Virology **292,** 247–257 (2002)
doi:10.1006/viro.2001.1271, available online at http://www.idealibrary.com on IDEAL®

Genetic Constraints and the Adaptive Evolution of Rabies Virus in Nature

Edward C. Holmes,*[,1] Christopher H. Woelk,* Raid Kassis,† and Hervé Bourhy†

*Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom; and †Unité de la Rage, Institut Pasteur, Paris Cedex 15, France

Received August 1, 2001; returned to author for revision September 28, 2001; accepted October 29, 2001

We used a molecular evolutionary approach to investigate the species adaptation of rabies virus in nature. A maximum likelihood analysis of selection pressures revealed that the nucleoprotein (N) and glycoprotein (G) genes of natural viral isolates were highly constrained, especially at nonsynonymous sites, in contrast to the higher rates of nonsynonymous evolution observed in viruses subject to laboratory passage. Positive selection was only found at a single amino acid site—position 183 in the ectodomain of the G gene. The low rate of nonsynonymous evolution in natural isolates of rabies virus may be due to constraints imposed by the need to replicate in multiple cell types within the host, which in turn facilitates cross-species transmission, or because viral proteins are not subject to immune selection. Using known dates in the epidemiologic history of European viral isolates, we estimated that overall rates of nucleotide substitution in rabies virus were similar to those observed in other RNA viruses. Assuming that the average rate of synonymous change does not vary among species, we estimated that the current genetic diversity in lyssavirus genotype 1 may have arisen only during the last 500 years. © 2002 Elsevier Science

Key Words: rabies virus; lyssavirus; adaptation; phylogeny; positive selection; cross-species transmission; maximum likelihood.

INTRODUCTION

Rabies virus (genus *Lyssavirus*, family Rhabdoviridae) is a RNA virus with a single-stranded, negative-sense genome of approximately 12 kb in length that infects a variety of vertebrate species. Although an effective vaccine is available, it is estimated that the human death toll due to rabies is still approximately 60,000 people each year (Meslin and Stöhr, 1997). Furthermore, bat-associated rabies viruses cause sporadic disease in humans and livestock species and major epidemics in terrestrial mammals are relatively commonplace (Childs *et al.,* 2000).

Rabies viruses form two types of association with their host species. In the first, the virus establishes a stable infection cycle within a particular mammalian species, with transmission occurring through infected saliva in bite wounds. Infections of this type are most notably observed in carnivorous mammals (dogs, foxes, raccoons, skunks) as well as a variety of bat species. Whether the virus always causes disease in these situations is unclear; although fatal rabies is common in infected dogs, foxes, and raccoons, the same does not always appear to be true of bats (Baer, 1991; Ronsholt *et al.,* 1998). The second form of virus–host interaction occurs when the virus jumps species boundaries to infect new hosts. Usually, such cross-species transmission events result in sporadic cases of disease without further transmission. The most obvious example of these "spill-over" infections is human rabies, which generally leads to a fatal outcome if symptoms arise, but where no subsequent transmission takes place. Occasionally, however, rabies viruses are able to establish productive infections in new host species (Tordo *et al.,* 1993; Nadin-Davis *et al.,* 1994; Smith *et al.,* 1995). An important example of such a successful host switch involved the transfer of the virus from dogs to the red fox (*Vulpes vulpes*) in Northeast Europe during the 1930s (Bourhy *et al.,* 1999). After the initial cross-species transmission event, rabies virus was able to spread rapidly westward and southward through European red fox populations in the subsequent 60 years (Anderson *et al.,* 1981; Bourhy *et al.,* 1999).

As cross-species transmission has been shown to initiate epidemics of rabies virus, it is important to determine why some host transfers are successful and others are not. In general, the factors that enable viruses to emerge in new species involve either ecological or genetic characteristics of the virus and host. For rabies viruses, a third category—behavioural factors—may be added, as naturally aggressive biting behavior will clearly facilitate transmission among canid mammals.

The ecological factors that control viral emergence can be broadly classified as those that change the proximity or density of host and recipient species. Changes in

[1] To whom correspondence and reprint requests should be addressed at Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K. Fax: +44 1865 310447. E-mail: Edward.Holmes@zoo.ox.ac.uk.

0042-6822/02 $35.00
© 2002 Elsevier Science
All rights reserved.

population density may be particularly important since more dense populations, which will have a more regular supply of susceptible hosts, are able to carry viruses with shorter durations of infection and higher virulence (Anderson and May, 1991). Such demographic changes are clearly implicated in the emergence of rabies in the raccoon dog (*Nyctereutes procyonoides*), a canid species that was introduced in large numbers into Northeast Europe for fur farming during the 1920s to 1950s (Nowak and Paradiso, 1983). The genetic factors that control viral emergence may involve either host susceptibility to infection or viral infectiousness. Because RNA viruses typically show high levels of genetic variation, it is likely that strains will differ in their ability to replicate in new hosts or are able to adapt quickly. Such a process has been documented in rabies virus *in vitro*, where substitutions in the viral glycoprotein (G) sequence which accumulate in cell culture can change the tropism for nervous tissues, thereby changing virulence (Morimoto *et al.,* 1996, 1998). That this is an adaptive process was highlighted in the study of Kissi *et al.* (1999), who observed substantial genetic variation in the G gene from viruses passaged through different host species, with greatly elevated rates of nonsynonymous (d_N) over synonymous (d_S) substitutions per site, indicative of positive selection.

Despite the laboratory evidence for selectively driven host adaptation in rabies virus, the molecular mechanisms controlling this process are poorly understood and there is little information about what genetic changes, if any, mediate host transfer in natural infections. Previously, we analysed the phylogenetic relationships and host species distribution of European rabies viruses (Bourhy *et al.,* 1999). This study focused on natural genetic variation in the nucleoprotein (N) and G gene sequences as these are likely to be important in host adaptation; the N protein is involved in the regulation of transcription and replication (Emerson, 1987), while the G protein, which reacts with host cell receptors, is the main target of the immune response and is also important in determining pathogenicity (Dietzschold *et al.,* 1983; Tuffereau *et al.,* 1998; Thoulouze *et al.,* 1998). Little evidence of positive selection was found. Only a small number of amino changes were observed among isolates, with few defining each species or regional population. Similarly, d_N was consistently less than d_S, even though some codons had elevated levels of nonsynonymous diversity.

In the current study we present a more comprehensive analysis of the evolutionary processes acting on rabies virus in nature. Until recently, most analyses of selection pressures using d_N and d_S (or more precisely the ratio d_N/d_S, also denoted ω) involved multiple pairwise comparisons. Although this approach is highly informative when selection pressure is relatively strong, it can miss positive selection that is localised to specific sites or lineages (Zanotto *et al.,* 1999). Consequently, more powerful methods have been developed that consider each codon and/or each branch separately (Yang *et al.,* 2000; Yang and Bielawski, 2000). These methods also have a more solid statistical basis as maximum likelihood inference is used to choose which of a series of models of codon substitution best fits the data in hand. Some of these models allow for positive selection (i.e., $\omega > 1$), whereas others do not ($\omega < 1$). Herein, we apply these methods to a large sample of G and N gene sequences, including some newly determined, to obtain a better understanding of the selection pressures acting on rabies virus in nature.

RESULTS

Evolutionary relationships of global rabies virus isolates inferred from G and N genes

To determine the evolutionary relationships among a worldwide sample of rabies viruses (lyssavirus genotype 1), we constructed maximum likelihood phylogenetic trees. The trees for 55 complete G and 80 complete N gene sequences, including a set of laboratory passaged strains, are presented in Figs. 1 and 2, respectively. A phylogenetic tree of 71 partial G gene sequences was essentially the same as that of the complete G gene sequences with the addition of a large sample of European red fox and raccoon dog isolates that cluster with the other European viruses as shown previously (Bourhy *et al.,* 1999; tree not shown, available on request). Of the G gene sequences newly described in this article, 9107MAR and 9147FRA cluster with other dog, red fox, and raccoon dog viruses from Europe and the Middle East, while Thai strain 8743THA is highly divergent, although sharing some evolutionary relationship with a Chinese street strain.

In general, these phylogenies reveal more clustering by geographical origin than host species, indicating that viruses are able to cross species boundaries fairly freely. The main exceptions to this were the viruses isolated from various American bat species, which show some species-specificity in the N gene phylogeny. However, even in this case geographical clusters containing viruses from multiple species are apparent (Arai *et al.,* 1997; Nadin-Davis *et al.,* 2001). Furthermore, the large cluster of principally dog-associated viruses, including a variety of vaccine strains, covers a wide geographical area encompassing Africa, Asia, Europe, the Middle East, and the Americas, indicating that strain movement has been widespread. Although more data from bat-associated rabies viruses are clearly needed to fully document the extent of viral traffic in nature, our phylogenetic analysis supports the theory that rabies viruses generally circulate in a series of epizootiological compartments—geographically discrete clusters where viral strains are able to infect a variety of species, even if they are most often found in a single species (Rupprecht and

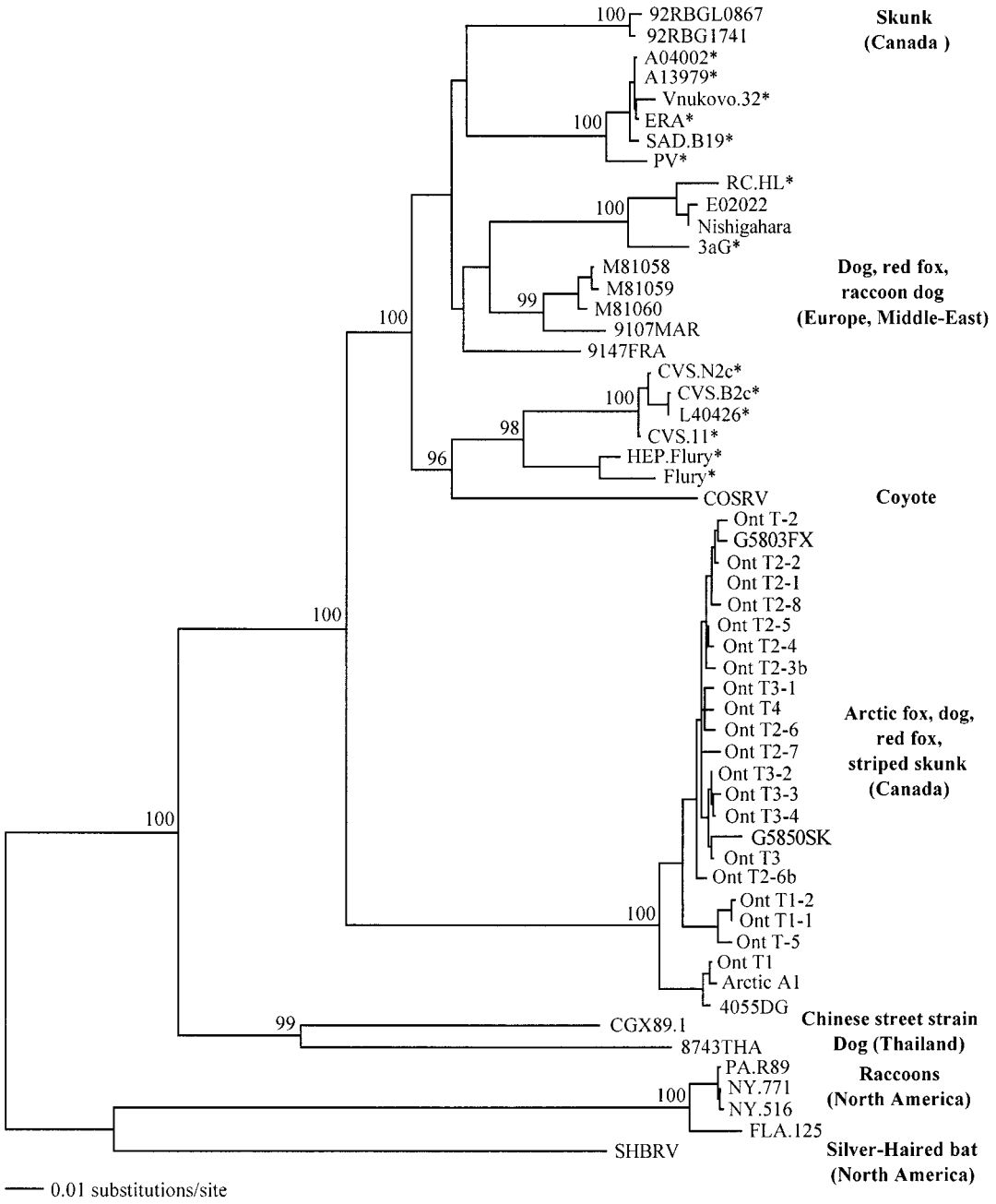

FIG. 1. Maximum likelihood phylogenetic tree of 55 sequences of the complete G gene of genotype 1 lyssaviruses. Neighbor-joining bootstrap values are shown next to key nodes only. The tree is mid-point rooted for clarity only and horizontal branch lengths are drawn to scale. The reservoir species (not necessarily the species from which the virus was isolated) and their geographical locations are indicated. GenBank accession numbers are given when strain names are not available. Heavily passaged isolates are marked by an asterisk.

Smith, 1994). However, when ecological conditions permit it, widespread viral movement is possible.

Selection pressures in the rabies virus G and N genes

The results of the maximum likelihood analysis of selection pressures acting on the G gene are presented in Tables 1 (complete sequences) and 2 (partial sequences). For the complete G gene sequences, excluding passaged strains, there was no evidence for positive selection acting at any site in this gene. Although the models of codon substitution which allow for positive selection (M2, M3, M8) were the most favoured, ω (d_N/d_S) values were generally low, indicating that this gene is subject to relatively strong selective constraints. In M3, for example, which assigns codons to three categories of site (p_0, p_1, p_2), each with a different ω value (ω_0, ω_1, ω_2), the vast majority of sites were highly constrained (p_1 =

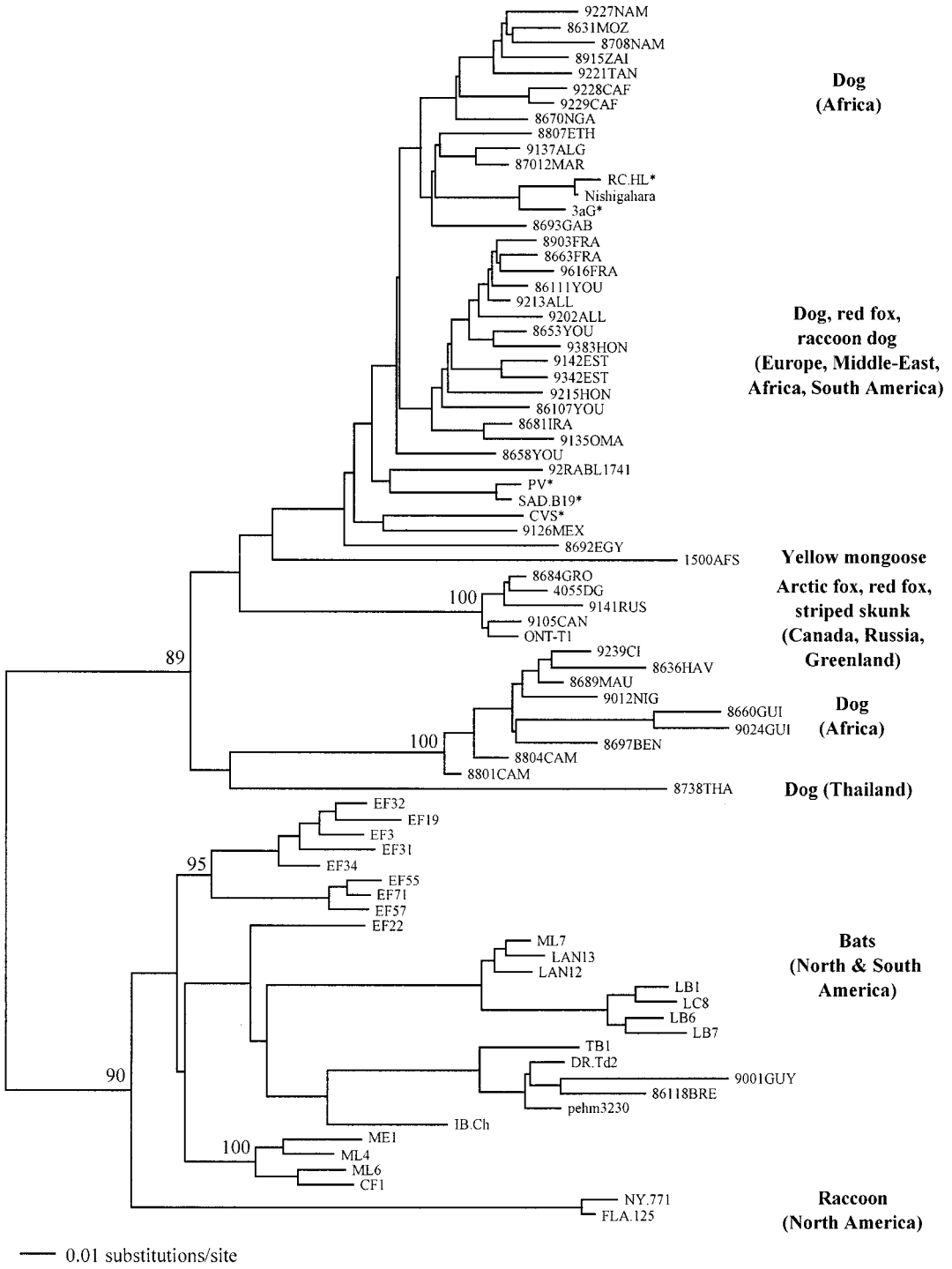

FIG. 2. Maximum likelihood phylogenetic tree of 80 sequences of the complete N gene of genotype 1 lyssaviruses. Neighbor-joining bootstrap values are shown for key nodes. The tree is mid-point rooted with horizontal branch lengths drawn to scale. The reservoir species and their geographical locations are indicated. GenBank accession numbers are given when strain names are not available. Heavily passaged isolates are marked by an asterisk.

0.933, $\omega_1 = 0.035$), a small group were seemingly subject to weak constraints ($p_0 = 0.063$, $\omega_0 = 0.633$), and an even smaller proportion fell into the ambiguous region between weak positive selection and neutral evolution ($p_2 = 0.004$, $\omega_2 = 1.555$). Very similar results were found when 14 passaged isolates were included in the analysis (results not shown, available from the authors on request), indicating that laboratory adaptation has not had a major effect on sequence diversity in this case.

In contrast, the selection analysis of the partial G gene sequences (excluding passaged strains), which includes a larger number of isolates of European origin, presents

TABLE 1

Summary of Selection Pressures in the Complete G Gene

Model	Site categories (p) and d_N/d_S (ω)	Likelihood test	χ^2	P
M0	$\omega = 0.1129$			
M1	$p_0 = 0.884$, $p_1 = 0.116$			
M2	$p_0 = 0.551$, $p_1 = 0.031$, $p_2 = 0.418$	M0 vs M2	243.946	<0.000
	$\omega_2 = 0.115$	M1 vs M2	191.248	<0.000
M3	$p_0 = 0.063$, $p_1 = 0.933$, $p_2 = 0.004$	M0 vs M3	245.136	<0.000
	$\omega_0 = 0.633$, $\omega_1 = 0.035$, $\omega_2 = 1.555$	M1 vs M3	192.438	<0.000
		M2 vs M3	1.190	0.552
M7	$p = 0.114$, $q = 1.205$			
M8	$p = 0.219$, $q = 3.247$	M7 vs M8	15.386	<0.000
	$p_{10} = 0.019$, $\omega_{10} = 1.106$			

some evidence for positive selection. In particular, M3 is favoured over competing models, although not quite significantly so over M2 ($P = 0.110$), and has one category of sites indicative of positive selection ($p_2 = 0.0004$; $\omega_2 = 3.290$). Closer inspection reveals that this site category is represented by just a single codon—amino acid 183 within the ectodomain of the G protein. Likewise, the M8 model, which also allows positive selection, was significantly favoured over M7, which does not, and also identified amino acid position 183 to be positively selected with greater than 95% probability, although with slightly weaker selection pressure ($\omega_{10} = 2.719$). Positive selection was also detected when passaged strains were included in the analysis (results not shown, available from the authors on request). In this case, both M3 and M8 detected two sites with a high probability of falling into the positively selected category—183 and 370—although the selection pressures at both M3 and M8 were weak, $\omega = 1.931$ and 1.776, respectively. Overall, these results provide the strongest evidence for positive selection acting at site 183, with marginal evidence for selection at site 370, and no important differences between natural and passaged isolates.

Unlike the G gene, there was no evidence for positive selection acting on any site in the N gene of rabies virus. Instead, the main evolutionary pattern documented was one of strong selective constraints (Table 3). Specifically, neither M3 nor M7 contained a category of sites with $\omega > 1.0$, and in both cases most sites were strongly constrained (for example, in M3 $p_1 = 0.879$ with $\omega_1 = 0.014$). Very similar results were obtained when the five heavily passaged isolates were included in the analysis.

Rates of nucleotide substitution in rabies virus

For the partial G gene sequence, the mean rates of synonymous and nonsynonymous substitution, per site, inferred from European viral isolates with known dates of sampling were estimated to be 4.10×10^{-4} ($\pm 0.30 \times 10^{-4}$) and 5.06×10^{-5} ($\pm 0.85 \times 10^{-5}$), respectively. The equivalent rates for the complete N gene sequences were 5.27×10^{-4} ($\pm 0.23 \times 10^{-4}$) and 2.85×10^{-5} ($\pm 0.265 \times 10^{-5}$), respectively. Hence, the synonymous rates are similar between genes, as expected if synonymous sites are evolving neutrally, while the nonsynonymous rate in G is almost twice that in N, again as expected if there is some localised positive selection in the former. The synonymous rate estimated for both genes is also at the low end of the range for other RNA viruses (range for 26 RNA viruses = 7.9×10^{-3} to $6.8 \times$

TABLE 2

Summary of Selection Pressures in the Partial G Gene

Model	Site categories (p) and d_N/d_S (ω)	Likelihood test	χ^2	P
M0	$\omega = 0.083$			
M1	$p_0 = 0.897$, $p_1 = 0.103$			
M2	$p_0 = 0.625$, $p_1 = 0.008$, $p_2 = 0.367$	M0 vs M2	67.56	<0.000
	$\omega_2 = 0.149$	M1 vs M2	163.416	<0.000
M3	$p_0 = 0.099$, $p_1 = 0.901$, $p_2 = 0.0004$	M0 vs M3	71.97	<0.000
	$\omega_0 = 0.358$, $\omega_1 = 0.029$, $\omega_2 = 3.290$	M1 vs M3	167.826	<0.000
		M2 vs M3	4.410	0.110
M7	$p = 0.194$, $q = 2.603$			
M8	$p = 0.257$, $q = 3.669$	M7 vs M8	11.770	0.003
	$p_{10} = 0.0007$, $\omega_{10} = 2.719$			

TABLE 3

Summary of Selection Pressures in the N Gene

Model	Site categories (p) and d_N/d_S (ω)	Likelihood test	χ^2	P
M0	$\omega = 0.057$			
M1	$p_0 = 0.926$, $p_1 = 0.074$			
M2	$p_0 = 0.669$, $p_1 = 0.005$, $p_2 = 0.326$	M0 vs M2	371.886	<0.000
	$\omega_2 = 0.098$	M1 vs M2	1111.518	<0.000
M3	$p_0 = 0.117$, $p_1 = 0.879$, $p_2 = 0.004$	M0 vs M3	377.034	<0.000
	$\omega_0 = 0.170$, $\omega_1 = 0.014$, $\omega_2 = 1.000$	M1 vs M3	1116.666	<0.000
		M2 vs M3	5.148	0.076
M7	$p = 0.150$, $q = 3.177$			
M8	$p = 0.296$, $q = 8.161$	M7 vs M8	74.444	<0.000
	$p_{10} = 0.005$, $\omega_{10} = 0.958$			

10^{-4}; Jenkins *et al.,* in press), although high enough to confirm that the basic rates of mutation and replication in rabies virus must be broadly similar to those in other RNA viruses, while the relatively low rate of nonsynonymous substitution again confirms that most protein-coding sites are subject to strong selective constraints.

Assuming that synonymous substitution rates are constant in all the viruses in our sample, which infect a variety of vertebrate hosts, it is possible to provide an approximate time scale for the evolution of genotype 1 lyssaviruses. In the N gene, for which most data are available, the most divergent pair of sequences on the tree (estimated by summing branch lengths) are NY771 from an American raccoon (*Procyon lotor*) and the African dog-associated strain 8708NAM, which have a synonymous distance of 0.3655. Given our average synonymous substitution rate estimated for the N gene, these sequences would have diverged around 347 years ago. Even if the lower synonymous rate from the G gene is used, these sequences would only have separated approximately 446 years ago. Hence, our analysis suggests that the genetic diversity in the currently available sample of genotype 1 lyssaviruses, including those from American bat species, might have only arisen within the last 500 years.

DISCUSSION

Evolutionary processes in rabies virus G and N genes

Our study provides evidence, albeit relatively weak, for the positive selection of amino acid replacements in rabies viruses in nature. Specifically, we identified amino site 183 as subject to moderate positive selection pressure ($d_N/d_S = 3.290$), with rather more marginal evidence for selection at codon position 370. There was no suggestion of positive selection acting on the N gene, nor on the various antigenic sites in the G gene, most notably site III, which appears to be an important determinant of neuropathogenesis (Dietzschold *et al.,* 1983; Flamand *et al.,* 1993), and which pairwise analyses of d_N/d_S have suggested might be subject to adaptive evolution (Badrane and Tordo, 2001). Further, we found no evidence for positive selection at positions 132 and 333, both of which have previously been shown to be involved in rabies pathogenesis (Tuffereau *et al.,* 1989) and which were well conserved in the sequences analysed here.

Closer inspection reveals that amino acid position 183 in the G gene is highly variable, with five different amino acids present in our sample of viruses. The most common amino acid at this position is Pro, which is found in isolates from a variety of species, while Val is present in some isolates from North American terrestrial mammals, although the majority carry a Leu, and Trp occurs in a single European fox isolate. Viruses from the divergent Silver-haired bat virus (SHBRV) and raccoons possess an Ala at site 183. Assuming the rooting of our G gene phylogenies is correct (Fig. 1), then a Pro → Leu replacement has occurred three times independently and it is this convergent evolution that clearly provides the signal of positive selection in these data. Indeed, the fact that this replacement has occurred twice within the European viruses alone is the reason selection was not detected until the partial sequences from the European red fox isolates were included in the analysis. Interestingly, the Pro → Leu change occurs twice more when passaged isolates are considered. This could mean that this replacement might also be involved in adaptation to passage conditions, although not all passaged strains possess this amino acid change (and this replacement was not seen in the study of Kissi *et al.,* 1999), or perhaps more likely that a Leu at position 183 was present in the natural ancestors of the passaged isolates.

Despite the evidence for positive selection at position 183, the mechanisms by which this selection occurs are unclear. Although position 183 does not fall into any of the antigenic sites described previously, it is immediately adjacent to antigenic site II (R184), is located within a neurotoxin-like region (Braci *et al.,* 1992; Donnelly-Roberts and Lentz, 1991), and is one which contains the putative attachment site for the nicotine acetylcholine

receptor. This region was also shown to be variable in a comparison of SHBRV and coyote street rabies (COSRV) strains (Morimoto *et al.,* 1996). Hence, it is possible that site 183 plays some role in determining cell tropism *in vivo*, although this is clearly an area for further study.

Visual inspection of the other putatively selected site, position 370, merely confirmed that it is difficult to unequivocally demonstrate selection at this site, although it is located within a pH-independent fusion domain (Durrer *et al.,* 1995; Kawai and Morimoto, 1994). By far the most common amino acid at this site was His, which is found in viruses isolated from a variety of species. A His → Asn change has occurred on three occasions within European viruses, although in only a small number of isolates, and SHBRV also possesses an Asn in this position. Finally, a single His → Arg change has occurred in a North American Arctic fox isolate. As with site 183, the His → Asn change has occurred twice more in viruses with long passage histories.

Although there is evidence for highly localised positive selection pressure in the G gene, the overall level of nonsynonymous variation exhibited in rabies virus is very low. While the relatively strong selective constraints in the N gene have been reported previously (Amengual *et al.,* 1997; Kissi *et al.,* 1995), the general conservation of the G gene is noteworthy as viral glycoproteins are often associated with high levels of nonsynonymous diversity and provide some of the best examples of positive selection in nature (Yang and Bielawski, 2000; Valarcher *et al.,* 2000). In particular, nonsynonymous rates up to two orders of magnitude higher than those seen in rabies virus have been documented in the glycoproteins of viruses thought to be under strong immune selection, such as HIV-1 (Leitner and Albert, 1999), hepatitis C virus (Power *et al.,* 1995), and influenza A virus (Bush *et al.,* 1999). Moreover, similar methods to those used here have identified more abundant positive selection in two other negative-strand RNA viruses—measles virus (Woelk *et al.,* 2001) and respiratory syncytial virus (Woelk and Holmes, 2001), and a third, vesicular stomatitis virus (VSV), has been suggested to undergo adaptive evolution in nature (Rodriguez *et al.,* 1996).

The constrained nonsynonymous evolution of rabies virus is even more striking given that laboratory studies reveal that genetic variation can be generated extremely rapidly. For example, Kissi *et al.* (1999) studied the generation of genetic diversity in a European fox isolate that was serially passaged in mice, dogs, cats, and cell culture. Nonsynonymous diversity appeared quickly, especially in the G gene, a process that appeared to be selectively driven. In particular, d_N/d_S ratios of 3.544 and 1.283 for G and N genes, respectively, were observed across the entire regions analysed. These ratios provide compelling evidence for positive selection as it must reflect elevated d_N levels at a large number of codons, although it is also possible that some of the nonsynonymous mutations observed during passaging are deleterious and will ultimately be removed by selection. Hence, the evolution of rabies virus can differ greatly between natural and laboratory environments, with the former dominated by purifying selection and the latter liable to strong positive selection.

If rabies virus can evolve quickly when the need arises, why is nonsynonymous evolution so constrained in nature? A relatively low rate of nonsynonymous substitution has also been observed in vector-borne RNA viruses (Jenkins *et al.,* in press; Weaver *et al.,* 1992), perhaps because the need to replicate in both vertebrate and invertebrate hosts imposes strong selective constraints (Weaver *et al.,* 1999), although this has been questioned (Novella *et al.,* 1999). Rabies viruses may represent an analogous example where genetic constraints are imposed by the need to replicate in very different cell types. Hence, although rabies virus has a strong neurotropism, replication *in vivo* does not only take place in neuronal cells. In particular, there is evidence that the virus replicates in muscle tissue at the site of inoculation before entering the peripheral and central nervous systems as well as the salivary glands and other nonnervous tissues (Charlton *et al.,* 1997). As a case in point, Morimoto *et al.* (1996) suggested that the ability to replicate in epidermal cells may have adaptive value for SHBRV as elevated viral load would increase the chance of finding a nerve fibre, an important requirement for bats where the infecting dose may be much less than infections involving canid mammals. If selection pressure for broad tropism is strong, so that rabies viruses are adapted to replicate in a diverse range of cell types, then it is also possible that this virus is *preadapted* to replicate in a wide range of species. In other words, we hypothesise that the constraints imposed by the need to replicate in a range of cell types mean that rabies virus can jump with relative ease to other species that have similar cell types. Hence, strong purifying and weak positive selection would be the norm in nature.

An alternative explanation for the relative lack of nonsynonymous variation is that the G glycoprotein is not subject to the strong immune selective pressures that characterise the envelope proteins of some other viruses. Specifically, the rapid dispersal of rabies virus into cells of the central nervous system (CNS) may be an adaptive strategy to evade host immune pressure as CNS cells are generally under weak immune surveillance (Charlton *et al.,* 1996, 1997) and a rapid spread of virus from the peripheral site of entry to the CNS with little prior replication has been reported (Ceccaldi *et al.,* 1989; Shankar *et al.,* 1991). Further, it is estimated that levels of immunity in raccoon populations experiencing a major rabies epizootic are very low (1–5%) (Childs *et al.,* 2000). Consequently, the envelope glycoprotein is not engaged in an evolutionary arms race with the host immune system. Finally, it is also possible that the evo-

lution of rabies viruses in nature is greatly affected by stochastic processes, including population bottlenecks which may occur among hosts during transmission, and within hosts as variants infect different cell types (Amengual *et al.,* 1997). The long-term outcome of these repeated sampling effects would be to dissipate the power of natural selection.

Age of genetic diversity in rabies virus

Our analysis of rates of nucleotide substitution suggests that the current genetic diversity in samples of lyssavirus genotype 1 from diverse geographical locations and different species may have only arisen within the last 500 years. If true, such a recent history has a number of important implications. First, it is clear that the virus itself is far older than this, with descriptions of the highly characteristic symptoms of rabies illness, often in dogs, dating back many millennia, including references in the Talmud and the works of Aristotle (Rupprecht and Hanlon, 1997). Consequently, such recent divergence times would mean that these ancient medical records are describing different viral lineages to those that are circulating today. More intriguing is that rabies was not described in the New World until the time of European colonisation, with the first good descriptions occurring in the 18th century (Rupprecht and Hanlon, 1997). Given our estimated times of divergence, the implication is that the viral lineages currently associated with New World bats first entered this continent at the time of European colonisation and have sustained their transmission there ever since. Of course, care must be taken in making such inferences as substitution rates may vary among species where the nature of the virus-host association is different, most notably bats. While more work is clearly needed to test the accuracy of the substitution rates and divergence times we estimate, it is noteworthy that our synonymous rate is broadly similar to those seen in a wide variety of RNA viruses and to that previously estimated for lyssaviruses (Badrane and Tordo, 2001). Moreover, the genetic diversification of rabies virus in populations of red fox and striped skunk (*Mephitis mephitis*) in Ontario has seemingly occurred during the last 40–50 years, which would not be the case if rabies were an unusually slowly evolving RNA virus (Nadin-Davis *et al.,* 1994, 1999). Finally, viral strains isolated from different species appear to be under similar selection pressures, suggesting that the overall substitution rate has not changed dramatically on the infection of new species. For example, in the N gene, for which a large number of isolates from different species are available, mean d_N/d_S values are similar among bat-associated viruses (0.053) to those obtained from all other mammalian species (0.062). The future investigation of rates of viral evolution in different species will evidently shed more light on the nature of species adaptation as well as on the ultimate origin of rabies virus.

MATERIALS AND METHODS

Sequence data

The G gene sequences of three rabies viruses were produced in this study. Two represented human associated strains; 8743THA was isolated in Thailand in 1983 while 9107MAR was isolated in Morocco in 1990. Dogs were the reservoir species in both cases. The third isolate, 9147FRA, was obtained from a red fox in France in 1991. RNA isolation, cDNA synthesis, PCR amplification, and cloning of these isolates was performed according to Kissi *et al.* (1995) using primers L, M1.1: 5′-CCTTGATGATATAGTTAAAGAGGC-3′ (nucleotides 2212–2235) and G6: 5′-CGTTGGTCACTGAAACTGCTAGAAG-3′ (nucleotides 5110–5096), according to Kissi *et al.* (1995) (positions relative to the PV rabies virus genome, Tordo *et al.,* 1986). Sequencing was performed as described previously (Kissi *et al.,* 1995).

All other rabies virus G and N gene sequences analysed were collected from GenBank. Our analysis was restricted to "classical" rabies viruses—genotype 1 of the genus *Lyssavirus*—because of the extensive diversity between viruses from different genotypes, especially at synonymous sites (Badrane *et al.,* 2001; Bourhy *et al.,* 1993, 1995). This meant that viruses from African, Australian, and European bats were excluded from the analysis, while those from American bats, which fall within the classical rabies viruses, were included.

Four different G gene data sets were constructed. Two covered the entire nucleotide sequence (1572 bp) and differed in that one contained all currently available sequences ($n = 55$), while the second excluded those strains known to have long passage histories ($n = 41$). Two further data sets were created containing partial G gene sequences (690 bp) that are available from a larger number of isolates, particularly those of European origin. Again, data sets containing all available isolates ($n = 71$) and with passaged isolates excluded ($n = 60$) were compiled. Long-term passaged isolates were removed as this process has been previously documented to produce spurious evidence for positive selection in other viruses (Woelk and Holmes, 2001). The isolates removed were as follows: all PV derivatives (e.g., ERA, SAD.B19, Vnukovo-32); all CVS derivatives; all FLURY derivatives; 3aG; and RC-HL.

Only complete gene sequences were analysed in the case of the N gene (1350 bp). Because of the very large number of sequences available and the computational intensity of the selection analysis, an initial tree of 121 isolates was constructed, which was then pruned by removing isolates that were closely related to others in the data set. This resulted in a data set of $n = 80$ that represented the full diversity of N gene sequences avail-

able, and $n = 75$ when passaged isolates were removed. A full list of all the virus isolates used in this study, as well as the sequence alignments, are available at http://evolve.zoo.ox.ac.uk/.

Phylogenetic analysis

Phylogenetic trees were estimated using a maximum likelihood (ML) method under the general time-reversible (GTR) model of nucleotide substitution, with the rate of each type of substitution estimated from the data. The ML base frequencies were also estimated from the data, as were the proportion of invariable sites and a gamma distribution of rate variation among sites. Parameter values are available from the authors on request. The starting tree in the search was found using neighbor-joining and this was followed by successive rounds of TBR branch-swapping, identifying the ML substitution parameters at each stage, until the tree of highest likelihood was found. The robustness of each node on the tree was assessed using the bootstrap resampling method, with all 1000 replicates estimated using the neighbor-joining procedure, but with the input genetic distances produced under the ML substitution model. All trees were estimated using the PAUP* package (Swofford, 2000).

Analysis of selection pressures

We used a maximum likelihood approach to investigate selection pressures in rabies virus. In this analysis the numbers of nonsynonymous (d_N) and synonymous (d_S) substitutions per site are determined for each codon using various models of codon substitution and taking into account the phylogenetic tree linking the sequences (Yang *et al.,* 2000). The models of codon substitution either fix or estimate d_N/d_S ratios, referred to here by the ω parameter. The model with the highest likelihood best explains the data and if this model shows $\omega > 1$ at any site, then positive selection is supported. The simplest model, M0, calculates a single ω value for all sites. M1 divides codons into two categories, one for invariant sites (p_0), with ω_0 fixed at 0, and second category (p_1), with ω_1 set to 1, thereby representing neutral sites. M1 is therefore a strictly neutral model of evolution. In contrast, M2 can account for positive selection because a third category of sites (p_2) is added at which ω_2 (estimated from the data), can be >1. In M3, the ω ratio is estimated separately from the data for three site classes. Because the ω value for any site class can be >1, this model also allows for positive selection. M7 and M8 both use a discrete beta distribution (with 10 categories) to model ω ratios among sites. The beta distribution can take a range of shapes described by parameters p and q, although M8 differs from M7 in that it estimates an 11th category of sites (p_{10}) at which ω can be >1. Models that are nested may be compared using a likelihood ratio test (LRT) in which twice their difference in log likelihood is compared to the value obtained under a χ^2 distribution. In particular, both M0 and M1 are nested with M2 and M3, and M7 is nested with M8. Bayesian methods can also be used to calculate the probability that a particular codon falls into the positively selected class, although only when this probability is >0.95 are sites considered to be selected in this analysis. Finally, the free ratio (FR) model estimates the ω ratio for the entire gene along each branch of the tree. This was used in the estimation of substitution rates. All these methods were implemented using the CODEML program of the PAML package (Yang, 1997).

Estimating rates of nucleotide substitution

Rates of nucleotide substitution for the G and N genes were estimated using information about the epidemiologic history of rabies virus in Europe. Specifically, records suggest that rabies virus first jumped from dogs to the red fox during the 1930s. This cross-species transmission most likely took place in Northeast Europe, with the virus spreading westward and southward from this point (Bourhy *et al.,* 1999). This spread is clearly visible in phylogenetic trees of the G and N genes; the deepest lineages associated with the red fox are from samples collected in Northeast Europe, while the most recent branches are from samples collected in Western Europe. Consequently, we conservatively assume that the root of the red fox part of the phylogenetic tree dates to 1930. We can therefore obtain estimates of nonsynonymous and synonymous rates for the red fox part of the tree by summing d_N and d_S values for the branches leading from the root point to each individual tip on the tree and then dividing by the total time depth of each pathway (i.e., 1930 to sampling date; mean = 57 years). This analysis was undertaken on 21 partial G genes sequences and 22 complete N gene sequences (excluding passaged strains). Although shared phylogenetic history will mean that not every estimate is independent, this method should provide a good general indication of the substitution rate in rabies virus. d_S and d_N values for each branch were estimated under the FR model in CODEML.

ACKNOWLEDGMENTS

We thank the Royal Society and the BBSRC for financial support and Dr. Susan Nadin-Davis for providing rabies virus N gene sequences prior to their availability on GenBank. Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. AF401285–AF401287.

REFERENCES

Amengual, B., Whitby, J. E., King, A., Serra Cobo, J., and Bourhy, H. (1997). Evolution of European bat lyssavirus. *J. Gen. Virol.* **78,** 2319–2328.

Anderson, R. M., Jackson, H. C., May, R. M., and Smith, A. M. (1981). Population dynamics of fox rabies in Europe. *Nature* **289,** 765–771.

Anderson, R. M., and May, R. M. (1991). "Infectious Diseases of Humans: Dynamics and Control." Oxford University Press, Oxford.

Arai, Y. T., Yamada, K., Kameoka, Y., Horimoto, T., Yamamoto, K., Yabe, S., Nakayama, M., and Tashiro, M. (1997). Nucleoprotein gene analysis and street rabies virus variants using RT-PCR. *Arch. Virol.* **142,** 1787–1796.

Badrane, H., Bahloul, C., Perrin, P., and Tordo, N. (2001). Evidence of two Lyssavirus phylogroups with distinct pathogenicity and immunogenicity. *J. Virol.* **75,** 3268–3276.

Badrane, H., and Tordo, N. (2001). Host switching in *Lyssavirus* history from the Chiroptera to the Carnivora orders. *J. Virol.* **75,** 8096–8104.

Baer, G. M. (1991). "The Natural History of Rabies." CRC Press, Boca Raton, FL.

Bourhy, H., Kissi, B., Audry, L., Smreczak, M., Sadkowska-Todys, M., Kulonen, K., Tordo, N., Zmudzinski, J. F., and Holmes, E. C. (1999). Ecology and evolution of rabies virus in Europe. *J. Gen. Virol.* **80,** 2545–2558.

Bourhy, H., Kissi, B., and Tordo, N. (1993). Molecular diversity of the Lyssavirus genus. *Virology* **194,** 70–81, doi:10.1006/viro.1993.1236.

Bracci, L., Lozzi, L., Rustici, M., and Neri, P. (1992). Binding of HIV-1 gp120 to the nicotinic receptor. *FEBS Lett.* **311,** 115–118.

Bush, R. M., Bender, C. A., Subbarao, K., Cox, N. J., and Fitch, W. M. (1999). Predicting the evolution of human influenza A. *Science* **286,** 1921–1925.

Ceccaldi, P. E., Gillet, J. P., and Tsiang, H. (1989). Inhibition of the transport of rabies virus in the central nervous system. *J. Neuropathol. Exp. Neurol.* **48,** 620–630.

Charlton, K. M., Casey, G. A., Wandeler, A. I., and Nadin-Davis, S. (1996). Early events in rabies virus infection of the central nervous system in skunks (*Mephitis mephitis*). *Acta Neuropathol.* **91,** 89–98.

Charlton, K. M., Nadin-Davis, S., Casey, G. A., and Wandeler, A. I. (1997). The long incubation period in rabies: Delayed progression of infection in muscle at the site of exposure. *Acta Neuropathol.* **94,** 73–77.

Childs, J. E., Curns, A. T., Dey, M. E., Real, L. A., Feinstein, L., Bjørnstad, O. N., and Krebs, J. W. (2000). Predicting the local dynamics of epizootic rabies among raccoons in the United States. *Proc. Natl. Acad. Sci. USA* **97,** 13666–13671.

Dietzschold, B., Wunner, W. H., Wiktor, T. J., Lopes, A. D., Lafon, M., Smith, C. L., and Koprowski, H. (1983). Characterization of an antigenic determinant of the glycoprotein that correlates with pathogenicity of rabies virus. *Proc. Natl. Acad. Sci. USA* **80,** 70–74.

Donnelly-Roberts, D. L., and Lentz, T. L. (1991). Structural and conformational similarity between synthetic peptides of curaremimetic neurotoxins and rabies virus glycoprotein. *Mol. Brain Res.* **11,** 107–113.

Durrer, P., Gaudin, Y., Ruigrok, R. W. H., Graf, R., and Brunner, J. (1995). Photolabeling identifies a putative fusion domain in the envelope glycoprotein of rabies and vesicular stomatitis viruses. *J. Biol. Chem.* **270,** 17575–17581.

Emerson, S. U. (1987). Transcription of vesicular stomatitis virus. *In* "The Rhabdoviruses" (R. R. Wagner, Ed.), pp. 245–269. Plenum Press, New York.

Flamand, A., Raux, H., and Ruigrok, R. W. H. (1993). Mechanisms of rabies virus neutralization. *Virology* **194,** 302–313, doi:10.1006/viro.1993.1261.

Jenkins, G. M., Rambaut, A., Pybus, O. G., and Holmes, E. C. Rates of molecular evolution in RNA viruses: A quantitative phylogenetic analysis. *J. Mol. Evol.,* in press.

Kawai, A., and Morimoto, K. (1994). Functional aspects of lyssavirus proteins. *In* "Current Topics in Microbiology and Immunology: Lyssaviruses" (C. E. Rupprecht, B. Dietzschold and H. Koprowski, Eds.), pp. 27–42. Springer-Verlag, New York.

Kissi, B., Badrane, H., Audry, L., Lavenu, A., Tordo, N., Brahimi, M., and Bourhy, H. (1999). Dynamics of rabies virus quasispecies during serial passages in heterologous hosts. *J. Gen. Virol.* **80,** 2041–2050.

Kissi, B., Tordo, N., and Bourhy, H. (1995). Genetic polymorphism in the rabies virus nucleoprotein gene. *Virology* **209,** 526–537, doi:10.1006/viro.1995.1285.

Leitner, T., and Albert, J. (1999). The molecular clock of HIV-1 unveiled through analysis of a known transmission history. *Proc. Natl. Acad. Sci. USA* **96,** 10752–10757.

Meslin, F.-X., and Stöhr, K. (1997). Prospects for immunization against rabies in developing countries. *In* "Rabies Control in Asia" (B. Dodet and F. X. Meslin, Eds.), pp. 15–18. Elsevier, Paris.

Morimoto, K., Hooper, D. C., Carbaugh, H., Fu, Z. F., Koprowski, H., and Dietzschold, B. (1998). Rabies virus quasispecies: Implications for pathogenesis. *Proc. Natl. Acad. Sci. USA* **95,** 3152–3156.

Morimoto, K., Patel, M., Corisdeo, S., Hooper, D. C., Fu, Z. F., Rupprecht, C. E., Koprowski, H., and Dietzschold, B. (1996). Characterization of a unique variant of bat rabies virus responsible for newly emerging human cases in North America. *Proc. Natl. Acad. Sci. USA* **93,** 5653–5658.

Nadin-Davis, S. A., Casey, G. A., and Wandeler, A. I. (1994). A molecular epidemiological study of rabies virus in central Ontario and western Quebec. *J. Gen. Virol.* **75,** 2575–2583.

Nadin-Davis, S. A., Huang, W., Armstrong, J., Casey, G. A., Bahloul, C., Tordo, N., and Wandeler, A. I. (2001). Antigenic and genetic divergence of rabies viruses from bat species indigenous to Canada. *Virus Res.* **74,** 139–156.

Nadin-Davis, S. A., Sampath, M. I., Casey, G. A., Tinline, R. R., and Wandeler, A. I. (1999). Phylogeographic patterns exhibited by Ontario rabies virus variants. *Epidemiol. Infect.* **123,** 325–336.

Novella, I. S., Hershey, C. L., Escarmis, C., Domingo, E., and Holland, J. J. (1999). Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells. *J. Mol. Biol.* **287,** 459–465, doi: 10.1006/jmbi.1999.2635.

Nowak, R. M., and Paradiso, J. L. (1983). "Walker's Mammals of the World. Vol. II," 4th ed. The John's Hopkins University Press, Baltimore, MD.

Nel, L., Jacobs, J., Jaftha, J., and Courteney, M. (1997). Natural spillover of a distinctly Canidae-associated biotype of rabies virus into an expanded wildlife host range in Southern Africa. *Virus Genes* **15,** 79–82.

Power, J. P., Lawlor, E., Davidson, F., Holmes, E. C., Yap, P. L., and Simmonds, P. (1995). Molecular epidemiology of an outbreak of infection with hepatitis C virus in recipients of anti-D immunoglobulin. *Lancet* **345,** 1211–1213.

Rodriguez, L. L., Fitch, W. M., and Nichol, S. T. (1996). Ecological factors rather than temporal factors dominate the evolution of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. USA* **93,** 13030–13035.

Ronsholt, L. Sorensen, K. J., Bruschke, C. J. M., Wellenberg, G. J., van Oirschot, J. T., Johnstone, P., Whitby, J. E., and Bourhy, H. (1998). Clinically silent rabies infection in (zoo) bats. *Vet. Res.* **142,** 519–520.

Rupprecht, C. E., and Hanlon, C. A. (1997). Rabies. *In* "Viral Infections of Humans. Epidemiology and Control," 4th ed. (A. S. Evans and R. A. Kaslow, Eds.), pp. 665–690. Plenum Medical Book Co., New York.

Rupprecht, C. E., and Smith, J. S. (1994). Raccoon rabies: The emergence of an epizootic in a densely populated area. *Semin. Virol.* **5,** 155–164.

Shankar, V., Dietzschold, B., and Koprowski, H. (1991). Direct entry of rabies virus into the central nervous system without prior local replication. *J. Virol.* **65,** 2736–2738.

Smith, J. S., Orciari, L. A., and Yager, P. (1995). Molecular epidemiology of rabies in the United States. *Semin. Virol.* **6,** 387–400.

Steck, F., and Wandeler, A. (1980). The epidemiology of fox rabies in Europe. *Epidemiol. Rev.* **2,** 71–95.

Swofford, D. L. (2000). Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.

Thoulouze, M.-I., Lafage, M., Schachner, M., Hartmann, U., Cremer, H., and Lafon, M. (1998). The neural cell adhesion molecule is a receptor for rabies virus. *J. Virol.* **72,** 7181–7190.

Tordo, N., Badrane, H., Bourhy, H., and Sacramento, D. (1993). Molecular epidemiology of lyssaviruses: Focus on the glycoprotein and pseudogenes. *Onderstepoort J. Vet. Res.* **60,** 315–323.

Tordo, N., Poch, O., Ermine, A., Keith, G., and Rougeon, F. (1986). Walking along the rabies genome: Is the large G-L intergenic region a remnant gene? *Proc. Natl. Acad. Sci. USA* **83,** 3914–3918.

Tuffereau, C., Bénéjean, J., Roque Alfonso, A. M., Flamand, A., and Fishman, M. C. (1998). Neuronal cell surface molecules mediate specific binding to rabies virus glycoprotein expressed by a recombinant baculovirus on the surfaces of lepidopteran cells. *J. Virol.* **72,** 1085–1091.

Tuffereau, C., Leblois, H., Bénéjean, J., Coulon, P., Lafay, F., and Flamand, A. (1989). Arginine or lysine in position 333 of ERA and CVS glycoprotein is necessary for rabies virulence in adult mice. *Virology* **172,** 206–212.

Valarcher, J. F., Schelcher, F., and Bourhy, H. (2000). Evolution of bovine respiratory syncytial virus. *J. Virol.* **74,** 10714–10728.

Weaver, S. C., Brault, A. C., Kang, W. L., and Holland, J. J. (1999). Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. *J. Virol.* **73,** 4316–4326.

Weaver, S. C., Rico-Hesse, R., and Scott, T. W. (1992). Genetic diversity and slow rates of evolution in new-world alphaviruses. *Curr. Top. Microbiol. Immunol.* **176,** 99–117.

Woelk, C. H., and Holmes, E. C. (2001). Variable immune driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV). *J. Mol. Evol.* **52,** 182–192.

Woelk, C. H., Jin, L., Holmes, E. C., and Brown, D. W. G. (2001). Immune and artificial selection in the hemagglutin (H) glycoprotein of measles virus. *J. Gen. Virol.* **82,** 2463–2474.

Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13,** 555–556.

Yang, Z., and Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15,** 496–502.

Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A. M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155,** 431–449.

Zanotto, P. M. de A., Kallas, E. G., de Souza, R. F., and Holmes, E. C. (1999). Genealogical evidence for positive selection in the *nef* gene of HIV-1. *Genetics* **153,** 1077–1089.