Evolutionary timescale of rabies virus adaptation to North American bats inferred from the substitution rate of the nucleoprotein gene

Gareth J. Hughes,† Lillian A. Orciari and Charles E. Rupprecht

Rabies Section, Centers for Disease Control and Prevention, 1600 Clifton Road, Mail-Stop G33, Atlanta, GA 30333, USA

Throughout North America, rabies virus (RV) is endemic in bats. Distinct RV variants exist that are closely associated with infection of individual host species, such that there is little or no sustained spillover infection away from the primary host. Using Bayesian methodology, nucleotide substitution rates were estimated from alignments of partial nucleoprotein (N) gene sequences of nine distinct bat RV variants from North America. Substitution rates ranged from 2×10\(-4\) to 1.38×10\(-3\) substitutions per site per year. A maximum-likelihood (ML) molecular clock model was rejected for only two of the nine datasets. In addition, using sequences from bat RV variants across the Americas, the evolutionary rate for the complete N gene was estimated to be 2×10\(-4\). This rate was used to scale trees using Bayesian and ML methods, and the time of the most recent common ancestor for current bat RV variant diversity in the Americas was estimated to be 1660 (range 1267–1782) and 1651 (range 1254–1773), respectively. Our reconstructions suggest that RV variants currently associated with infection of bats from Latin America (Desmodus and Tadarida) share the earliest common ancestor with the progenitor RV. In addition, from the ML tree, times were estimated for the emergence of the three major lineages responsible for bat rabies cases in North America. Adaptation to infection of the colonial bat species analysed (Eptesicus fuscus, Myotis spp.) appears to have occurred much quicker than for the solitary species analysed (Lasionycteris noctivagans, Pipistrellus subflavus, Lasiusa borealis, Lasiusa cinereus), suggesting that the process of virus adaptation may be dependent on host biology.

INTRODUCTION

Rabies virus (RV) is the type member of the genus Lyssavirus, family Rhabdoviridae. RV has a negative-sense RNA genome of ~12 kb and consists of five genes encoding the nucleoprotein (N), phosphoprotein, matrix protein, glycoprotein and polymerase (Tordo et al., 1986). The N encapsidates the viral RNA, forming the template for transcription and replication (Tordo & Kouknetzoff, 1993). RNA viruses are characterized by a high mutation rate during replication due to the lack of proofreading and post-replication error correction by RNA polymerases (Domingo & Holland, 1994). This genetic diversity seems to provide an adaptive potential for RV that can vary according to natural history (Morimoto et al., 1998; Kissi et al., 1999).

Indigenous bat rabies was first recognized in the USA in 1953 (Scatterday & Galton, 1954), although the exact age of its emergence remains uncertain. During 2001, bats were responsible for 17% of reported wildlife rabies cases and all three cases of human rabies in the USA (Krebs et al., 2002). In North America, rabid bats have been found for almost every species that has been thoroughly investigated (Constantine, 1979). The genetic diversity of current bat RV variants suggests that up to 30 different lineages of RV may exist in the USA alone (Smith, 2002). A more conservative assessment of RV diversity in the Americas suggests 13 distinct RV variants (Nadin-Davis et al., 2001). In the USA during 2001, bat rabies where the host could be identified at species level was dominated by two host species: Eptesicus fuscus (big brown bat; 47·1%) and Tadarida brasiliensis (Mexican free-tailed bat; 28·3%). In addition, small numbers of cases were attributed to Lasiusa cinereus (hoary bat; 5·5%), Lasiusa borealis (red bat; 4·3%), Myotis lucifugus (little brown bat; 3·7%), Lasionycteris noctivagans (silver-haired bat; 3·0%) and Pipistrellus subflavus (eastern pipistrelle; 1·2%). In Canada, with the exception of T. brasiliensis, the species of bats most often found rabid are identical to the USA (Nadin-Davis et al., 2001). The emergence of the lineage of RV associated with infection of Lasionycteris noctivagans and P. subflavus, which has caused a disproportionately high
number of human cases, has been attributed to an increased viral infectivity (Morimoto et al., 1996; Messenger et al., 2003), although this hypothesis remains largely unproven.

Timing the divergence of viral lineages relies on an accurate estimation of the rate of nucleotide substitution and subsequent application of a molecular clock. Computational methods to estimate substitution rates are problematic (Drummond et al., 2003) and the validity of applying clock-like models to viral sequence data is contentious (Holmes, 2003). Recently, Bayesian techniques using Markov Chain Monte Carlo (MCMC) methods have been developed for estimating substitution rates from dated sequences (Drummond et al., 2002) and have been successfully applied to RV where estimations of evolutionary rate and subsequent divergence times were shown to be both comparable to maximum-likelihood methods and supported by field prevalence data (Hughes et al., 2004).

Here, we have applied computational methods to nucleotide sequence data from the N gene of RV. We used distinct datasets, which represented all of the major bat RV variants in the USA and Canada, to estimate rates of molecular evolution for individual variants. In turn, using complete N gene sequences from North American bat RV variants, together with bat species associated with Latin America, we reconstructed the timescale for the evolution of bat RV variants in North America.

**METHODS**

**Datasets.** We used nine sets of partial N gene sequences (all of which could be differentiated by either host species or geography) and an alignment of complete N gene sequences, all of which were bat-associated RV variants from North, South and Central America (Table 1). Within each dataset, identical sequences with the same year of isolation were removed. Partial N gene sequences from the USA were obtained directly from the Centers for Disease Control and Prevention (CDC) RV archival database (sequences used in this study are to be submitted to GenBank) and corresponded to nt 1107–1350 (244 nt) of the Pasteur RV N gene (GenBank accession no. M13215; Tordo et al., 1986). An alignment of RV sequences from Desmodus rotundus (vampire bat) covering the same region of the N gene as the US sequences was generated using data obtained from GenBank (accession nos AB083804, AB083805, AB083807–AB083809, AB083811, AB083813, AB083814, AB083818, AF045166, AF351847, AF351852, U22478, U22479). Partial N gene sequences from Canada were obtained from GenBank and have been described in detail elsewhere (Nadin-Davis et al., 2001). Canadian RV sequences corresponded to nt 287–670 (384 nt) of the Pasteur RV N gene. Complete N gene sequences were taken from GenBank (accession nos AF045166, AF351827–AF351829, AF351831–AF351848, AF351851–AF351862, AF394872–AF394875, AF394879–AF394886, AY039224, AY039226–AY039229, U22478).

**Table 1. Datasets used in this study**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Primary host species</th>
<th>Years*</th>
<th>( n^\dagger )</th>
<th>Sequence length (nt)</th>
<th>( \omega^\dagger )</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desmodus</td>
<td>Desmodus rotundus</td>
<td>14</td>
<td>14</td>
<td>244</td>
<td>0.0516</td>
<td>GenBank</td>
</tr>
<tr>
<td>US Tadarida</td>
<td>Tadarida brasiliensis</td>
<td>23</td>
<td>19</td>
<td>244</td>
<td>0.0833</td>
<td>CDC archival database</td>
</tr>
<tr>
<td>US Eptesicus</td>
<td>Eptesicus fuscus</td>
<td>20</td>
<td>69</td>
<td>244</td>
<td>0.0937</td>
<td>CDC archival database</td>
</tr>
<tr>
<td>US RAC</td>
<td>Procyn lotor</td>
<td>17</td>
<td>40</td>
<td>244</td>
<td>0.1382</td>
<td>CDC archival database</td>
</tr>
<tr>
<td>US L-clade</td>
<td>Lasiurus borealis, Lasiurus cinereus</td>
<td>21</td>
<td>87</td>
<td>244</td>
<td>0.0563</td>
<td>CDC archival database</td>
</tr>
<tr>
<td>US LnPs</td>
<td>Lansionycyctes noctivagans, Pipistrellus subflavus</td>
<td>41</td>
<td>33</td>
<td>244</td>
<td>0.0701</td>
<td>CDC archival database</td>
</tr>
<tr>
<td>CN Eptesicus</td>
<td>Eptesicus fuscus</td>
<td>26</td>
<td>39</td>
<td>384</td>
<td>0.0194</td>
<td>Nadin-Davis et al. (2001)</td>
</tr>
<tr>
<td>US CASK</td>
<td>Mephitis mephitis</td>
<td>27</td>
<td>14</td>
<td>244</td>
<td>0.0545</td>
<td>CDC archival database</td>
</tr>
<tr>
<td>CN Myotis</td>
<td>Myotis spp.</td>
<td>8</td>
<td>14</td>
<td>384</td>
<td>0.0677</td>
<td>Nadin-Davis et al. (2001)</td>
</tr>
<tr>
<td>Bat complete N</td>
<td>North, Central and South American bats</td>
<td>27</td>
<td>53</td>
<td>1353</td>
<td>0.0647</td>
<td>GenBank</td>
</tr>
</tbody>
</table>

*Total number of years over which the sequences were isolated.
†Number of sequences within each dataset.
‡Ratio of non-synonymous substitutions per non-synonymous site (\( d_s \)) to synonymous substitutions per synonymous site (\( d_{st} \)).
Evolution of RV adaptation to North American bats

requires no assumptions be made regarding tree topology (Drummond et al., 2002). Previously, we have employed this method for RV sequences where the results were found to be comparable to ML-based methods (Hughes et al., 2004). In addition, for feline immunodeficiency virus, the MCMC method has been shown to produce shorter confidence limits than those based on ML (Biek et al., 2003).

In each case, an input file for BEAST was generated using the BEAUti program (available at http://www.evolv.zoo.ox.ac.uk) with sequences dated with the year of isolation. Analysis was performed using the HKY85 substitution model (Hasegawa et al., 1985) without site rate heterogeneity. A coalescent model (prior) of constant population size was used. MCMC analysis was optimized using the criteria suggested in the program’s documentation. This included an operator acceptance probability of $\sim 25\%$ and an effective sample size of $>100$. BEAST output was assessed using the Tracer program (available at http://www.evolv.zoo.ox.ac.uk). The distribution shape for estimated values of $\mu$ was checked to ensure that adequate sampling of the chain had taken place. For each estimation, three replicate BEAST runs were performed to test the repeatability of the analysis.

To test for possible variation in estimated substitution rates along the N gene, we used a sliding window analysis to assess region-specific rates and corresponding values of sequence diversity. Nucleotide diversity ($\pi$) was calculated from the alignment of 53 complete N gene sequences of bat RV variants using the computer program DnaSP (Rozas et al., 2003). For estimation of both $\mu$ and $\pi$, a sliding window of 250 nt with a step size of 100 was used. Values of $\mu$ were then calculated from individual alignments of 250 nt as described above.

**Molecular clock testing and estimation of divergence times.**

The hypothesis of a molecular clock (i.e. a constant rate of nucleotide substitution across the whole tree) was tested using ML trees generated by the program TipDate (Rambaut, 2000). In each case, an ML tree was generated from PHYLIP and used as input. For each dataset, an ML tree was generated in TipDate under the assumption of a single rate of nucleotide substitution using dated tips (SRDT model) and the likelihood of this tree was compared with that of an ML tree generated without the assumption of a molecular clock (where the substitution rate is allowed to vary for individual branches; DR model). Trees were generated using the HKY85 model of nucleotide substitution with transition to transversion ratios estimated directly from the data. For SRDT trees, rates of substitution estimated from BEAST analysis were used to scale trees and the ML root estimated by TipDate. The likelihoods of the SRDT and DR models were then compared using a likelihood ratio test as described elsewhere (Rambaut, 2000). The assumption of the molecular clock was not rejected if the likelihood of the SRDT tree was not significantly worse ($P<0.05$) than that of the DR tree. For the complete N gene alignment, SRDT trees were generated using the mean and upper and lower 95% confidence limits (CL) of the BEAST-estimated values of $\mu$. SRDT trees scaled according to time were used to estimate times of the most recent common ancestor (TMRCA) and divergence times of major lineages. An additional estimate of the TMRCA was made using BEAST by performing the MCMC analysis without dated tips using the mean and upper and lower 95% CL of the values of $\mu$ to scale trees.

**Simulations.** In order to assess the ability of the MCMC method to accurately infer rates, we simulated nucleotide evolution with predefined rates and tested the precision of the MCMC method using varying numbers of taxa and ranges of isolation dates. The evolution of DNA sequences using defined substitution rates was simulated using the program Seq-Gen (Rambaut & Grassly, 1997). Phylogenetic trees generated by the program COMPONENT (available at http://taxonomy.zoology.gla.ac.uk) using pre-defined numbers of taxa were used as input for Seq-Gen. Two sets of simulations were performed to assess the precision of the MCMC method to estimate substitution rates: (i) to test the effect of the period of isolation (time between first isolation and last) for a fixed number of taxa ($n=50$) with varying substitution rate; and (ii) to test the effect of the number of taxa over a fixed isolation period (25 years) with varying substitution rate. For all simulations, sequences of 250 nt in length were generated using the HKY85 model of nucleotide substitution.

To test the effect of varying the number of taxa, sequences were removed at random from the COMPONENT tree to ensure that topological relationships between sequences were retained (although the process of random removal meant that overall topology was not identical). Where differing isolation periods were required (but with the same number of taxa), trees were rescaled according to the timescale required for simulation. Seq-Gen-generated alignments were then used as input for BEAST as described for the biological datasets.

**RESULTS**

**Positive selection analysis**

A mean value for $\omega$ (taken from M0) for each dataset is shown in Table 1. Although there was considerable variation in $\omega$ between datasets (from 0.02 for CN Eptesicus to 0.14 for US RAC), there was no evidence of positive selection for any dataset using codon substitution models in CODEML.

**Simulations of sequence evolution**

The results of BEAST analysis of simulated datasets are shown in Table 1. Although there was considerable variation between datasets (from 0.02 for CN Eptesicus to 0.14 for US RAC), there was no evidence of positive selection for any dataset using codon substitution models. In addition, for simulations of sequence evolution, the ability of the method to determine rates of viral evolution, but did show that better-suited data (an optimal number of taxa, i.e. that is not too computationally expensive, and a large range of isolation dates) dramatically increased the certainty of the mean value.

**Estimation of the rate of molecular evolution and molecular clock testing**

Estimated values of $\mu$ for each dataset, with 95% CL, are shown in Fig. 2. For seven of the nine partial N gene datasets, the assumption of a molecular clock was not significantly rejected (Table 2). Rejection of the molecular
clock for US Eptesicus, CN Eptesicus and bat complete N was highly significant \((P < 0.001)\). In all three cases where the SRDT model was significantly rejected by the DR model, the likelihood of the SRDT model was higher than that using a single rate without dated tips model (SR) (results not shown). The improved likelihood of the SRDT model over the SR model has been suggested as an indication that rates and divergence times inferred from the SRDT model reflect the mean values for the dataset and are appropriate to use for further analysis (Jenkins et al., 2002; Twiddy et al., 2003). In addition, it has been demonstrated that rejection of the molecular clock model due to low-rate heterogeneity can still yield reliable estimates of evolutionary rate (Jenkins et al., 2002) and divergence times similar to those produced by a relaxed molecular clock model (Lemey et al., 2004).

The results of sliding window analysis along the N gene are shown in Fig. 3. Although the analysis showed variation in both \(\pi\) and \(\mu\) along the N gene, there was no significant correlation between \(\pi\) and \(\mu\) (Spearman’s rank correlation coefficient, \(r_s = 0.098, n = 12, P = 0.762\)), \(\pi\) and the 95% confidence interval range of \(\mu\) (\(r_s = 0.150, n = 12, P = 0.700\)), \(\pi\) and midpoint position (\(r_s = 0.070, n = 12, P = 0.829\)) and \(\mu\) and midpoint position (\(r_s = -0.091, n = 12, P = 0.779\)). The lack of a significant correlation between \(\pi\) and \(\mu\) suggests that estimating evolutionary rates using the MCMC method is not influenced by the variability within the gene (or gene fragment) but by the phylogenetic position of individual taxa. No value of \(\mu\) was greatly different to the estimate for the complete N gene, and although seven of the 12 sliding window values were greater than the upper 95% CL of the complete N gene value of \(\mu\), three alignments

---

**Fig. 1.** Variation in the range of estimated substitution rates for simulated data using pre-defined values of \(\mu\). Data were generated using varying ranges of isolation dates with a fixed number of taxa \((n = 50)\) (a) and varying numbers of taxa using a fixed period of isolation (25 years) (b). Individual data points show the 95% confidence interval deviation from the pre-defined value of \(\mu\). Missing data points indicate failure of BEAST to estimate values of \(\mu\) due to the limited amount of variation generated by low substitution rates.

**Fig. 2.** Estimated rates of molecular evolution. Substitution rates were calculated using an MCMC-based method and are shown on a logarithmic scale. Datasets are arranged in descending order of mean substitution rate. Open circles indicate that the assumption of a molecular clock was significantly rejected \((P < 0.001)\). Error bars show 95% CL of the mean.

---

**Table 2.** Likelihood ratio testing for the assumption of a molecular clock

<table>
<thead>
<tr>
<th>Dataset</th>
<th>(\chi^2) *</th>
<th>Degrees of freedom</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desmodus</td>
<td>17.25</td>
<td>11</td>
<td>0.101</td>
</tr>
<tr>
<td>US Tadarida</td>
<td>15.31</td>
<td>16</td>
<td>0.502</td>
</tr>
<tr>
<td>US Eptesicus</td>
<td>126.96</td>
<td>66</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>US RAC</td>
<td>20.31</td>
<td>37</td>
<td>0.326</td>
</tr>
<tr>
<td>US L-clade</td>
<td>85.26</td>
<td>84</td>
<td>0.441</td>
</tr>
<tr>
<td>US LnPs</td>
<td>40.49</td>
<td>30</td>
<td>0.096</td>
</tr>
<tr>
<td>CN Eptesicus</td>
<td>116.82</td>
<td>36</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>US CASK</td>
<td>5.28</td>
<td>11</td>
<td>0.917</td>
</tr>
<tr>
<td>CN Myotis</td>
<td>8.84</td>
<td>11</td>
<td>0.636</td>
</tr>
<tr>
<td>Bat complete N</td>
<td>137.50</td>
<td>50</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

*Twice the difference in likelihood between the SRDT model (clock) and the DR model (no clock).
estimated extremely small lower CL, together suggesting that the mean value for the complete N gene represented a good estimation of $\mu$ (and its variability) across the whole gene.

**TMRCA and the timescale of bat RV variant evolution**

The SRDT tree for the complete N gene alignment, generated using the mean value of $\mu$ estimated by BEAST ($2.3 \times 10^{-5}$), is shown in Fig. 4. The TMRCA for current bat RV variant diversity in the Americas (based on the complete N gene alignment) was estimated to be 1651 (range 1254–1773) from the ML SRDT tree and 1660 (range 1267–1782) from MCMC-based analysis (BEAST). The inclusion of rate heterogeneity in the HKY model gave an identical mean value of $\mu$ for the complete N gene dataset and did not significantly alter the estimated TMRCA (results not shown).

The divergence times of individual lineages were estimated from the ML SRDT tree from TipDate. The first branching from the parental virus is predicted to have been the divergence of an RV that gave rise to the current variants associated with infection of *T. brasiliensis* and *D. rotundus*, i.e. bats geographically associated with Latin America and the southern USA. It appears that adaptation to *T. brasiliensis* and *D. rotundus* was a relatively long process; current

**Fig. 3.** Sliding window analysis using an alignment of 53 complete bat RV variant N genes (window size 250 nt, step size 100). Closed circles show substitution rates and 95% CL estimated from BEAST. Open circles indicate where lower CL were exceptionally small. Open squares show the nucleotide diversity calculated from DnaSP. Broken lines show the mean and 95% CL of the substitution rate calculated using the complete N gene sequences (1353 nt).

**Fig. 4.** ML SRDT tree for complete bat N gene sequences scaled to time using the mean value of $\mu$ estimated from BEAST. Sequence names have been altered to include the year of isolation.
diversity of RV isolates associated with these two bat species has a TMRCA of 1837 (range 1856–1892), suggesting that the progenitor viruses for the RV variants associated with infection of these two species circulated for ~160 years before being displaced by the evolution of variants better adapted to species-specific transmission.

Within North America, compartmentalization of RV into lineages associated with infection of solitary bat species (hosts for the LnPs and L-clade viruses) and colonial bat species (E. fuscus, Myotis spp.) is estimated to have occurred 29 years (range 18–62) after the divergence of the Tadarida/Desmodus lineage. In turn, each of the two North American lineages underwent further division to become the four extant RV variants that have been analysed in this study and which together represent the majority of current bat extant RV variants that have been analysed in this study. The establishment of RV variants associated with E. fuscus and Myotis spp. occurred much earlier, around 1768 (range 1507–1847), indicating a much shorter period of adaptation than was required for the Tadarida/Desmodus and LnPs/L-clade lineages.

**DISCUSSION**

Currently, the validity of applying a molecular clock to RNA virus evolution is unclear (Holmes, 2003). Saturation of synonymous mutations (leading to potential underestimations of substitution rates; Holmes, 2003), recombination (reviewed by Awadalla, 2003), RNA secondary structure (Simmonds & Smith, 1999) and selection pressures can undoubtedly all contribute to misleading estimates of evolutionary rate. The lack of recombination found for complete bat N genes (Haydon et al., 2004), the fact that the RV genome is complexed with the N (Tordo & Kouknetzoff, 1993) and thus not subject to constraints on RNA secondary structure, and the strong evidence for purifying selection found in this study and elsewhere (Kissi et al., 1995; Holmes et al., 2002; Kuzmin et al., 2004) imply that the RV N gene is suitable for estimations of evolutionary rate and subsequent analysis of evolutionary history. Here, we assumed that the substitution rate of the RV N gene has remained constant over the course of the timescale measured (349–746 years). Although RV has undoubtedly undergone periods of intense host adaptation, the strong evidence for purifying selection suggests that fitness fluctuations should not have affected the evolutionary dynamics of the N gene.

In this study we limited grouping of sequences (as much as was possible given the lack of sequence data available for a number of variants) to those that represented, as accurately as could be determined, distinct lineages of RV (Smith, 2002). The lack of sustained spillover between RV variants (Smith, 2002; Krebs et al., 2002) ensured that our partial N gene datasets represented distinct RV transmission networks. Of the nine partial N gene datasets (each representing a distinct RV variant), seven produced trees under the assumption of a molecular clock (SRDT model) that were not significantly less likely than trees generated using the non-clock model (DR), i.e. the model of a molecular clock was not rejected. The biological significance of datasets for which the clock was significantly rejected is unclear, although both are RV variants associated with infection of the same bat species (E. fuscus).

The relative consistency of μ estimations between datasets supports the hypothesis of a molecular clock governing evolution of the RV N gene, irrespective of the host species. Although rates do vary by host species, there is considerable overlap in the CL for individual values. In fact, the similarity of estimated values of μ (range 2.32 × 10^{-4}–1.38 × 10^{-3}) suggested that the number of replication cycles per unit of time was similar for the RV variants studied here. Additionally, our estimates of substitution rate (i.e. both d_s and d_e combined) were close to previously estimated d_e rates for the N gene (5.27 × 10^{-4}; Holmes et al., 2002) and the glycoprotein gene (3.1 × 10^{-4}–5.5 × 10^{-4}; Badrane & Tordo, 2001). This was no doubt due to the lack of positively selected codons along the N gene found here and elsewhere (Holmes et al., 2002; Kuzmin et al., 2004).

Using complete bat RV variant N gene sequences, we have been able to reconstruct the evolutionary history of bat rabies in the Americas. Although the molecular clock model for the complete N gene alignment was rejected (probably in part due to both the amalgamation of RV variants with subtly different substitution rates and minor variation in μ along the N gene), we have used the SRDT tree for evolutionary timescaling as we believe the mean μ for this alignment to represent the best available approximation of substitution rate for the N gene. Additionally, there is good concordance between TMRCA estimates for the complete N gene alignment generated using ML (SRDT trees) and the MCMC-based methods.

There is some variation in previous estimates of the TMRCA of existing worldwide RV diversity. The establishment of non-chiropteran rabies (globally) and the development of RV variants in the Americas have been dated to have occurred between 542 and 1113 (Badrane & Tordo, 2001). Another study estimated the TMRCA of current global RV diversity to be within the last 500 years, implying that RV variants associated with New World bats originated from RV introduced during European colonization (Holmes et al., 2002), a hypothesis not refuted by our estimation of the TMRCA of current bat RV diversity in the Americas.

The results of our TMRCA analysis suggest that RV associated with infection of Latin American bats (D. rotundus and the migratory T. brasiliensis) share an earlier common ancestor with the progenitor RV responsible for seeding bat rabies in North America compared with those variants associated with non-Tadarida North American bat species. The presence of bat rabies in the Americas prior to European
colonization may have been due to infection with extinct lineages of RV, the diversity of which is no longer evident, although this hypothesis is not testable. Whether or not current North American bat rabies is directly descended although this hypothesis is not testable. Whether or not lineages of RV, the diversity of which is no longer evident, colonization may have been due to infection with extinct

While rabies in carnivores was well appreciated in the Old World for thousands of years, most of the earliest explorers remarked upon its notable absence in ‘New Spain’, at least in dogs, until the early 1700s (Baer et al., 1996). Several non-mutually exclusive mechanisms have been proposed for the arrival of rabies in the New World. Ancestral lyssaviruses may already have been present in the rudiments of various warm-blooded hosts in the super continent of Pangea, as various rhabdovirus lineages diverged and radiated among plants, invertebrates and vertebrates (Shope, 1975; Rupprecht et al., 1991). Much more recently, the disease may have entered through terrestrial animal migrations via the ancient Bering land bridge (Crandell, 1975). During viral incubation periods that can encompass months, infected animals may have rafted along coastlines on storm debris or flotsam, stowed away on primitive human conveyances or been brought intentionally for translocated domestic purposes (Smith & Seidel, 1993). In addition, rabies may have entered directly by flight from infected bats, given their unique and considerable aerodynamic abilities (Constantine, 2003). However, these pre-historic contributions of bats to rabies in the Americas and associations with particular fauna are largely speculative, with the exception of evidence based on fossil records of vampire bats.

Based on the fossil record, several species of vampire bats were once widely distributed throughout North America. During the Quaternary period, many suitable large-bodied and abundant mammalian species may have served as suitable prey for vampire bats but perished during the Pleistocene extinctions. As such, barring any major impact of climate change or a shortage of suitable roosts such as caves, a post-Pleistocene ecological bottleneck may have been created due to a relative shortage in blood supply for vampires during pre-Columbian times (Ray et al., 1988). With the colonization of the New World, the introduction of cattle should have allowed the local expansion of bat populations. For example, in modern times, Desmodus populations are larger in livestock areas of Argentina than in non-cattle-rearing regions (Delpietro & Russo, 1993). Although not identified as rabies at the time, Spanish colonists recognized a link between bat bites and death in humans and animals as early as the 1500s and throughout the intervening centuries, but such observations were not attributed to rabies until the 1900s in Brazil and elsewhere (Baer & Smith, 1991). Extant vampires do not hibernate, do not migrate and are currently limited in distribution from Mexico to Argentina, approximated by a 10 °C winter isotherm (McNab, 1973). The entry of RV into colder latitudes would have required a more tolerant migratory species, such as Tadarida. The association of the Desmodus/Tadarida lineage and subsequent formation of the bat RV variants circulating today is supported by the results of this study. Of course, other bat species may have been equally or more important. The prominence of considerable cattle losses through blood feeding by rabid vampires provided an incentive worthy of study and notation, whereas rabies in insectivorous or frugivorous species may have gone largely undetected, except by accident.

Following the introduction of RV into North American bats, our analysis suggests that there was rapid adaptation to new host species. The evolution of RV variants associated with colonial species (E. fuscus, Myotis spp.) occurred first (~1768), approximately 70 years before the compartmentalization of Tadarida/Desmodus-associated variants and 80 years before the compartmentalization of LnPs/L-clade-associated variants. Perhaps the biology of colonial bat species (a higher density of susceptible contacts) ensured a greater number of replication cycles with time and thus a speedier adaptation process than for solitary species (where contacts with susceptible hosts would be less frequent). Of course, if such a situation were true, it would ensure higher levels of nucleotide substitution during the adaptive period, in turn suggesting that the sampling period for isolates used in this study is beyond the period of intensive adaptation. Whether or not such a process occurs in actuality will only be testable should nucleotide sequence data be available over the course of an analogous period of adaptation. Given the current epidemiological situation of bat rabies in North America and its apparent recent endemic stability among chiropteran hosts, the development of such an extant scenario appears remote, barring major environmental perturbations on a continental scale, such as more dramatic climatic change or enhanced habitat degradation.

ACKNOWLEDGEMENTS

The authors thank Jean Smith for retrieval of sequence data, Andrew Rambaut for continued guidance with sequence analysis software and Dan Haydon for helpful comments on the manuscript. G. J. H. was funded by an American Society for Microbiology/National Center for Infectious Diseases post-doctoral fellowship.

REFERENCES


