MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 2120–2136 doi: 10.1111/j.1365-294X.2010.04630.x

Spatial and temporal dynamics of rabies virus variants in big brown bat populations across Canada: footprints of an emerging zoonosis

SUSAN A. NADIN-DAVIS,* YUQIN FENG,* DELPHINE MOUSSE,† ALEXANDER I. WANDELER* and STÉPHANE ARIS-BROSOU†‡

*Centre of Expertise for Rabies, Ottawa Laboratory (Fallowfield), Canadian Food Inspection Agency, Ottawa, Ontario, Canada, †Department of Biology, University of Ottawa, Ottawa, Ontario, Canada, ‡Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, Canada

Abstract

Phylogenetic analysis of a collection of rabies viruses that currently circulate in Canadian big brown bats (*Eptesicus fuscus*) identified five distinct lineages which have emerged from a common ancestor that existed over 400 years ago. Four of these lineages are regionally restricted in their range while the fifth lineage, comprising two-thirds of all specimens, has emerged in recent times and exhibits a recent demographic expansion with rapid spread across the Canadian range of its host. Four of these viral lineages are shown to circulate in the US. To explore the role of the big brown bat host in dissemination of these viral variants, the population structure of this species was explored using both mitochondrial DNA and nuclear microsatellite markers. These data suggest the existence of three subpopulations distributed in British Columbia, mid-western Canada (Alberta and Saskatchewan) and eastern Canada (Quebec and Ontario), respectively. We suggest that these three bat subpopulations may differ by their level of female phylopatry, which in turn affects the spread of rabies viruses. We discuss how this bat population structure has affected the historical spread of rabies virus variants across the country and the potential impact of these events on public health concerns regarding rabies.

Keywords: bat subpopulations, big brown bat, DNA barcoding, microsatellite loci, rabies virus variants

Abbreviations for Canadian provinces and US states AB,: Alberta; BC, British Columbia; MB, Manitoba; NB, New Brunswick; ON, Ontario; QC, Quebec; SK, Saskatchewan; AL, Alabama; AZ, Arizona; CA, California; CO, Colorado; CT, Connecticut; NY, New York; PA, Pennsylvania; TX, Texas; WA, Washington.

Received 19 March 2009; revision received 25 February 2010; accepted 8 March 2010

Introduction

Lyssaviruses are single-stranded RNA viruses with negative sense genomes of the family Rhabdoviridae, and are the agents responsible for eliciting rabies-like disease in a wide range of mammals. The lyssavirus genome comprises five genes (reviewed by Wunner 2007), of which three, N, G and P, have been extensively characterized. The conserved N gene, which encodes the nucleoprotein that regulates viral transcription and replication through genome encapsidation, has been used for genotype discrimination (Kuzmin *et al.* 2005). The less conserved G gene encodes the surface glycoprotein that is responsible for viral attachment to host cells and critical to viral pathogenicity (Badrane & Tordo 2001). The P gene sequence encodes a highly variable multi-

Correspondence: Susan A. Nadin-Davis,
E-mail: susan.nadin-davis@inspection.gc.ca

functional phosphoprotein that permits finely detailed epidemiological studies (Nadin-Davis *et al.* 2002).

Currently the *Lyssavirus* genus is divided into seven distinct genotypes (GTs 1-7; Tordo *et al.* 2004). Many species of chiroptera figure prominently as reservoir hosts for all but one genotype (GT 3; Badrane & Tordo 2001) while as yet unclassified lyssaviruses have been recovered from various Eurasian bat species (Kuzmin *et al.* 2003, 2005). Classical rabies virus (RABV), which constitutes GT 1, is globally distributed in various carnivore reservoirs but associated with chiroptera only on the American continent (Kissi *et al.* 1995; Nadin-Davis *et al.* 2002). Since 1950s, when North American bats were first identified as RABV vectors (reviewed in Baer & Smith 1991), passive surveillance has identified several hundreds of bat rabies cases annually. Moreover, bat rabies has emerged as a significant public health issue since, in recent years, a high percentage of human rabies cases indigenously acquired in North America have been due to exposure to bat RABVs (Messenger *et al.* 2002). Antigenic and genetic typing methods differentiate several distinct RABV strains, each associated with particular chiropteran species (Smith *et al.* 1995; Nadin-Davis *et al.* 2001).

In Canada and in the United States (US), the nonmigratory big brown bat (*Eptesicus fuscus*), a member of the Vespertilionidae family in the suborder Microchiroptera, is the bat species most frequently reported as rabies positive (Pybus 1986; Rosatte 1987; Mondul *et al.* 2003). The role of the big brown bat as a rabies vector does however decrease from north (e.g., NY; Childs *et al.* 1994) to south (e.g. AL, TX), where other bat species are increasingly important in this regard (Rohde *et al.* 2004; Hester *et al.* 2007). While prior genetic characterization of RABVs associated with big brown bats in the US and Canada suggested that these viruses are relatively heterogeneous and can be divided into at least two quite distinct groupings that differ in their regional distribution (Smith *et al.* 1995; Nadin-Davis *et al.* 2001; Jackson *et al.* 2008), a comprehensive study of these RABV variants and their range across North America has never been described.

The big brown bat is the only representative of the *Eptesicus* genus indigenous to the US and Canada where its range extends throughout the US and into most southern regions of Canada (Kurta & Baker 1990). Based on morphological variation and differences in reproductive and ecological traits across its North American range, several regionally distributed subspecies of *E. fuscus* have been proposed of which three occur in distinct regions of Canada (Kurta & Baker 1990). However, genetic support for this species subdivision in Canada is lacking and information about the population structure of big brown bats across the country is scant. The nonmigratory and relatively sedentary lifestyle of these bats, together with their small home ranges, suggest that the population may be divided into several discrete groups with limited inter-group gene flow; such a population structure could play a major role in limiting transmission and spread of rabies.

DNA barcoding has recently emerged as a genetic tool that purports to facilitate eukaryote species identification (Hebert *et al.* 2003) and, more controversially, to use sequence data to identify cryptic species (Hebert *et al.* 2004). Certain characteristics of mitochondrial genetics, such as its maternal inheritance pattern, variation in the mutation rate of the targeted cytochrome oxidase I (*COX*1) gene, mitochondrial introgression, heteroplasmy and interference from nuclear pseudogenes could complicate and/or preclude accurate species assignment and identification in certain situations (Frézal & Leblois 2008). However, for many eukaryotic taxa, this approach appears to have performed well, especially when ecological characteristics of the specimens under study have been considered as, for example, in a study of Lepidoteran specimens (Hajibabaei *et al.* 2006). DNA barcoding has yielded novel insights into bat species identification (Clare *et al.* 2007; Mayer *et al.* 2007). For more detailed population studies, nuclear microsatellite loci have provided great insights into the biology, ecology and evolution of several members of the chiroptera (see Kerth *et al.* 2002; Neubaum *et al.* 2007; Ngamprasertwong *et al.* 2008).

To better understand the factors contributing to rabies virus transmission in this bat host in Canada, we have undertaken studies to: (i) better define the spatial distribution of all rabies viral variants that are harboured by the big brown bat, (ii) explore the bat's species substructure using DNA barcoding methods, (iii) examine the big brown bat population structure across the country using several newly identified polymorphic microsatellite loci. As such, this combined analysis of viral phylogeography and host genetic variation sheds light on the phylodynamics (Grenfell *et al.* 2004) of this pathogen–host relationship with new insights into the emergence and spread of rabies in this host.

Materials and methods

Bat sampling

This study employed a collection of 388 big brown bats drawn from submissions received at the two laboratories, Nepean, ON and Lethbridge, AB, responsible for rabies diagnosis across Canada. Specimens were designated as follows: a two digit number indicating year of submission, a two letter code representing the province of origin, a four or five digit specimen identifier

followed by the two letter suffix BB to indicate the species as big brown bat. A few specimens of other bat species, included for comparative viral analysis, were identified to species using the following suffices: HB, hoary bat (*Lasiurus cinereus*), LB, little brown bat *(Myotis lucifugus*), RB, red bat (*Lasiurus borealis*) and SH, silver-haired bat (*Lasionycteris noctivagans*). Specimens spanned the years 1972–2007, although the majority were recovered in the years 2003–2004, and represent the range of the big brown bat across the country. Few specimens were recovered from the eastern provinces adjacent to the Atlantic Ocean because of the species' rarity in this region. Specimens were identified to species based on their morphological characters and brain tissue from each sample was tested for the presence of rabies virus by Direct Fluorescent Antibody testing (Webster & Casey 1988). Both rabies-positive and rabies-negative bats were selected for host genome microsatellite scoring to increase the geographical coverage of the population analysis. A smaller subgroup of these specimens was employed for partial *COX*1 locus characterization. The use of this collection for all three analyses, grouped according to the province of origin of the specimens, is summarized in Table 1.

Rabies-infected brain tissue was used as a source of viral RNA; where possible tissue was recovered from the original bat specimen but when material was in short supply the virus was passaged once through suckling mice according to Canadian Council of Animal Care guidelines. Total brain RNA was extracted using TRIzol reagent (Invitrogen, Burlington, ON, Canada) according to the supplier's directions, dissolved in RNase-free water and stored at −80 °C. For microsatellite or *COX*1 gene analysis, DNA was recovered from approximately 50 mg bat brain or lung tissue by grinding in a hexadecyl trimethyl ammonium bromide solution followed by phase separation with the addition of chloroform and alcohol precipitation as described previously (Desloire *et al.* 2006). The final dried DNA pellet was dissolved in TLE buffer (10 mM Tris–HCl, pH 8.0, 0.1 mM EDTA) and stored at −20 °C. Nucleic acid concentration was determined spectrophotometrically.

Table 1 Provincial distribution of numbers of specimens employed for each analysis on all 388 bat specimens

Province	RABV P sequencing	Microsatellite scoring	Barcoding (COX1)
AB	16	25	10
BC	55	82	34
NB	3	0	2
ON	125	125	39
QC	14	37	11
SK	19	26	11
Total	232	295	107

Virus sequencing and phylogeny

Total rabies-infected brain RNA was used as template to amplify the complete viral P gene using standard RT-PCR methods (Nadin-Davis 1998) with the RabP-for/RabPrev primer pair (see Table S1, Supporting Information). Thermal cycling, performed in a GeneAmp 9700 thermocycler (Applied Biosystems, Foster City, CA, USA) used the profile: 94 °C, 2 min; 94 °C, 0.5 min, 50 °C, 0.5 min, 72 °C, 2 min × 35 cycles; 72 °C, 3 min; 4 °C hold. For samples yielding little to no visible product, a second round of PCR was performed using a nested primer pair, RabP967/Rabp995 (see Table S1, Supporting Information) to generate a 664 bp amplicon; the thermocycling profile was similar to that of the first round except that it employed only 30 cycles with an extension time of 1 min. Amplicons, purified using a Wizard PCR preps purification system (Promega, Madison, WI, USA), were subjected to thermocycling sequencing reactions using a Thermosequenase kit (GE Healthcare, Baie d'Urfé, QC, Canada) and IR-dye labelled primers corresponding to the nested PCR primers. Nucleotide sequences were resolved on a NEN 4200L automated system (Li-Cor Biosciences, Lincoln, NE, USA), manually edited using E-seq ver. 2 software (Li-Cor) and downloaded in FASTA format for subsequent alignment using Clustal-X version 1.8 (Thompson *et al.* 1997). An alignment of 582 nucleotides for 237 RABV specimens, including five reference samples of other RABV strains, was employed for phylogenetic analysis using both the neighbour joining (NJ) algorithm implemented in PHYLIP version 3.63 with 1000 bootstrap replicates (Felsenstein 2005) and subsequently by a maximum likelihood (ML) method using PhyML version 3.0 with 1000 bootstrap replicates (Guindon *et al.* 2009). Of the 232 big brown bat RABV strain sequences, five were removed from the data set, as they were found to result from spillover transmission from other bat reservoirs (see section 'Results'). The final data set of 227 sequences was then used to reconstruct the phylogenetic tree with MrBayes version 3.1.2 (Ronquist & Huelsenbeck 2003). Using the DNA substitution model GTR + Γ_4 + I selected with the Akaike Information Criterion as implemented in ModelTest (Posada & Crandall 1998), two independent runs were performed for 100 million generations with sampling every 1000 generations; convergence of the two runs was observed by 8 000 000 generations, as determined by comparing time-series plots. A Maximum Clade Credibility tree was constructed using the TreeAnnotator software of the BEAST package (Drummond *et al.* 2006). Finally,

neutrality was assessed with the Tajima test (Tajima 1989), as implemented in the R package pegas (Paradis 2010); in all rigor, codon models cannot be used with our data since these models assume that nonsynonymous differences are fixed, which is not the case at the population level.

Across border comparison was not possible based on the viral P gene, as few submissions from the US for this locus are deposited in GenBank, so an N gene sequence data set was used to explore whether the Canadian situation extended south into the US. Using selected sequences generated in a previous study (Nadin-Davis *et al.* 2001), together with sequence data on US isolates recovered from GenBank, NJ and ML trees were generated. This analysis assumes, probably correctly (Chare *et al.* 2003), that no recombination occurs between the N and P genes. This assumption was tested on N and P gene sequences arranged contiguously by analysing for recombination using several local statistical methods including: Gene Conversion (Geneconv: Padidam *et al.* 1999), the Maximum chi-square test (MaxChi: Maynard Smith 1992), maximum mismatch chi-square (Chimaera: Posada & Crandall 2001), the RDP method (Martin & Rybicki 2000) and 3Seq (Boni *et al.* 2007), with use of Bootscan (Martin *et al.* 2005) and sister-scanning (SiScan: Gibbs *et al.* 2000) reserved for confirmatory testing, as implemented in the recombination detection program (RDP) version 3b41 (Heath *et al.* 2006) available at http://darwin.uvigo.es/rdp/rdp.html.

Maps of the spatial distribution of all viral variants were constructed using the software package ESRI ArcGIS ArcView, version 8.

Bayesian skyline plots of viral populations

The final data set of 227 viral sequences was further analysed in a Bayesian framework with BEAST version 1.4.8 (Drummond *et al.* 2006), first to reconstruct a rooted phylogeny from serially sampled sequences, second to estimate the divergence times of these sequences and third, to infer the past demographics of the sampled sequences. Two Markov chain Monte Carlo (MCMC) samplers were run to check convergence, and each sampler included 100 million steps with a thinning of 1000; the first 10% of each chain was removed as burn-in, as determined by time–series plots. The same model of evolution, GTR + Γ_4 + I, was used as above. Rates of evolution were assumed to follow an uncorrelated lognormal distribution (Drummond *et al.* 2006), while speciation was assumed to follow one of three demographic prior processes: the basic coalescent (constant population size), a coalescent with exponential growth and a coalescent with piecewise population size (the piecewise-constant Bayesian skyline model) with $m = 10$ and $m = 100$ groups (Drummond *et al.* 2005). Results were analysed with Tracer (tree.bio.ed.ac.uk/software/tracer). To test for the effect of potential population structure, we also split the original 227 sequence data set according to the five groups identified by our phylogenetic analysis (see section 'Results') and ran the same Bayesian skyline models as described above (with $m = 10$ only and 100 million steps in the MCMC) on each of these five groups. The strict molecular clock was tested in a maximum likelihood framework with baseml in PAML version 4.2b (Yang 2007) using the ML tree estimated above, as well as in a Bayesian framework with BEAST to take into account the fact that sequences were sampled at different dates.

Bat COX1 sequencing

Big brown bat DNA was used as template for amplification of a 707 bp 5′ proximal region of the mitochondrial *COX*1 gene with primer pair COXF2/COXR1 (see Table S1, Supporting Information). These primers, originally designated as FishF1 and FishR1 (Ward *et al.* 2005), are useful for a broad range of vertebrates. The thermocycling profile employed was 94 °C, 4 min followed by 30 cycles of 94 °C for 15 s, 45 °C for 35 s, 72 °C for 1.5 min, 72 °C for 5 min. Amplicon purification and automated nucleotide sequencing on both strands were performed as detailed above using IR-dye labelled versions of the *COX* PCR primers and the reads resolved using AlignIR software. Multiple sequence alignment was performed with Clustal-X and phylogenetic analyses were performed with PHYLIP under NJ and with PhyML version 3 under ML as described. Neutrality was assessed with the Tajima test as described above.

Bat microsatellite typing

Microsatellite loci based on dinucleotide (GT) repeats were isolated using established protocols (Glenn & Schable 2005). Clones generated in the pCR-TOPO vector (Invitrogen) were initially characterized by restriction endonuclease analysis and those having inserts in the 400–900 bp size range were sequenced. A total of nine clones containing suitable microsatellite loci were identified as detailed elsewhere (Feng 2006) and amplification primers flanking each locus (summarized in Table S2, Supporting Information) were designed using Primer 3 software (Rozen & Skaletsky 2000). The forward primer of each microsatellite primer pair included the M13 reverse primer sequence at its 5′ end to facilitate sizing and scoring as detailed previously (Kovar

et al. 2005). Microsatellite PCRs were performed in 10 μl reactions containing 0.2 mM dNTPs, 0.5 pmol each microsatellite primer, 1 pmol IR 700 labelled M13 reverse primer, 1 unit of *Taq* DNA polymerase and 20 ng of genomic DNA. The thermocycling profile was 95 °C, 2 min; 95 °C, 30 s, 55 °C, 40 s, 65°C, 1 min × 30 cycles; 65 °C, 7 min; 4 °C hold. Certain reactions could be multiplexed due to substantial differences in size of the loci under study. Products were analysed by denaturing gel electrophoresis run through a 0.25 mm 6% polyacrylamide gel on a NEN 4200L sequencing system and sized using an IR 700 dye-labelled 50-350 DNA size standard electrophoresed in multiple wells across the gel in each run. SAGA generation 2 software (Li-Cor) was employed for automatic allele size scoring.

Analysis of bat population structure

GenePop was used to test the microsatellite data for Hardy–Weinberg equilibrium (HWE) and for linkage disequilibrium. *P*-values were estimated with 1000 permutations with Genetix version 4.05.2 (Belkhir *et al.* 1996, 2004). GenePop was also used to test for null alleles, whose impact was assessed by generating a secondary data set with FreeNA (Chapuis & Estoup 2007), a program that removes the effect of null alleles.

Evidence for selection was assessed with LOSITAN (Antao *et al.* 2008), which is built upon FDIST (Beaumont & Nichols 1996). The method consists in using the relationship between F_{ST} and heterozygosity to detect loci that are potentially under balancing or positive selection. Confidence intervals are generated by coalescent simulations under a simple island model with constant population size. Here, we performed 100 000 replicates. The program was run under the two mutation models implemented, the infinite allele and the stepwise models. The neutral F_{ST} was estimated in two steps as in Antao *et al.* (2008) by checking the options 'neutral mean F_{ST}' and 'force mean F_{ST}'.

The population structure predicted from these data was examined by four different approaches. First, principal components were extracted with R (cran.r-project.org) from the microsatellite loci. This analysis does not assume anything about the population structure, which was estimated by means of the Partitioning Around Medoids algorithm run on the first two principal components. The optimal number of populations or clusters was determined according to the Median Split Silhouette (Pollard & van der Laan 2002). Briefly, silhouettes measure the homogeneity of each principal component that is grouped in a cluster compared to the situation where this component is left out. The median of these measures taken over all components and over all clusters is a measure of the homogeneity of each cluster, and hence constitutes an objective function to find the most appropriate number of clusters. This objective function is a measure of cluster heterogeneity that can be minimized to estimate the number of clusters.

Second, fixation indices (F_{ST} and R_{ST}) between pairs of samples were estimated with GenePop version 4.0 (Raymond & Rousset 1995). Numbers of migrants were estimated using the estimate first derived by Wright (1943), under the assumptions of the infinite allele model and neglecting quadratic terms in the expression of F_{ST} at equilibrium (see Hartl & Clark 2007). Third, we used the population genetics-based models implemented in Structure version 2.3.3 (Pritchard *et al.* 2000) both to infer the number of clusters (or subpopulations) and to assign individuals to these subpopulations. We either assumed the presence of admixture, or not, and the analysed loci were considered to be independent. Each MCMC sampler was run for 1 000 000 steps and the first 250 000 steps were discarded as burn-in periods, as very conservatively determined by time–series plots. Convergence was carefully checked as with the MCMC samplers run above on the DNA alignments. The number of clusters was estimated by taking inspiration from the gap statistic (Tibshirani *et al.* 2001), which is the actual justification of the method used by Evanno *et al.* (2005). Briefly, Structure was run for K subpopulation for K varying from 1 to 10 and K = 15 and 20. Each MCMC sampler was run twice under each K to check for convergence. The harmonic mean of the log-likelihood values sampled from the posterior distribution, an estimate of the marginal log-likelihood (Newton & Raftery 1994) was then plotted as a function of K. The optimal number $K^*$of populations was taken at the largest gap of marginal log-likelihood between two consecutive values of K.

Fourth, the genetic analysis was combined with the geographic information about exact sampling locations with Geneland version 3.1.5 (Guillot *et al.* 2008), which clusters genetic samples into subpopulations so that each achieve HWE and each loci are in linkage equilibrium. The spatial distribution of subpopulations is assumed to follow the coloured Poisson–Voronoi tessellation model. Allele frequencies are assumed to be independent and to follow a Dirichlet prior distribution. Under this general setup, two MCMC samplers were run for 10^6 steps and a thinning of 100. The first 2×10^5 steps were discarded as burn-in periods, as determined by time-series plots. First, the number of subpopulations was estimated by setting a uniform prior $U(0,50)$ and running the samplers described above. Then, following Guillot *et al.* (2008), the same samplers were run by fixing the number of subpopulations to that estimated in the first series of runs. Two models were run,

either assuming that allele frequencies are correlated or uncorrelated.

The nucleotide sequence data generated during the course of this study have been deposited to GenBank and assigned the following accession numbers: GU207610–GU207834 for the rabies virus P gene, and GU207502–GU207609 for the big brown bat COX1 gene.

Results

RABV phylogeny and spatial distribution of variants

Phylogenetic analysis of 237 RABV specimens, characterized over a 582 bp coding region of the P gene, was initially performed by a NJ analysis (Fig. S1A, Supporting Information). To allow comparison with prior results (Nadin-Davis *et al.* 2001), this collection included previously characterized viruses from both the big brown bat host and single representatives of the viruses known to circulate in other bat species. The resulting tree identified six main clades, all of which were supported with high bootstrap values, and a separate branch for the little brown bat RABV specimen. The most divergent clade (designated as 'spillover' group) comprised eight samples including five big brown bat specimens together with the two Lasiurus variants and the silver-haired variant. This grouping illustrates that five specimens recovered from big brown bats represented viral strains normally associated with other bat hosts: Four were of the silver-haired bat type and one of the hoary bat type. These five sequences, which are unrepresentative of big brown bat viruses, were removed from all subsequent studies. The remaining five clades, which represent viruses normally associated exclusively with big brown bats, were designated BB1-5 (Fig. S1A, Supporting Information). The ML analysis returned a very similar tree topology (Fig. S1B, Supporting Information).

Virus population structure and dynamics

The final data set of 227 RABV sequences was then examined by Bayesian methods. First the Bayesian generated a consensus tree with a clustering pattern virtually identical to that observed by NJ (not shown). Second, the strict molecular clock was severely rejected ($2\Delta l = 588.04$, $P < 0.0001$; $\log_{10}$ Bayes factor in favour of a relaxed clock = 4.85, which is 'substantial' evidence [Kass & Raftery 1995;]), thereby justifying the use of a relaxed molecular clock model. Third, using the Bayes factor as implemented in BEAST (see Suchard *et al.* 2005), the three coalescent prior distributions (constant population size, exponential growth and piecewise coalescent) were compared. The estimated marginal $\log_{10}$-likelihoods of the three models were −3039.667, −3038.397 and −3036.268, respectively, hereby suggesting that a piecewise coalescent prior model is strongly favoured. All the results below are therefore based on this model. The Bayesian phylogenetic analysis generated a topology in excellent agreement with that of the NJ and ML analyses performed above and suggests that all big brown bat rabies viruses can be traced back to a series of relatively ancient diversification events, but that a variant of more recent origin is most widespread (Fig. 1). Tajima's D was found to be slightly negative ($D = -0.243$), but not significantly different from 0 ($P = 0.860$, with the assumption that D follows a beta distribution under the null hypothesis of neutrality). The spatial distribution of each of these five groups across Canada is illustrated in Fig. 2.

The most recent common ancestor (TMRCA) for the entire RABV lineage currently circulating in big brown bats was estimated to have emerged in 1573 (95% HPD: 1763–1338) and five main clades have subsequently emerged from that original progenitor. Clade BB1, which has emerged independently from the rest of the other groups is exclusively represented in the southern regions of the BC mainland and around Vancouver Island. BB2 is part of a lineage that diversified in 1634 (1802–1484), and is currently only found in southern QC, eastern ON, the Niagara region of ON immediately adjacent to the US border and NB; the viruses harboured within the latter two regions, which are geographically segregated from the other members of this clade, form a distinct sub-branch (designated BB2a) of the BB2 clade. The progenitor of the lineage that spawned clades BB3, BB4 and BB5 emerged around 1845 (1885–1704) and diverged shortly thereafter to generate the viruses that would yield the closely related BB3 and BB4 clades which are almost exclusively represented today in BC; a single BB3 isolate was recovered from SK. All the remaining sampled diversity belongs to clade BB5, which emerged recently, from 1936 (1939–1838) and which is widely distributed from BC to ON; the BB5a subclade, which diversified around 1991 (1993–1972), is found only in eastern Canada.

The Bayesian skyline analysis reveals that a dramatic change in population size occurred approximately a decade before 2007 (Fig. 3A). From the past 250 years at least, the ancestral effective population size of infected bats scaled to time ($N_e\ \tau$) was stable around 250 individuals. Then from about 1950 to 1995, a small decline to $N_e\ \tau \sim 160$ can be detected, followed by a dramatic increase of almost an order of magnitude to ~1000 infected individuals. However, since about year 2000, viral population sizes appear to have stabilized. Similar results, with larger credibility intervals, were obtained when assuming $m = 100$ groups in the skyline

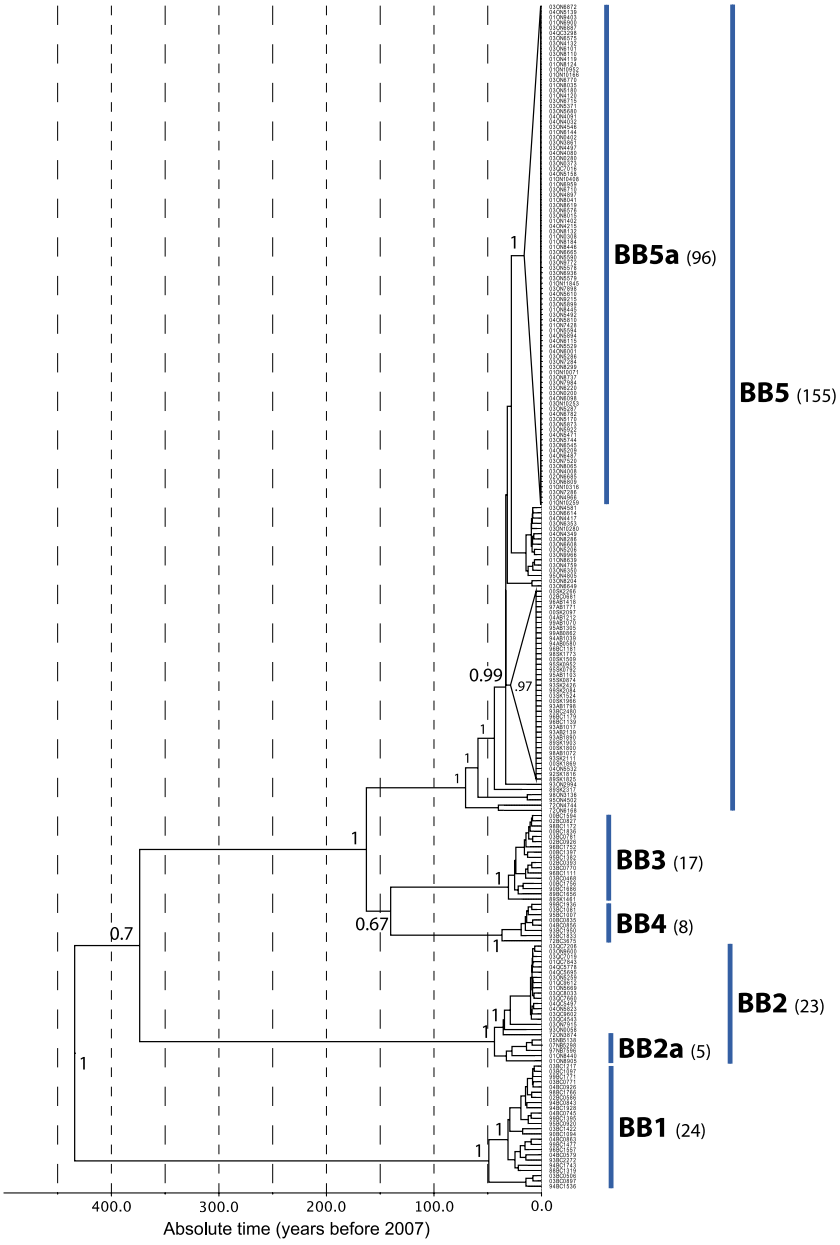

Fig. 1 Viral P gene phylogeny scaled to time. Branch lengths are in units of absolute time (years before 2007). Numbers at internal nodes represent clade posterior probabilities.

analysis (not shown). Under the assumption that the clades detected in the phylogenetic study (Fig. 1) represent a genuine population structure, a clade-by-clade skyline analysis shows that clades BB1, BB2 and BB3/BB4 show stable effective population sizes throughout the depth of the subtrees (Fig. 3B,C). Only clade BB5 exhibits a recent dramatic increase in effective population size (Fig. 3C) such that clade BB5 appears to be solely responsible for the recent dramatic increase in effective population size. Taken altogether, the phylogenetic and the skyline analyses suggest that a recent event triggered a rapid diversification and spread of rabies viruses belonging to clade BB5 across Canada.

RABV variant comparison with US

To put into context the results obtained above for Canada, the distribution of these RABV variants in the US was also explored. Using N gene sequences for specimens representative of each BB clade, comparison to 11 RABVs recovered from several US states was performed by a ML analysis (Fig. 4). This phylogeny supports the following distribution of variants within the US: viruses belonging to clade BB1 were found in specimens from AZ, CA and CO; viruses of clades BB2 were found in PA and CT; a single BB3 virus was recovered from WA, no US specimen clustered within clade BB4 while some CO and PA specimens grouped within BB5,

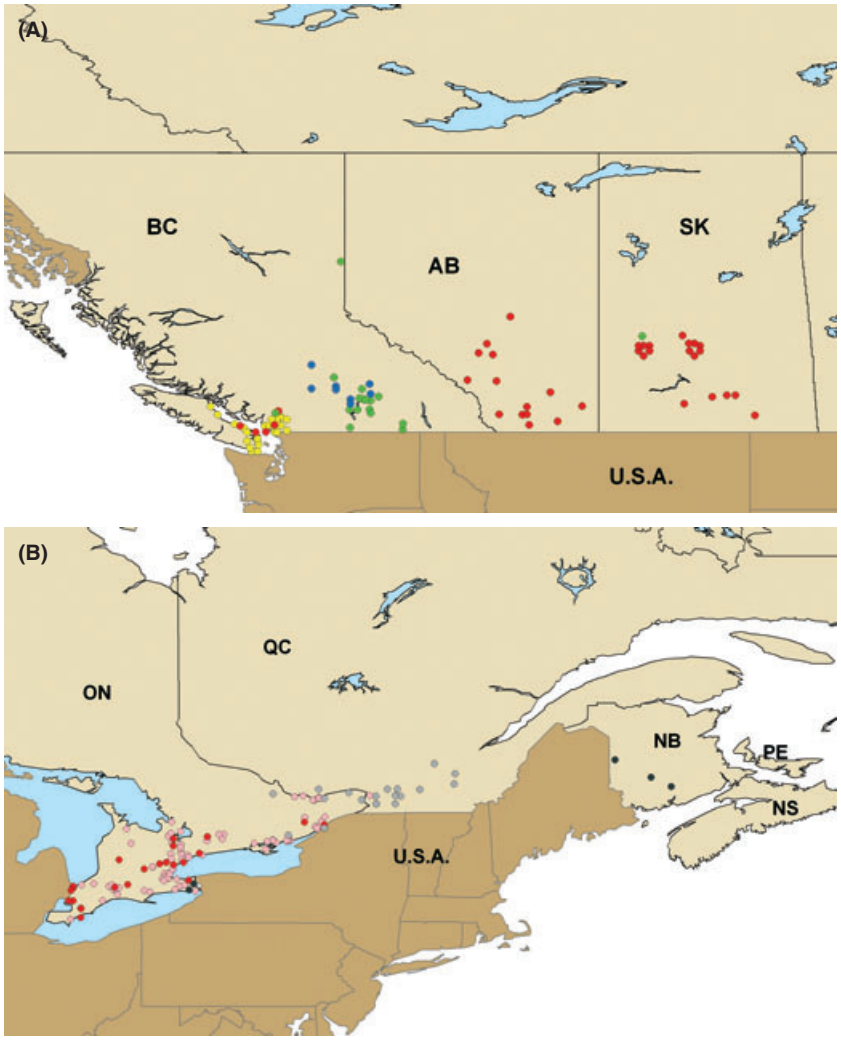

Fig. 2 Maps showing the distribution of all characterized RABVs according to their variant designation in western (A) and eastern Canada (B). Viral variants are colour-coded as follows: BB1, yellow; BB2, grey; BB2a, black; BB3, green; BB4, blue; BB5, red; BB5a, pink.

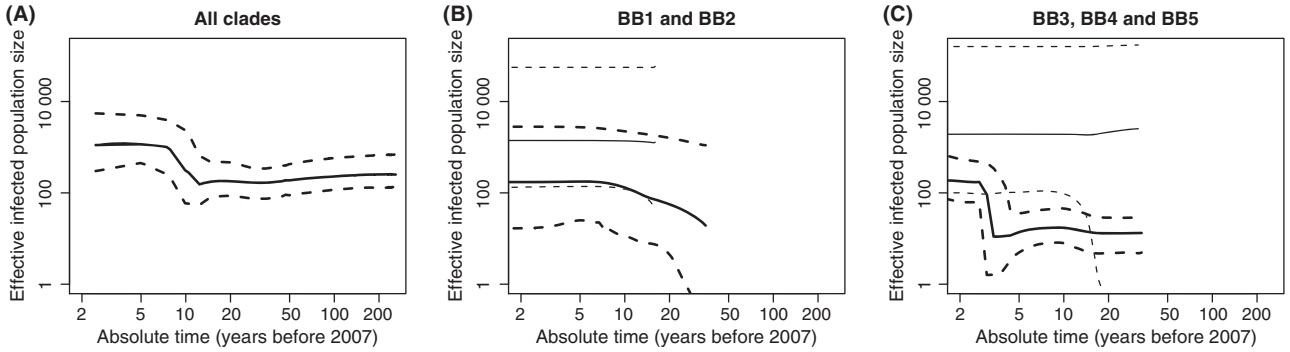

Fig. 3 Bayesian skyline plot for all sampled P genes. (A) all data analysed simultaneously; (B) individual analyses of the BB1 (thin lines) and BB2 clades (thick lines); (C) individual analyses of the BB3 plus BB4 (thin lines) and BB5 clades (thick lines). The x axes are in years before 2007, and the y axes are equal to $N_e\,\tau$, the product of the effective population size and of the generation time in years. The 95% HPD envelop (upper and lower curves) are represented in each case.

consistent with a widespread distribution of clade BB5 across North America. To confirm that these results were not affected by recombination events between the N and P genes, a data set of 12 samples (1930 bases), for which sequences at both targets were available, was analysed for recombination using the RDP software.

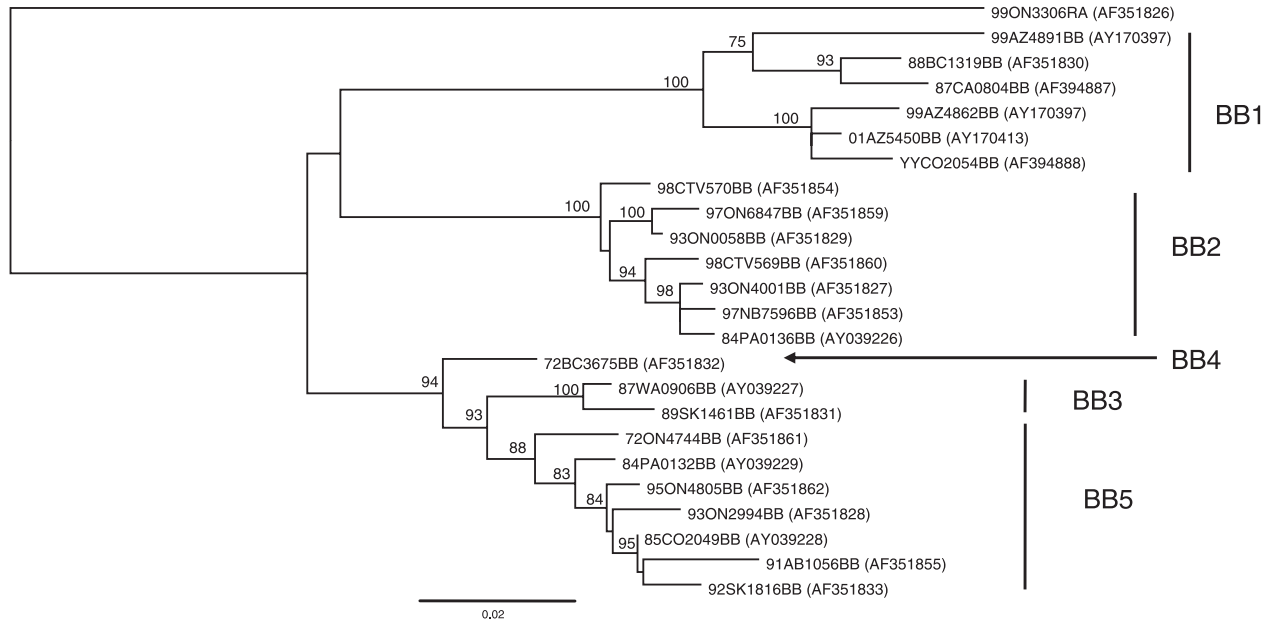

Fig. 4 Phylogeny of 23 North American big brown bat viral N gene sequences by ML analysis. Each sample designation is followed in brackets by its corresponding GenBank accession number. The group designation is indicated to the right of each clade. Branch lengths represent genetic distance based on the scale at bottom. Numbers at internal nodes represent bootstrap values as percentages for the clade above or to the right. A raccoon strain of rabies (99ON3306RA) was used as an outgroup.

Using a P value cut-off of 0.01 with Bonferroni correction no recombination event was identified by the independent detection methods used. Similarly nonsignificant results were obtained for the 227 big brown bat P sequences.

COX1 gene characterization of bat phylogeny

As DNA barcoding has yielded novel insights into bat species identification, we explored its use to examine the subspecies structure of big brown bats across Canada. Partial *COX*1 gene sequences were determined for 107 big brown bats collected from across the country; just over half (63) of these specimens were rabies-positive. Alignment of these data established that there was significant sequence variation within the group as confirmed by a NJ analysis (data not shown) and a ML analysis (Fig. 5) which divided the sequences into two groups (A and B) comprising 74 and 33 individuals, respectively. Maximal genetic divergence within the two groups ranged between 0–0.614% (group A) and 0–0.153% (group B) while intergroup genetic divergence was 7.842–8.185%. Group B bats all originated from BC while all group A bats, except for one that originated from the BC/AB border east of the Rocky Mountain range, came from other parts of the country, from AB to NB. The 20 rabies-positive group B bats from BC harboured all four RABV variants (6, BB1; 5, BB3; 6, BB4 and 3, BB5) recovered from the province. The single BC bat that clustered with the COX A group was also unusual in that it harboured a BB3 RABV variant located well away from the locations of other isolates of this variant (see Fig. 2). Tajima's D was found to be slightly negative ($D = -0.392$), but not significantly different from 0 ($P = 0.740$), so that this marker behaves neutrally.

Bat population structure: the view from the nuclear genome

A total of nine microsatellite loci were identified and characterized as detailed in section 'Materials and methods'. Attempts to score 295 bat specimens at all nine loci resulted in an overall success rate of 98.64%. All loci were highly polymorphic and yielded a total of 338 distinct alleles; the number of alleles per locus ranged from 13 to 39. All loci appeared to be in linkage equilibrium (P values between 0.0808 and 1.0000), while the HWE across all loci and all sampled populations is severely rejected ($P < 0.0001$). Similar results for linkage and Hardy–Weinberg expectations were obtained after accounting for null alleles (see below).

Neutrality was assessed based upon the method implemented in FDIST under two mutation models. At the 99.5% level, the infinite allele model suggested that locus 7 was putatively evolving under positive selection and that loci 3, 5 and 9 were under balancing selection, while the stepwise mutation model suggested that locus 7 evolves adaptively and locus 5 is under balancing selection (Fig. S2, Supporting Information). These

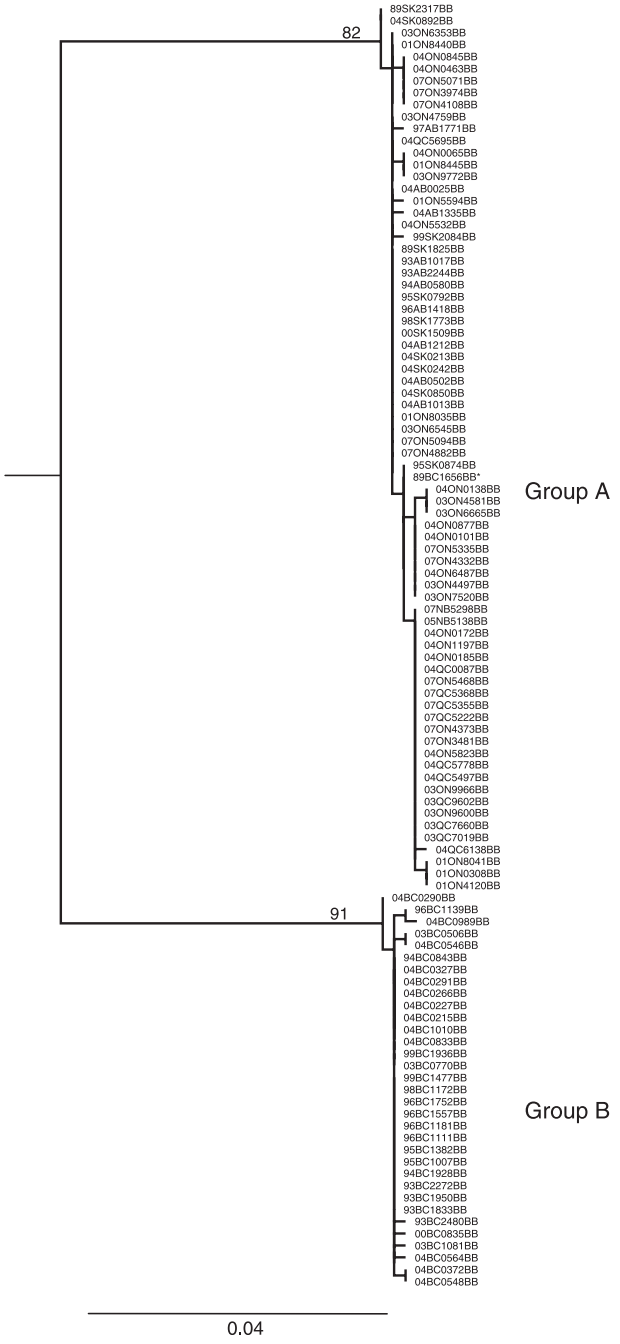

Fig. 5 Phylogeny of 107 bats by ML analysis of mitochondrial *COX*1 gene sequences. Branch lengths represent genetic distance based on the scale at bottom; the tree was rooted using a silver-haired bat sequence (06AB0448SH) as outgroup (not shown). Numbers at internal nodes represent bootstrap values as percentages for the clade to the right. The two groups A and B are identified. The single BC specimen clustering in group A is identified by *.

results therefore suggest that at least loci 7 and 5 may not be neutral. All subsequent analyses were performed without these two loci. The HWE across all loci and all sampled populations is still severely rejected ($P < 0.0001$), which shows that selection is not solely responsible for this departure from HWE.

As detailed below, four distinct methods of microsatellite loci analysis identified two big brown bat subpopulations designated as eastern and western. Both subpopulations show very high and highly significant ($P < 0.0001$) F_{IS} values on the original raw data, 0.4498 for the western subpopulation and 0.4024 for the eastern subpopulation with very similar F_{IS} values ($\sigma^2(F_{IS}) = 0.0266$) determined for all analysed (7) loci. While this might have suggested a high level of inbreeding, evidence for null alleles was found for all analysed loci. After adjusting the data for null alleles using the software package FreeNA, all evidence for inbreeding disappeared (Table 2), so that the large F_{IS} values represent solely the effect of null alleles.

Four approaches were applied to these data to examine the bat population structure. The simple principal component analysis (PCA), for which the first three components explain 21.4% of the total variance, strongly suggests the presence of two subpopulations. Those can be identified visually with respect to the origin of the samples: the eastern (QC and ON) and the western (SK, AB and BC) subpopulations (Fig. 6). The silhouette analysis confirms that the optimal number of clusters (subpopulations) is two.

This initial result is also supported by the estimation of pairwise F_{ST} values (Table 3), which show a low and nonsignificant differentiation within eastern (QC and ON) and within western (SK, AB and BC) subpopulations ($F_{ST} < 0.93\%$), and much larger and significant indices between these two subpopulations ($F_{ST} > 3.90\%$). This latter F_{ST} bound still translates into approximately five migrants per generation, which is > 1 and therefore considered to be large enough for population homogenization under the standard island and infinite site models. Similar values were estimated for pairwise R_{ST} (not shown), which suggest that our results are robust to the maker type and to the mutation model used to analyse these microsatellite data.

Table 2 F_{IS} values of sampled populations averaged over all seven loci evolving neutrally

	F_{IS} (raw data)	F_{IS} (FreeNA data)
AB	**0.4038**	−0.0027
BC	**0.4843**	−0.0404
ON	**0.3666**	−0.0148
QC	**0.4196**	−0.0489
SK	**0.3722**	−0.0250

Significant values ($\alpha = 5\%$) are in bold face.

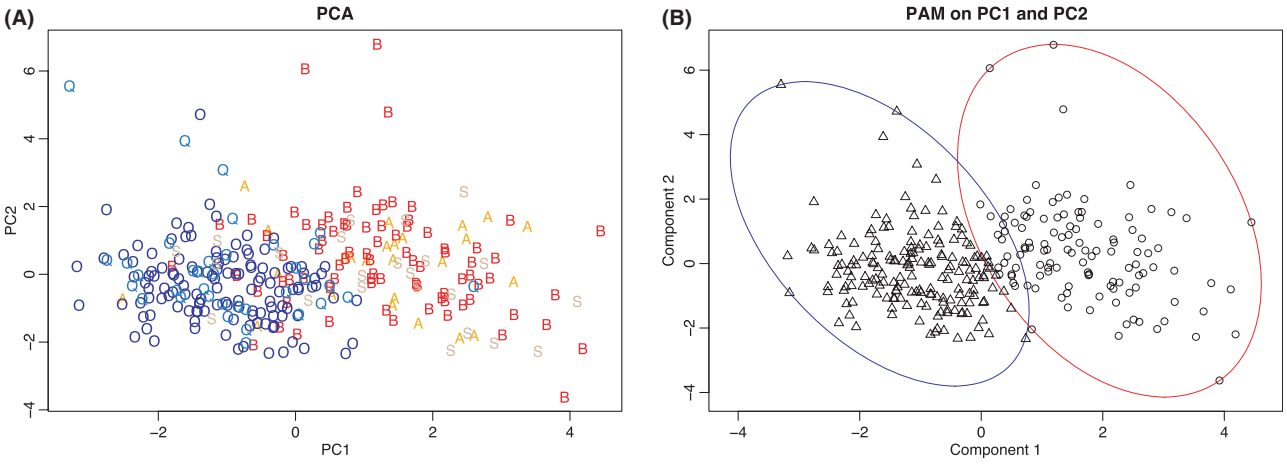

Fig. 6 Principal component analysis of bat microsatellites. (A) Projection of allelic data on the first two axes of the PCA. Note that the province of origin is represented by just a single letter as follows: A, AB; B, BC; O, ON; Q, QC; S, SK. B: Partition Around Medoid cluster analysis on the first two axes of the PCA where the number of clusters was identified by Median Split Silhouette. The symbols (Δ) and (O) represent members of the eastern and western samples, respectively.

Table 3 Pairwise F_{ST} values between pairs of sampled populations

	AB	BC	ON	QC	SK
AB	–	0.0087	**0.0416**	**0.0483**	0.0054
BC	0.0093	–	**0.03889**	**0.0440**	0.0056
ON	**0.0495**	**0.0502**	–	0.00034	**0.0335**
QC	**0.0563**	**0.0561**	0.0021	–	**0.0384**
SK	0.0074	0.0049	**0.0390**	**0.0455**	–

F_{ST} values on raw data are above the diagonal; F_{ST} values on the FreeNA data, where the effect of null alleles is removed, are below the diagonal. Significant values ($\alpha = 5\%$) are in bold face.

The more detailed analysis based on the population genetics model implemented in Structure showed that, in spite of the large number of migrants suggested under the simple island model with the assumption of the infinite allele model, there is a strong signal supporting the two subpopulations detected by PCA (Fig. S3A,B, Supporting Information; solid lines). Furthermore, this signal is independent of the level of admixture (Fig. S3A,B, Supporting Information; broken lines). This number of subpopulations $K = 2$ is also confirmed by the Geneland analysis, that incorporates geographic information (Fig. S4, Supporting Information). Although Fig. S4 (Supporting Information) suggests that the posterior probability for $K = 2$ and $K = 3$ subpopulations are relatively close (42% vs. 33%), additional analyses where K was fixed to 2, 3, 4 and 5 showed that all individuals were still assigned to two subpopulations (Fig. S5, Supporting Information). Similar results were obtained under a model where loci were correlated (not shown).

The integration of these results with the geographic information about sampling locations with Geneland suggests again a strong signal for the existence of two subpopulations of bats, one eastern (QC and ON) and one western (SK, AB and BC), well differentiated genetically and spatially (Fig. 7). Irrespective of the correlation model assumed for the loci, the relationship between these two subpopulations is identical in that the map of posterior probabilities of subpopulation membership shows an apparently smooth gradient between the eastern and the western subpopulations (Fig. 7). However, as no further structure could be detected by Geneland within the western and the eastern subpopulations, it is possible that central Canada represents a severe barrier to gene flow between eastern and western Canada, so that the apparently smooth gradient (Fig. 7) is actually very steep. A greater sampling effort should focus on Central Canada, particularly MB and western ON, in order to confirm the complete genetic structure of bat populations in Canada.

Discussion

Viral phylodynamics

The relatively high mutation rate of pathogens, in particular of viruses, can reveal the population structure and recent demographics of their host but the type of information revealed by such studies depends on the extent to which the evolutionary histories of pathogens and hosts correlate (reviewed by Wirth *et al.* 2005). In a study of raccoon RABV spread in the mid-Atlantic region of the US, discrete viral lineages radiated out

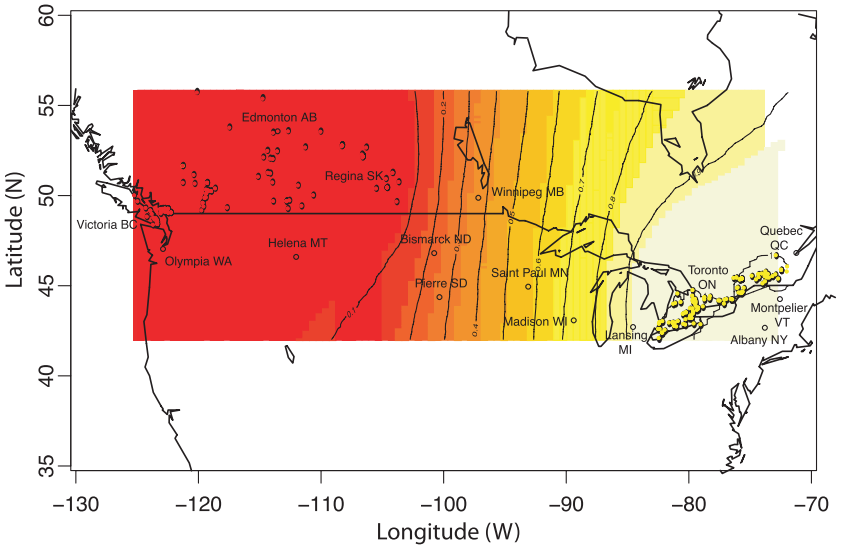

Fig. 7 Map of posterior probabilities to belong to the eastern (ON, QC) population across the entire sampling range. Sampled individuals are coloured according to their assigned population: red for the western and yellow for the eastern population. Provincial and select northern state capitals are indicated (open circles). The number of population K is assumed to follow a uniform prior distribution $U(1,50)$ and loci are assumed to be uncorrelated.

from the outbreak origin to generate spatially restricted virus groups which persisted throughout the epizoonosis (Biek *et al.* 2007). That study identified periods of exponential viral population growth during northwards progression of the epizoonosis interspersed with periods of constant population size, a phenomenon likely linked to habitat and host density; mountain ranges acted as major barriers to rabies spread due to reduced host density in forested regions of higher elevation. A more global study of the phylogeography of RABVs in carnivores found that the virus' phylogenetic structure is dominated by population subdivision rather than gene flow such that geographically separate groups of viruses with relatively limited inter-group contact have emerged around the world (Bourhy *et al.* 2008). To compare these phylodynamic patterns observed for RABVs associated with terrestrial mammals with those for viruses associated with a mammal having a distinct aerial lifestyle, we have conducted an extensive population genetics study of the big brown bat in Canada and a phylogenetic analysis of its associated RABVs.

Integral to this analysis was an estimation of the time scale over which rabies has circulated in this bat host. The Bayesian analysis of 227 RABV P gene sequences estimated the nucleotide substitution rate for this gene at 1.77×10^{-4} per site per year, a value slightly below the range (2.32×10^{-4} to 1.38×10^{-3}) previously reported for the N gene of American bat RABVs (Hughes *et al.* 2005), but within the range estimated for the N gene of viruses of European and arctic foxes (Kuzmin *et al.* 2008). Under a relaxed molecular clock, we estimated that the rabies virus variants currently associated with these bats in Canada evolved at approximately 1573 (95% HPD: 1763–1338), a significantly earlier date than the 1825 date suggested for North American big brown bats by Hughes *et al.* (2005). It is unclear if the differences in substitution rates estimated in the two studies can account for this difference in time scaling of the respective evolutionary trees, as Hughes *et al.* (2005) also used a simplistic constant population size coalescent prior which was strongly rejected by our data in favour of the skyline model. Besides, this time scale difference is unlikely to reflect an earlier evolution of the viruses in Canada as US isolates form a continuum with those found in Canada, and it is likely that several of the viruses currently circulating in Canada emerged in the US (see below). It is thus apparent that the introduction of the current rabies virus lineage into this host species is, in evolutionary terms, a very recent event given that the fossil record suggests that the big brown bat was widely distributed in North America during the Pleistocene period (reviewed by Kurta & Baker 1990). The possibility remains that a distinct rabies virus lineage had been associated with this species in the past but has been replaced with the viruses presently in circulation.

Phylogenetic analyses show that RABVs in big brown bats of Canada form five clades that have distinct geographic locations and past population dynamics. On each coast, the BB1/BB3/BB4 (West) and BB2 (East) clades are relatively ancient, dating back over 400 years, and these clades show no evidence of any significant change in their population dynamics over the past 200 years. These observations would appear to be in accord with the patterns of virus spread observed in nonflying mammals (Bourhy *et al.* 2008). In contrast, the relatively cosmopolitan clade BB5, which is found from BC to ON, and which constitutes the majority of our passive sample collection (68% of sampled rabies-positive bats), shows evidence of a dramatic increase in

prevalence between 2002 and 2004, with a subsequent stabilization. The deepest branches within the BB5 clade were all sampled from ON and only in more recent times were specimens of this type recovered in the more central and western provinces of SK, AB and BC (Fig. 1). If we assume that human translocation of infected bats is unlikely, then this evidence from the skyline analyses would suggest that the BB5 clade first emerged in the east, where this lineage grew very rapidly, followed by a rapid spread to western regions either by a direct route, through central Canada, or a more southern route through the central states of the US. However, in the absence of samples collected from central Canada, these two hypotheses cannot currently be teased apart. Additionally a significant and recent (around 1991) branching event within clade BB5 generated the BB5a subtype that appears to be restricted to southern and eastern ON and bordering areas of QC.

To further assess the hypothesis of a continuum of viral populations across North America, we included RABV sequences sampled from the US in our study. Similarly to the situation in Canada, US members of the BB1 clade were only sampled from western states of CA, CO and AZ, and BB3 viruses were found in WA. On the other side of the continent, the US BB2/2a viruses were recovered from the eastern states of PA and CT, probably forming a continuum with the Canadian BB2 variants from ON, QC and NB. Representatives of the central clade, BB5, in the US were also found in PA and CO. Despite the limited number of available US sequences, this general distribution pattern of US rabies viruses suggests that the phylodynamic patterns found in Canada reflect the situation throughout North America, and in particular that the spread of the BB5 clade to the west coast was an event that occurred throughout the entire North American continent. Moreover, the recovery of BB1 and BB2 variants only from areas of Southern Canada close to the US border, together with the more widespread circulation of these variants in the US, suggests to us that these variants probably emerged in the US and spread northwards into Canada.

Host population structure

In parallel to the elucidation of the viral phylodynamics, the host mitochondrial *COX*1 analysis suggests a significant divide between the big brown bats of BC and those from the rest of Canada. Indeed, the approximately 8% inter-group genetic divergence compared to a maximum intra-group distance of 0.614% is significant given that intra-species variation at the *COX*1 locus is normally not greater than 2–3% (Hebert *et al.* 2003) and a 10-fold difference in genetic distances between inter-species and intra-species groupings has been proposed as a threshold for species differentiation (Hebert *et al.* 2004). However, no further delineation of subspecies is suggested by our *COX*1 phylogeny.

In contrast to this *COX*1 divide for the host between BC subpopulations and those from the rest of Canada, the microsatellite data show evidence for two major Canadian bat subpopulations, one in the west (BC, AB and SK) and one in the east (QC and ON). On the basis of both the *COX*1 and microsatellite data we thus distinguish three bat groupings from our studies: bats of BC (Cox B, MSAT West), bats of the remaining western provinces (Cox A, MSAT West) and the bats of eastern Canada (Cox A, MSAT East). The identification of these three groups is entirely consistent with the range of three subspecies previously identified in Canada (Kurta & Baker 1990): the BC big brown bats with *E. fuscus bernardinus* (Rhoads 1902) which is found in the far west (primarily the province of BC), the remaining western bats with *E. fuscus pallidus* (Young 1908) which is located in Alberta, Saskatchewan and Manitoba and all eastern bats with *E. fuscus fuscus* (Palisot de Beauvois 1796) which occupies the eastern half of this bat's Canadian range. Further exploration of the correlation of our genetic studies with these subspecies assignments is in order. Based on our microsatellite studies the western and eastern subpopulations are either linked by a smooth migration gradient, or separated by an area in central Canada that is devoid of *E. fuscus* hosts (Fig. 7). Again, a more extensive sampling from this region would be necessary to assess these hypotheses.

However, because the patterns of divergence of the nuclear microsatellite markers and the maternally-inherited mitochondrial genome are highly contrasted, we posit that a high level of female philopatry is countered by strong male-biased gene dispersal, in particular in the BC subpopulation. Although it remains possible that other mechanisms could explain our data, female philopatry has been invoked to explain rather similar observations made for microsatellite and mitochondrial loci in *Myotis bechsteinii* in Europe (Kerth *et al.* 2002) and in Daubenton's bat in Scotland (Ngamprasertwong *et al.* 2008). Moreover, a capture–recapture study of European Bat Lyssavirus type 1 prevalence in colonies of *E. isabellinus* in Spain found evidence for highly philopatric female behaviour (Vázquez-Morón *et al.* 2008). Thus the two *COX*1 host groupings characterized in this study may be interpreted by BC females (Cox B group) with a limited migration range and limited migration habits, perhaps due to mobility restrictions caused by the Rocky Mountain range that runs down the eastern side of the province, while the Cox A group, distributed across the rest of Canada, may be interpreted by less philopatric females.

Joint bat/RAVB phylodynamics

The BC subpopulation carries several rabies lineages (BB1/BB3/BB4) distinct from those (BB2/BB5) that circulate in the east. Only BB5, a lineage that has emerged relatively recently, is widely distributed across much of the country. Compared to the older rabies lineages, BB5 appears to have a higher fitness that explains its larger prevalence today. The high spatial correlation of the BC RABV variants with their distinctive *COX*1 locus is suggestive of a strong sex-biased pattern of viral transmission in which virus spread occurs to a significant degree in nurseries where males are largely excluded. In contrast, the relatively rapid cross-continental spread of the BB5 variant would appear to be more likely due to transmission by less philopatric males. Further investigation of the ecological and biological factors that influence RABV transmission in this species might shed some light onto this apparent inconsistency and explore the possibility that the emergence of the BB5 variant has occurred in some measure due to changes in bat to bat transmission patterns.

Few prior studies have explored the links between North American big brown bat population structure and the circulation of discrete RABV variants. Using several (but linked) mtDNA markers to identify multiple big brown bat lineages, Neubaum *et al*. (2007) showed that two of these lineages circulate in CO. However, there was no correlation between host lineage and circulation of two RABV variants, corresponding to groups BB1 and BB5 of this study (see Fig. 4; Shankar *et al*. 2005; Neubaum *et al*. 2008). It was suggested that although bats of the area were separated into two distinct subpopulations several thousand years ago, over the last two centuries human activity has facilitated greater bat comingling during a period of RABV variant emergence. This situation contrasts with that in Canada where the two identified mtDNA lineages show clear phylogeographic structure.

To conclude, although there is little evidence for increased case incidence, the data presented in this work indicate that rabies in big brown bats shows almost all the characteristics of a truly emerging zoonosis in Canada and more generally in North America since the most common lineage now extant in this species has evolved recently. In this regard, the timescale of emergence of the viruses associated with other bat reservoirs in Canada would be interesting to explore. Particularly notable is the importance of bat strains of rabies virus from a public health standpoint since the last four cases of human rabies in Canada (in 1985, 2000, 2003 and 2007) were all due to bat variants. Rapidly changing human demographics and the reported recent northward range expansion and growing population densities of the big brown bat, facilitated by increasing availability of heated attics which can be used by bats as hibernacula (Agosta 2002), increase the likelihood of human–bat contacts. Thus, this emergence of rabies in big brown bats in Canada will be a matter of increasing public health concern. Furthermore, the occasional spillover of bat strains into wild and domestic animals could, in rare instances, potentially establish new viral-host associations, as has occurred in skunks in Flagstaff, AZ (Leslie *et al*. 2006). In this regard our identification of five spillover events involving viruses normally harboured by other chiropteran hosts is notable. Improved surveillance together with better epidemiological information on bat-associated rabies outbreaks will help to identify emerging threats to both human and animal health due to this deadly disease.

Acknowledgments

We acknowledge the excellent technical assistance of E. Guerrero and K. Knowles in the *COX*1 gene characterization. We acknowledge financial support to YF from the Ontario Ministry of Natural Resources during the course of this study. We thank A. Clark for producing the maps of the viral variant distributions. We also thank Alexander Borisenko (University of Guelph) and Leslie Real (Emory University) for helpful comments during the preparation of this manuscript. This work was supported by the Natural Sciences and Engineering Research Council (SAB).

References

Agosta SJ (2002) Habitat use, diet and roost selection by the Big Brown Bat (*Eptesicus fuscus*) in North America: a case for conserving an abundant species. *Mammal Review*, **32**, 179–198.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.

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

Baer GM, Smith JS (1991) Rabies in nonhematophagus bats. In: *The Natural History of Rabies* (ed. Baer GM), pp. 341–366, 2nd edn. CRC Press, Boca Raton.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, B Biological Science*, **263**, 1619–1626.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) *GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*, **176**, 1035–1047.

Bourhy H, Reynes J-M, Dunham EJ *et al*. (2008) The origin and phylogeography of dog rabies virus. *Journal of General Virology*, **89**, 2673–2681.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution.*, **24**, 621–631.

Chare ER, Gould EA, Holmes EC (2003) Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *Journal of General Virology*, **84**, 2691–2703.

Childs JE, Trimarchi CV, Krebs JW (1994) The epidemiology of bat rabies in New York State, 1988–92. *Epidemiology and Infection*, **113**, 501–511.

Clare EL, Lim BK, Engstrom MD, Eger JL, Hebert PDN (2007) DNA barcoding of neotropical bats: species identification and discovery within Guyana. *Molecular Ecology Notes*, **7**, 184–190.

Desloire S, Moro CV, Chauve C, Zenner L (2006) Comparison of four methods of extracting DNA from *D. gallinae* (Acari: Dermanyssidae). *Veterinary Research*, **37**, 725–732.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular and Biological Evolution*, **22**, 1185–1192.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**, e88.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Felsenstein J (2005) *PHYLIP (Phylogeny Inference Package) version 3.6*. Distributed by the author Department of Genome Sciences, University of Washington, Seattle.

Feng Y (2006) *Molecular epidemiological analysis of rabies viruses associated with population structure of bat hosts*. MSc thesis, University of Ottawa, Ottawa, Ontario, Canada.

Frézal L, Leblois R (2008) Four years of DNA barcoding: current advances and prospects. *Infection, Genetics and Evolution*, **8**, 727–736.

Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*, **16**, 573–582.

Glenn TC, Schable NA (2005) Isolating microsatellite DNA loci. *Methods in Enzymology*, **395**, 202–222.

Grenfell BT, Pybus OG, Gog JR *et al.* (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, **303**, 327–332.

Guillot G, Santos F, Estoup A (2008) Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics*, **24**, 1406–1407.

Guindon S, Delsuc F, Dufayard JF, Gascuel O (2009) Estimating maximum likelihood phylogenies with PhyML. *Methods in Molecular Biology.*, **537**, 113–137.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences, USA*, **103**, 968–971.

Hartl DL, Clark AG (2007) *Principles of Population Genetics*, 4th edn. Sinauer Associates, Inc, Sunderland, MA.

Heath L, van der Walt E, Varsani A, Martin DP (2006) Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *Journal of Virology*, **80**, 11827–11832.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society London, Series B*, **270**, 313–322.

Hebert PDN, Stoeckle MY, Zemlakl TS, Francis CM (2004) Identification of birds through DNA barcodes. *PLOS Biology*, **2**, 1657–1663.

Hester LC, Best TL, Hudson MK (2007) Rabies in bats from Alabama. *Journal of Wildlife Diseases*, **43**, 291–299.

Hughes GJ, Orciari LA, Rupprecht CE (2005) Evolutionary timescale of rabies virus adaptation to North American bats inferred from the substitution rate of the nucleoprotein gene. *Journal of General Virology*, **86**, 1467–1474.

Jackson FR, Turmelle AS, Farino DM *et al.* (2008) Experimental rabies virus infection of big brown bats (*Eptesticus fuscus*). *Journal of General Virology*, **44**, 612–621.

Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.

Kerth G, Mayer F, Petit E (2002) Extreme sex-biased dispersal in the communally breeding, nonmigratory Bechstein's bat (*Myotis bechsteinii*). *Molecular Ecology*, **11**, 1491–1498.

Kissi B, Tordo N, Bourhy H (1995) Genetic polymorphism in the rabies virus nucleoprotein gene. *Virology*, **209**, 526–537.

Kovar J, Walker J, Steffens D, Harford J, Qiu J (2005) *Bovine Microsatellite Multiplexing for Herd Evaluation and Parentage*. Li-Cor Report, 520, Lincoln, NE.

Kurta A, Baker RH (1990) *Eptesicus fuscus*. *Mammalian Species*, **356**, 1–10.

Kuzmin IV, Orciari LA, Arai YT *et al.* (2003) Bat lyssaviruses (Aravan and Khujand) from Central Asia: phylogenetic relationships according to N, P and G gene sequences. *Virus Research*, **97**, 65–79.

Kuzmin IV, Hughes GJ, Botvinkin AD, Orciari LA, Rupprecht CE (2005) Phylogenetic relationshsips of Irkut and West Caucasian bat viruses within the Lyssavirus genus and suggested quantitative criteria based on the N gene sequence for lyssavirus genotype definition. *Virus Research*, **111**, 28–43.

Kuzmin IV, Hughes GJ, Botvinkin AD, Gribencha SG, Rupprecht CE (2008) Arctic and arctic-like rabies viruses: distribution, phylogeny and evolutionary history. *Epidemiology and Infection*, **136**, 509–519.

Leslie MJ, Messenger S, Rohde RE *et al.* (2006) Bat-associated rabies virus in skunks. *Emerging Infectious Diseases*, **12**, 1274–1277.

Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, **16**, 562–563.

Martin DP, Posada D, Crandell KA, Williamson C (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research in Human Retroviruses*, **21**, 98–102.

Mayer F, Dietz C, Kiefer A (2007) Molecular species identification boosts bat diversity. *Frontiers in Zoology*, **4**, 4. doi:10.1186/1742-9994-4-4.

Maynard Smith J (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, **34**, 126–129.

Messenger SL, Smith JS, Rupprecht CE (2002) Emerging epidemiology of bat-associated cryptic cases of rabies in humans in the United States. *Clinical Infectious Diseases*, **35**, 738–747.

Mondul AM, Krebs JW, Childs JE (2003) Trends in national surveillance for rabies among bats in the United States

(1993-2000). *Journal of the American Veterinary Medicine Association*, **222**, 633–639.

Nadin-Davis SA (1998) Polymerase chain reaction protocols for rabies virus discrimination. *Journal of Virological Methods*, **75**, 1–8.

Nadin-Davis SA, Huang W, Armstrong J *et al.* (2001) Antigenic and genetic divergence of rabies viruses from bat species indigenous to Canada. *Virus Research*, **74**, 139–156.

Nadin-Davis SA, Abdel-Malik M, Armstrong J, Wandeler AI (2002) Lyssavirus P gene characterisation provides insights into the phylogeny of the gene and identifies structural similarities and diversity within the encoded phosphoprotein. *Virology*, **298**, 286–305.

Neubaum MA, Douglas MR, Douglas ME, O'Shea TJ (2007) Molecular ecology of the big brown bat (*Eptesicus fuscus*): genetic and natural history variation in a hybrid zone. *Journal of Mammalogy*, **88**, 1230–1238.

Neubaum MA, Shankar V, Douglas MR *et al.* (2008) An analysis of correspondence between unique rabies virus variants and divergent big brown bat (*Eptesicus fuscus*) mitochondrial DNA lineages. *Archives of Virology*, **153**, 1139–1142.

Newton MA, Raftery AE (1994) Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistics Society B*, **56**, 3.

Ngamprasertwong T, Mackie IJ, Racey PA, Piertney SB (2008) Spatial distribution of mitochondrial and microsatellite DNA variation in Daubenton's bat within Scotland. *Molecular Ecology*, **17**, 3243–3258.

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology*, **265**, 218–225.

Palisot de Beauvois AMFJ (1796) *Catalogue raisonné du muséum, de Mr. C. W. Peale*. Parent, Philadelphia, 42 pp.

Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, **26**, 419–420.

Pollard KS, van der Laan MJ (2002) Statistical inference for simultaneous clustering of gene expression data. *Mathematical Biosciences*, **176**, 99–121.

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences, USA*, **98**, 13757–13762.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Pybus MJ (1986) Rabies in insectivorous bats of western Canada. *Journal of Wildlife Diseases*, **22**, 307–313.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rhoads SN (1902) On the common brown bats of peninsular Florida and southern California. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **53**, 618–619.

Rohde RE, Mayes BC, Smith JS, Neill SU (2004) Bat rabies, Texas, 1996-2000. *Emerging Infectious Diseases*, **10**, 948–952.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Rosatte RC (1987) Bat rabies in Canada: history, epidemiology, and prevention. *Canadian Veterinary Journal*, **28**, 754–756.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386. Humana Press, Totowa, NJ.

Shankar V, Orciari LA, de Mattos C *et al.* (2005) Genetic divergence of rabies viruses from bat species of Colorado, USA. *Vector Borne Zoonotic Diseases*, **5**, 330–341.

Smith JS, Orciari LA, Yager PA (1995) Molecular epidemiology of rabies in the United States. *Seminars in Virology*, **6**, 387–400.

Suchard MA, Weiss RE, Sinsheimer JS (2005) Models for estimating Bayes factors with applications to phylogeny and tests of monophyly. *Biometrics*, **61**, 665–673.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.

Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistics Society B*, **63**, 411–423.

Tordo N, Benmansour A, Calisher C *et al.* (2004) Rhabdoviridae. In: *Virus Taxonomy, VIIIth Report of the ICTV (International Committee on Taxonomy of Viruses)* (eds Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA), pp. 623–644. Elsevier Academic Press, London.

Vázquez-Morón S, Juste J, Ibáñez C *et al.* (2008) Endemic circulation of European bat lyssavirus type 1 in serotine bats, Spain. *Emerging Infectious Diseases*, **14**, 1263–1266.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 1847–1857.

Webster WA, Casey GA (1988) Diagnosis of rabies infection. In: *Rabies* (eds Campbell JB, Charlton KM), pp. 201–222. Kluwer Academic Publishers, Boston, MA.

Wirth T, Meyer A, Achtman M (2005) Deciphering host migrations and origins by means of their microbes. *Molecular Ecology*, **14**, 3289–3306.

Wright S. (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Wunner WH (2007) Rabies virus. In: *Rabies*, 2nd edn (eds Jackson AC, Wunner WH), pp. 23–68. Academic Press, London.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution.*, **24**, 1586–1591.

Young RT (1908) Notes on the distribution of Colorado mammals, with a description of a new species of bat (*Eptesicus pallidus*) from Boulder. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **60**, 403–409.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Primers used for PCR in this study.

Table S2 Details of clones used to design primer pairs used for MSAT amplification and scoring.

Figure S1: Phylogenetic tree of 237 viral P gene sequences.

Figure S2 FDIST analysis of the microsatellite loci.

Figure S3 Determination of the optimal number of populations with Structure.

Figure S4 Posterior distribution of the number of populations with Geneland.

Figure S5 Map of posterior probabilities to belong to the eastern (ON, QC) population across the entire sampling range.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.