Published in final edited form as:

Infect Genet Evol. 2009 September ; 9(5): 827–831. doi:10.1016/j.meegid.2009.04.016.

Relaxation of Purifying Selection on the SAD Lineage of Live Attenuated Oral Vaccines for Rabies Virus

Austin L. Hughes

Department of Biological Sciences, University of South Carolina, Columbia SC 29205

Abstract

Analysis of patterns of nucleotide sequence diversity in wild-type rabies virus (RABV) genomes and in the SAD live attenuated oral vaccine lineage was used to test for the relaxation of purifying selection in the latter and provide evidence regarding the genomic regions where such relaxation of selection occurs. The wild-type sequences showed evidence of strong past and ongoing purifying selection both on non-synonymous sites in coding regions and on non-coding regions, particularly the start and end and 5' UTR regions. SAD vaccine sequences showed a relaxation of purifying selection at nonsynonymous sites in coding regions, resulting a substantial number of amino acid sequence polymorphisms at sites that were invariant in the wild-type sequences. Moreover, SAD vaccine sequences showed high levels of mutation accumulation in the non-coding regions that were most conserved in the wild-type sequences. Understanding the biological effects of the unique mutations accumulated in the vaccine lineage is important because of their potential effects on antigenicity and effectiveness of the vaccine.

Rabies virus (RABV), which causes an acute and generally fatal illness of humans and domestic animals, is the oldest recorded human viral pathogen and remains an important source of mortality in many parts of the world (Fu 1997; Warrell and Warrell 2004). In Europe, while control measures have greatly reduced human cases of rabies, wild mammals, particularly the red fox (*Vulpes vulpes*), continue to provide a reservoir of rabies infection (Lontai 1997). As a consequence, the main control strategy has involved oral vaccination of foxes through baits treated with an attenuated live vaccine (Geue et al. 2008). All of the commercially available oral rabies vaccines for wildlife are derived from a single common ancestral isolate known as Street Alabama Dufferin (SAD), derived from an infected dog in North America in 1935 (Fenje 1960; Sacramento et al. 1992).

A number of different vaccine strains have been developed from the original SAD strain by passaging under a variety of protocols (Geue et al. 2008). As in other live attenuated virus vaccines, mutations during passaging have resulted in a certain degree of nucleotide sequence polymorphism within the SAD lineage (Geue et al. 2008). Understanding the evolution of live vaccines is important because evolutionary changes may affect the biological properties of the vaccine, including its antigenicity and effectiveness, as well as its potential to reacquire virulence (Hughes 2009). For example, attenuated live vaccines may show a pattern of relaxation of purifying selection, as was observed in three live attenuated vaccine strains of viruses belonging to the family *Paramyxoviridae* (Hughes 2009). In the case of the SAD

Austin L. Hughes, Ph.D., Department of Biological Sciences, University of South Carolina, Coker Life Sciences Bldg., 700 Sumter St., Columbia SC 29208 USA, Tel: 1-803-777-9186, Fax: 1-803-777-4002, Email: E-mail: austin@biol.sc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

vaccine strains, there have been reports of SAD vaccine-associated rabies in wildlife, in spite of evidence of the stability of the vaccine in experimental passaging studies (Beckert et al. 2009).

Purifying selection – that is, selection against deleterious variants – is the most prevalent form of natural selection at the molecular level (Nei 1987; Hughes 1999, 2008). Evidence for the prevalence of purifying selection on coding regions is provided by the observation that the number of synonymous substitutions per synonymous site (d_S) generally exceeds the number of nonsynonymous (amino acid-altering) substitutions per nonsynonymous site (d_N). This pattern evidently occurs because most nonsynonymous mutations are deleterious to protein function and thus are eliminated by purifying selection. In fact, there are two distinct aspects of purifying selection, revealed by different methods of analysis: (1) selection against strongly deleterious mutations, which are eliminated quickly; and (2) selection against slightly deleterious mutations, which may be relatively inefficient if the effective population size is small or recombination is limited (Ohta 1973; Hughes 2008).

RABV is a single stranded RNA negative-strand virus belonging to the genus *Lyssavirus* (family *Rhabdoviridae*). As in other members of the family, there are five genes, designated *N*, *P*, *M*, *G*, and *L*, each including a 5' untranslated region (UTR) and 3' UTR (Conzelmann 1998; Tordo et al. 1986). These are illustrated in Figure 1 according to their order in the RNA antigenome (Conzelmann 1998). There are also start, end, and intergenic regions; the longest intergenic region (between *G* and *L*) may represent the location of a remnant gene that has lost functionality (Tordo et al. 1986).

The purpose of the present paper is to use the tools of molecular evolutionary genetics to understand the evolutionary factors at work in the evolution of the SAD rabies vaccine lineage, in particular the role of purifying selection. By examining the patterns of nucleotide sequence polymorphism in wild-type genomic sequences, I provide evidence for the strength of both past and ongoing purifying selection on different genomic regions. By comparing these patterns with those seen in the SAD vaccine and in SAD vaccine-derived sequences isolated from red foxes, I both test for the relaxation of purifying selection in the latter and provide evidence regarding the genomic regions where such relaxation of selection occurs.

Methods

Sequence analyses were based on 35 rabies virus (RABV) sequences, 13 wild-type; 14 from SAD vaccine strains; and 6 vaccine-derived strains isolated from red foxes in Europe (for accession numbers, see figure 2). Genomic sequences were aligned using CLUSTALW (Thompson et al. 1997). In computing pairwise distances among a set of sequences, any site at which the alignment postulated a gap in any of the sequences was excluded from all comparisons ("complete deletion," Nei and Kumar 2000). Complete deletion ensures that the same set of sites is compared in each pairwise comparison (Nei and Kumar 2000).

Phylogenetic analyses were conducted including the above-mentioned 35 sequences plus two other sequences belonging to the attenuated strains Ni-CE, RC-HL, and RV-97 (for accession numbers, see Figure 2). A number of different methods of phylogenetic analysis were applied to both complete genome sequences and to coding regions only, all of which showed similar results (not shown). The phylogenetic tree shown below was constructed by the neighbor-joining (NJ) method (Saitou and Nei 1987) on the basis of the maximum composite likelihood (MCL) distance (Tamura et al. 2007). The reliability of clustering patterns in the tree was assessed by bootstrapping (Felsenstein 1985); 1000 bootstrap samples were used. The tree was rooted following the phylogenetic analyses of Wu et al. (2007). Ancestral genomic sequences

were reconstructed based on the NJ tree by the maximum parsimony method (Swofford 2003).

The nucleotide usage in the RABV sequences analyzed here was remarkably even. At third positions in coding regions, the G+C content was 49.5%, while in non-coding regions it was 41.1%. By the MCL method, the transition:transversion ratio at non-coding sites and at third positions of codons was estimated at 5.4:1. Because of this strong transitional bias, the number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous (amino acid-altering) substitutions per nonsynonymous site (d_N) were estimated by Li's (1993) method, which takes into account transitional bias in affecting the relative probabilities of synonymous and nonsynonymous mutations at two-fold and three-fold degenerate sites. The number of nucleotide substitutions per site (d) in non-coding regions was estimated by the MCL method. For coding regions, we computed the mean of d_S for all pairwise comparisons (termed synonymous nucleotide diversity and symbolized π_S) and the mean of d_N for all pairwise comparisons (termed nonsynonymous nucleotide diversity and symbolized π_N). Likewise, in non-coding regions, we computed mean d for all pairwise comparisons (nucleotide diversity or π). Standard errors of π_S , π_N , and π were computed by the bootstrap method (Nei and Kumar 2000); 1000 bootstrap samples were used. Separate estimates of π_S and π_N were obtained for each of the five protein-coding genes (Figure 1), as well as for the combined coding regions. In the case of non-coding regions, π was estimated separately for four separate categories of sites: (1) the start and end sites (upstream to the 5' UTR of *N* and downstream to the 3' UTR of *L*); (2) the 5' UTR regions of the protein-coding genes; (3) the 3' UTR regions of the protein-coding genes; and (4) the intergenic regions (Figure 1).

Gene diversity ("heterozygosity") at individual polymorphic nucleotide sites (SNP sites) was estimated by the formula:

$$1 - \sum_{i=1}^n x_i^2$$

where n is the number of alleles and x_i is the population frequency of the i^{th} allele (Nei 1987, p. 177). In coding regions, single nucleotide polymorphisms were classified either as synonymous or nonsynonymous depending on their effect of the encoded nucleotide sequence. Ambiguous sites were excluded from these analyses; i.e., sites at which both synonymous and nonsynonymous variants occurred or at which the polymorphism could be considered synonymous or nonsynonymous depending on the pathway taken by evolution.

I computed π_S , π_N , and π as well as gene diversity at polymorphic sites in both coding and non-coding regions separately for the following sets of sequences: (1) the wild-type isolates ($N = 13$), not including SAD or any other vaccine strain or vaccine-derived sequences; (2) the SAD vaccine sequences ($N = 14$); and (3) the SAD vaccine-derived sequences ($N = 6$). In testing for differences in gene diversity among categories (synonymous, nonsynonymous, and non-coding) of polymorphic sites, randomization tests were used. In each test, 1000 pseudo-data sets were created by sampling (with replacement) from the data; a difference between two categories was considered significant at the α level if it was greater than the absolute value of 100(1- α) % of the differences observed between the same categories in the pseudo-data sets.

Results

Phylogenetic Analysis

In the NJ tree of complete rabies virus genomes, the SAD lineage formed a separate clade, with 100% bootstrap support, which was distinct from the other attenuated RABV strains (Figure 2). The SAD vaccine-derived wild isolates formed part of the clade with the SAD vaccine strains (Figure 2). (Topology only of the same tree is illustrated in Supplementary Figure S1). There were much shorter branch length within the SAD clade than in the rest of the tree (Figure 2), reflecting a lower level of sequence divergence within the SAD clade. The phylogenetic tree provided strong support for the hypothesis that SAD vaccine-derived wild strains originated independently at least twice. The vaccine-derived wild sequences EU886631 and EU886636 clustered with the SAD P5/88(pro) vaccine sequence EF206715 (the SAD P5/88 vaccine strain; Geue et al. 2008) with 100% bootstrap support (Supplementary Figure S1). Other vaccine and vaccine-derived sequences fell outside the latter cluster (Supplementary Figure S1).

Nucleotide Diversity

In wild-type RABV sequences, π_S was significantly greater than π_N in each of the five protein-coding genes (Table 1). By contrast, in the case of SAD vaccine genomes, π_S was not significantly greater π_N in any of the five genes or in the combined coding regions of all five genes (Table 1). In the case of the SAD vaccine-derived genomes, π_S was significantly greater than π_N in the *L* gene but not in any of the other genes or in the combined coding regions of all five genes (Table 1). For each gene, the ratio of $\pi_N:\pi_S$ was much higher in SAD vaccine genomes than in the wild type (Table 1), and this ratio was over 1.0 in the *G* gene and close to 1.0 in the *M* gene (Table 1). For the five genes, mean $\pi_N:\pi_S$ for SAD vaccine genomes (0.77 ± 0.11) was significantly greater than that for the wild-type (0.06 ± 0.02) by a paired t-test ($P = 0.002$). Although the mean $\pi_N:\pi_S$ for SAD vaccine-derived sequences (0.27 ± 0.12) was greater than that for wild-type sequences, the difference was not significant. These results thus support the hypothesis that the five protein-coding genes of RABV have been subject to strong purifying selection on the encoded amino acid sequences but that there has been relaxation of this selection in the SAD vaccine genomes.

In wild-type RABV, π in non-coding regions was significantly lower than π_S at synonymous sites in all coding regions (Table 2). The overall π in non-coding regions (0.3160) was just 43.1% of overall π_S (0.7330; Tables 1 and 2). This pattern indicates purifying selection on non-coding regions. This purifying selection appeared to be particularly strong in the case of 5' UTR, which showed a significantly lower π than did either the 3' UTR or intergenic regions (Table 2). In the case of SAD vaccine genomes, on the other hand, the start and end regions, the 5' UTR, the 3' UTR, and the intergenic regions all showed values of π that were not significantly different from π_S in all coding regions (Table 2). In fact, in the case of the SAD vaccine genomes, π in all non-coding regions was significantly greater than π_S in all coding regions, a reversal of the pattern seen in wild-type RABV (Table 2). In the SAD vaccine-derived genomes, there were no polymorphisms in the 5'UTR and 3'UTR, resulting in significantly lower π in these regions than π_S in all coding regions (Table 2). However, overall π in all non-coding regions of the SAD vaccine-derived genomes was not significantly different from overall π_S (Table 2).

Polymorphic Sites

Patterns of gene diversity were examined in order to test for ongoing purifying selection at polymorphic sites in coding and non-coding regions. In the wild-type sample, mean gene diversity at polymorphic nonsynonymous sites (0.336) was significantly less than that at either polymorphic synonymous sites (0.369) or polymorphic non-coding sites (0.356; Figure 3A).

In the SAD vaccine sequences, there was no significant difference in mean gene diversity among nonsynonymous (0.282), synonymous (0.285), and noncoding (0.299) polymorphic sites (Figure 3B). Likewise, in the SAD vaccine-derived sequences, there was no significant difference in mean gene diversity among nonsynonymous (0.333), synonymous (0.328), and noncoding (0.389) polymorphic sites, although in this case the numbers of polymorphic sites in the three categories (6, 10, and 3, respectively) were very small.

There were 4218 sites in coding regions that were polymorphic in the wild-type sequences but not in the SAD vaccine sequences. Of these, 1862 (44.1 %) were nonsynonymous. By contrast, of 65 sites in coding regions that were polymorphic in SAD but not in the wild-type sequences, 48 (73.8 %) were nonsynonymous. The difference between proportions was highly significant ($\chi^2 = 22.9$; 1 d.f.; $P < 0.001$). Likewise, of 50 sites in coding regions that were polymorphic in both wild-type and SAD, only 12 (24.0 %) were nonsynonymous. This pattern also differed significantly from that seen at sites polymorphic only in the wild-type sequences ($\chi^2 = 8.1$; 1 d.f.; $P < 0.001$) and from that seen at sites polymorphic only in SAD ($\chi^2 = 28.1$; 1 d.f.; $P < 0.001$).

Polymorphism vs. Divergence

Maximum parsimony was used to reconstruct the nucleotide sequence changes along the branch ancestral to the SAD clade (Figure 2). Along this branch, there were 263 (64.3%) synonymous changes; 76 (18.6%) nonsynonymous changes; and 70 (17.1%) changes in noncoding sites. Along this branch, d_S was 0.0847 ± 0.0058 , d_N was 0.0107 ± 0.0010 ; and d in non-coding regions was 0.0653 ± 0.0081 . There was a significant difference between d_S and d_N ($P < 0.001$), but not between d_S and d in non-coding regions. Thus, there was evidence along the branch ancestral to the SAD clade of strong purifying selection on the RABV coding regions, but a relaxation of purifying selection on non-coding regions.

Within the population of SAD vaccine sequences, there were 36 (29.8%) synonymous polymorphic sites; 58 (47.9%) nonsynonymous polymorphic sites; and 27 (22.3%) noncoding polymorphic sites. These proportions differed strikingly from those of changes along the branch ancestral to the SAD clade, and the difference in proportions was highly significant ($\chi^2 = 53.0$; 2 d.f.; $P < 0.001$). This difference was due mainly to the great excess of nonsynonymous polymorphisms (47.9%) relative to nonsynonymous divergent changes (18.6%).

Discussion

Examination of the patterns of nucleotide sequence polymorphism was used to test for evidence for the relaxation of purifying selection on the SAD strain of live attenuated RABV vaccine, and to identify the regions of the genome where such relaxation is most apparent. Wild-type RABV sequences showed strong evidence of both past and ongoing purifying selection. This selection was focused mainly on nonsynonymous sites in the five protein-coding genes and on sites in the non-coding portions of the genome, particularly the 5' UTR of genes.

In wild-type genomes, the nonsynonymous nucleotide diversity (π_N) was significantly lower than synonymous nucleotide diversity (π_S) in each of the genes, and π_N was overall only 5% as great as π_S (Table 1). These results support the hypothesis that purifying selection has acted to eliminate a substantial fraction of nonsynonymous mutations occurring in RABV genomes. Nucleotide diversity (π) in non-coding regions was also significantly reduced compared to π_S , although the reduction was not so great as in the case of nonsynonymous sites, indicating that noncoding sites have also been subject to purifying selection. The nucleotide diversity was lowest in the 5' UTR of genes and significantly less than in either the 3' UTR or the intergenic regions (Table 2), indicating that the 5' UTR have been the focus of particularly strong past

purifying selection. In addition, mean gene diversity was significantly reduced at nonsynonymous sites in comparison to both synonymous and non-coding sites, implying that ongoing purifying selection is acting to reduce the frequency of certain nonsynonymous variants. The evidence for both past and ongoing purifying selection on RABV is consistent with evidence from other RNA viruses, which in general are subject to highly effective purifying selection (Hughes and Hughes 2007).

SAD vaccine sequences showed a strikingly different pattern of nucleotide sequence polymorphism, indicating an overall relaxation of purifying selection. None of the genes showed π_N significantly less than π_S (Table 1), and none of the non-coding regions showed π significantly less than π_S (Table 2). In each of the five genes, the ratio $\pi_N:\pi_S$ was substantially greater in the SAD vaccine sequences than in the wild-type in the SAD vaccine sequences; π_S and π_N were nearly identical in two of the genes (*M* and *G*), with π_N actually slightly higher than π_S in *G* (Table 1).

The level of polymorphism within the SAD vaccine sequences was much lower than in the wild-type sequences, and it might be argued that the lack of a significant difference between π_S and either π_N or π merely reflects the difficulty to detecting such differences statistically when the number of substitutions is small. However, there are several lines of evidence against this interpretation. First, when the same analysis was applied to 6 SAD vaccine-derived sequences isolated from red foxes, there was an even lower level of sequence diversity than the SAD vaccine sequences; yet in the case of the SAD vaccine-derived sequences, π_S was significantly greater than π_N in the *L* gene (Table 1). Likewise in the SAD vaccine-derived sequences, π in the 5' UTR and 3' UTR was significantly less than π_S in coding regions (Table 2). Furthermore, in the case of the SAD vaccine sequences, overall π in the non-coding regions was actually significantly greater than π_S was significantly, a reversal of the pattern seen in the wild-type sequences.

Relaxation of purifying selection at nonsynonymous sites was seen likewise in analyses of gene diversity at polymorphic sites in the SAD vaccine sequences. In contrast to the wild-type, mean gene diversities at synonymous and nonsynonymous sites were not significantly different in the case of SAD vaccine sequences (Figure 3B). Moreover, sites polymorphic in SAD vaccine sequences but not in the wild-type were predominately nonsynonymous. Likewise, the SAD vaccine sequences showed an excess of nonsynonymous polymorphisms in comparison to divergent changes that occurred along the branch ancestral to the SAD clade.

Since there were more than 400 reconstructed nucleotide changes along the branch ancestral to the SAD clade, it is not possible to determine which changes are responsible for loss of virulence. It is worth noting that there was strong purifying selection at nonsynonymous sites in coding regions, but not at non-coding sites, along this branch. This suggests the possibility that the attenuated phenotype may arise due to changes in non-coding regions rather than amino acid sequence changes. However, in order to reconstruct the changes that occurred during attenuation of the SAD vaccine, it will be necessary to obtain an outgroup sequence closer to the SAD ancestor than are any currently available sequences.

The present results support the hypothesis that the SAD vaccine lineage is characterized by a substantial relaxation of purifying selection and therefore a relaxation of functional constraint. In order to understand the factors involved in maintaining vaccine effectiveness, it is important to understand the biological effects of such changes in functional constraint (Hughes 2009). Because the SAD vaccine is administered to wildlife in baits, its effectiveness in the field can generally be determined only indirectly, through its impact on rabies prevalence (Lontai 1997). In this case, it is particularly important to monitor mutational changes occurring in the vaccine strain in order to ensure continued vaccine effectiveness.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grant GM43940 from the National Institutes of Health.

References

- Beckert A, Geue L, Vos A, Neubert A, Freuling C, Müller T. Genetic stability (in vivo) of the attenuated oral rabies virus vaccine SAD B19. *Microbiol Immunol* 2009;53:16–21. [PubMed: 19161553]
- Conzelmann K-K. Nonsegmented negative-strand RNA viruses: genetics and manipulation of viral genomes. *Annu Rev Genet* 1998;32:123–162. [PubMed: 9928477]
- Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985;39:783–791.
- Fenje P. Propagation of rabies virus in cultures of hamster cells. *Can J Microbiol* 1960;6:479–484. [PubMed: 13698836]
- Fu ZF. Rabies and rabies research: past, present and future. *Vaccine* 1997;15(Suppl):S20–S24. [PubMed: 9218287]
- Geue L, Schares S, Schnick S, Kliemt J, Beckert A, Freuling C, Conraths FJ, Hoffmann B, Zanoni R, Marston D, McElhinney L, Johnson N, Fooks AR, Tordo N, Müller T. Genetic characterization of attenuated SAD rabies virus strains used for oral vaccination of wildlife. *Vaccine* 2008;26:3227–3235. [PubMed: 18485548]
- Hughes AL. Adaptive Evolution of Genes and Genomes. Oxford University Press; New York: 1999.
- Hughes AL. Near neutrality: leading edge of the neutral theory of molecular evolution. *Ann N Y Acad Sci* 2008;1133:162–179. [PubMed: 18559820]
- Hughes AL. Relaxation of purifying selection on live attenuated vaccine strains of the family *Poxviridae*. *Vaccine* 2009;27:1685–1690. [PubMed: 19195493]
- Hughes AL, Hughes MA. More effective purifying selection in RNA viruses than in DNA viruses. *Gene* 2007;404:117–125. [PubMed: 17928171]
- Li W-H. Unbiased estimates of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* 1993;36:96–99. [PubMed: 8433381]
- Lontai I. The current status of rabies prevention in Europe. *Vaccine* 1997;15(Suppl):S16–S19. [PubMed: 9218286]
- Nei, M. Molecular Evolutionary Genetics. Columbia University Press; New York: 1987.
- Nei, M.; Kumar, S. Molecular Evolution and Phylogenetics. Oxford University Press; New York: 2000.
- Ohta T. Slightly deleterious mutant substitutions in evolution. *Nature* 1973;246:96–98. [PubMed: 4585855]
- Sacramento D, Bedrane H, Bourhy H, Tordo N. Molecular epidemiology of rabies virus in France: comparison with vaccine strains. *J Gen Virol* 1992;73:1149–1158. [PubMed: 1588319]
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–425. [PubMed: 3447015]
- Swofford, DL. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates; Sunderland, Massachusetts: 2003.
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596–1599. [PubMed: 17488738]
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Diggins DG. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25:4876–4882. [PubMed: 9396791]
- Tordo N, Poch O, Ermine A, Keith G, Rougeon F. Walking along the rabies genome: is the large G-L intergenic region a remnant gene? *Proc Natl Acad Sci USA* 1986;83:3914–3918. [PubMed: 3459163]

- Warrell MJ, Warrell DA. Rabies and other lyssavirus diseases. *Lancet* 2004;363:959–969. [PubMed: 15043965]
- Wu X, Franka R, Velasco-Villa A, Rupprecht CE. Are all lyssavirus genes equal for phylogenetic analyses? *Virus Research* 2007;129:91–103. [PubMed: 17681631]

Figure 1.

Schematic diagram of the RABV genome, illustrated in the order (5' to 3') of the antigenome RNA (Conzelmann 1998), showing the five genes *N*, *P*, *M*, *G*, and *L*. The lengths of regions are drawn approximately proportional to those in accession NC_001542 (M13215).

A)

B)

Figure 3.

Mean gene diversity at polymorphic synonymous (S), nonsynonymous (N) and noncoding (NC) sites in (A) wild-type RABV; and (B) SAD vaccine genomes. Numbers of sizes in each category are indicated. Randomization tests of the equality of mean gene diversity at synonymous and noncoding sites with that at nonsynonymous sites: ** $P < 0.01$; *** $P < 0.001$).

Table 1
Synonymous (π_S) and nonsynonymous (π_N) nucleotide diversity ($\pm$ S.E.) in SAD and wild-type RABV genomes.

	SAD Vaccine (N = 16)			SAD Vaccine-derived (N = 6)			Wild-type (N = 13)		
	π_S	π_N	$\pi_N:\pi_S$	π_S	π_N	$\pi_N:\pi_S$	π_S	π_N	$\pi_N:\pi_S$
Gene (no. codons)									
N (450)	0.0042 $\pm$ 0.0019	0.0023 $\pm$ 0.0010	0.55	0.0011 $\pm$ 0.0010	0.0006 $\pm$ 0.0006	0.55	0.6771 $\pm$ 0.0352	0.0198 $\pm$ 0.0024***	0.03
P (297)	0.0079 $\pm$ 0.0035	0.0059 $\pm$ 0.0020	0.75	0.0011 $\pm$ 0.0010	0.0009 $\pm$ 0.0008	0.00	0.7023 $\pm$ 0.0384	0.0759 $\pm$ 0.0082***	0.11
M (202)	0.0112 $\pm$ 0.0047	0.0104 $\pm$ 0.0033	0.93	0.0000 $\pm$ 0.0000	0.0009 $\pm$ 0.0008	--	0.6987 $\pm$ 0.0521	0.0310 $\pm$ 0.0047***	0.04
G (523)	0.0022 $\pm$ 0.0013	0.0024 $\pm$ 0.0007	1.09	0.0010 $\pm$ 0.0009	0.0004 $\pm$ 0.0004	0.40	0.7392 $\pm$ 0.0320	0.0578 $\pm$ 0.0048***	0.08
L (2127)	0.0021 $\pm$ 0.0006	0.0011 $\pm$ 0.0003	0.52	0.0013 $\pm$ 0.0005	0.0002 $\pm$ 0.0001*	0.15	0.7597 $\pm$ 0.0173	0.0232 $\pm$ 0.0013***	0.03
All (3599)	0.0034 $\pm$ 0.0006	0.0024 $\pm$ 0.0004	0.71	0.0011 $\pm$ 0.0004	0.0003 $\pm$ 0.0001	0.27	0.7330 $\pm$ 0.0122	0.0325 $\pm$ 0.0014***	0.04

Z-tests of the hypothesis that $\pi_S = \pi_N$:

* $P < 0.05$;
*** $P < 0.001$.

Table 2

Nucleotide diversity ($\pi \pm$ S.E., with numbers of aligned sites *in brackets*) in non-coding regions of SAD and wild-type RABV genomes.

Region	SAD vaccine	SAD Vaccine-derived	Wild-type
Start + end	0.0319 $\pm$ 0.0110 [128]	0.0084 $\pm$ 0.0064 [107]	0.2626 $\pm$ 0.1202 ^c [70]
5' UTR	0.0050 $\pm$ 0.0037 [113]	0.0000 $\pm$ 0.0000 ^b [113]	0.1426 $\pm$ 0.0248 ^c [113]
3'UTR	0.0058 $\pm$ 0.0022 [452]	0.0000 $\pm$ 0.0000 ^b [452]	0.3523 $\pm$ 0.0208 ^{c,d} [435]
Intergenic	0.0036 $\pm$ 0.0015 [435]	0.0012 $\pm$ 0.0012 [435]	0.3554 $\pm$ 0.0215 ^{c,d} [435]
All non-coding	0.0077 $\pm$ 0.0015 ^a [11128]	0.0013 $\pm$ 0.0007 [11128]	0.3160 $\pm$ 0.0127 ^c [1053]

Z-tests of the hypothesis that π equals π_S in all coding regions:

^a $P < 0.05$;

^b $P < 0.01$;

^c $P < 0.001$

Z-tests of the hypothesis that π equals the corresponding value for 5'UTR:

^d $P < 0.001$.