prime and 5 prime regions of AY987478 were clustered with putative parents with a bootstrap value of 100%, indicating little difference between the two sequences in the two regions. By checking the sequences, there are regions of about 450 bases long that are identical between the recombinant and the corresponding parent, which is

Table 6: Detection of selection pressure on glycoprotein gene using methods implemented in the Datamonkey website.

Dataset	Mean *dN/dS*			Positive selection sites			Negative selection sites			Codon (*P*-Value)
	SLAC	FEL	REL	SLAC	FEL	REL	SLAC	FEL	REL	
Dataset I	0.1226		0.1278	0	0	0	397	418	0	
Dataset II	0.1231		0.1274	0	0	0	391	417	0	
Dataset III	0.1214		0.1233	0	1	0	386	416	0	416 (.0999)

Dataset I: The whole 53 nucleotide sequences. Dataset II: AY987478 was excluded. Dataset III: the six putative recombinants were excluded.

rare considering the high mutation rate in RNA viruses. The homologous recombination rate in negative-sense RNA virus was found to be low [46], which is supported by a recent report that homologous recombination is very rare or absent in influenza A virus [17]. Further experimentation is needed to prove that the recombinant AY987478 is not an artifact.

In summary, we did not find significant support for positive selection pressure on G gene in lyssavirus isolates from different rabies hosts and vaccine strains that cover 70 years of evolution in 21 countries. The recombination analysis suggested an orphan event that needs further investigation. It appears that evolution of the G gene may not play a major role in lyssavirus adaptation. It is surprising considering the functions of glycoprotein in lyssavirus infection. It has been reported that host switching from chiropters to carnivores has occurred in lyssavirus evolution history [7, 9]. Spillovers of lyssaviruses from chiropters to other animals may have happened repeatedly and still occur [8]. Transmission of European bt lyssavirus 1 (EBLV-1) was reported in sheep [51], stone marten [52], and cats [53]. For a successful spillover and subsequent adaptation, there must be effective cross-species viral exposure and compatibility between the virus and the new host to allow replication and transmission. Lyssavirus infections are typically transmitted by the virus-laden saliva of a rabid animal via a bite or scratch, which can facilitate cross-species viral exposures. The initial viral interaction with cells of a new host plays a critical role in determining host specificity and host shift [45]. For example, feline virus acquired the ability to infect dogs through changes in its capsid protein that binds to canine transferrin receptor on canine cells [54]. Lyssavirus G is a surface glycoprotein responsible for receptor recognition and membrane fusion [7–9, 55]. It is reasonable to expect that the protein is under positive selection pressure in the viral adaptation to the new host. The lack of positive selection in the G glycoprotein suggests that the virus is not subject to strong immune selection [25]. The G gene may escape the immunity of the host since lyssaviruses migrate from the peripheral to the central nervous systems [7]. Recent investigation demonstrated that diminishing frequencies of both cross-species transmission and host shifts were found with increasing phylogenetic distance between bat species [9], indicating the virus, thus the G gene, is subject to less selection pressure in a similar host and cellular environment [7, 25]. However, the G gene might have been under relative low positive selection that was not detected by current computational methods. More sensitive method or properly relaxed statistical significance stringency with experimental verification may help identify the role of the G gene in lyssavirus adaptation.

Acknowledgments

The authors thank Jan Pohl, Elizabeth Neuhaus, and Charles E. Rupprecht for support in this investigation. They also thank Kathryn Kellar and Scott Sammons for helpful suggestions to the paper. Use of trade names and commercial sources are for identification only and do not imply endorsement by the U S Department of Health and Human Services. The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the funding agency.

References

[1] T. Lefébure and M. J. Stanhope, "Evolution of the core and pan-genome of Streptococcus: positive selection, recombination, and genome composition," *Genome Biology*, vol. 8, no. 5, pp. R71.1–R71.16, 2007.

[2] D. A. Steinhauer, E. Domingo, and J. J. Holland, "Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase," *Gene*, vol. 122, no. 2, pp. 281–288, 1992.

[3] K. Kirkegaard and D. Baltimore, "The mechanism of RNA recombination in poliovirus," *Cell*, vol. 47, no. 3, pp. 433–443, 1986.

[4] M. Worobey and E. C. Holmes, "Evolutionary aspects of recombination in RNA viruses," *Journal of General Virology*, vol. 80, no. 10, pp. 2535–2543, 1999.

[5] D. M. Pfukenyi, D. Pawandiwa, P. V. Makaya, and U. Ushewokunze-Obatolu, "A retrospective study of wildlife rabies in Zimbabwe," *Tropical Animal Health and Production*, vol. 41, no. 4, pp. 565–572, 2009.

[6] A. I. Wandeler, S. A. Nadin-Davis, R. R. Tinline, and C. E. Rupprecht, "Rabies epidemiology: some ecological and evolutionary perspectives," in *Lyssaviruses*, C. E. Rupprecht, B. Dietzchold, and H. Koprowski, Eds., pp. 297–324, Springer, Berlin, Germany, 1994.

[7] H. Badrane and N. Tordo, "Host switching in Lyssavirus history from the chiroptera to the carnivora orders," *Journal of Virology*, vol. 75, no. 17, pp. 8096–8104, 2001.

[8] L. K. Crawford-Miksza, D. A. Wadford, and D. P. Schnurr, "Molecular epidemiology of enzootic rabies in California," *Journal of Clinical Virology*, vol. 14, no. 3, pp. 207–219, 1999.

[9] D. G. Streicker, A. S. Turmelle, M. J. Vonhof, I. V. Kuzmin, G. F. McCracken, and C. E. Rupprecht, "Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats," *Science*, vol. 329, no. 5992, pp. 676–679, 2010.

[10] B. Dietzschold, W. H. Wunner, T. J. Wiktor et al., "Characterization of an antigenic determinant of the glycoprotein that correlates with pathogenicity of rabies virus," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 80, no. 1, pp. 70–74, 1983.

[11] X. Yan, P. S. Mohankumar, B. Dietzschold, M. J. Schnell, and Z. F. Fu, "The rabies virus glycoprotein determines the distribution of different rabies virus strains in the brain," *Journal of Neurovirology*, vol. 8, no. 4, pp. 345–352, 2002.

[12] R. K. Bradley, A. Roberts, M. Smoot et al., "Fast statistical alignment," *PLos Computational Biology*, vol. 5, no. 5, Article ID e1000392, 2009.

[13] K. Tamura, J. Dudley, M. Nei, and S. Kumar, "MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1596–1599, 2007.

[14] T. C. Bruen, H. Philippe, and D. Bryant, "A simple and robust statistical test for detecting the presence of recombination," *Genetics*, vol. 172, no. 4, pp. 2665–2681, 2006.

[15] I. B. Jakobsen and S. Easteal, "A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences," *Computer Applications in the Biosciences*, vol. 12, no. 4, pp. 291–295, 1996.

[16] J. M. Smith, "Analyzing the mosaic structure of genes," *Journal of Molecular Evolution*, vol. 34, no. 2, pp. 126–129, 1992.

[17] M. F. Boni, D. Posada, and M. W. Feldman, "An exact nonparametric method for inferring mosaic structure in sequence triplets," *Genetics*, vol. 176, no. 2, pp. 1035–1047, 2007.

[18] S. L. Kosakovsky Pond, D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. W. Frost, "GARD: a genetic algorithm for recombination detection," *Bioinformatics*, vol. 22, no. 24, pp. 3096–3098, 2006.

[19] S. L. Kosakovsky Pond and S. D. W. Frost, "Datamonkey: rapid detection of selective pressure on individual sites of codon alignments," *Bioinformatics*, vol. 21, no. 10, pp. 2531–2533, 2005.

[20] K. S. Lole, R. C. Bollinger, R. S. Paranjape et al., "Full-length human immunodeficiency virus type 1 genomes from subtype C- infected seroconverters in India, with evidence of intersubtype recombination," *Journal of Virology*, vol. 73, no. 1, pp. 152–160, 1999.

[21] Z. Yang, "PAML 4: phylogenetic analysis by maximum likelihood," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1586–1591, 2007.

[22] Z. Yang, R. Nielsen, N. Goldman, and A. M. K. Pedersen, "Codon-substitution models for heterogeneous selection pressure at amino acid sites," *Genetics*, vol. 155, no. 1, pp. 431–449, 2000.

[23] Z. Yang, W. S. W. Wong, and R. Nielsen, "Bayes empirical Bayes inference of amino acid sites under positive selection," *Molecular Biology and Evolution*, vol. 22, no. 4, pp. 1107–1118, 2005.

[24] H. Badrane, C. Bahloul, P. Perrin, and N. Tordo, "Evidence of two Lyssavirus phylogroups with distinct pathogenicity and immunogenicity," *Journal of Virology*, vol. 75, no. 17, pp. 3268–3276, 2001.

[25] K. J. Guyatt, J. Twin, P. Davis et al., "A molecular epidemiology study of Australian bat lyssavirus," *Journal of General Virology*, vol. 84, no. 2, pp. 485–496, 2003.

[26] Q. Tang, L. A. Orciari, C.E. Rupprecht, and X. Zhao, "Sequencing and positional analysis of the glycoprotein gene of four Chinese rabies viruses," *Zhongguo Bingduxue*, vol. 15, no. 1, pp. 22–33, 2000 (Chinese).

[27] T. Hemachudha, S. Wacharapluesadee, B. Lumlertdaecha et al., "Sequence analysis of rabies virus in humans exhibiting encephalitic or paralytic rabies," *Journal of Infectious Diseases*, vol. 188, no. 7, pp. 960–966, 2003.

[28] H. Ito, N. Minamoto, T. Watanabe et al., "A unique mutation of glycoprotein gene of the attenuated RC-HL strain of rabies virus, a seed virus used for production of animal vaccine in Japan," *Microbiology and Immunology*, vol. 38, no. 6, pp. 479–482, 1994.

[29] S. Agrawal, A. K. Shasany, and S. P. S. Khanuja, "Plant transformation vectors having street rabies virus (Indian strain) glycoprotein gene," *Journal of Plant Biochemistry and Biotechnology*, vol. 14, no. 2, pp. 81–87, 2005.

[30] B. H. Hyun, K. K. Lee, IN. J. Kim et al., "Molecular epidemiology of rabies virus isolates from South Korea," *Virus Research*, vol. 114, no. 1-2, pp. 113–125, 2005.

[31] S. L. Meng, J. X. Yan, GE. L. Xu et al., "A molecular epidemiological study targeting the glycoprotein gene of rabies virus isolates from China," *Virus Research*, vol. 124, no. 1-2, pp. 125–138, 2007.

[32] X. Bai, C. K. Warner, and M. Fekadu, "Comparisons of nucleotide and deduced amino acid sequences of the glycoprotein genes of a Chinese street strain (CGX89-1) and a Chinese vaccine strain (3aG) of rabies virus," *Virus Research*, vol. 27, no. 2, pp. 101–112, 1993.

[33] E. Yelverton, S. Norton, J. F. Obijeski, and D. V. Goeddel, "Rabies virus glycoprotein analogs: biosynthesis in Escherichia coli," *Science*, vol. 219, no. 4585, pp. 614–619, 1983.

[34] A. Benmansour, M. Brahimi, C. Tuffereau, P. Coulon, F. Lafay, and A. Flamand, "Rapid sequence evolution of street rabies glycoprotein is related to the highly heterogeneous nature of the viral population," *Virology*, vol. 187, no. 1, pp. 33–45, 1992.

[35] S. A. Nadin-Davis, G. Allen Casey, and A. I. Wandeler, "A molecular epidemiological study of rabies virus in central Ontario and western Quebec," *Journal of General Virology*, vol. 75, no. 10, pp. 2575–2583, 1994.

[36] S. A. Nadin-Davis, M. I. Sampath, G. A. Casey, R. R. Tinline, and A. I. Wandeler, "Phylogeographic patterns exhibited by Ontario rabies virus variants," *Epidemiology and Infection*, vol. 123, no. 2, pp. 325–336, 1999.

[37] S. A. Nadin-Davis, W. Huang, and A. I. Wandeler, "The design of strain-specific polymerase chain reactions for discrimination of the racoon rabies virus strain fron indigenous rabies viruses of Ontario," *Journal of Virological Methods*, vol. 57, no. 1, pp. 1–14, 1996.

[38] K. Morimoto, M. Patel, S. Corisdeo et al., "Characterization of a unique variant of bat rabies virus responsible for newly emerging human cases in North America," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 11, pp. 5653–5658, 1996.

[39] E. C. Holmes, C. H. Woelk, R. Kassis, and H. Bourhy, "Genetic constraints and the adaptive evolution of rabies virus in nature," *Virology*, vol. 292, no. 2, pp. 247–257, 2002.

[40] H. Bourhy, J. M. Reynes, E. J. Dunham et al., "The origin and phylogeography of dog rabies virus," *Journal of General Virology*, vol. 89, no. 11, pp. 2673–2681, 2008.

[41] D. K. Clarke, E. A. Duarte, A. Moya, S. F. Elena, E. Domingo, and J. Holland, "Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses," *Journal of Virology*, vol. 67, no. 1, pp. 222–228, 1993.

[42] M. Oberle, O. Balmer, R. Brun, and I. Roditi, "Bottlenecks and the maintenance of minor genotypes during the life cycle of Trypanosoma brucei," *PLos Pathogens*, vol. 6, no. 7, Article ID e1001023, 2010.

[43] L. M. Wahl, P. J. Gerrish, and I. Saika-Voivod, "Evaluating the impact of population bottlenecks in experimental evolution," *Genetics*, vol. 162, no. 2, pp. 961–971, 2002.
[44] M. Poss, A. Idoine, H. A. Ross, J. A. Terwee, S. VandeWoude, and A. Rodrigo, "Recombination in feline lentiviral genomes during experimental cross-species infection," *Virology*, vol. 359, no. 1, pp. 146–151, 2007.
[45] C. R. Parrish, E. C. Holmes, D. M. Morens et al., "Cross-species virus transmission and the emergence of new epidemic diseases," *Microbiology and Molecular Biology Reviews*, vol. 72, no. 3, pp. 457–470, 2008.
[46] E. R. Chare, E. A. Gould, and E. C. Holmes, "Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses," *Journal of General Virology*, vol. 84, no. 10, pp. 2691–2703, 2003.
[47] L. Geue, S. Schares, C. Schnick et al., "Genetic characterisation of attenuated SAD rabies virus strains used for oral vaccination of wildlife," *Vaccine*, vol. 26, no. 26, pp. 3227–3235, 2008.
[48] A. Flockerzi, J. Maydt, O. Frank et al., "Expression pattern analysis of transcribed HERV sequences is complicated by ex vivo recombination," *Retrovirology*, vol. 4, no. 39, pp. 1–12, 2007.
[49] G. Luo and J. Taylor, "Template switching by reverse transcriptase during DNA synthesis," *Journal of Virology*, vol. 64, no. 9, pp. 4321–4328, 1990.
[50] A. Meyerhans, J. P. Vartanian, and S. Wain-Hobson, "DNA recombination during PCR," *Nucleic Acids Research*, vol. 18, no. 7, pp. 1687–1691, 1990.
[51] H. Bourhy, L. Dacheux, C. Strady, and A. Mailles, "Rabies in Europe in 2005," *Euro Surveillanc*, vol. 10, no. 11, pp. 213–216, 2005.
[52] K. Tjørnehøj, A. R. Fooks, J. S. Agerholm, and L. Rønsholt, "Natural and experimental infection of sheep with european bat lyssavirus type-1 of danish bat origin," *Journal of Comparative Pathology*, vol. 134, no. 2-3, pp. 190–201, 2006.
[53] L. Dacheux, F. Larrous, A. Mailles et al., "European bat lyssavirus transmission among cats, Europe," *Emerging Infectious Diseases*, vol. 15, no. 2, pp. 280–284, 2009.
[54] K. Hueffer, J. S. L. Parker, W. S. Weichert, R. E. Geisel, J. Y. Sgro, and C. R. Parrish, "The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor," *Journal of Virology*, vol. 77, no. 3, pp. 1718–1726, 2003.
[55] P. Durrer, Y. Gaudin, R. W. H. Ruigrok, R. Graf, and J. Brunner, "Photolabeling identifies a putative fusion domain in the envelope glycoprotein of rabies and vesicular stomatitis viruses," *Journal of Biological Chemistry*, vol. 270, no. 29, pp. 17575–17581, 1995.