

# 

**Citation:** Hayman DTS, Fooks AR, Marston DA, Garcia-R JC (2016) The Global Phylogeography of Lyssaviruses - Challenging the 'Out of Africa' Hypothesis. PLoS Negl Trop Dis 10(12): e0005266. doi:10.1371/journal.pntd.0005266

**Editor:** Remi Charrel, Aix Marseille University, Institute of Research for Development, and EHESP School of Public Health, FRANCE

Received: July 24, 2016

Accepted: December 15, 2016

Published: December 30, 2016

**Copyright:** © 2016 Hayman et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files (all accession numbers are in S2-S6 Figs and S1 Table).

**Funding:** DTSH acknowledges funding from the Royal Society of New Zealand Marsden Fund (MAU1503) and the Research and Policy for Infectious Disease Dynamics program of the Science and Technology Directorate, Department of Homeland Security, Fogarty International Center, National Institutes of Health; JCGR and DTSH from **RESEARCH ARTICLE** 

# The Global Phylogeography of Lyssaviruses -Challenging the 'Out of Africa' Hypothesis

#### David T. S. Hayman<sup>1</sup>\*, Anthony R. Fooks<sup>2,3</sup>, Denise A. Marston<sup>2</sup>, Juan C. Garcia-R<sup>1</sup>

1 Molecular Epidemiology and Public Health Laboratory, Hopkirk Research Institute, Massey University, Palmerston North, New Zealand, 2 Wildlife Zoonoses and Vector-borne Diseases Research Group, Animal and Plant Health Agency (APHA), Weybridge-London, United Kingdom, 3 Department of Clinical Infection, Microbiology & Immunology, Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom

\* d.t.s.hayman@massey.ac.nz, davidtshayman@gmail.com

## Abstract

Rabies virus kills tens of thousands of people globally each year, especially in resource-limited countries. Yet, there are genetically- and antigenically-related lyssaviruses, all capable of causing the disease rabies, circulating globally among bats without causing conspicuous disease outbreaks. The species richness and greater genetic diversity of African lyssaviruses, along with the lack of antibody cross-reactivity among them, has led to the hypothesis that Africa is the origin of lyssaviruses. This hypothesis was tested using a probabilistic phylogeographical approach. The nucleoprotein gene sequences from 153 representatives of 16 lyssavirus species, collected between 1956 and 2015, were used to develop a phylogenetic tree which incorporated relevant geographic and temporal data relating to the viruses. In addition, complete genome sequences from all 16 (putative) species were analysed. The most probable ancestral distribution for the internal nodes was inferred using three different approaches and was confirmed by analysis of complete genomes. These results support a Palearctic origin for lyssaviruses (posterior probability = 0.85), challenging the 'out of Africa' hypothesis, and suggest three independent transmission events to the Afrotropical region, representing the three phylogroups that form the three major lyssavirus clades.

### Author Summary

Rabies virus kills tens of thousands of people globally each year and causes indescribable misery and family disturbance, especially in developing countries. Yet in much of the world there are related viruses, called lyssaviruses, which circulate among bats without causing conspicuous outbreaks. The greater diversity of African lyssaviruses has led to the hypothesis that Africa is the origin of these viruses. To test this hypothesis, the genetic data from 153 representative viruses from 16 available lyssavirus species from across the world dated between 1956 and 2015 were analysed. Statistical models were used to reconstruct the historical processes that lead to the contemporary distribution of these viruses.

the New Zealand Ministry of Health; and ARF was financially supported by the UK Department for Environment, Food and Rural Affairs (Defra), Scottish Government and Welsh Government by grants SEV3500 and SE0431, The Research and Policy for Infectious Disease Dynamics program of the Science and Technology Directorate, Department of Homeland Security, Fogarty International Center, National Institutes of Health and from the EU H2020-funded Research Infrastructure Grant 'European Virus Archive Global (EVAg)' [H2020 - grant agreement n°653316]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

Our results support a Palearctic origin for lyssaviruses, not Afrotropic, and suggest three independent transmission events to Africa from the Palearctic region.

#### Introduction

Determining the evolutionary history of viruses is fundamental to our understanding of the patterns and processes occurring during viral emergence and spread. Emergence and spread of viral diseases is a permanent threat in animal and public health and special attention has been given to fast-evolving RNA viruses due to the high mortality rates recorded worldwide. The family *Rhabdoviridae* contains a diverse variety of RNA viruses that replicate in vertebrates, invertebrates and plants. The vast majority of rhabdoviruses have invertebrate vectors that play a role in the transmission to plants, fishes or mammals. The lyssaviruses, which cause the disease rabies, are unique within these negative-sense, single-stranded RNA viruses because they do not require arthropod vectors and are well-adapted to their mammalian hosts [1].

The prototypic virus within the genus *Lyssavirus* is rabies virus (RABV) [2]. RABV has a global distribution (with the exception of Australasia, Antarctica, and some islands). The principal reservoir of RABV is the domestic dog (*Canis familiaris*), although RABV is enzootic in a number of wildlife Carnivora, including fox, mongoose, raccoon, and skunk populations. RABV variants undergo genetic adaptation to each particular host, resulting in new clades or biotypes relating to the local fauna [3–11]. Well-established wildlife Carnivora reservoirs for RABV are apparently absent in South America and Australasia. However, lyssaviruses are present in both of these regions and throughout the rest of the world in bat hosts (Chiroptera).

All lyssaviruses have been isolated from bats, with the exception of Mokola virus (MOKV) and Ikoma lyssavirus (IKOV), and phylogenetic analyses suggest all lyssaviruses have bat origins [12–16]. RABV has only been identified in bats in the Americas and is the only lyssavirus detected circulating in American bat species. This observation is in direct contrast to the rest of the world, where RABV has not been detected in bat populations. The greatest diversity of lyssaviruses occurs in bats in Eurasia and on the African continent. The species are divergent enough that sera raised against specific virus species often do not neutralize other virus species and are divided into phylogroups (PG) [14, 17, 18].

Rabies virus is a member of phylogroup 1 (PG1). European bat lyssavirus-1 (EBLV-1) and 2 (EBLV-2), Bokeloh bat lyssavirus (BBLV), Irkut (IRKV), Aravan (ARAV) and Khujand (KHUV) viruses all belong to PG1. These PG1 viruses have all been isolated from bats in Eurasia, as has West Caucasian Bat Virus (WCBV) from PG3. Lleida bat lyssavirus (LLEBV) is a tentative new species yet to be recognised by the International Committee on Taxonomy of Viruses (ICTV). LLEBV has been identified from a bat in Spain and is most closely related to the African Ikoma lyssavirus (IKOV) genetically in PG3 [19–21]. Gannoruwa bat lyssavirus (GBLV), also awaiting ICTV recognition as a new species, has recently been isolated from fruit bats in Sri Lanka, and phylogenetic analysis indicated that it is most closely related to RABV (PG1) [22]. In Australasia the only lyssavirus identified is Australian bat lyssavirus (ABLV), also a PG1 virus, which circulates within the Australian bat populations. In Africa, however, all three phylogroups are represented: Duvenhage virus (DUVV) from PG1, Lagos bat virus (LBV), MOKV and Shimoni bat virus (SHIBV) from PG2 [13], and IKOV [23] from PG3.

The observation that the greatest genetic diversity of lyssaviruses is in Africa has led to the hypothesis that Africa is the continent where lyssaviruses originated, most likely from an African bat reservoir [24]. This hypothesis is founded on sound observations, though does not

address the lack of known reservoir(s) for MOKV, and was proposed despite a lack of understanding of the ecology of MOKV and many other lyssavirus species. Here, the hypothesis that lyssaviruses had their origins in Africa was tested by using a probabilistic phylogeographical approach, which provides additional insights into the historical biogeography of lyssaviruses.

#### Methods

We used 153 nucleoprotein (N) gene sequences from the 14 recognized and 2 putative lyssavirus species (LLEBV, GBLV) in these analyses (final dataset, see below), along with complete genome sequences from all 16 (putative) species when appropriate for confirmatory analyses. Lyssavirus sequences were mostly derived from bats, with exception of MOKV and IKOV (see Introduction). Using bat-derived sequences was particularly important for RABV because the evolutionary history was less likely to be confounded by the global spread of RABV by human movement of terrestrial carnivores and the post-war RABV epidemics in wildlife [25–27]. Sequences represent serially sampled data; the earliest sequence from 1956 and the last from 2015, including the most recently available sequences from GBLV and LLEBV, spanning a 59-year period.

The dataset was aligned in ClustalX2.1 [28] and inspected by eye. Bayesian Markov Chain Monte Carlo (MCMC) implemented in BEAST software v.1.8.3 [29] was used for phylogenetic analysis and estimation of divergence times. A codon partition strategy was implemented with a general time reversible (GTR) model of substitution with gamma distributed variation in rates amongst sites and a proportion of sites assumed to be invariant according to the Akaike criterion in Modeltest [30]. The lower number of substitutions per site in EBLV-1 (S1 Fig) [31] compared across the tree can potentially cause problems for the estimation of the posterior probabilities and other parameters. To mitigate this issue, the number of EBLV-1 sequences was reduced in the final dataset (S1 Table). Divergence times were estimated using a strict clock model in BEAST assuming an underlying coalescent process with a constant population size. To be more conservative in our estimates of the divergence times and assuming that purifying selection has removed deleterious mutations from rate estimates in short timeframes [32–34], a 2.3 x 10<sup>-4</sup> substitution rate estimated by Bourhy and colleagues [35] was used. An uncorrelated lognormal relaxed clock was also considered because it assumes that the substitution rate along branches is not correlated.

Parameter effective sample sizes were visualized in Tracer v.1.5 (http://tree.bio.ed.ac.uk/ software/tracer/). The results from two independent runs with 5x10<sup>7</sup> MCMC length of chain were combined. The first 10% of maximum clade credibility (MCC) trees were discarded as burn-in in TreeAnnotator v1.8.3 [36]. Final trees were visualized in FigTree v.1.4.2 [37]. The fit of each analysis (strict and lognormal) were evaluated with Bayes Factors in Tracer v.1.5 (http://tree.bio.ed.ac.uk/software/tracer/).

Ancestral distributions were first performed in BEAST as part of the analysis for the estimations of the divergence times. The analysis assumed the forward and reverse rates to be symmetrical (Mk1). For comparison, our sampling was reduced to contain a single complete genome representative of each taxon, including GBLV and LLEBV, and we performed a RAxML analysis [38] to obtain a Maximum Likelihood phylogenetic tree. This phylogenetic tree was used to infer the most probable ancestral distribution for the internal nodes with Likelihood and Parsimony approaches and restriction of equal probability for all state changes with the Mk1 model using Mesquite v3.04 [39]. For ancestral state analyses viruses were categorized by the terrestrial ecozone from which they were isolated [35]. The major ecozones were Afrotropical, Palearctic, Oriental and Nearctic and Neotropic (both Americas) (Fig 1). Note that we use the term Oriental for the often named Indomalaya region but do not partition



Fig 1. Evolutionary relationships between lyssaviruses. The time-scaled phylogeny was generated from 153 nucleoprotein gene sequences and inferred with a Lognormal relaxed-clock Bayesian analysis using BEAST. Branch colours correspond to ecozones shown on the inset map. Support values corresponding to Bayesian posterior probabilities (above branches) and states probabilities from the different assigned ecozones (below branches) are indicated for key nodes. The time scale in years is shown. Phylogroups 3 (green, top) and 2 (blue) are shaded and key nodes discussed in the text labelled A-D. Virus names are *Mokola virus* (MOKV); *Australian bat lyssavirus* (ABLV); *European bat lyssavirus-1* (EBLV-1); *European bat lyssavirus-2* (EBLV-2), *Irkut* (IRKV), *Aravan* (ARAV), *Khujand* (KHUV); *West Caucasian Bat Virus* (WCBV); *Lagos bat virus* (LBV); *Duvenhage virus* (DUVV); *Shimoni bat virus* (SHIBV); *Bokeloh bat lyssavirus* (BBLV); *Ikoma virus* (IKOV); *Lleida virus* (LLEBV); *Gannoruwa bat lyssavirus* (GBLV); *Rabies virus* (RABV).

doi:10.1371/journal.pntd.0005266.g001

the Palearctic into Palearctic, Saharo-Arabian and Sino-Japanese regions [40] because no viruses were available from these latter two regions. Because RABV can be identified in more than one ecozone, its distribution category was treated as uncertain for the Likelihood analysis whilst polymorphic for the Parsimony analysis. Summary trees are presented in the main text

and detailed trees with additional information, such as with all tip labels, are presented as figures in the supplementary information.

#### Results

Using a panel of lyssavirus N genes with global distributions there was strong support for the overall tree topology (Fig 1). The topology was confirmed through complete genome analysis (Fig 2). The uncorrelated relaxed clock (-25893.0, Fig 1) outperformed the strict clock model (-25999.2, S2 Fig). The results were supportive of a Palearctic origin for the lyssaviruses (posterior probability (PP) = 0.85, Fig 1, PP = 0.86 S2 Fig, strict clock). The results suggest there were three independent transmission events from the Palearctic to the African region, one each from the three putative phylogroups (Fig 1). One event led to the presence of IKOV in Africa (node B). Another event led to the distinct PG2 virus clade (SHIBV, LBV and MOKV) having their current African distribution (node A). It has been proposed that EBLV-1 had its origins in Africa, being closely related to DUVV, whereas our analysis suggested there was greater support for DUVV being a subsequent introduction into the Afrotropical region from the Palearctic (PP = 0.96, node C). Likelihood and Parsimony analyses of individual viral species genomes both provided support for these results (Fig 2). Our analysis supports an easterly spread of lyssaviruses to Australasia, the Oriental realm and the Americas. The inclusion of GBLV into our dataset supports previous findings that GBLV shares a most recent common ancestors (MRCA) with RABV (PP = 1), but support for this ancestral state node was the weakest (PP = 0.36). ABLV (previously the most closely related lyssavirus species to RABV) shares a MRCA with GBLV, although support for this ancestral state is also weak (PP = 0.42) (Fig 1).

Though not the aim of our analysis, we also estimate the time to the MRCA (tMRCA). The uncertainty around these estimates is large; however they suggest that these events probably took place tens of thousands of years ago. The three median tMRCA for the branching events relating to the Palearctic to Afrotropical region clades are 20820 years ago (95% highest posterior density, HPD, 3995 to 166820) for the PG1/PG2 ancestors (node A), 9676 year ago (1102 to 83408 95% HPD) for IKOV and LLEBV MRCA (node B) and 9048 years ago (1405 to 77181 95% HPD) for the DUVV and EBLV-1 MRCA (node C). The results also suggest that once RABV entered the Americas there was widespread dispersal of RABV between the Neotropical and Nearctic regions (Fig 1). Estimated tMRCA for the earliest RABV in our dataset is 3726 years ago (593 to 31478 95% HPD, node D).

#### Discussion

Our analyses support a Palearctic origin for the lyssaviruses (PP = 0.85). This is despite the high diversity of lyssaviruses found in the Africotropical region. The support for most state probabilities is high, suggesting there is strong geographic and temporal structure to lyssavirus evolution as previously demonstrated by RABV in non-volant Carnivora species [41, 42]. Our estimation of the temporal origins of extant lyssaviruses varies depending on whether we used a fixed rate of evolution or estimated the rate using a relaxed clock. This does not, however, affect the phylogeographic inferences from this analysis because of the accurate estimation of evolutionary relationships among species (Figs 1 and 2). These analyses reject the hypothesis that lyssaviruses emerged as viruses of bats in Africa and suggest three distinct emergence events from the Palearctic region into the Afrotropical region (Fig 1).

There is topological support for the current phylogroups (PG1-3), each with monophyletic origins (Fig 1). A PG 3, with lower support, contains species IKOV, LLEBV and WCBV. PG 2 has a single Afrotropical ecozone range composed by SHIBV, MOKV and LBV. PG 1 divides into two major lineages, one which contains IRKV, EBLV-1 and DUVV, and another lineage





Fig 2. Ancestral state reconstruction using complete genomes of the 16 *Lyssavirus* species. The reconstruction was based on the Maximum Likelihood tree and using the parsimony (A) and likelihood (B) models in Mesquite v.3.0.4. Coloured pie-charts represent proportions generated from the different assigned states of the character (see colour legends). Grey terminal pie-chart in likelihood analysis indicated a polymorphic state that was coded as uncertain in the data matrix for RABV species. Support values are indicated above branches and correspond to bootstrap and posterior probabilities, respectively. Virus names are detailed in Fig 1.

doi:10.1371/journal.pntd.0005266.g002

that includes the Palearctic (ARAV, BBLV, KHUV and EBLV-2), Australian (ABLV), Oriental (GBLV) and American (RABV) lyssavirus species. A possible mechanism for the distribution of these viruses may be provided by considering the hosts of WCBV and LLEBV (PG3) and potential hosts of EBLV-1 and DUVV (PG1). Each of these viruses has been detected in the same bat genus, Miniopterus, thus these bats may have played a role in the inter-continental spread of lyssaviruses. This bat genus occurs in the Afrotropical and Palearctic ecozones, though species such as *Miniopterus schreibersi* previously thought to be distributed across both ecozones are now recognized to have more complex taxonomies [43]. However, the Palearctic species, Miniopterus schreibersi, from which WCBV was isolated in the Russian Caucasus and LLEBV and EBLV-1 were identified in Spain, occurs in both North Africa and Eurasia. WCBV is yet to be detected in Africa, but specific antibodies against WCBV have been detected in African Miniopterus in Kenya, providing possible evidence of a bat lyssavirus species circulating in both Africa and Europe [14, 44, 45]. There is no evidence of cross-neutralization between WCBV and IKOV (most closely related lyssavirus to WCBV), suggesting the antibodies detected in the *Miniopterus* spp in Kenya were WCBV specific [46]. The genetically close relationship between EBLV-1 and DUVV has been used as evidence of viruses from Africa

entering Europe [47–49]. However, in our analysis, a greater support for transmission of EBLV-1 from the Palearctic to the Afrotropical zone was observed. The observation that both EBLV-1 and DUVV share a common ancestor with IRKV (Palearctic) strengthens this finding. The isolation of SHIBV from *Hipposideros vittatus* in Kenya [13] is the first lyssavirus isolation from a bat of the genus *Hipposideros*. *Hipposideros* has a broad distribution in the Old World from tropical Africa through to China, although like *Miniopterus* each species has a more specific distribution. Further sampling within this taxon will determine if SHIBV has crossed from the Palearctic to Afrotropical ecozones, as may be the case from the limited data available from WCBV [13] and our analyses (Fig 1).

Our results were confirmed through complete genome analyses with single representatives from each species. Despite this strong support for our conclusion that lyssaviruses have spread from the Palearctic region, it should be recognized that two of the most divergent viruses, WCBV and LLEBV, isolated from Russia [14] and Spain [20] respectively, may influence these findings [50, 51] and the ancestors of the *Lyssavirus* genus themselves may have originated from outside the Palearctic region, indeed even within the Afrotropical ecozone. The relationships of RNA viruses can be influenced by the different processes, such as through high evolutionary rates and subsequent purifying selection [52, 53]. We observed that EBLV-1 had reduced nucleotide substitutions per site compared to other viruses. The earliest EBLV-1 sequence is from 1968 and the sequences are taken from a relatively large geographic area (ranging from Ukraine to Spain), however EBLV-1 has a restricted host range. EBLV-1 has only ever been detected in the serotine bat (Eptesicus serotinus) outside Spain. In Spain, it has been reported from six European insectivorous bat species through active survey, although the majority were E. serotinus [47, 54]. Therefore, future studies should determine if host restriction is reducing the rate of nucleotide substitution and by what mechanism. The data from Spain may be evidence of this virus expanding into new niches and undergoing a bottleneck that reduced genetic diversity during the process. However, of the relatively well studied bat lyssaviruses, EBLV-2 has even greater known host restriction to *M. daubentonii*, and narrower geographic range, so alternative mechanisms may be responsible for the reduced substitution rate we estimated. As more full length genomes from more viruses become available, especially from bats, better inference can be made regarding evolutionary relationships for lyssaviruses [22]. In particular, we suggest that future studies aim to discover more viruses from Africa, Asia and the eastern Palearctic region.

Other significant dispersal events observed from this dataset are from the Palearctic to Australasia and between the Nearctic and Neotropical ecozones in both directions. These analyses provide support for the spread of lyssaviruses, in an easterly direction, from the Palearctic to Australasia and subsequent colonization into the Americas. Australia is free of RABV within its wildlife population with only occasional cases of imported human rabies [55-58], however, isolation of ABLV from sick bats [59], people [60–63], and through surveillance in bats [64– 66] suggest ABLV is well established. Indeed, two distinct ABLV lineages apparently circulate, one now isolated from all four species of Australian Pteropodidae and another from an insectivorous bat species, Saccolaimus flaviventris [59, 64, 65, 67]. The isolation of GBLV from Pteropus medius in Sri Lanka in 2015 was significant because it is more closely related to RABV than any of the other Old World lyssaviruses currently identified. The host reservoir of GBLV (*P. medius*) is interesting for a number of reasons when considering spread of lyssaviruses between ecozones. Firstly, ABLV has been isolated from all four Pteropus spp in Australia. Secondly, members of the *Pteropus* genus are present throughout Asia and Australasia, providing a possible mechanism for transmission between the two ecozones. Thirdly, serosurveillance of bat populations in Asia has detected lyssavirus-specific antibodies, yet no virus had been

isolated [16]. However, the ABLV host *S. flaviventris* has also been reported in Papua New Guinea [68] suggesting other pathways may exist.

Our analyses of the RABV data provide strong support for transmission between the American ecozones (Fig 1). This may be due to two, not mutually exclusive mechanisms. One is that despite recent evidence of host phylogeny constraining inter-species virus transmission with the USA [69], RABV has been isolated from over 23 bat species in the USA alone. Host signatures for species variants exist, but adaptation is not sufficient to prevent cross-species transmission [69]. Therefore, mutations in the RABV genome may have led to reduced host restriction and RABV being able to spread more easily between bats in both North and South American locations. An additional mechanism for this finding may be the hosts themselves. There are highly sociable, numerous, and/or migratory bats which occur throughout the Americas, such as *Tadarida brasiliensis, Lasiurus cinereus*, and *Eptesicus fuscus*. These migratory and widespread bat species of the Americas may have rapidly disseminated RABV between ecozones, enabling the promiscuous RABV to rapidly exploit unoccupied niches. Furthermore, the presence of *L. cinereus* in both Hawaii and British Columbia (http://www. iucnredlist.org/details/11345/0) demonstrates how a migratory species may be a potential vector of RABV or a RABV ancestor from the Palearctic to the Americas.

Our molecular clock analyses provide support for the hypothesis that RABV was circulating in the bat populations of the Americas before the arrival of Europeans in the late  $15^{\text{th}}$  Century. Previously it has been claimed that Spanish conquistadors reported attacks by bats on humans and that native Americans knew that cauterisation may prevent disease development [70]. The median tMRCA for GBLV and RABV is >7,000 years ago. The median tMRCA for the first internal branch in RABV is 3726 years ago (593 to 31478 95% HPD). We are cautious when interpreting the ages of RNA viruses using molecular clock analysis because of the impacts of purifying selection on RNA viruses [52, 71], however purifying selection should push our dates further into the past rather than bring the estimates forward in time. Thus, we are confident that our analysis provides support for the probable presence of RABV in bats, and likely in the Americas, before the arrival of Europeans in 1492CE, because our lowest 95% HPD RABV tMRCA date is 1423CE and the median estimate is 3726 years ago in 1790 BCE.

The origins of RABV in dogs is debated and phylogenetic analyses have questioned the reports of RABV from ancient Greek references [70]. Our analyses demonstrate that lyssaviruses were almost certainly circulating in Palearctic bats at this time. Similarly, whether estimating evolutionary rates using relaxed clocks as in other analyses [35] or fixing the evolutionary rate, there appear to be extant RABV circulating in bats in the Americas at this time. Thus, it may be possible that the reports from 23<sup>rd</sup> Century BC are due to an "extant" RABV if spillover had occurred from bats to terrestrial carnivores already. Future analyses of all extant lyssaviruses accounting for purifying selection may help elucidate these relationships further [52].

Lastly, extant Chiroptera bats likely originated in Asia/Europe and young clades are found in the Americas [72]. This biogeographic reconstruction reflects a similar pattern as the one found in the *Lyssavirus* genus. However, the time of divergence for extant Chiroptera is in Millions of years and there is no information to suggest co-speciation between bats and lyssavirus reflects a possible ancient origin of these viruses, as has been found in other groups (e.g. coronaviruses [71] and papillomaviruses [73]).

In conclusion, our analyses provide support for the monophyletic, Palearctic origins of lyssaviruses with dispersal from there to the rest of the world. And while three dispersal events have been from the Palearctic to the Afrotropical regions, arguably the dispersal events that led to the greatest impact on animal and human health are those eastward, where the RABV species appears to have evolved and dispersed globally from, leading to 23,000–93,000 human deaths a year [74]. Understanding why this lyssavirus, but not others, has emerged globally will provide insights into the processes that drive viral emergence.

## **Supporting Information**

**S1 Table. Partial nucleoprotein (N) or complete (C) lyssavirus genome sequences from GenBank used in the analysis.** Virus names are as Fig 1. (DOCX)

S1 Fig. Phylogeny of a panel of lyssaviruses showing the slow rate of nucleotide substitutions for European bat lyssavirus-1 (green) compared to the other viruses (black). (PDF)

S2 Fig. Evolutionary relationships between lyssaviruses based on the nucleoprotein (N) gene with a strict clock analysis using a fixed evolutionary rate of  $2.3 \times 10^{-4}$  (according to [35]) in BEAST. Support values corresponding to Bayesian posterior probabilities are indicated. Tips are labelled with the following information: Ecozone, follow by the\_species name, country where it was isolated, GenBank accession number and year of isolation. Virus names and other details are as Fig 1. (PDF)

**S3 Fig. Evolutionary relationships between lyssaviruses.** The phylogenetic tree was generated from 153 nucleoprotein gene sequences and inferred with a Lognormal relaxed-clock Bayesian analysis using BEAST. Support values corresponding to Bayesian state probabilities from the different assigned ecozones are indicated. Tips are labelled with the following information: Ecozone, follow by the species name, country where it was isolated, GenBank accession number and year of isolation. Virus names and other details are as <u>Fig 1</u>. (PDF)

**S4 Fig. Evolutionary relationships between lyssaviruses.** The phylogenetic tree was generated from 153 nucleoprotein gene sequences and inferred with a Lognormal relaxed-clock Bayesian analysis using BEAST showing divergence times in years with 95% credible intervals. Branch colours correspond to ecozones shown on the map. The time scale is in years. Virus names are as Fig 1.

(PDF)

**S5 Fig. Ancestral state reconstruction using complete genomes of the 16** *Lyssavirus* **species based on the Maximum Likelihood tree using likelihood models in Mesquite v.3.0.4.** Coloured pie-charts represent proportions generated from the different assigned states of the character (see colour legends). The grey terminal pie-chart indicated a polymorphic state that was coded as uncertain in the data matrix for RABV species. Support values are indicated above branches and correspond to bootstrap and posterior probabilities, respectively. Virus names are as Fig 1.

(PDF)

**S6 Fig. Ancestral state reconstruction using complete genomes of the 16** *Lyssavirus* **species based on the Maximum Likelihood tree and using the parsimony models in Mesquite v.3.0.4.** Coloured pie-charts represent proportions generated from the different assigned states of the character (see colour legends). Support values are indicated above branches and correspond to bootstrap and posterior probabilities, respectively. Virus names are as Fig 1. (PDF)

#### Acknowledgments

Thank you to Dr. Jessica Hedge, University of Oxford, for early useful discussions and two anonymous reviewers for their thoughtful reviews.

#### **Author Contributions**

Conceptualization: DTSH.

Data curation: JCGR DAM.

Formal analysis: DTSH JCGR.

Funding acquisition: DTSH ARF.

Investigation: DTSH JCGR.

Methodology: DTSH JCGR.

Resources: ARF DAM.

Software: DTSH JCGR DAM.

Supervision: DTSH.

Validation: ARF DAM.

Visualization: DTSH JCGR.

Writing - original draft: DTSH.

Writing - review & editing: DTSH JCGR DAM ARF.

#### References

- Bourhy H, Cowley JA, Larrous F, Holmes EC, Walker PJ. Phylogenetic relationships among rhabdoviruses inferred using the L polymerase gene. Journal of General Virology. 2005; 86(Pt 10):2849–58. doi: 10.1099/vir.0.81128-0 PMID: 16186241
- Afonso CL, Amarasinghe GK, Banyai K, Bao Y, Basler CF, Bavari S, et al. Taxonomy of the order Mononegavirales: update 2016. Archives of Virology. 2016; 161(8):2351–60. doi: <u>10.1007/s00705-016-2880-1</u> PMID: <u>27216929</u>
- 3. Real LA, Russell C, Waller L, Smith D, Childs J. Spatial dynamics and molecular ecology of North American rabies. Journal of Heredity. 2005; 96(3):253–60. doi: 10.1093/jhered/esi031 PMID: 15677743
- 4. Hass CC, Dragoo JW. Rabies in hooded and striped skunks in Arizona. Journal of Wildlife Diseases. 2006; 42(4):825–9. doi: 10.7589/0090-3558-42.4.825 PMID: 17255450
- Velasco-Villa A, Orciari LA, Souza V, Juarez-Islas V, Gomez-Sierra M, Castillo A, et al. Molecular epizootiology of rabies associated with terrestrial carnivores in Mexico. Virus Research. 2005; 111(1):13–27. doi: 10.1016/j.virusres.2005.03.007 PMID: 15896399
- Rupprecht CE, Smith JS, Fekadu M, Childs JE. The ascension of wildlife rabies: a cause for public health concern or intervention? Emerging Infectious Diseases. 1995; 1(4):107–14. doi: <u>10.3201/</u> eid0104.950401 PMID: 8903179
- von Teichman BF, Thomson GR, Meredith CD, Nel LH. Molecular epidemiology of rabies virus in South Africa: evidence for two distinct virus groups. The Journal of General Virology. 1995; 76 (Pt 1)(1):73– 82.
- Nel LH, Thomson GR, Von Teichman BF. Molecular epidemiology of rabies virus in South Africa. Onderstepoort Journal of Veterinary Research. 1993; 60(4):301–6. PMID: 7777315
- 9. Swanepoel R, Barnard BJ, Meredith CD, Bishop GC, Bruckner GK, Foggin CM, et al. Rabies in southern Africa. Onderstepoort Journal of Veterinary Research. 1993; 60(4):325–46. PMID: 7777317
- Davis PL, Rambaut A, Bourhy H, Holmes EC. The evolutionary dynamics of canid and mongoose rabies virus in Southern Africa. Archives of Virology. 2007; 152(7):1251–8. doi: 10.1007/s00705-007-0962-9 PMID: <u>17401615</u>

- Childs J, Real AL. Epidemiology. In: Jackson AC, Wunner WH, editors. Rabies. 2nd ed. London: Elsevier; 2007. p. 123–200.
- Badrane H, Tordo N. Host switching in Lyssavirus history from the Chiroptera to the Carnivora orders. Journal of Virology. 2001; 75(17):8096–104. doi: 10.1128/JVI.75.17.8096-8104.2001 PMID: 11483755
- Kuzmin IV, Mayer AE, Niezgoda M, Markotter W, Agwanda B, Breiman RF, et al. Shimoni bat virus, a new representative of the Lyssavirus genus. Virus Research. 2010; 149(2):197–210. doi: 10.1016/j. virusres.2010.01.018 PMID: 20138934
- Kuzmin IV, Hughes GJ, Botvinkin AD, Orciari LA, Rupprecht CE. Phylogenetic relationships 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. 2005; 111(1):28–43. doi: 10. 1016/j.virusres.2005.03.008 PMID: 15896400
- Kuzmin IV, Orciari LA, Arai YT, Smith JS, Hanlon CA, Kameoka Y, et al. Bat lyssaviruses (Aravan and Khujand) from Central Asia: phylogenetic relationships according to N, P and G gene sequences. Virus Research. 2003; 97(2):65–79. PMID: <u>14602198</u>
- Banyard AC, Hayman D, Johnson N, McElhinney L, Fooks AR. Bats and lyssaviruses. Advances in Virus Research. 2011; 79:239–89. doi: 10.1016/B978-0-12-387040-7.00012-3 PMID: 21601050
- Fooks A. The challenge of new and emerging lyssaviruses. Expert Review of Vaccines. 2004; 3 (4):333–6. doi: 10.1586/14760584.3.4.333 PMID: 15270628
- Horton DL, McElhinney LM, Marston DA, Wood JLN, Russell CA, Lewis N, et al. Quantifying Antigenic Relationships among the Lyssaviruses. Journal of Virology. 2010; 84(22):11841–8. doi: 10.1128/JVI. 01153-10 PMID: 20826698
- Freuling CM, Beer M, Conraths FJ, Finke S, Hoffmann B, Keller B, et al. Novel lyssavirus in Natterer's bat, Germany. Emerging Infectious Diseases. 2011; 17(8):1519–22. doi: 10.3201/eid1708.110201 PMID: 21801640
- Ceballos NA, Moron SV, Berciano JM, Nicolas O, Lopez CA, Juste J, et al. Novel Lyssavirus in Bat, Spain. Emerging Infectious Diseases. 2013; 19(5):793–5. doi: <u>10.3201/eid1905.121071</u> PMID: 23648051
- Marston DA, Ellis RJ, Wise EL, Aréchiga-Ceballos N, Freuling CM, Banyard AC, et al. Complete genomic sequence of Lleida bat lyssavirus. Genome Announcements 2017; 5(2):e01427–16.
- 22. Gunawardena PS, Marston DA, Ellis RJ, Wise EL, Karawita AC, Breed AC, et al. Lyssavirus in Indian Flying Foxes, Sri Lanka. Emerging Infectious Diseases. 2016; 22(8):1456–9. doi: 10.3201/eid2208. 151986 PMID: 27434858
- Marston DA, Ellis RJ, Horton DL, Kuzmin IV, Wise EL, McElhinney LM, et al. Complete genome sequence of Ikoma lyssavirus. Journal of Virology. 2012; 86(18):10242–3. doi: 10.1128/JVI.01628-12 PMID: 22923801
- Nel LH, Rupprecht CE. Emergence of lyssaviruses in the Old World: the case of Africa. Current Topics in Microbiology and Immunology. 2007; 315:161–93. PMID: 17848065
- Velasco-Villa A, Reeder SA, Orciari LA, Yager PA, Franka R, Blanton JD, et al. Enzootic rabies elimination from dogs and reemergence in wild terrestrial carnivores, United States. Emerging Infectious Diseases. 2008; 14(12):1849–54. doi: 10.3201/eid1412.080876 PMID: 19046506
- Childs JE, Curns AT, Dey ME, Real AL, Rupprecht CE, Krebs JW. Rabies epizootics among raccoons vary along a North-South gradient in the Eastern United States. Vector-borne and Zoonotic Diseases. 2001; 1(4):253–67. doi: 10.1089/15303660160025895 PMID: 12653126
- Flamand A, Coulon P, Lafay F, Kappeler A, Artois M, Aubert M, et al. Eradication of rabies in Europe. Nature. 1992; 360(6400):115–6. doi: 10.1038/360115a0 PMID: 1436089
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23(21):2947–8. doi: <u>10.1093/bioinformatics/btm404</u> PMID: <u>17846036</u>
- Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology. 2007; 7:214. doi: 10.1186/1471-2148-7-214 PMID: 17996036
- Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics. 1998; 14 (9):817–8. PMID: 9918953
- Davis PL, Holmes EC, Larrous F, Van der Poel WH, Tjornehoj K, Alonso WJ, et al. Phylogeography, population dynamics, and molecular evolution of European bat lyssaviruses. Journal of Virology. 2005; 79(16):10487–97. doi: 10.1128/JVI.79.16.10487-10497.2005 PMID: 16051841
- Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. Journal of Virology. 2010; 84(19):9733–48. doi: 10.1128/JVI.00694-10 PMID: 20660197

- Holmes EC. Error thresholds and the constraints to RNA virus evolution. Trends in Microbiology. 2003; 11(12):543–6. PMID: 14659685
- **34.** Holmes EC. Molecular clocks and the puzzle of RNA virus origins. Journal of Virology. 2003; 77 (7):3893–7. doi: 10.1128/JVI.77.7.3893-3897.2003 PMID: 12634349
- Bourhy H, Reynes JM, Dunham EJ, Dacheux L, Larrous F, Huong VT, et al. The origin and phylogeography of dog rabies virus. The Journal of General Virology. 2008; 89(Pt 11):2673–81. doi: 10.1099/vir. 0.2008/003913-0 PMID: 18931062
- 36. Rambaut A, Drummond A. TreeAnnotator v1. 7.5. 2013.
- Rambaut A. FigTree v. 1.4. Molecular evolution, phylogenetics and epidemiology. Edinburgh, UK: University of Edinburgh, Institute of Evolutionary Biology. 2012.
- Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22(21):2688–90. doi: <u>10.1093/bioinformatics/btl446</u> PMID: <u>16928733</u>
- 39. Maddison W, Maddison DR. Mesquite: a modular system for evolutionary analysis. 2001.
- Holt BG, Lessard J-P, Borregaard MK, Fritz SA, Araújo MB, Dimitrov D, et al. An update of Wallace's zoogeographic regions of the world. Science. 2013; 339(6115):74–8. doi: 10.1126/science.1228282 PMID: 23258408
- Biek R, Henderson JC, Waller LA, Rupprecht CE, Real LA. A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. Proceedings of the National Academy of Sciences U S A. 2007; 104(19):7993–8.
- 42. Real LA, Henderson JC, Biek R, Snaman J, Jack TL, Childs JE, et al. Unifying the spatial population dynamics and molecular evolution of epidemic rabies virus. Proceedings of the National Academy of Sciences U S A. 2005; 102(34):12107–11.
- Miller-Butterworth CM, Eick G, Jacobs DS, Schoeman MC, Harley EH. Genetic and phenotypic differences between South African long-fingered bats, with a global miniopterine phylogeny. Journal of Mammalogy. 2005; 86(6):1121–35.
- Kuzmin IV, Niezgoda M, Franka R, Agwanda B, Markotter W, Beagley JC, et al. Possible emergence of West Caucasian bat virus in Africa. Emerging Infectious Diseases. 2008; 14(12):1887–9. doi: 10.3201/ eid1412.080750 PMID: 19046512
- Botvinkin AD, Poleschuk EM, Kuzmin IV, Borisova TI, Gazaryan SV, Yager P, et al. Novel lyssaviruses isolated from bats in Russia. Emerging Infectious Diseases. 2003; 9(12):1623–5. doi: 10.3201/eid0912. 030374 PMID: 14720408
- Horton DL, Banyard AC, Marston DA, Wise E, Selden D, Nunez A, et al. Antigenic and genetic characterization of a divergent African virus, Ikoma lyssavirus. Journal of General Virology. 2014; 95(Pt 5):1025–32. doi: 10.1099/vir.0.061952-0 PMID: 24496827
- Serra-Cobo J, Amengual B, Abellan C, Bourhy H. European bat lyssavirus infection in Spanish bat populations. Emerging Infectious Diseases. 2002; 8(4):413–20. doi: <u>10.3201/eid0804.010263</u> PMID: 11971777
- Schneider LG, Cox JH. Bat lyssaviruses in Europe. Current Topics in Microbiology and Immunology. 1994; 187:207–18. PMID: 7859491
- Amengual B, Whitby JE, King A, Cobo JS, Bourhy H. Evolution of European bat lyssaviruses. The Journal of general virology. 1997; 78 (Pt 9):2319–28.
- Crisp MD, Cook LG. Do early branching lineages signify ancestral traits? Trends in Ecology and Evolution. 2005; 20(3):122–8. doi: 10.1016/j.tree.2004.11.010 PMID: 16701355
- Losos JB. Commentaries—Uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. Animal Behaviour. 1999; 58(6):1319–24. doi: 10.1006/anbe.1999.1261 PMID: 10600155
- Wertheim JO, Pond SLK. Purifying Selection Can Obscure the Ancient Age of Viral Lineages. Molecular Biology and Evolution. 2011; 28(12):3355–65. doi: 10.1093/molbev/msr170 PMID: 21705379
- Revell LJ, Harmon LJ, Collar DC. Phylogenetic signal, evolutionary process, and rate. Systematic Biology. 2008; 57(4):591–601. doi: 10.1080/10635150802302427 PMID: 18709597
- Echevarria JE, Avellon A, Juste J, Vera M, Ibanez C. Screening of active lyssavirus infection in wild bat populations by viral RNA detection on oropharyngeal swabs. Journal of Clinical Microbiology. 2001; 39 (10):3678–83. doi: 10.1128/JCM.39.10.3678-3683.2001 PMID: 11574590
- McColl KA, Gould AR, Selleck PW, Hooper PT, Westbury HA, Smith JS. Polymerase chain reaction and other laboratory techniques in the diagnosis of long incubation rabies in Australia. Australian Veterinary Journal. 1993; 70(3):84–9. PMID: 8476363

- Faoagali JL, De Buse P, Strutton GM, Samaratunga H. A case of rabies. Medical Journal of Australia. 1988; 149(11–12):702–7. PMID: 3200197
- Johnson N, Fooks A, McColl K. Human rabies case with long incubation, Australia. Emerging Infectious Diseases. 2008; 14(12):1950–1. doi: 10.3201/eid1412.080944 PMID: 19046531
- McColl KA, Chamberlain T, Lunt RA, Newberry KM, Middleton D, Westbury HA. Pathogenesis studies with Australian bat lyssavirus in grey-headed flying foxes (Pteropus poliocephalus). Australian Veterinary Journal. 2002; 80(10):636–41. PMID: 12465817
- 59. Fraser GC, Hooper PT, Lunt RA, Gould AR, Gleeson LJ, Hyatt AD, et al. Encephalitis caused by a Lyssavirus in fruit bats in Australia. Emerging Infectious Diseases. 1996; 2(4):327–31. doi: 10.3201/ eid0204.960408 PMID: 8969249
- Warrilow D. Australian bat lyssavirus: a recently discovered new rhabdovirus. Current Topics in Microbiology and Immunology. 2005; 292:25–44. PMID: 15981466
- Warrilow D, Smith IL, Harrower B, Smith GA. Sequence analysis of an isolate from a fatal human infection of Australian bat lyssavirus. Virology. 2002; 297(1):109–19. PMID: 12083841
- Hanna JN, Carney IK, Smith GA, Tannenberg AE, Deverill JE, Botha JA, et al. Australian bat lyssavirus infection: a second human case, with a long incubation period. Medical Journal of Australia. 2000; 172 (12):597–9. PMID: <u>10914106</u>
- Allworth AM, Murray K, Morgan J. A human case of encephalitis due to a lyssavirus recently identified in fruit bats. Communicable Diseases Intelligence. 1996; 20:325.
- Gould AR, Kattenbelt JA, Gumley SG, Lunt RA. Characterisation of an Australian bat lyssavirus variant isolated from an insectivorous bat. Virus Research. 2002; 89(1):1–28. PMID: 12367747
- Gould AR, Hyatt AD, Lunt R, Kattenbelt JA, Hengstberger S, Blacksell SD. Characterisation of a novel lyssavirus isolated from Pteropid bats in Australia. Virus Research. 1998; 54(2):165–87. PMID: 9696125
- 66. Hooper PT, Lunt RA, Gould AR, Samaratunga H, Hyatt AD, Gleeson LJ, et al. A new lyssavirus—the first endemic rabies-related virus recognized in Australia. B I Pasteur. 1997; 95(4):209–18.
- Guyatt KJ, Twin J, Davis P, Holmes EC, Smith GA, Smith IL, et al. A molecular epidemiological study of Australian bat lyssavirus. The Journal of General Virology. 2003; 84(Pt 2):485–96. doi: 10.1099/vir.0. 18652-0 PMID: 12560583
- 68. Flannery TF. Mammals of New Guinea: Reed; 1995.
- Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. Science. 2010; 329 (5992):676–9. doi: 10.1126/science.1188836 PMID: 20689015
- Baer G. The History of Rabies. In: Jackson AC, Wunner WH, editors. Rabies. 2nd ed. London: Academic Press; 2007.
- Wertheim JO, Chu DK, Peiris JS, Kosakovsky Pond SL, Poon LL. A case for the ancient origin of coronaviruses. Journal of Virology. 2013; 87(12):7039–45. doi: 10.1128/JVI.03273-12 PMID: 23596293
- 72. Teeling EC, Springer MS, Madsen O, Bates P, O'Brien S J, Murphy WJ. A molecular phylogeny for bats illuminates biogeography and the fossil record. Science. 2005; 307(5709):580–4. doi: 10.1126/science. 1105113 PMID: 15681385
- Rector A, Lemey P, Tachezy R, Mostmans S, Ghim SJ, Van Doorslaer K, et al. Ancient papillomavirushost co-speciation in Felidae. Genome Biology. 2007; 8(4):R57. doi: 10.1186/gb-2007-8-4-r57 PMID: 17430578
- Knobel DL, Cleaveland S, Coleman PG, Fevre EM, Meltzer MI, Miranda ME, et al. Re-evaluating the burden of rabies in Africa and Asia. Bulletin of the World Health Organisation. 2005; 83(5):360–8.