 doi:10.20944/preprints202210.0151.v1

Geographic range overlap rather than phylogenetic distance explains rabies virus transmission among closely related bat species

AUTHORS AND AFFILIATIONS:

Maude Jacquot[1,#,*], Megan A. Wallace[1,##], Daniel G. Streicker[1,2] & Roman Biek[1,*]

1- School of Biodiversity, One Health and Veterinary Medicine, Boyd Orr Centre for Population and Ecosystem Health, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom

2- MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

Current Addresses:

#- Ifremer, RBE-ASIM, Station La Tremblade F-17390 La Tremblade, France

##- Centre for Ecology & Conservation, Biosciences, College of Life and Environmental Sciences, University of Exeter, Penryn Campus, Penryn, Cornwall

*__Corresponding authors:__ maude.jacquot@ifremer.fr ; Roman.Biek@glasgow.ac.uk

Keywords: host shifts; cross-species transmission; rabies; genetic divergence; host-pathogen interaction; niche overlap; range overlap

Running title: Rabies virus cross-species transmission within a host genus

 doi:10.20944/preprints202210.0151.v1

ABSTRACT

Most pathogens are capable of infecting multiple host species and such cross-species transmission (CST) can have dramatic consequences, as highlighted by recent disease emergence events affecting human, animal and plant health. Understanding the ecological and evolutionary factors that constrain or facilitate the ability of disease agents to infect and establish in a novel host is therefore a timely and important research area. Previous work across different pathogens, including rabies virus (RABV), found that increased evolutionary distance between hosts reduces the frequency of cross-species transmission and of permanent host shifts. However, whether this effect of host relatedness still holds for transmission among recently diverged hosts, and the importance of this effect relative to other predictors, is not well understood. We addressed this question by quantifying the CST frequency of RABV between North American bat species within the genus *Myotis*, using a multi-decade data set containing 128 nucleoprotein (N) sequences from ten host species. For comparison, we also conducted an equivalent analysis of a RABV dataset from North American bat species comprising nine genera. We found that at the within genus scale, host relatedness failed to explain the frequency of CST events. However, CST frequency increased with overlap in species' host range. Moreover, we found evidence of CST occurring among a higher proportion of species, and CST more frequently resulting in sustained transmission in the novel host in the *Myotis* dataset compared to the multi-genus dataset. Our results suggest that among recently diverged species, the ability to infect a novel host is no longer restricted by physiological barriers but instead is limited by physical contact. Our results improve predictions of where future CST events for RABV might occur and clarify the relationship between host divergence and pathogen emergence.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202210.0151.v1

INTRODUCTION

The molecular biology of RNA viruses makes them prone to cross-species transmission (CST). These CST events result in either a dead end infection with no onward transmission in the new host species (R_0=0), a stuttering, unsuccessful chain of infections (0<R_0<1), or establishment in a novel host – a host shift (R_0>1) (Holmes 2010; Longdon *et al.* 2014). RNA virus host shifts from animal reservoirs have spawned many emerging infectious diseases with large health and economic impacts, such as HIV/AIDs, H5N1 avian influenza and most recently, SARS-CoV-2 (Andersen, Rambaut, Lipkin, Holmes, & Garry, 2020; and see reviews; Morens, Folkers, & Fauci, 2008; Parrish et al., 2008; Wasik et al., 2019). Between host species of different genera, evidence from both experimental and field data suggests that the probability of an RNA virus successfully establishing in a novel host increases with genetic relatedness between the donor and recipient host species (eg. Streicker *et al.* 2010; Longdon *et al.* 2011).

However, what is unclear is whether this phylogenetic distance effect holds over different degrees of relatedness and whether it might become irrelevant at very fine taxonomic scales. There are reasons to doubt its relevance at both extremes of genetic distance. Among extremely divergent species all host shifts may become equally improbable, and conversely, among very closely related species, physiological differences among hosts may become relatively inconsequential. Thus, we don't yet know if high host relatedness still increases the probability of a host shift between more recently diverged hosts, and the importance of this effect relative to ecological predictors. This is important as, among mammals, the majority of CST is predicted to occur among closely related host species (Albery *et al.* 2020). If reduced physiological barriers to sustained transmission mean that phylogenetic constraints are less relevant at this level, our ability to predict patterns of CST in nature at this scale might instead depend on alternative predictors.

Rabies virus (RABV) can infect a broad range of species, providing an opportunity to investigate what drives patterns of RNA virus host shifts in wild, multi-host communities (see Fisher et al. 2018). RABV (genus: *Lyassavirus,* family: *Rhabdoviridae*) causes the zoonosis called rabies, a disease with one of the highest known fatality rates in humans (Singh *et al.* 2017). The virus is capable of infecting all mammals, but paradoxically is maintained in distinct

 doi:10.20944/preprints202210.0151.v1

host species-associated transmission cycles, usually in carnivores or bats. RABV phylogenetic history is littered with evidence for CST (Kuzmin *et al.* 2012), and its ability to shift hosts is demonstrated by the spread of the cosmopolitan dog rabies lineage into grey foxes, coyotes, and other wild mesocarnivores during European colonization of the Americas (Velasco-Villa *et al.* 2017). More recent CST events have created new reservoirs of sustained RABV transmission in species of mongoose and ferret-badgers (Fisher *et al.* 2018). This ability to establish in novel host species presents a significant danger to domestic animals and humans, and a challenge to RABV control (Wallace *et al.* 2014).

The probability of RABV CST events initially occurring, as well as resulting in sustained host shifts, is influenced by both ecological and evolutionary factors (Mollentze *et al.* 2014, 2020) but disentangling their relative effects is challenging. Ecological factors' major role in mediating RABV CST events was postulated after studies observed host shifts of RABV strains which were seemingly 'pre-adapted' to novel hosts, through standing genetic variation (Kuzmin *et al.* 2012; Borucki *et al.* 2013). As a result, contact rates – within or between species – are essential to its onward transmission. Ecological factors, such as non-overlaps in host range, or differences in foraging or roosting niche, which reduce inter-species contact rates, could therefore be important in maintaining RABV variants in their host-specific cycles. However, for some host groups, the non-random clustering of RABV reservoirs on the phylogeny suggests that host-specific evolutionary factors also constrain RABV CST events (Huang *et al.* 2014; and see review Longdon *et al.* 2014). The evolutionary history of RABV has also revealed a likely role of adaptive evolution in facilitating host shifts (Streicker et al, 2012). Indeed, it has been hypothesised that high genetic relatedness between closely related species facilitates CST due to the lower number of adaptive changes necessary for virus exchange (Streicker *et al.* 2010).

Previous works have examined the importance of both ecology and evolution in determining the success of RABV host shifts across highly divergent bat species (nine genera and two of the four North American bat families) (Streicker *et al.* 2010; Faria *et al.* 2013). In these studies, higher host genetic relatedness was associated with a significantly higher frequency of both CST (external & internal nodes of the virus phylogeny) and host shifts (internal nodes only). Larger overlaps in species' geographic range were also associated with significantly more frequent CST, though the strength of this effect was lower than that of genetic relatedness.

 doi:10.20944/preprints202210.0151.v1

In contrast, no association was detected between CST frequency and other ecological predictors in these data. However, it remains unknown whether host relatedness and range overlap, the primary determinants of CST frequency between highly-divergent species, can still predict RABV CST between closely related species.

Here, we estimate rates of RABV CST events over the virus' phylogenetic history within a single bat genus (*Myotis*), based on data from ten North American species collected over a 25 year period. *Myotis* (family Vespertilionidae, subfamily Myotinae) are small (most < 10g), generally insectivorous bats with variable roosting habits and colony sizes, estimated to have diverged from other North American Vespertilionidae 20-25 MYA (Miller-Butterworth *et al.* 2007; Bray *et al.* 2013; Kumar *et al.* 2017). Rabies virus has emerged independently multiple times in the *Myotis* genus and a small proportion of variants form suspected multi-species transmission cycles out-with *Myotis* (Streicker *et al.* 2010). We examine whether between-host genetic divergence, as opposed to range overlap or other ecological predictors, can best explain the rate of overall CST within *Myotis*, and events which led to successful host shifts. We compare our estimates of CST frequency within *Myotis* to those for a previously published dataset containing nine bat genera (Streicker *et al.* 2010), hypothesising that the lower average divergence among *Myotis* hosts presents fewer barriers to CST and host shifts in comparison. We therefore predicted that compared to more divergent species, the *Myotis* RABV data would yield *i*) higher CST rates overall; *ii*) a higher proportion of CST events on internal branches, indicative of ongoing transmission in the new host *i.e.* host shifts (Faria *et al.* 2013); *iii*) evidence for CST in a greater proportion of host species examined.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202210.0151.v1

MATERIAL AND METHODS

Compilation of *Myotis* associated RABV sequences

We compiled a dataset of 128 RABV nucleoprotein (N) gene sequences (1350 bp) from *Myotis* bats collected across Canada and the USA. Most of these sequences (n = 112), were generated through a passive surveillance programme at the Canadian Food Inspection Agency's Centre for Expertise for Rabies, between 1990 and 2015 (Nadin-Davis *et al.* 2017). RABV sequences used in the analysis were collected from individuals identified to *Myotis* host species through Cytochrome oxidase subunit 1 (CO1) barcoding. We only included RABV variants that fell within a monophyletic clade associated with the bat genus *Myotis* (MYCAN clade); strains normally associated with other bat genera were excluded from the analysis (Nadin-Davis *et al.* 2017). The collected sequences were supplemented by *Myotis* associated RABV sequences (n = 16) from the USA and Canada, available on Genbank as of May 2022.

The final dataset included RABV N gene sequences collected from ten *Myotis* host species: the South eastern bat *M. austroriparius* (n = 3), the California bat *M. californicus* (n = 24), the Eastern small footed bat *M. leibii* (n = 1), the Keens bat *M. keenii* (n = 8), the little brown bat *M. lucifugus* (n = 32), the Western long eared bat *M. evotis* (n = 28), the long-legged bat *M. volans* (n = 1), the Northern long-eared bat *M. septentrionalis* (n = 11), the western small footed bat *M. ciliolabrum* (n = 1) and the Yuma bat *M. yumanensis* (n = 19). The sequences used in this analysis can be found in Genbank under the accession numbers summarised in **Table S1**.

CST quantification within the *Myotis* genus in comparison to bats of multiple genera

We modelled host species as a discrete trait for each RABV sequence over the genealogy by ancestral state inference using a discrete asymmetric phylogenetic diffusion model (Lemey *et al.* 2009) in BEAST v1.8.3 (Drummond *et al.* 2012). This approach estimates the probability of the internal nodes and branches being associated with a specific host, based on information about host states of the samples at the branch tips. A Bayesian stochastic search variable selection procedure (Lemey *et al.* 2009) was employed to allow for CST between specific host pairs to be included or excluded from the model. Because CST events are likely to be rare, we quantified their frequency using a robust Markov jump (MJ)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202210.0151.v1

157 counting procedure that determines the posterior expectations of the number of CST along
158 the branches of the tree (Minin & Suchard 2008). For each host species pair, we divided
159 the expected number of CST events (MJ count) by the sum of the tree branch lengths (in
160 years) and by the combined number of sequences available for the two species composing
161 the pair in order to obtain the mean number of CST events per year and per capita, hereafter
162 referred to as the CST rate.

163 Analyses were performed under a general time reversible model of nucleotide substitution,
164 with invariant sites, gamma distributed categories of rate variation and empirical base
165 frequencies (Guindon & Gascuel 2003; Darriba *et al.* 2012). A lognormal relaxed molecular
166 clock was used to calibrate the RABV phylogeny (Drummond *et al.* 2006), along with a flexible
167 Gaussian Markov Random Field (GMRF) coalescent model, the skyride (Minin et al. 2008). A
168 MCMC chain of 100 million steps was run and sub-sampled every 10,000 generations. The
169 BEAGLE library was used to increase computational speed (Ayres *et al.* 2012; Drummond *et*
170 *al.* 2012).

171 Whether a stationary distribution was reached, and sufficient mixing (*i.e.* effective sample
172 size >200) for all parameter estimates were checked in Tracer after removing the initial 10%
173 of the samples as burn-in. Maximum clade credibility (MCC) trees were generated from tree
174 output files from BEAST in TreeAnnotator v1.8.3 and annotated in the FigTree v1.4.3 graphical
175 user interface (available at http://tree.bio.ed.ac.uk/software/figtree/).

176 BEAST log files were processed to obtain and visualise estimates of the number of inferred
177 CST events between each host species pair over the complete history of the RABV phylogeny
178 as well as the CST rate.

179 As a reference point for the amount of CST to expect between less closely related species, we
180 also quantified the frequency of CST among North American bats of multiple genera by re-
181 analysing the data from (Streicker *et al.* 2010) using the same methodology as described
182 above. This dataset consisted of 372 RABV N gene sequences from 17 bat species or species
183 complexes. This includes species from the families Vespertilionidae and Molossidae, and the
184 genera *Eptesicus* (n = 118), *Antrozous* (n = 3), *Lasiurus* (n = 121), *Lasionycteris* (n = 17), *Myotis*
185 (n = 33), *Nycticeius* (n = 3), *Parastrellus* (n = 5), *Perimyotis* (n = 14), *and Tadarida* (n = 56) and
186 2 additional sequences not attributed to a host species. Among the *Myotis* genera within this
187 data set, four discrete species states were defined: three of these represented single species

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202210.0151.v1

(*M. californicus*, n = 10; *M. yumanensis,* n = 11*; M. austroriparius,* n = 2) whereas the fourth grouped sequences derived from *M. lucifugus* (n = 5), *M. evotis* (n = 4) and *M. thysanodes* (n = 1), which are thought to form a species complex. Here, two MCMC chains of 800 million steps were run and sub-sampled every 80,000 generations. Convergence to a stationary distribution was checked in Tracer after removing the initial 10% of the samples as burn-in.

Genetic and ecological host predictors of CST within the *Myotis* genus

For each *Myotis* host species pair, we considered four host variables, as well as sample size, as potential predictors of RABV CST events:

Host Genetic Distance – We estimated the phylogenetic distance among host species (or host species clusters, see below) based on the average pairwise differences between mitochondrial CO1 sequences. In total, 171 CO1 sequences, representing all ten *Myotis* host species, were compiled from Genbank. Where possible, CO1 sequences were selected that matched the geographic origin of the RABV samples. Sequences selected were aligned using MUSCLE v3.8.31 (Edgar 2004; Gouy *et al.* 2010). A HKY + I + G model of nucleotide substitution was used to calculate the genetic divergence between CO1 sequences, selected using jModelTest v2.1.8 (Darriba *et al.* 2012) that generated cAIC values, *i.e.* AIC corrected for small sample sizes (Guindon & Gascuel 2003; Darriba *et al.* 2012). A distance matrix was calculated using the TREE PUZZLE analysis tool (Schmidt *et al.* 2002), and the R package ape (Paradis *et al.* 2004).

Range Overlap – Geographic range overlap was calculated as the size of overlapping area between the range of the each of the *Myotis* host species in the USA and Canada (**Figure 1A**). Range overlaps were calculated from NatureServe shape files (NatureServe 2016) using R packages maptools, spatstat, sp and raster (Bivand & Lewin-Koh; Pebesma & Bivand 2005; Baddeley *et al.* 2015; Hijmans 2016).

Foraging niche difference - Three morphological measurements; wing aspect ratio (wing span^2/wing area), wing loading capacity (body weight/wing and tail membrane area) and body length, were collected for each of the ten *Myotis* species from the literature (Krutzsch 1954; Farney, John & Fleharty 1969; Streicker *et al.* 2010) (Table S2). Foraging niche difference was calculated as the mean Euclidean distance between these three morphological

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202210.0151.v1

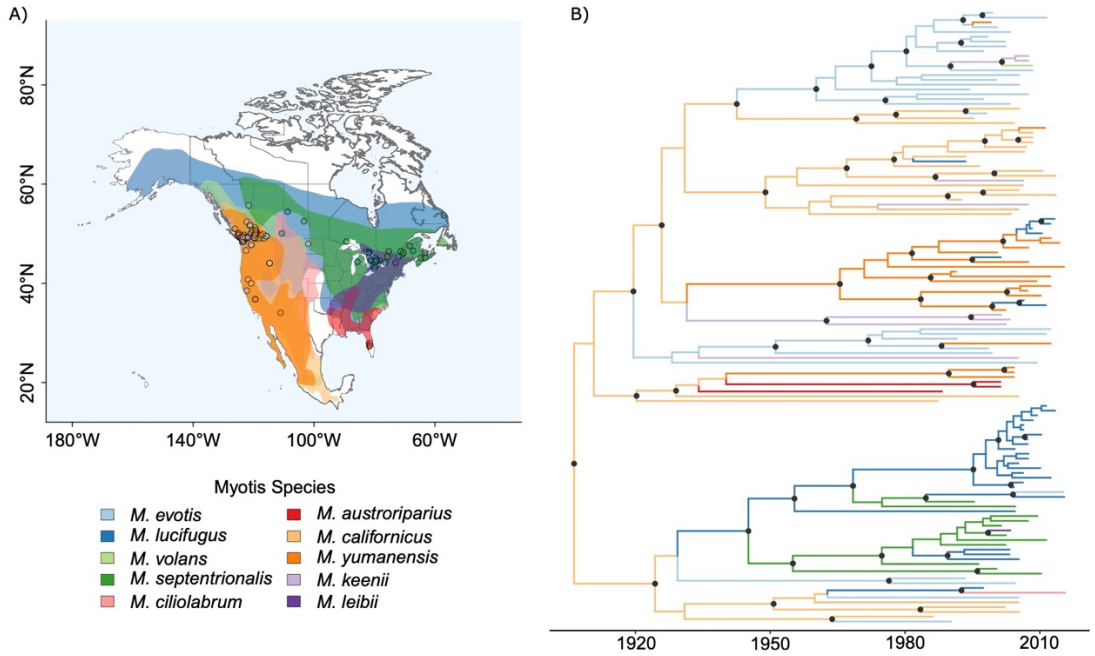
measurements, which can be used as a proxy for similarity in the foraging niche of bat species (Streicker *et al.* 2010; Faria *et al.* 2013).

Roosting Niche Overlap – Observational data on the summer roost types for the *Myotis* host species was gathered from the literature. For each species, we noted which of five types (tree hollows, man-made structures, cave, tree bark, or rock crevice) of summer roosting locations it was reported to use. The number of overlapping roost types (0-5) was used as a measure of similarity between the roosting niches of the species, and therefore their likelihood of co-roosting (Table S2).

Sample size - Sample size, which can particularly impact the results for MJ, was included as an additional predictor. This predictor consisted of a symmetric matrix of the Manhattan distance between the number of RABV sequences representing each discrete species state (Lemey *et al.* 2014).

Each predictor vector, originating from a converted matrix, was log-transformed and standardized using the R package gdata (Warnes *et al.* 2017) to limit the impact of extreme values and to make variables as comparable as possible.

Figure 1. Location and host species origins of samples used in this study A) Map of *Myotis* species sample locations and host ranges, colours indicate host species range and points show the location of collection of a *Myotis* bat infected with rabies virus (RABV). B) Maximum Clade Credibility tree of *Myotis* bat RABV nucleoprotein gene sequences. Branches are coloured by associated host species as specified in the map legend. Node circles indicate nodes with posterior support >0.9.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202210.0151.v1

Testing predictors of RABV CST frequency with the Myotis genus

We assessed the ability of our four host traits, and sample size, to explain variance in the CST frequency between *Myotis* host species using two approaches.

First, we assessed the evidence for a correlation between a matrix of the estimated CST rates between species (MJ count/branch lengths sum/sum of sample sizes of the two species composing a pair) and each of the predictor matrices. We did this by performing mantel tests based on Spearman's rank correlation values, with 10,000 matrix permutations, using the R package ecodist (Goslee & Urban 2007). Similarly, we also assessed these correlations using multiple regression on matrices (MRM), where predictor variables were removed from the model in reverse order of significance until only significant variables remained. From permutation tests we generated regression coefficients, R-squared values, and F-statistics for lack of fit, along with associated p-values.

Secondly, we applied a phylogenetic diffusion approach that simultaneously tests and quantifies potential predictors in a Generalized Linear Model (GLM) framework (Lemey *et al.* 2014). Estimated MJ counts, among the fixed number of host species are parameterized as a linear function of one or multiple predictors. We also employed a branch partitioning approach in order to estimate numbers of CST events for the internal and tip branches of the RABV phylogeny separately (Faria et al. 2013). This allowed us to calculate the inclusion probability of predictors in two distinct GLMs and to use this to distinguish between predictors of sustained host shifts (internal branches), and un-sustained spill-over (tip branches). BEAST MCMC chains of 800 million steps were run and sub-sampled every 80,000 generations for both analyses. Bayes Factors (BF) - computed by the comparison of the posterior and prior odds that a particular predictor is required to explain the diffusion process - were used as a measure of support for the inclusion of predictors (Lemey *et al.* 2012). Support for the inclusion of a predictor was considered substantial when BF > 3, strong if BF > 10, and decisive if >100 (Jeffreys 1961).

Clustering of sequences associated with *Myotis* species

In preliminary analyses of the CO1 sequences from the ten *Myotis* host species, we noticed overlap in the within- and between-species host genetic divergences and non-monophyletic

 doi:10.20944/preprints202210.0151.v1

clustering in the phylogenetic trees. This is unsurprising given evidence that gene flow exists between some species in this genus (Morales & Carstens 2018). To test whether poorly resolved species impacted our results, we repeated our analysis, employing a hierarchical clustering approach to combine such species into species clusters. Successive application of the R function hclust, the Ward minimum variance clustering method, and Ward's clustering criterion (Murtagh & Legendre 2014) to the previously calculated genetic distance matrix, grouped the sequences into seven clusters: sequences collected from *M. californicus, M. leibii* and *M. ciliolabrum* bats (total n = 26) were grouped into one monophyletic group, whilst sequences from *M. keenii* and *M. evotis* (n = 38) made up the other monophyletic multi-species cluster. We then repeated the MJ analysis and simultaneous predictor testing using these seven host species clusters as states (Table S1) in BEAST, as well as the MRM analysis.

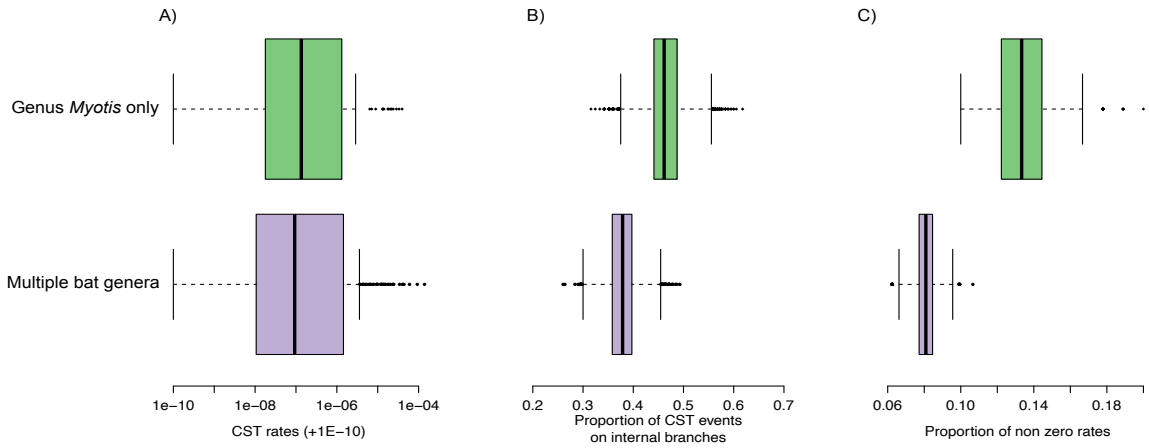
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202210.0151.v1

RESULTS

RABV evolution and CST within North American *Myotis* compared to a multi-genera dataset

Based on our set of N gene sequences from *Myotis* bats, RABV was estimated to be evolving at a rate of 2.76 x 10^{-4} (95% HPD interval: 1.83 – 3.66 x 10^{-4}) nucleotide substitutions per site per year, similar to the previously estimated rate for RABV in North American bats overall (Faria *et al.* 2013). The most recent common ancestor (MRCA) of *Myotis* associated bat rabies dated back to 1904 (95% HPD interval: 1866 – 1941) (**Figure 1B**). The estimated mean number of CST events over the time period since the MRCA was 40.84 (95% HPD: 34-48). This equates to an expected mean of 3.12 x 10^{-04} CST events per year per capita among all *Myotis* species and, on average, 3.47 x 10^{-06} events per year per capita between each possible pairwise combination of *Myotis* species. The latter estimate was indistinguishable from one we obtained for the independent dataset containing multiple bat genera (3.39 x 10^{-06} events per year per capita) (**Figure 2A**). However, consistent with our predictions, the proportion of CST events on internal branches was higher in the *Myotis* dataset (on average 46.53%) compared to the multi-genera dataset (on average 37.91%), although the 95% HPD intervals were overlapping (**Figure 2B**). We also found a greater proportion of species pairs showing evidence of CST in the *Myotis* data compared to the multi-genus dataset (on average 11.96/90 = 13.29% vs 21.61/272 = 7.95%) with no overlap among 95% HPD intervals (**Figure 2C**).

Figure 2. Frequency and characteristics of rabies virus cross-species transmission (CST) for ten North American bat species within the genus *Myotis*, compared to a dataset containing multiple genera of North American bats. Posterior distributions of A) the mean rates (+1E^{-10}) of inferred CST events per year per capita for all pairs of species on a log scaled x axis; B) the proportion of CST events on internal branches (indicative of sustained transmission and host shifts); C) the proportion of species pairs with non-zero rates.



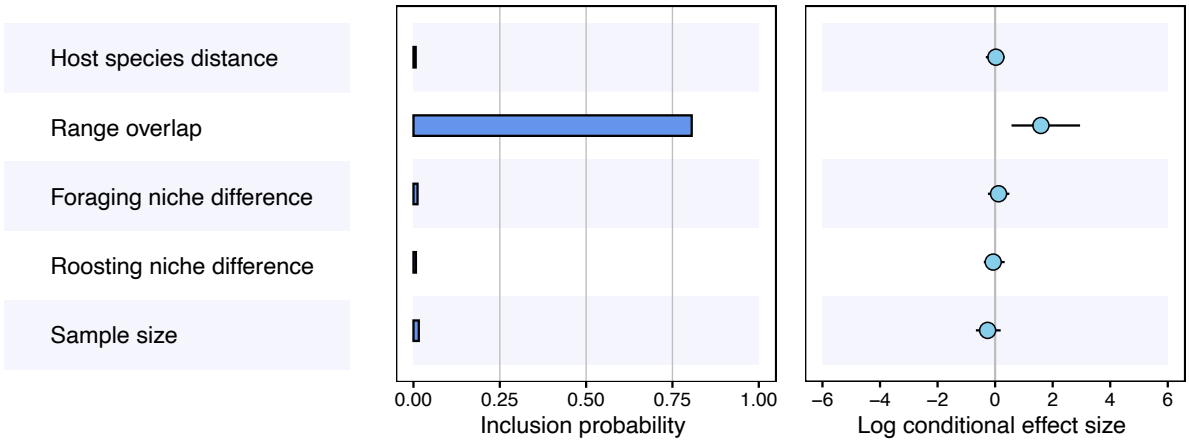
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202210.0151.v1

The vast majority of inferred CST events among *Myotis* bats were estimated to have occurred in recent decades (**Figure S1**). Across these inferred CST events between *Myotis* species, some species showed a particular propensity to being the origin or recipient of RABV CST events. The highest CST rates – in descending order – *originated* from *M. californicus*, *M. evotis*, *M. septentrionalis* and *M. lucifugus*. Meanwhile, *M. lucifugus*, *M. keenii*, *M. evotis* and *M. yumanensis received* the highest number of per year per capita CST events from all other species. RABV moving from and to *M. californicus* and all other species made up 41.7% and 11.8% respectively of the inferred CST numbers (**Figure S2**), a large proportion of which occurred within the last 30 years. CST from and to *M. lucifugus* represented 13.4% and 22.1% of inferred events.

Testing genetic and ecological host predictors of CST in *Myotis* bat species

Phylogenetic distance between *Myotis* species did not correlate with CST rates in the Mantel test (r = 0.013, p-value = 0.45), indicating that virus exchange between species was not constrained by their relatedness. Positive association was seen for species range overlap (r = 0.38, p-value = 0.003) but none of the predictor matrices were retained in the model in the MRM analysis. Considering species clusters instead of species, none of the correlations were significant and none of the predictor matrices were retained in the MRM analysis.

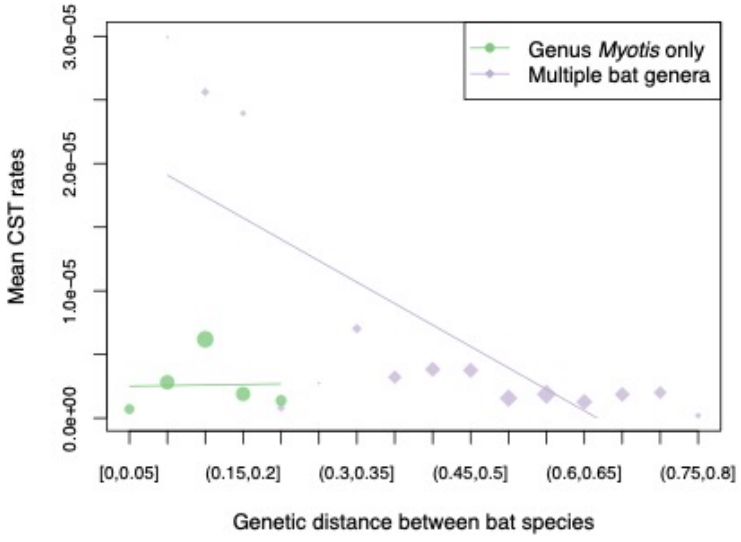
Figure 3. Support and contribution of potential predictors in the phylogenetic GLM of rabies CST events within the *Myotis* genus only. For each potential predictor, support is represented by an inclusion probability and a relative contribution indicated for log scale GLM coefficients conditional on the predictor being included in the model (posterior mean and 95% Bayesian CI). Range overlap was the only supported predictor (BF = 27.9).



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202210.0151.v1

We also tested the predictors of the frequency of RABV CST among *Myotis* bat species using a GLM approach in BEAST. Range overlap had a high probability (> 0.8) of being included as a predictor in the model and was found to be significant in explaining the number of CST events among the different *Myotis* species (BF = 27.9: strong support) (**Figure 3**). Similar results were obtained when considering clusters of species specified as discrete states (BF = 5.93: substantial support) (**Table S3**). We also found range overlap to significantly correlate with the number of CST events on the internal nodes of the *Myotis* bat RABV phylogeny, representing sustained transmission *i.e.*, host shifts (BF = 27.6: strong support) (**Table S4**). None of the other tested factors were supported as a predictor of CST among *Myotis* (BF < 3), including genetic distance between hosts. Indeed, within the *Myotis* genus, genetic distances were not correlated to the number of CST events per year per capita (correlation coefficient = -0.05, p-value = 0.67) whereas a significant negative correlation was found in the dataset with multiple bat genera (correlation coefficient = -0.30, p-value = 5.50×10^{-07}) (**Figure 4**). In the multi-genera dataset, we inferred especially high numbers of expected CST events from *Lasiurus cinereus* to *Eptesicus fuscus* (mean 7.84 CST events), as well as from *Lasiurus borealis* into *Lasiurus seminolus* (mean 6.79). High numbers of CST also occurred within the *Myotis* genus especially from *M. californicus* to the *M. lucifugus* species complex (mean 6.02 CST events), and from *M. lucifugus* species complex into *Myotis yumanensis* (mean 5.64).

Figure 4. Frequency of inferred cross-species transmission events per year per capita among bat species as function of the genetic distances between species. Intervals of genetic distances values were delimited according to bin widths of 0.5 and means of mean number of expected CST events per year for each bin are shown. Point sizes are log-proportional to sample size (number of pairs of species) within each bin.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202210.0151.v1

DISCUSSION

Isolating the factors which influence CST frequency is key to increasing our understanding of pathogen emergence and predicting future outbreaks. Previous studies, both in the lab and the field, have found that the probability of CST and host shifts in RNA viruses decreases with phylogenetic distance between hosts (Streicker *et al.* 2010; Longdon *et al.* 2011; Faria *et al.* 2013). Here, we assessed the relative importance of host relatedness and ecological factors in explaining the frequency of CST events among relatively recently diverged species within the same genus. We show that, when host relatedness is high, CST is no longer predictable by host genetic relatedness, but only by the extent of geographic overlap among species.

We found no significant difference between the rates of rabies virus CST within the *Myotis* genus and that of a wider phylogeny of North American bats **(Figure 2).** This may be because some species pairs in the wider phylogeny exhibit much higher rates of RABV CST than *Myotis* species pairs. Specifically, the exceptionally high rates of CST among *Lasiurus borealis* and *Lasiurus seminolus* raise the multiple genera average as a whole. At the same time, levels of RABV CST for *Myotis* species may be near the maximum level allowed for by ecological opportunity. RABV dynamics within hibernating bats are extremely seasonal, with hibernation thought to maintain RABV infections until the arrival of new susceptibles in the 'birth pulse' (George *et al.* 2011). The shortness of this transmission window may put an upper limit on the contact rate of infected *Myotis* bats. Indeed, an analysis of substitution rates across bat species found that RABVs associated with temperate (mostly hibernating) bat species evolved more slowly per year, suggesting pauses in transmission which would presumably also reduce CST (Streicker *et al.* 2012b).

When compared with a wider phylogeny of multiple genera of bats, we found some indication for a higher proportion of CST on internal nodes in the *Myotis* RABV phylogeny (**Figure 2**), though the posteriors overlapped. This suggests that when a virus is passed among host species within the same genus, the proportion of CST events which lead to onward transmission (those found on the internal nodes of the phylogeny), is increased. This is consistent with the idea that viruses are more pre-adapted to establish in novel hosts that are less divergent from the donor host, presumably due to finding a similar physiological and immunological environment (Mollentze *et al.* 2014). This onward transmission can lead to

 doi:10.20944/preprints202210.0151.v1

384 either a permanent host shift, or the formation of a true multi-species clade of the virus –
385 both of which seem possible in *Myotis*-associated RABV from the viral phylogeny (**Figure 1B**).
386 Thus, it seems that between closely related host species, it is not the rate of CST which is
387 changed, but the outcome.

388 We also found that a greater proportion of species within the *Myotis* genus were donors or
389 recipients of CST events compared to North American bats in general, despite the latter group
390 containing other genera with high CST rates (*e.g.* Lasiurus). This pattern parallels the
391 phylogenetic clustering in susceptibility previously observed in insect Rhabdoviruses
392 (Longdon *et al.* 2011). In short, when compared with the larger phylogeny, RABV variants
393 infecting more closely related species show similar CST rates, but these CST events involve a
394 greater proportion of host species and are more likely to result in a successful host shift. These
395 observations suggest that between closely related species, fewer barriers to successful RABV
396 host shifts exist, consistent with our predictions.

397 We also used a phylogenetic diffusion approach to simultaneously reconstruct the ancestral
398 phylogeny of RABV infections within the *Myotis* genus, and to test for possible predictors of
399 the number of CST events along this phylogeny. Previous studies of North American bats, in
400 multiple genera (Streicker *et al.* 2010; Faria *et al.* 2013) found the frequency of both sustained
401 RABV transmission, and spillover, to increase with increased host species relatedness. In
402 contrast, when we implemented this approach within the single genus *Myotis*, we found no
403 significant relationship between host species relatedness and the frequency of CST events.
404 The *Myotis* species have been diverging from their most recent common ancestor for 12.2 ±
405 2.0 MY, a relatively short period of evolutionary time (Stadelmann *et al.* 2007). Low inter-
406 species divergence could allow for less constrained transmission of RABV variants between
407 species through two non-exclusive mechanisms. First, the shorter divergence time, and
408 resulting similarity in genetic background, between very closely related host species might
409 limit the number of adaptive RABV substitutions needed for one of these variants to invade a
410 new host species and increase the probability of sustained transmission. Standing genetic
411 variation in RABV-infected individuals, in the form of maintained rare variants, would provide
412 the raw material for this adaptation (Borucki *et al.* 2013). Indeed, within a single genus, the
413 time between CST and emergence of a new RABV variant has been found to be shorter if
414 fewer positively selected substitutions are needed for RABV variants to invade a new host

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202210.0151.v1

(Streicker *et al.* 2012a). Second, closely related host species may also show similar levels of susceptibility to the viral variants due to the loss or gain of common cellular, physiological or immunological components in their phylogenetic history (Longdon *et al.* 2011). The evolution of these key determinants of susceptibility may have occurred prior to the diversification of the *Myotis* genus, allowing for less restrained movement of RABV variants. Alternatively, niche partitioning among closely related sympatric species might diminish ecological opportunities for CST despite their higher likelihood of infection after exposure.

Our results suggest that between closely related species, ecological factors become more relevant to explain variation in the frequency of CST than host genetic relatedness. Specifically, we found the geographic range overlap of species to explain a significant proportion of variation in the number and rate of CST events between species (**Figure 3**). This suggests that at this taxonomic scale, the frequency of CST events is primarily driven by the rate of physical contacts between species. If this is the case, higher rates of CST might be found in regions where a diversity of habitats, and climatic variance, support an increased diversity of host species, as previously suggested for pathogen sharing between primates (Pedersen & Davies 2009). In our dataset, this might be the case for the province of British Columbia, which has diverse habitats that support up to nine *Myotis* species (Nadin-Davis *et al.* 2017). In contrast, we found no consistent evidence that the frequency of CST between *Myotis* bat species is dependent on similarities in their foraging or roosting niches.

It is unlikely that geographic range overlap is the only important ecological factor in explaining CST of RABV among closely related species. However, accurately modelling the complex ecological niches inhabited by bat species is a challenge. For example, we calculated the predictor of foraging niche difference from three morphological measurements as a proxy for niche similarity, a method which may struggle to represent features such as temporal foraging niche differentiation (Streicker *et al.* 2010). Additionally, our roosting niche overlap predictor, calculated through the number of overlapping summer roost types, relied on sparse observational literature for some species and might not sufficiently reflect actual ecological interactions to recover a statistical signal. Less social bat species are also often under-represented by passive sampling, hampering the description of their role as RABV reservoirs (Nadin-Davis *et al.* 2001; Condori-Condori *et al.* 2013). The recent documentation of the

 doi:10.20944/preprints202210.0151.v1

Myotis genus supporting its own specific RABV variants (Nadin-Davis *et al.* 2017), further highlights the need for more detailed ecological data from less frequently sampled species.

In this study we investigated the low end of a wide continuum of host genetic divergence and its effect on the probability of a pathogen infecting a novel host. This contrasts with the much larger divergences considered by previous studies of RABV in North American bats. While we show that effects qualitatively differ along this continuum, the specific shape of the relationship between host genetic divergence and CST frequency, which may well be nonlinear, remains elusive. One of the challenges for studying this relationship is the difficulty of meaningfully quantifying the frequency of CST from genetic data given that the detection of events will depend on available sequences and temporal scales captured by the virus phylogeny. However, these limitations are unlikely to qualitatively alter our findings given that neither sample size nor taxonomic classification had a discernible effect on our results. It is also important to emphasise that host divergence or relatedness must be considered proxies for a whole range of host-specific factors that will influence pathogen host range. Ideally, future investigations would seek to replace host relatedness with a more mechanistic way of predicting which viruses are able to establish in a novel host species (eg. Mollentze *et al.* 2020). Ultimately, the relationship between CST and host relatedness will differ between pathogen groups, and will be dependent on the biological mechanisms underlying host specificity in each case. Mapping this relationship as we have done here for a specific group of RABV variants and hosts could thus provide a model for similar studies in other RABV variants and RNA viruses more generally.

ETHICS

This research project did not require any ethical approval.

DATA ACCESSIBILITY

Sample information and GenBank accession numbers of the rabies sequences used in this study are summarized **Table S1.** Sources of ecological data used to generate predictors can be found in **Table S2**.

 doi:10.20944/preprints202210.0151.v1

AUTHORS CONTRIBUTIONS

R.B. and D.S. conceived the study. M.J. and M.W. conducted data analysis. M.W., M.J. and R.B. primarily drafted the manuscript and all authors contributed to finalize it. All authors gave final approval for publication.

COMPETING INTERESTS

The authors declare having no conflict of interests.

FUNDING

MW was supported during the preparation of this manuscript by a NERC Doctoral Training Partnership grant (NE/L002558/1). DS was supported by a Wellcome Senior Fellowship (217221/Z/19/Z).

ACKNOWLEDGEMENTS

We thank Susan Nadin-Davis for making the data used in this study publicly available and for providing helpful clarifications regarding the metadata.

 doi:10.20944/preprints202210.0151.v1

REFERENCES

Albery, G.F., Eskew, E.A., Ross, N. & Olival, K.J. (2020). Predicting the global mammalian viral sharing network using phylogeography. *Nat. Commun.*, 11, 1–9.

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C. & Garry, R.F. (2020). The proximal origin of SARS-CoV-2. *Nat. Med.*, 26, 450–452.

Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O., *et al.* (2012). BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. *Syst. Biol.*, 61, 170–173.

Baddeley, A., Rubak, E. & Turner, R. (2015). *Spatial Point Patterns: Methodology and Application with R.* London: Chapman and Hall/CRC Press.

Bivand, R. & Lewin-Koh, N. (2011). maptools: Tools for Reading and Handling Spatial Objects, R package version 0.8-39.

Borucki, M.K., Chen-Harris, H., Lao, V., Vanier, G., Wadford, D.A., Messenger, S., *et al.* (2013). Ultra-Deep Sequencing of Intra-host Rabies Virus Populations during Cross-species Transmission. *PLoS Negl. Trop. Dis.*, 7, e2555.

Bray, T.C., Mohammed, O.B. & Alagaili, A.N. (2013). Phylogenetic and Demographic Insights into Kuhl's Pipistrelle, Pipistrellus kuhlii, in the Middle East. *PLoS One*, 8.

Condori-Condori, R.E., Streicker, D.G., Cabezas-Sanchez, C. & Velasco-Villa, A. (2013). Enzootic and Epizootic Rabies Associated with Vampire Bats, Peru. *Emerg. Infect. Dis.*, 19, 1463–1469.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods*, 9, 772–772.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J. & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol.*, 4, 699–710.

Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.*, 29, 1969–1973.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32, 1792–1797.

518 Faria, N.R., Suchard, M. a, Rambaut, A., Streicker, D.G. & Lemey, P. (2013). Simultaneously
519 reconstructing viral cross-species transmission history and identifying the underlying
520 constraints. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 368, 20120196.

521 Farney, John & Fleharty, E.. (1969). Aspect Ratio, Loading, Wing Span, and Membrane Areas
522 of Bats, 71, 3–6.

523 Fisher, C.R., Streicker, D.G. & Schnell, M.J. (2018). The spread and evolution of rabies virus:
524 conquering new frontiers. *Nat. Rev. Microbiol.*, 16, 241–255.

525 George, D.B., Webb, C.T., Farnsworth, M.L., O'Shea, T.J., Bowen, R.A., Smith, D.L., *et al.* (2011).
526 Host and viral ecology determine bat rabies seasonality and maintenance. *Proc. Natl.*
527 *Acad. Sci. U. S. A.*, 108, 10208–13.

528 Goslee, S. & Urban, D. (2007). The ecodist Package: Dissimilarity-based functions for
529 ecological analysis. *J. Stat. Softw.*, 22, 1–19.

530 Gouy, M., Guindon, S. & Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user
531 interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.*, 27,
532 221–4.

533 Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large
534 phylogenies by maximum likelihood. *Syst. Biol.*, 52, 696–704.

535 Hijmans, R.. (2016). raster: Geographic Data Analysis and Modeling, R package version 2.5-8.

536 Holmes, E.C. (2010). The comparative genomics of viral emergence. *Proc. Natl. Acad. Sci.*, 107,
537 1742 LP – 1746.

538 Huang, S., Bininda-Emonds, O.R.P., Stephens, P.R., Gittleman, J.L. & Altizer, S. (2014).
539 Phylogenetically related and ecologically similar carnivores harbour similar parasite
540 assemblages. *J. Anim. Ecol.*, 83, 671–680.

541 Jeffreys, H. (1961). *Theory of Probability*. *Theory Probab.*

542 Krutzsch, P.H. (1954). Notes of the Habits of the Bat, Myotis californicus. *J. Mammal.*, 35, 539–
543 545.

544 Kumar, S., Stecher, G., Suleski, M. & Blair Hedges, S. (2017). TimeTree: A Resource for
545 Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.*, 34, 1812–1819.

546 Kuzmin, I. V, Shi, M., Orciari, L.A., Yager, P.A., Velasco-Villa, A., Kuzmina, N.A., *et al.* (2012).
547 Molecular Inferences Suggest Multiple Host Shifts of Rabies Viruses from Bats to
548 Mesocarnivores in Arizona during 2001–2009. *PLOS Pathog.*, 8, e1002786.

549 Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., *et al.* (2012). The seasonal
550 flight of influenza : a unified framework for spatiotemporal hypothesis testing, 1–16.

551 Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., *et al.* (2014). Unifying viral
552 genetics and human transportation data to predict the global transmission dynamics of
553 human influenza H3N2. *PLoS Pathog.*, 10.

554 Lemey, P., Rambaut, A., Drummond, A. & Suchard, M. (2009). Bayesian phylogeography finds
555 its roots. *PLoS Comput. Biol.*, 5.

556 Longdon, B., Brockhurst, M. a., Russell, C. a., Welch, J.J. & Jiggins, F.M. (2014). The Evolution
557 and Genetics of Virus Host Shifts. *PLoS Pathog.*, 10, e1004395.

558 Longdon, B., Hadfield, J.D., Webster, C.L., Obbard, D.J. & Jiggins, F.M. (2011). Host Phylogeny
559 Determines Viral Persistence and Replication in Novel Hosts. *PLoS Pathog.*, 7, e1002260.

560 Miller-Butterworth, C.M., Murphy, W.J., O'Brien, S.J., Jacobs, D.S., Springer, M.S. & Teeling,
561 E.C. (2007). A family matter: Conclusive resolution of the taxonomic position of the long-
562 fingered bats, Miniopterus. *Mol. Biol. Evol.*, 24, 1553–1561.

563 Minin, V.N., Bloomquist, E.W. & Suchard, M.A. (2008). Smooth Skyride through a Rough
564 Skyline: Bayesian Coalescent-Based Inference of Population Dynamics. *Mol. Biol. Evol.*,
565 1459–1471.

566 Mollentze, N., Biek, R. & Streicker, D.G. (2014). The role of viral evolution in rabies host shifts
567 and emergence. *Curr. Opin. Virol.*, 8, 68–72.

568 Mollentze, N., Streicker, D.G., Murcia, P.R., Hampson, K. & Biek, R. (2020). Virulence
569 mismatches in index hosts shape the outcomes of cross-species transmission. *Proc. Natl.*
570 *Acad. Sci. U. S. A.*, 117, 28859–28866.

571 Morales, A.E. & Carstens, B.C. (2018). Evidence that Myotis lucifugus "Subspecies" are Five
572 Nonsister Species, Despite Gene Flow. *Syst. Biol.*, 67, 756–769.

573 Morens, D.M., Folkers, G.K. & Fauci, A.S. (2008). Emerging infections: a perpetual challenge.
574 *Lancet Infect. Dis.*, 8, 710–719.

575 Murtagh, F. & Legendre, P. (2014). Ward ' s Hierarchical Agglomerative Clustering Method :
576 Which Algorithms Implement Ward ' s Criterion ? *J. Classif.*, 31, 274–295.

577 Nadin-Davis, S., Alnabelseya, N., Knowles, M.K., Streicker, D. & Lemey, P. (2017). The
578 phylogeography of Myotis bat-associated rabies viruses across Canada. *PLoS Negl. Trop.*
579 *Dis.*, 11, e0005541.

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

583 NatureServe. (2016). *Natureserve Web Service*.

584 Parrish, C.R., Holmes, E.C., Morens, D.M., Park, E.-C., Burke, D.S., Calisher, C.H., *et al.* (2008).
585 Cross-species virus transmission and the emergence of new epidemic diseases.
586 *Microbiol. Mol. Biol. Rev.*, 72, 457–470.

587 Pebesma, E.. & Bivand, R.. (2005). Classes and methods for spatial data in R, R News 5 (2).

588 Pedersen, A.B. & Davies, T.J. (2009). Cross-species pathogen transmission and disease
589 emergence in primates. *Ecohealth*, 6, 496–508.

590 Singh, R., Singh, K.P., Cherian, S., Saminathan, M., Kapoor, S., Manjunatha Reddy, G.B., *et al.*
591 (2017). Rabies – epidemiology, pathogenesis, public health concerns and advances in
592 diagnosis and control: a comprehensive review. *Vet. Q.*, 37, 212–251.

593 Stadelmann, B., Lin, L.K., Kunz, T.H. & Ruedi, M. (2007). Molecular phylogeny of New World
594 Myotis (Chiroptera, Vespertilionidae) inferred from mitochondrial and nuclear DNA
595 genes. *Mol. Phylogenet. Evol.*, 43, 32–48.

596 Streicker, D.G., Altizer, S.M., Velasco-Villa, A. & Rupprecht, C.E. (2012a). Variable evolutionary
597 routes to host establishment across repeated rabies virus host shifts among bats. *Proc.*
598 *Natl. Acad. Sci. U. S. A.*, 109, 19715–20.

599 Streicker, D.G., Lemey, P., Velasco-villa, A. & Rupprecht, C.E. (2012b). Rates of Viral Evolution
600 Are Linked to Host Geography in Bat Rabies, 8.

601 Streicker, D.G.G., Turmelle, A.S.S., Vonhof, M.J.J., Kuzmin, I.V. V., McCracken, G.F.F. &
602 Rupprecht, C.E.E. (2010). Host phylogeny constrains cross-species emergence and
603 establishment of rabies virus in bats. *Science (80-.).*, 329, 676–679.

 doi:10.20944/preprints202210.0151.v1

604 Velasco-Villa, A., Mauldin, M.R., Shi, M., Escobar, L.E., Gallardo-Romero, N.F., Damon, I., *et al.*
605 (2017). The history of rabies in the Western Hemisphere. *Antiviral Res.*, 146, 221–232.

606 Wallace, R.M., Gilbert, A., Slate, D., Chipman, R., Singh, A., Wedd, C., *et al.* (2014). Right place,
607 wrong species: a 20-year review of rabies virus cross species transmission among
608 terrestrial mammals in the United States. *PLoS One*, 9, e107539–e107539.

609 Warnes, G.R., Bolker, B., Gorjanc, G., Grothendieck, G., Korosec, A., Lumley, T., *et al.* (2017).
610 gdata: Various R Programming Tools for Data Manipulation. R package version 2.18.0.

611 Wasik, B.R., De Wit, E., Munster, V., Lloyd-Smith, J.O., Martinez-Sobrido, L. & Parrish, C.R.
612 (2019). Onward transmission of viruses: How do viruses emerge to cause epidemics after
613 spillover? *Philos. Trans. R. Soc. B Biol. Sci.*, 374.

614

 doi:10.20944/preprints202210.0151.v1

SUPPLEMENTARY DATA

Table S1. Rabies samples used in this study.

Table S2. Table of morphological measurements used to compute the foraging niche predictor for each *Myotis* species in the analysis, along with the observed summer roost types used to compute the roosting niche overlap predictor.

Table S3. Posterior inclusion probabilities, Bayes factor support and the mean conditional effect size for the 5 predictors of CST in internal and external nodes of the RABV phylogeny, between *Myotis* species clusters.

Table S4. Posterior inclusion probabilities, Bayes factors and the mean conditional effect size (with upper and lower quantiles) for the 5 predictors of host shifts of RABV between *Myotis* species (species as discrete states).

Figure S1. Density of RABV CST events within *Myotis* genus over time.

Figure S2. Mean number of RABV CST events between each pair of *Myotis* species.