

Article

Evaluating the Impact of Anthropogenic Factors on the Dissemination of Contemporary Cosmopolitan, Arctic, and Arctic-like Rabies Viruses

Andrei A. Deviatkin ^{1,2,*} , Yulia A. Vakulenko ^{3,4} , Mariia A. Dashian ⁵ and Alexander N. Lukashev ³

¹ Laboratory of Molecular Biology and Biochemistry, Institute of Molecular Medicine, Sechenov First Moscow State Medical University, 119435 Moscow, Russia

² The National Medical Research Center for Endocrinology, 117036 Moscow, Russia

³ Martsinovskiy Institute of Medical Parasitology, Tropical and Vector Borne Diseases, Sechenov First Moscow State Medical University, 119435 Moscow, Russia; vjulia94@gmail.com (Y.A.V.); alexander_lukashev@hotmail.com (A.N.L.)

⁴ Department of Virology, Faculty of Biology, Lomonosov Moscow State University, 119234 Moscow, Russia

⁵ Faculty of Biomedicine, Pirogov Medical University, 117997 Moscow, Russia; ria.dashyan@gmail.com

* Correspondence: andreideviatkin@gmail.com

Abstract: Rabies is a globally prevalent viral zoonosis that causes 59,000 deaths per year and has important economic consequences. Most virus spread is associated with the migration of its primary hosts. Anthropogenic dissemination, mainly via the transportation of rabid dogs, shaped virus ecology a few hundred years ago and is responsible for several current outbreaks. A systematic analysis of aberrant long-distance events in the steppe and Arctic-like groups of rabies virus was performed using statistical (Bayesian) phylogeography and plots of genetic vs. geographic distances. The two approaches produced similar results but had some significant differences and complemented each other. No phylogeographic analysis could be performed for the Arctic group because polar foxes transfer the virus across the whole circumpolar region at high velocity, and there was no correlation between genetic and geographic distances in this virus group. In the Arctic-like group and the steppe subgroup of the cosmopolitan group, a significant number of known sequences (15–20%) was associated with rapid long-distance transfers, which mainly occurred within Eurasia. Some of these events have been described previously, while others have not been documented. Most of the recent long-distance transfers apparently did not result in establishing the introduced virus, but a few had important implications for the phylogeographic history of rabies. Thus, human-mediated long-distance transmission of the rabies virus remains a significant threat that needs to be addressed.

Keywords: rabies virus; Bayesian; transfer; distances; cosmopolitan RABV; steppe RABV; Arctic RABV; Arctic-like RABV



Citation: Deviatkin, A.A.; Vakulenko, Y.A.; Dashian, M.A.; Lukashev, A.N. Evaluating the Impact of Anthropogenic Factors on the Dissemination of Contemporary Cosmopolitan, Arctic, and Arctic-like Rabies Viruses. *Viruses* **2022**, *14*, 66. <https://doi.org/10.3390/v14010066>

Academic Editor: Charles E. Rupprecht

Received: 7 December 2021

Accepted: 29 December 2021

Published: 30 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rabies virus (RABV) is a negative-sense ssRNA virus belonging to the genus *Lyssavirus*. A number of lyssaviruses can cause a lethal disease called rabies [1]. However, RABV is the only lyssavirus adapted to long-term circulation, not only in bats but also in other mammals, mainly carnivores. Such adaptation is the reason why RABV causes the vast majority of human cases of rabies, with around 59,000 deaths annually [2].

RABV consists of bat-associated and carnivore-associated phylogenetic groups. Bat-associated RABV circulates only among the New World bats. Carnivore-associated RABV may be found in every continent except Australia and the Antarctic [3] and consists of several phylogenetic groups: Cosmopolitan, Africa-3, Arctic-related, Africa-2, Asian, and Indian [3–5].

It has been hypothesized that intercontinental transportation of rabid animals by humans provided the global distribution of cosmopolitan RABV, likely over the last few

hundreds of years [6,7]. However, it is unclear how systematic human-assisted transfer occurred at different times. Only a few cases of recent long-distance anthropogenic transfers have been reported recently. For example, a virus isolated from a bear in the Russian Far East was genetically most close to viruses collected in European Russia [8]; a sequence of human RABV found in the USA grouped with Brazilian viruses on the phylogenetic tree [9].

Expansion of natural RABV reservoirs is relatively slow. The mean annual virus spread rate depends on the species of the reservoir animal and the ecological region [10]. The rate of rabies spread by rabid foxes in Central Europe was estimated to be 30–60 km per year [11]. The speed of epizootic expansion in Western Siberia steppes was significantly higher, 160–513 km per year [12]. The highest virus dissemination rates were observed in the Arctic region and provided by the long-distance migration of Arctic foxes over the frozen ocean in winter [13–15]. This unique ecology of the reservoir host provides a highly dynamic RABV reservoir with virus spread rates of thousands of kilometers per year and almost instant intercontinental transmission of the virus. For example, a rabies virus found in Franz Josef Land (Russia) had a sequence nearly identical to viruses isolated in Alaska (USA) [16]. The origin of this circumpolar Arctic RABV [17,18] reservoir is probably associated with the dissemination of Arctic-like RABV to the Arctic region [19,20]. Currently, Arctic-like RABV is distributed over vast territories in Eurasia, including China, India, Pakistan, Afghanistan, Iran, and Iraq.

Understanding natural virus dissemination patterns is critical for planning preventive measures. Vaccination of humans and domestic animals is an effective measure against the disease. Furthermore, wildlife vaccination can be used to eradicate virus reservoirs in a large territory. Indeed, some parts of the world, e.g., western Europe, have been declared rabies-free areas [21]. Simultaneously, combating against rabies has been less successful in Asia and Africa [2,22]. A single long-distance virus transfer (for example, human mediated) can stall long-term eradication efforts or lead to significant healthcare consequences (reviewed in [23]). However, a systematic exploration of such aberrant transmission events has not been performed.

Historically, finding an RNA virus with a sequence nearly identical to those of viruses from a distant location unequivocally indicates that a long-distance transfer has occurred. In the case of the rabies virus, the simplest explanation is human involvement in pathogen dissemination. This approach may be called the gold standard in the epidemiological investigation of isolated imported cases. However, it is more challenging to analyze the role of anthropogenic dissemination in pathogen evolution and spreading history as a whole, especially as the number of known sequences grows exponentially and goes beyond the capacity of classical analysis tools.

There are several possible approaches to assess the virus spread rate. The first one is based on the dependence between evolutionary distances and the geographical spacing of virus sequences. In that case, all studied objects are analyzed in all possible pairs. Furthermore, evolutionary distances (the proportion of nucleotide sites at which the two compared sequences are different) and geographic distances are calculated for every pair [24]. The results are shown as a scatterplot (referred to as gene-geo plot below). Unlike phylogenetic trees, gene-geo plots do not establish the relationship between individual groups of viruses but show the fingerprint of the geographic and evolutionary distances correspondence. The distribution of dots in this plot may indicate isolated virus transfers or general phylogeographical patterns in the evolution of a virus group as a whole [25].

Another approach, the Bayesian phylogeographic framework, infers virus dissemination velocity within different phylogenetic tree branches. A significant deviation of the inferred spread rates at tree branches would suggest the role of an anthropogenic factor or another possible explanation. This method has two variations: discrete phylogeography that uses isolation locations expressed as fixed categories and continuous phylogeography that uses actual sampling coordinates [26]. A variation of this approach has been applied to characterize RABV spread in North Africa [27]. Continuous phylogeography is highly

applicable to distinct RABV subgroups that have been circulating in one main host with a relatively constant ecological pattern [28] but may be less reliable when there are discrete location groups (such as continents) and various hosts with specific ecological traits.

In the current study, we have implemented both the genetic distance-based approach and continuous phylogeography to several RABV groups, i.e., cosmopolitan, Arctic and Arctic-like, to study the role of anthropogenic virus spread on a global scale.

2. Materials and Methods

Prior to analysis, sequences represented in GenBank as of June 2020 and annotated as belonging to the species *Rabies lyssavirus* (Taxonomy ID: 11292) were extracted ($n = 22,628$). To provide the best balance of sequence length and number of sequences in the data set, 7616 entries aligning with genome positions 170–1150 in the reference sequence NC_001542 were selected (see below) (Figure S1). The data set was generated using in-house scripts. Then, sequences were aligned using MAFFT v7.310 [29].

Data sets for RABV genome coverage visualization were generated using open-source tools (Supplementary File S1). Then, artificial sequences and sequences with ambiguous positions in the selected 981 nt region were omitted, leaving 7565 genome fragments.

Maximum likelihood (ML) phylogenetic inference (Figure 1) was performed using IQ-TREE [30] under the GTR+G4 model (alignment file available at https://raw.githubusercontent.com/AndreiDeviatkin/gene_geo/main/data/fasta/txid11292_260620_950%2B_210-1140_cut_al_final.fas, assessed on 28 December 2021).

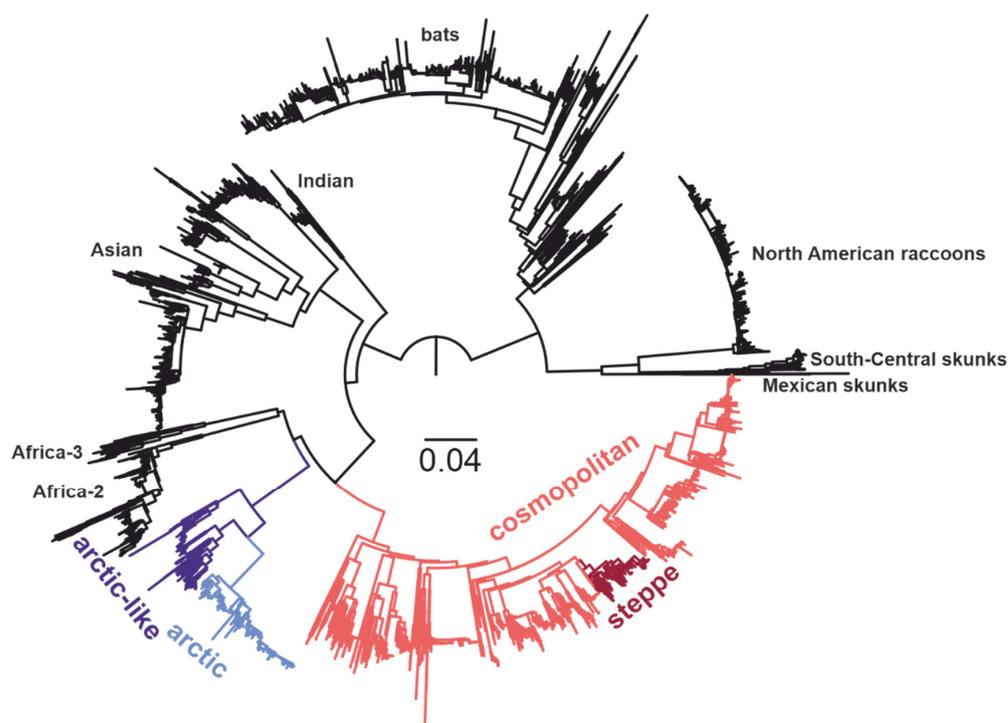


Figure 1. Maximum likelihood tree for RABV (alignment length = 981 nucleotides, number of sequences = 7565). RABV groups analyzed here are indicated by color: cosmopolitan group—pale red; steppe subgroup of the cosmopolitan group—dark red; Arctic-like group—dark blue; Arctic group—pale blue.

Then, cosmopolitan ($n = 2423$) and Arctic-related (Arctic and Arctic-like groups, $n = 720$) RABV sequences were extracted from the whole data set. Furthermore, vaccine-associated sequences were omitted. Arctic and Arctic-like groups were analyzed separately due to suggestively different ecological traits. The steppe subgroup of cosmopolitan viruses

was also selected for a separate analysis as one of the best-explored subgroups with a known anthropogenic transfer event.

Finally, only sequences with known dates and locations of sample collection were selected, resulting in the following data sets: cosmopolitan ($n = 1824$), steppe subgroup of cosmopolitan ($n = 205$), Arctic ($n = 453$) and Arctic-like ($n = 201$). Identical sequences with an identical place of sample collection were omitted in the Arctic group before Bayesian analysis, so only 292 out of 453 sequences were used.

Bayesian continuous phylogeographic analysis was performed for the Arctic-like and steppe data sets using the Markov chain Monte Carlo (MCMC) approach implemented in BEAST 1.10.4 [31]. The BEAGLE library v3.1.0 was used to accelerate computation [32]. Prior to analysis, the data sets were tested with TempEst v. 1.5.3 [33] for the presence of a temporal signal.

Maximum clade credibility (MCC) trees were inferred under the relaxed random walk model with a Cauchy distributed diffusion rate variation among branches, which suggests heterogeneity in branch velocities. We used a GTR substitution model with the uncorrelated relaxed log-normal molecular clock assumption [34] and Bayesian skyline population size. MCMC chains were run for 700 and 713 million generations, sampling every 50,000 steps for the steppe and Arctic-like data sets, respectively. MCMC performance was inspected for convergence using Tracer v.1.7 [35]. Maximum clade credibility (MCC) trees were annotated with TreeAnnotator v.1.10.8 using a burn-in of 250 and 71 million generations for the steppe and Arctic-like data sets, respectively. Trees were visualized with FigTree v.1.4.4 [36]. Velocities for MCC tree branches were calculated and then visualized using in-house python script (Supplementary File S2). The same procedure was implemented for control data sets without obvious signs of long-distance transfers.

A map with the locations of sample collection (Figure 2D) was generated in the R environment. The precise coordinates of virus collection sites were not specified in the annotations of most sequences. Therefore, the coordinates of capitals of countries smaller than 1,000,000 sq km were set as places of virus collection. Countries and territories with an area of more than 1,000,000 sq km, i.e., Russia, China, Mongolia, Kazakhstan, Canada, Greenland, USA, Brazil, India, were subdivided into smaller regions according to their administrative regions to infer virus sampling coordinates.

An interactive map with labeled markers is available at <https://rpubs.com/andreideviatkin/RABVdistribution> (assessed on 28 December 2021). In order to avoid the overlap of isolates from the same region, a slight random noise was introduced into the coordinates used in this map.

Genetic and geographic distance concordances (Gene-geo plots) were visualized in the R environment. Gene-geo plots have not demonstrated consistent statistical patterns for tested data sets. The source code is available at https://github.com/AndreiDeviatkin/gene_geo (assessed on 28 December 2021). This code was implemented in the R shiny webserver as an online tool available at <https://andreideviatkin.shinyapps.io/gene-geo/> (assessed on 28 December 2021).

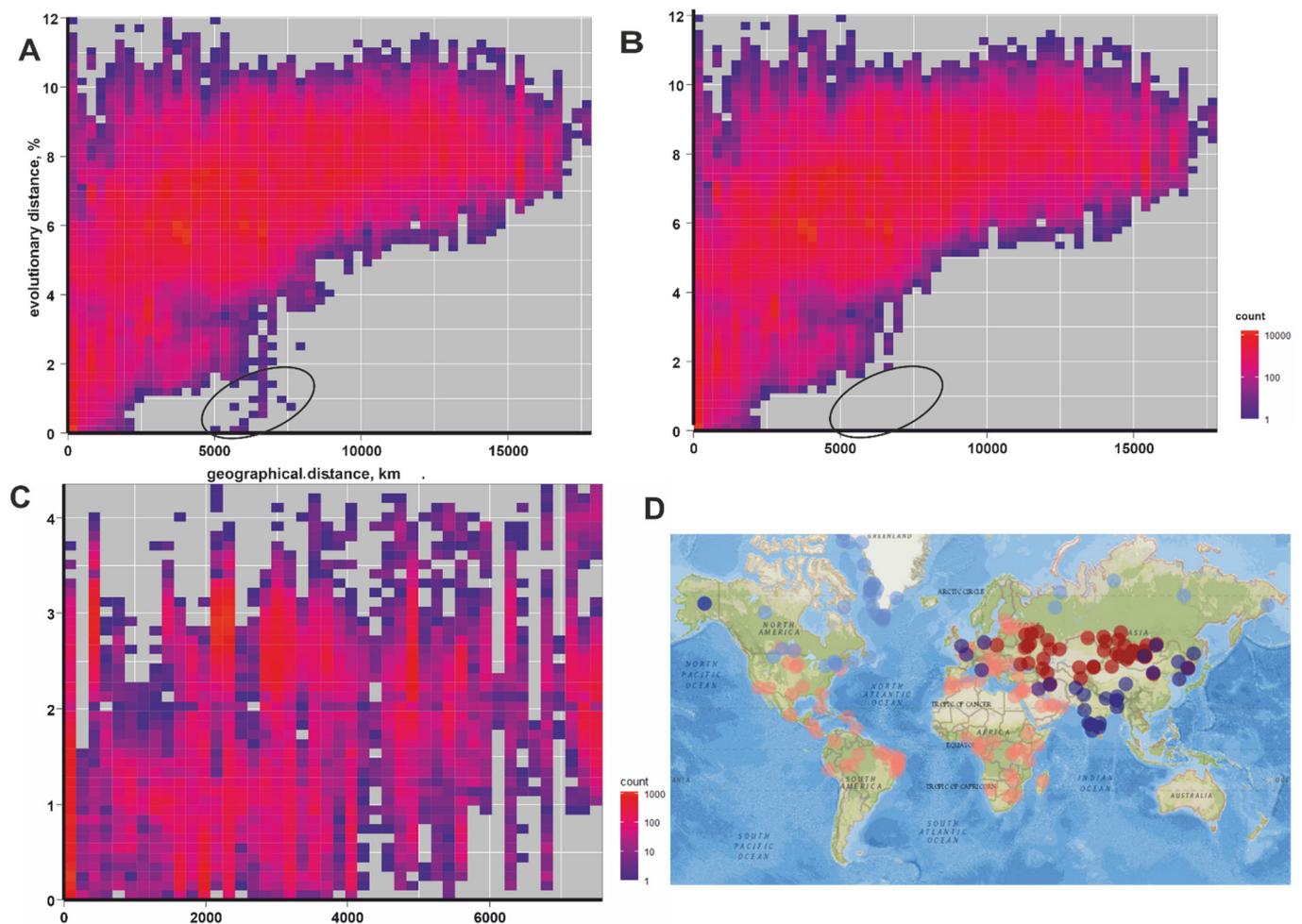


Figure 2. (A) Gene-geo plot for cosmopolitan RABV group ($n = 1824$). (B) Gene-geo plot for cosmopolitan RABV group without two sequences of viruses transferred between distant regions (#KP997032, #KC737850). (C) Gene-geo plot for the Arctic RABV group ($n = 453$). (D) The distribution of the cosmopolitan group (pale red), the steppe subgroup of cosmopolitan group (dark red), the Arctic-like group (dark blue), and the Arctic group (pale blue) in the global map. The interactive map with labeled markers is available at <https://rpubs.com/andreideviatkin/RABVdistribution> (assessed on 28 December 2021).

3. Results

The genome coverage plot indicated that RABV genome fragments were not evenly represented in GenBank (Figure S1). A 981 nt genome region encoding nucleoprotein N (genome positions 170–1150 in the reference sequence NC_001542) was chosen to provide maximum coverage at a maximum length. Four groups were selected for analysis: two relatively small related groups with different main hosts and suggestively different transfer patterns (Arctic, $n = 292$ and Arctic-like, $n = 201$), the globally prevalent and well-explored cosmopolitan group ($n = 1824$), and its steppe subgroup ($n = 205$), which are well-represented in GenBank and have well-characterized virus trafficking rates [12] and a known long-distance transmission event [16].

Calculation of a Bayesian phylogeographical framework grows with the number of sequences and is not practical for such large data sets. Furthermore, the accuracy of Bayesian phylogeographic models for large data sets remains unclear, as these models tend to perform best at intermediate sequence data set sizes [37]. Thus, the whole cosmopolitan RABV group was analyzed only for the correspondence of geographic and genetic distances (“gene-geo” plot). In addition, the Bayesian phylogeographic framework failed to produce

converged parameters for the Arctic RABV data set even after billions of state generations. Thus, this group was also analyzed only by the gene-geo plot.

A gene-geo plot of cosmopolitan RABV demonstrated a fair correspondence between genetic and geographic distances. This relationship was not linear, likely due to mutation saturation at about 10% nucleotide sequence distance and the impossibility of gradual virus spread between continents. The plot contained several sequence pairs that obviously deviated from the general distribution of pairwise values (Figure 2A, circled). These dots correspond to just two putative long-distance transfer events. One was the isolation of a Central Russian steppe lineage RABV from a bear in the Russian region of Primorye (Far East), likely following introduction by a hunting dog [8]. The second one was the isolation of a Brazilian RABV from a patient that immigrated from Brazil to the United States [9]. No other obvious long-distance trafficking events could be suggested from the whole cosmopolitan RABV data set (Figure 2B, the same as Figure 2A, but the two aforementioned “transferred” sequences were omitted).

The gene-geo plot for Arctic RABV demonstrated a limited concordance between genetic and geographic distances distribution patterns (Figure 2C). This lack of a “geographic signal” explains the failure of Bayesian phylogeographic inference in this RABV group. There was evidence of multiple long-distance transfers, i.e., many nearly identical sequence pairs with collection sites separated by thousands of kilometers.

The distance-based method (gene-geo plot) was tested in parallel with the Bayesian phylogeographic approach for the other data sets, i.e., the steppe subgroup (a group of cosmopolitan RABV, which also included the aforementioned long-distance virus transmission across Russia) (Figure 3) and the Arctic-like group, which also features several suggestive long-distance trafficking events (Figure 4).

The mean nucleotide substitution rate inferred using the N-gene data set for steppe subgroup ($n = 205$) was 4.13×10^{-4} (95% HPD 3.24×10^{-4} – 5.00×10^{-4}) nucleotide substitutions/site/year (s/s/y). This is concordant with the estimates of RABV N-gene substitution rates made for other subgroups of carnivore RABV: 3.89×10^{-4} s/s/y (2.88×10^{-4} – 4.96×10^{-4}) [38], 2.44×10^{-4} (2.10×10^{-4} – 2.80×10^{-4}) s/s/y [3], 3.82×10^{-4} (2.62×10^{-4} – 5.02×10^{-4}) s/s/y [39], 2.47×10^{-4} (1.82×10^{-4} – 3.12×10^{-4}) s/s/y [16], 5.23×10^{-4} (3.94×10^{-4} – 6.68×10^{-4}) s/s/y [40]. This confirmed validity of the Bayesian framework provided a rough estimate for the whole cosmopolitan RABV group substitution rate, suggesting that 1% of substitutions accumulate over 20–50 years.

The gene-geo plots for both the Arctic-like and steppe groups were compatible with recent long-distance virus transfers evident as dots with low genetic and high geographic distances (Figures 3A and 4A). The phylogeographic analysis also found tree branches with velocities over 500 km/year in both RABV groups (Figure 3C,E and Figure 4C,E). To further highlight the capacity of the two methods, known and suggestive “transferred” sequences were omitted from the data sets by excluding sequences identical in over 99.7% nucleotide positions and collected at distances of over 500 km. Additionally, Arctic-like RABV sequences obviously transferred between distant regions were also excluded (e.g., #MG011654 collected in France shared 98% identical nucleotides with #AB699219 collected in Bangladesh). The results of both approaches, i.e., gene-geo plots (Figures 3B and 4B) and the velocity rate distribution (Figure 3D,F and Figure 4D,F), were in concordance. For Arctic-like RABV, this resulted in the absence of “recent transfers” on gene-geo plots and branches with velocity rates greater than 500 km/year upon phylogeographic analysis (Figure 3D,F). However, for steppe RABV, three high-speed (over 500 km/year) branches remained in the phylogenetic analysis of the data sets with manually removed discordant gene-geo sequences (Figure 4D,F). Two branches corresponded to viruses isolated in two large regions, i.e., Russian Zabaykalsky Krai (432,000 sq km) and Chinese Inner Mongolia (1,183,000 sq km). These sequences (#KX533959 and #KY243236) shared 99.6% identical nucleotides (977 out of 981) and thus were not omitted by the artificial threshold. These high-speed branches could be artifacts due to data set limitations because precise isolation locations were unknown and inferred from the coordinates

of administrative capitals of these bordering regions. One more high-speed branch that persisted in the purified steppe RABV data set was a deep branch that led to a group of viruses collected in the steppe zone to the east from the Ural Mountains (Russian Southern Siberia, Russian Far East, Mongolia, Northern China) (Figure 4F, indicated by an asterisk). This finding suggests a human-assisted transfer of steppe RABV to the east from the Ural Mountains that occurred between 1958 and 1961 (95% highest posterior density (HPD): 1944–1978). This suggestive human-assisted transfer was not evident in the gene-geo plot (Figure 4B).

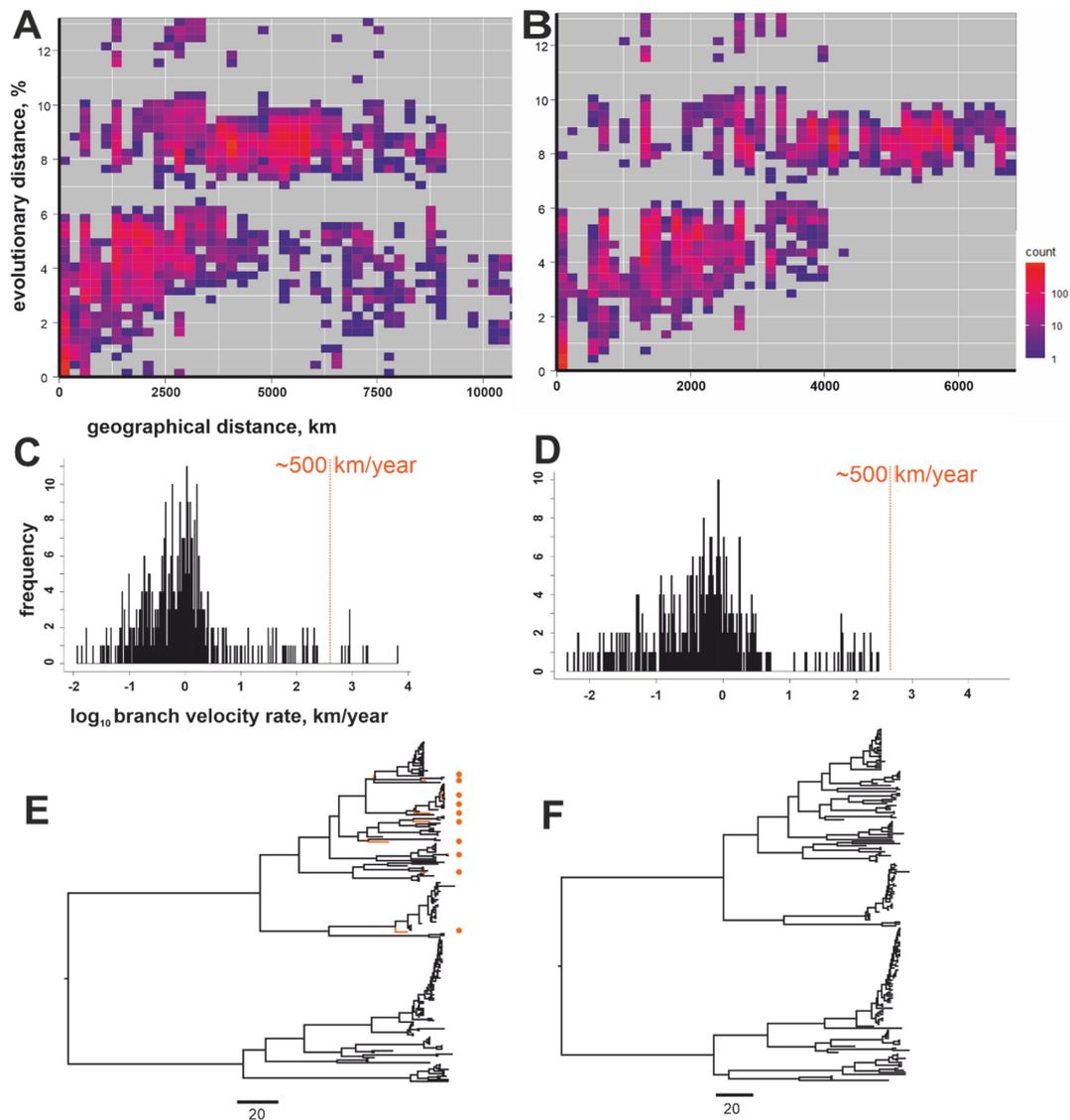


Figure 3. (A) Gene-geo plot for the Arctic-like RABV group ($n = 201$). (B) Gene-geo plot for Arctic-like RABV group without sequences of viruses apparently transferred between distant regions (#EU836832, #MG011654, #AY956319, #KF154996, #KU963488, #FJ228546, #KY982922, #EF611869, #KY860613). These nine sequences and 16 more sequences that were identical in more than 99.7% of nucleotides and whose collection sites were at a distance of at least 500 km were omitted from the alignment, yielding 176 sequences. (C,D) The distribution of the inferred velocity rates at the branches of MCC (maximum clade credibility) phylogenetic trees (E,F) in log-scale for all Arctic-like RABV group sequences ((C,E) $n = 201$) and Arctic-like RABV group sequences without a putative long-distance transfer history ((D,F) $n = 176$). The orange line indicates a velocity of 500 km/year. Tree branches with velocity rates over 500 km/year are shown in orange and marked with orange circles.

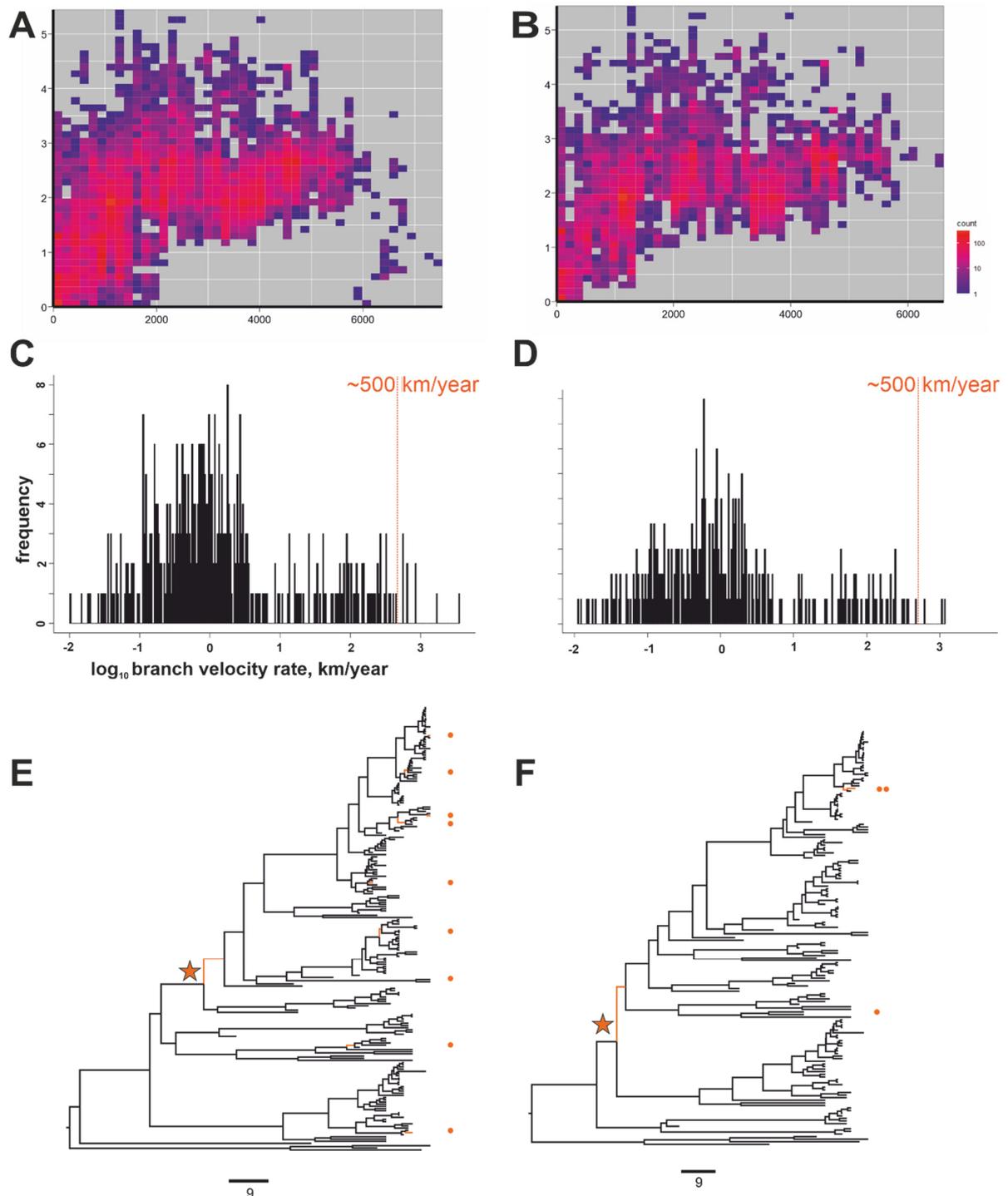


Figure 4. (A) Gene-geo plot for the steppe RABV group ($n = 205$). (B) Gene-geo plot for the steppe RABV group without sequences that were identical in more than 99.7% of nucleotides and whose collection sites were separated by at least 500 km were removed from the corresponding alignments ($n = 161$). (C,D) The distribution of velocity rates in the branches of MCC (maximum clade credibility) phylogenetic trees (E,F) in log-scale for all steppe RABV group sequences ((C) $n = 205$) and excluding sequences with a putative long-distance transfer history ((D) $n = 161$). The orange line indicates a velocity rate of 500 km/year. (E,F) MCC phylogenetic trees for all steppe RABV group sequences ((E) $n = 205$) and after omitting sequences with a putative long-distance transfer history ((F) $n = 161$). Branches with velocity rates over 500 km/year are colored orange and marked with orange circles. The star indicates a transfer from the South Urals region to the east, as discussed in the text.

The most vivid examples of sequences classified as “cosmopolitan” or “arctic-like” that had evidence of long-distance transfers upon gene-geo analysis are summarized in Table S1. As long-distance transfers of the Arctic group viruses are apparently systematic, data on individual sequences was not shown in Table S1. Roughly half of these long-distance events have been reported previously. In two cases, the branch velocity was not notably affected, likely because it could be spread among several branches by the statistical algorithm. It is noteworthy that a stringent distance cut-off was used; thus, less prominent long-distance transfers could have been overlooked. In addition, in this work, a high-throughput approach was used without an in-depth analysis of individual cases and without consideration of geographical and political barriers (state borders, water bodies, transportation networks, etc.); therefore, these findings should be treated with care. In two more cases, long-distance transfers were detected by the gene-geo plots but could not be verified by Bayesian phylogeography because the non-reduced cosmopolitan data set was too large.

4. Discussion

In general, RABV is considered a typical object for phylogeographic studies [41]. However, in the Arctic region, virus transmission speed was strikingly different from this classical pattern and was so high that it precluded a statistical phylogeography analysis. The case of Arctic RABV indicates that a phylogeographic signal should be checked to ensure correlation between geographic and genetic distances, and thus the applicability of a data set for a phylogeographic analysis. Gene-geo plot does so similarly to classical verification of a temporal signal prior to doing a statistical phylogenetics analysis involving a molecular clock. This is conventionally performed by assaying the correlation of root-to-tip distances in a sequence-based phylogenetic tree with isolation dates, for example, using TempEst (formerly Path-O-Gene) software [33]. One key difference in the verification of a temporal and geographic signal in a sequence data set is that an object cannot go back in time, but host migration trajectories are complex, and a virus can be carried back to its place of origin. Thus, the formalization of “geographic signal” in genomic data requires further studies.

While the Bayesian frameworks converged for other RABV groups, potential human-mediated transfers were a ubiquitous factor to be considered according to the gene-geo plots. We used a stringent cut-off of 500 km/year and a comparable genetic distance criterion to define likely anthropogenic transfers. This approach detected several known events and a comparable number of undescribed ones (Table S1). Overall, about 10–20% of all sequences bore signs of long-distance transfers. The actual fraction of affected viruses could be even higher because a 500 km/year cut-off was chosen arbitrarily, and there are even more sequences with branch rates above 300 km/year, which are also poorly compatible with a natural spread of steppe RABV. Of course, this number does not represent the actual proportion of human-assisted virus transfers because sequences from cases of medical and veterinary significance are overrepresented in GenBank. However, it is clear that in both Arctic-like and steppe RABV, unnatural long-distance transfers remain a significant factor in rabies epidemiology. In addition, some RABV strains demonstrated resistance to neutralizing mAbs and vaccine-induced antibodies [42].

It has been discussed previously that the history of RABV was significantly shaped by human activity. The worldwide dissemination of the carnivore-associated group of RABV was associated with colonial expansion and international trade accompanied by uncontrolled dog transportation that occurred in the 15th–19th centuries [3]. The role of long-distance transfers in recent RABV evolution is less clear. The predominantly gradual virus spread confined by natural barriers was described in both a global data set [4] and in country-wide studies, for example, in China [43]. The gene-geo plots suggest just three intercontinental transfers in both the cosmopolitan and Arctic-like RABV groups. Two events, one in the cosmopolitan group (Brazil-USA) and one in the Arctic-like group (India-Alaska), had been described previously [9,44]. The third case, to the best of our knowledge,

has not been discussed yet. Isolate with GenBank #KU963488, collected in Alaska, USA, and #KX434489, collected from buffalo's brain in Andhra Pradesh, India, shared 98.8% identical nucleotides (969 out of 981). Other related sequences were found in Asia; thus, the direction of this transfer was obviously from Asia to Alaska. Isolate #KU963488 was found in a wild animal (Arctic fox). This means that the anthropogenic transfer of RABV from India to Alaska led to spill-over to a novel host. At the same time, there were no other sequences of that Arctic-like subgroup in Alaska. Considering that #KU963488 was collected in 2007, it is likely that the Arctic-like group of Indian origin was not fixed in Arctic foxes after this particular introduction. It is noteworthy, though, that the Arctic group suggestively originated from the Arctic-like group upon virus introduction to circumpolar regions hundreds of years ago [19].

In the cosmopolitan group, all virus pairs spaced by over 8000 km also differed by at least 5% in the nucleotide sequence. Extrapolating RABV substitution rates identified here and elsewhere (see above), this corresponds to 100–250 years of circulation after the most recent common ancestor. Thus, virus control measures in the last century were generally efficient to prevent virus spread between the Old and New World, and intercontinental RABV transfer is not a major factor in virus evolution nowadays.

The majority of recent long-distance virus trafficking found in this study occurred within Eurasia (Table S1). Moreover, there could be more such events because our definition of long-distance transfers in terms of gene-geo distances and inferred branch velocities was arbitrary and could even underestimate their impact. One such event, suggested by statistical phylogeographic analysis and dating about 60 years ago, could have led to virus spread in East Siberia. This finding is concordant with long-distance transfers involving the steppe and Arctic-like groups that were noted in China but presumably occurred over 100 years ago, according to the presented phylogenetic trees [43]. One historical example that shows a possible mechanism of such human-assisted virus transfer was the translocation of thousands of raccoons from Florida, USA, to Virginia, USA in 1977–1981, which could have resulted in the emergence of raccoon rabies in the middle Atlantic states of the USA [23,45]. There was no evidence that recent (less than 30 years ago) virus transfers resulted in the establishment of the introduced lineages of cosmopolitan and Arctic-like RabV in Eurasia. In addition, few to none such recent long-distance (over 1000 km) transfers could be suggested from the phylogenetic analysis provided for Iran and China [42,46]. However, human-mediated RABV spread could have shaped the epizooty in Brazil at the end of the 20th century [47], is common nowadays in Tanzania [48], and was the major factor driving the current epizooty in Indonesia [49]. In North Africa, recent long-distance virus transfers have been noted but were hardly the main route of virus spread [27]. Thus, intracontinental long-distance transfers remained an important factor of RABV evolution in the mid-20th century in many locations. Currently, virus trafficking control is efficient in some countries but lacking in others. Moreover, the current data set is relevant only for countries with sufficient resources to isolate and sequence the virus. In contrast, less affluent countries may remain entirely out of the scope of such studies.

The two methods for the detection of aberrant long-distance virus transfers produced comparable results but had significant differences. Statistical phylogeography found many known and unreported long-distance transfer events but failed in the cases with an absent phylogeographic signal (Arctic RABV) or a too high number of sequences (cosmopolitan group). Gene-geo plots were applicable in both these cases to detect either a few long-distance transfers in the cosmopolitan group (Figure 2A) or to characterize the Arctic data set as a whole (Figure 2C). Data set reduction [50] can be used to circumvent the data set size limitation, but this should be performed with care because it may introduce unexpected biases in statistical phylogenetic calculations. However, the gene-geo plot failed to detect non-recent events, as exemplified by the putative cross-Urals transfer detected in the steppe group only by statistical phylogenetics (Figure 4E,F). Gene-geo plots do not provide statistical significance of the findings. However, the statistical support from Bayesian frameworks may be misleading if improper priors were used. In the absence of

an independent reference method, there is no way to assess the validity of the predictions. It can be thus suggested that using methods complementing each other may produce more precise results. Long-distance movements are rare or even non-existent in some RABV groups but common in others. Anthropogenic transfers have been investigated epidemiologically on many aforementioned occasions. Herein we evaluated different methods for analyzing them and providing an overview of their effect on the evolution of distinct RABV groups.

A major limitation of the large-scale phylogeographic analysis is the lack of precise virus sampling coordinates in GenBank records. Dedicated studies can use exact sampling location data; however, most published sequences have just country or, at best, province data. Many regions are hundreds of kilometers across, and virus isolation coordinates had to be approximated. Transfers between such regions could appear more rapid than they actually were. On a global scale, we observed a fair correlation between genetic and geographic distances. However, a more precise analysis, such as inferring virus trafficking rates, may not be reliable with the available data. In addition, for this reason, we abstained from analyzing less obvious potential long-distance transfers. Clearly, precise sampling coordinates would significantly increase the value of sequence data.

Another general limitation to ecological studies using RABV genomic data comes from the over-representation of cases of medical or veterinary importance in sequence databases. These events generally represent endemic viruses in the area but are, in most cases, dead-end events in virus evolution. Indeed, most long-distance virus transfers described here did not result in the establishment of novel virus lineages in distant regions. Only one suggestive transfer that occurred less than 70 years ago actually significantly impacted steppe RABV phylogeography (indicated by a star in Figure 4E,F).

5. Conclusions

The human-mediated spread of rabies is a significant threat (reviewed in [23]) and requires preventive countermeasures. Molecular surveillance for rabies allows for detecting long-distance virus transfers even in areas that are already endemic. A simple and easily automatable gene-geo analysis demonstrated suitable correspondence with sophisticated phylogeographic methods and may be a valuable tool for routine surveillance and a supplementary method for dedicated phylogeographic studies.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v14010066/s1>, Figure S1: RABV genome layout and RABV genome fragments longer than 950 nucleotides represented in GenBank. Purple lines indicate the genome region used for this study.; Table S1: Selected closely related viruses found at distant locations; File S1: in-house script for RABV genome coverage visualization, File S2: in-house python script for MCC tree branches velocities calculation and visualization.

Author Contributions: Conceptualization, A.A.D. and A.N.L.; Data curation, A.A.D. and M.A.D.; Formal analysis, A.A.D. and Y.A.V.; Funding acquisition, A.N.L.; Project administration, A.A.D. and A.N.L.; Software, A.A.D. and Y.A.V.; Visualization, A.A.D.; Writing—original draft, A.A.D. and A.N.L.; Writing—review and editing, A.A.D., Y.A.V. and A.N.L. All authors have read and agreed to the published version of the manuscript.

Funding: The development of software for “gene-geo” approach (A.A.D.), data curation (A.A.D., Y.A.V.), bayesian calculation (Y.A.V., A.A.D., A.N.L.) was supported by the Russian Science Foundation (grant # 19-15-00055). Data vizualization at RPubS server and selection of the most represented in GenBank RABV genome fragment (A.A.D.) was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-15-2020-899).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and source code presented in this study are openly available in the GitHub repository https://github.com/AndreiDeviatkin/gene_geo (assessed on 28 December 2021). This code was implemented in the R shiny webserver as an online tool available at andreideviatkin.shinyapps.io/gene-geo/ (assessed on 28 December 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hayman, D.T.S.; Fooks, A.R.; Marston, D.; Garcia-R, J.C. The Global Phylogeography of Lyssaviruses—Challenging the ‘Out of Africa’ Hypothesis. *PLoS Negl. Trop. Dis.* **2016**, *10*, e0005266. [[CrossRef](#)] [[PubMed](#)]
- World Health Organization. *WHO Expert Consultation on Rabies: Third Report*; WHO Technical Report Series 1012; WHO: Geneva, Switzerland, 2018.
- Troupin, C.; Dacheux, L.; Tanguy, M.; Sabeta, C.; Blanc, H.; Bouchier, C.; Vignuzzi, M.; Duchene, S.; Holmes, E.; Bourhy, H. Large-Scale Phylogenomic Analysis Reveals the Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts. *PLoS Pathog.* **2016**, *12*, e1006041. [[CrossRef](#)] [[PubMed](#)]
- Bourhy, H.; Reynes, J.-M.; Dunham, E.J.; Dacheux, L.; Larrous, F.; Huong, V.T.Q.; Xu, G.; Yan, J.; Miranda, M.E.G.; Holmes, E. The origin and phylogeography of dog rabies virus. *J. Gen. Virol.* **2008**, *89*, 2673–2681. [[CrossRef](#)] [[PubMed](#)]
- Fischer, S.; Freuling, C.M.; Müller, T.; Pfaff, F.; Bodenhofer, U.; Höper, D.; Fischer, M.; Marston, D.A.; Fooks, A.R.; Mettenleiter, T.C.; et al. Defining objective clusters for rabies virus sequences using affinity propagation clustering. *PLoS Negl. Trop. Dis.* **2018**, *12*, e0006182. [[CrossRef](#)] [[PubMed](#)]
- Velasco-Villa, A.; Mauldin, M.R.; Shi, M.; Escobar, L.E.; Gallardo-Romero, N.F.; Damon, I.; Olson, V.A.; Streicker, D.G.; Emerson, G. The history of rabies in the Western Hemisphere. *Antivir. Res.* **2017**, *146*, 221–232. [[CrossRef](#)]
- Mey, C.; Metlin, A.; Duong, V.; Ong, S.; In, S.; Horwood, P.F.; Reynes, J.-M.; Bourhy, H.; Tarantola, A.; Buchy, P. Evidence of two distinct phylogenetic lineages of dog rabies virus circulating in Cambodia. *Infect. Genet. Evol.* **2016**, *38*, 55–61. [[CrossRef](#)]
- Shchelkanov, M.Y.; Deviatkin, A.; Ananiev, V.Y.; Frolov, E.V.; Dombrovskaya, I.E.; Dedkov, V.G.; Ardashev, A.V.; Kolomeets, S.A.; Korotkova, I.P.; Lyubchenko, E.N.; et al. Isolation and complete genome sequencing of rabies virus strain isolated from a brown bear (*Ursus arctos*) that attacked a human in Primorsky krai (November 2014). *Probl. Virol.* **2016**, *61*, 180–186. [[CrossRef](#)]
- Boland, T.A.; McGuone, D.; Jindal, J.; Rocha, M.; Cumming, M.; Rupprecht, C.E.; Barbosa, T.F.S.; Oliveira, R.D.N.; Chu, C.J.; Cole, A.J.; et al. Phylogenetic and epidemiologic evidence of multiyear incubation in human rabies. *Ann. Neurol.* **2013**, *75*, 155–160. [[CrossRef](#)]
- Turcitu, M.; Barboi, G.; Vuta, V.; Mihai, I.; Boncea, D.; Dumitrescu, F.; Codreanu, M.; Johnson, N.; Fooks, A.; Müller, T.; et al. Molecular epidemiology of rabies virus in Romania provides evidence for a high degree of heterogeneity and virus diversity. *Virus Res.* **2010**, *150*, 28–33. [[CrossRef](#)]
- Anderson, R.; Jackson, H.C.; May, R.M.; Smith, A.M. Population dynamics of fox rabies in Europe. *Nat. Cell Biol.* **1981**, *289*, 765–771. [[CrossRef](#)]
- Kuzmin, I.V.; Botvinkin, A.D.; McElhinney, L.; Smith, J.S.; Orciari, L.A.; Hughes, G.J.; Fooks, A.R.; Rupprecht, C.E. Molecular epidemiology of terrestrial rabies in the former soviet Union. *J. Wildl. Dis.* **2004**, *40*, 617–631. [[CrossRef](#)]
- Fuglei, E.; Tarroux, A. Arctic fox dispersal from Svalbard to Canada: One female’s long run across sea ice. *Polar Res.* **2019**, *38*, 3512. [[CrossRef](#)]
- Nadin-Davis, S.A.; Falardeau, E.; Flynn, A.; Whitney, H.; Marshall, H.D. Relationships between fox populations and rabies virus spread in northern Canada. *PLoS ONE* **2021**, *16*, e0246508. [[CrossRef](#)]
- Joly, K.; Gurarie, E.; Sorum, M.S.; Kaczensky, P.; Cameron, M.D.; Jakes, A.F.; Borg, B.L.; Nandintsetseg, D.; Hopcraft, J.G.C.; Buuveibaatar, B.; et al. Longest terrestrial migrations and movements around the world. *Sci. Rep.* **2019**, *9*, 15333. [[CrossRef](#)]
- Deviatkin, A.A.; Lukashov, A.N.; Poleshchuk, E.; Dedkov, V.G.; Tkachev, S.; Sidorov, G.N.; Karganova, G.G.; Galkina, I.V.; Shchelkanov, M.Y.; Shipulin, G. The phylodynamics of the rabies virus in the Russian Federation. *PLoS ONE* **2017**, *12*, e0171855. [[CrossRef](#)]
- Mansfield, K.; Racloz, V.; McElhinney, L.; Marston, D.; Johnson, N.; Rønsholt, L.; Christensen, L.; Neuvonen, E.; Botvinkin, A.; Rupprecht, C.; et al. Molecular epidemiological study of Arctic rabies virus isolates from Greenland and comparison with isolates from throughout the Arctic and Baltic regions. *Virus Res.* **2006**, *116*, 1–10. [[CrossRef](#)] [[PubMed](#)]
- Hanke, D.; Freuling, C.M.; Fischer, S.; Hueffer, K.; Hundertmark, K.; Nadin-Davis, S.; Marston, D.; Fooks, A.R.; Bøtner, A.; Mettenleiter, T.C.; et al. Spatio-temporal Analysis of the Genetic Diversity of Arctic Rabies Viruses and Their Reservoir Hosts in Greenland. *PLoS Negl. Trop. Dis.* **2016**, *10*, e0004779. [[CrossRef](#)] [[PubMed](#)]
- Kuzmin, I.V.; Hughes, G.J.; Botvinkin, A.D.; Gribencha, S.G.; Rupprecht, C.E. Arctic and Arctic-like rabies viruses: Distribution, phylogeny and evolutionary history. *Epidemiol. Infect.* **2008**, *136*, 509–519. [[CrossRef](#)] [[PubMed](#)]
- Nadin-Davis, S.A.; Turner, G.; Paul, J.P.V.; Madhusudana, S.N.; Wandeler, A.I. Emergence of Arctic-like Rabies Lineage in India. *Emerg. Infect. Dis.* **2007**, *13*, 111–116. [[CrossRef](#)]
- Freuling, C.M.; Hampson, K.; Selhorst, T.; Schröder, R.; Meslin, F.X.; Mettenleiter, T.C.; Müller, T. The elimination of fox rabies from Europe: Determinants of success and lessons for the future. *Philos. Trans. R. Soc. B Biol. Sci.* **2013**, *368*, 20120142. [[CrossRef](#)]

22. Johnson, N.; Un, H.; Fooks, A.R.; Freuling, C.; Müller, T.; Aylan, O.; Vos, A. Rabies epidemiology and control in Turkey: Past and present. *Epidemiol. Infect.* **2009**, *138*, 305–312. [[CrossRef](#)] [[PubMed](#)]
23. Fèvre, E.M.; Bronsvoort, M.; Hamilton, K.A.; Cleaveland, S. Animal movements and the spread of infectious diseases. *Trends Microbiol.* **2006**, *14*, 125–131. [[CrossRef](#)]
24. Hutchison, D.W.; Templeton, A.R. Correlation of Pairwise Genetic and Geographic Distance Measures: Inferring the Relative Influences of Gene Flow and Drift on the Distribution of Genetic Variability. *Evolution* **1999**, *53*, 1898. [[CrossRef](#)]
25. Deviatkin, A.A.; Kholodilov, I.S.; Belova, O.A.; Bugmyrin, S.V.; Bespyatova, L.A.; Ivannikova, A.Y.; Vakulenko, Y.A.; Lukashev, A.N.; Karganova, G.G. Baltic Group Tick-Borne Encephalitis Virus Phylogeography: Systemic Inconsistency Pattern between Genetic and Geographic Distances. *Microorganisms* **2020**, *8*, 1589. [[CrossRef](#)]
26. Faria, N.R.; Suchard, M.A.; Rambaut, A.; Lemey, P. Toward a quantitative understanding of viral phylogeography. *Curr. Opin. Virol.* **2011**, *1*, 423–429. [[CrossRef](#)]
27. Talbi, C.; Lemey, P.; Suchard, M.A.; Abdelatif, E.; Elharrak, M.; Jalal, N.; Faouzi, A.; Echevarría, J.E.; Morón, S.V.; Rambaut, A.; et al. Phylodynamics and Human-Mediated Dispersal of a Zoonotic Virus. *PLoS Pathog.* **2010**, *6*, e1001166. [[CrossRef](#)]
28. Dellicour, S.; Rose, R.; Faria, N.R.; Vieira, L.F.P.; Bourhy, H.; Gilbert, M.; Lemey, P.; Pybus, O. Using Viral Gene Sequences to Compare and Explain the Heterogeneous Spatial Dynamics of Virus Epidemics. *Mol. Biol. Evol.* **2017**, *34*, 2563–2571. [[CrossRef](#)]
29. Yamada, K.D.; Tomii, K.; Katoh, K. Application of the MAFFT sequence alignment program to large data—Reexamination of the usefulness of chained guide trees. *Bioinformatics* **2016**, *32*, 3246–3251. [[CrossRef](#)]
30. Nguyen, L.-T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)]
31. Suchard, M.A.; Lemey, P.; Baele, G.; Ayres, D.L.; Drummond, A.J.; Rambaut, A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **2018**, *4*, vey016. [[CrossRef](#)] [[PubMed](#)]
32. Ayres, D.L.; Darling, A.; Zwickl, D.J.; Beerli, P.; Holder, M.; Lewis, P.O.; Huelsenbeck, J.P.; Ronquist, F.; Swofford, D.L.; Cummings, M.P.; et al. BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics. *Syst. Biol.* **2011**, *61*, 170–173. [[CrossRef](#)]
33. Rambaut, A.; Lam, T.T.; Max Carvalho, L.; Pybus, O.G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2016**, *2*, vew007. [[CrossRef](#)]
34. Drummond, A.J.; Ho, S.Y.W.; Phillips, M.J.; Rambaut, A. Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol.* **2006**, *4*, e88. [[CrossRef](#)] [[PubMed](#)]
35. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. [[CrossRef](#)] [[PubMed](#)]
36. FigTree v1.4.4. Available online: <https://github.com/rambaut/figtree/releases> (accessed on 28 December 2021).
37. Vaiente, M.A.; Scotch, M. Going back to the roots: Evaluating Bayesian phylogeographic models with discrete trait uncertainty. *Infect. Genet. Evol.* **2020**, *85*, 104501. [[CrossRef](#)] [[PubMed](#)]
38. McElhinney, L.M.; Marston, D.; Freuling, C.M.; Cragg, W.; Stankov, S.; Lalošević, D.; Müller, T.; Fooks, A.R. Molecular diversity and evolutionary history of rabies virus strains circulating in the Balkans. *J. Gen. Virol.* **2011**, *92*, 2171–2180. [[CrossRef](#)]
39. Talbi, C.; Holmes, E.; De Benedictis, P.; Faye, O.; Nakouné, E.; Gamatié, D.; Diarra, A.; Elmamy, B.O.; Sow, A.; Adjogoua, E.V.; et al. Evolutionary history and dynamics of dog rabies virus in western and central Africa. *J. Gen. Virol.* **2009**, *90*, 783–791. [[CrossRef](#)]
40. Guo, Z.; Tao, X.; Yin, C.; Han, N.; Yu, J.; Li, H.; Liu, H.; Fang, W.; Adams, J.; Wang, J.; et al. National Borders Effectively Halt the Spread of Rabies: The Current Rabies Epidemic in China Is Dislocated from Cases in Neighboring Countries. *PLoS Negl. Trop. Dis.* **2013**, *7*, e2039. [[CrossRef](#)]
41. Biek, R.; Henderson, J.C.; Waller, L.A.; Rupprecht, C.E.; Real, L.A. A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7993–7998. [[CrossRef](#)]
42. Wang, W.; Ma, J.; Nie, J.; Li, J.; Cao, S.; Wang, L.; Yu, C.; Huang, W.; Li, Y.; Yu, Y.; et al. Antigenic variations of recent street rabies virus. *Emerg. Microbes Infect.* **2019**, *8*, 1584–1592. [[CrossRef](#)]
43. Wang, L.; Wu, X.; Bao, J.; Song, C.; Du, J. Phylodynamic and transmission pattern of rabies virus in China and its neighboring countries. *Arch. Virol.* **2019**, *164*, 2119–2129. [[CrossRef](#)] [[PubMed](#)]
44. Velasco-Villa, A.; Reeder, S.A.; Orciari, L.A.; Yager, P.A.; Franka, R.; Blanton, J.D.; Zuckero, L.; Hunt, P.; Oertli, E.H.; Robinson, L.E.; et al. Enzootic Rabies Elimination from Dogs and Reemergence in Wild Terrestrial Carnivores, United States. *Emerg. Infect. Dis.* **2008**, *14*, 1849–1854. [[CrossRef](#)] [[PubMed](#)]
45. Jenkins, S.R.; Winkler, W.G. Descriptive epidemiology from an epizootic of raccoon rabies in the middle Atlantic states, 1982–1983. *Am. J. Epidemiol.* **1987**, *126*, 429–437. [[CrossRef](#)]
46. Horton, D.; McElhinney, L.M.; Freuling, C.M.; Marston, D.; Banyard, A.C.; Goharriz, H.; Wise, E.; Breed, A.; Saturday, G.; Kolodziejek, J.; et al. Complex Epidemiology of a Zoonotic Disease in a Culturally Diverse Region: Phylogeography of Rabies Virus in the Middle East. *PLoS Negl. Trop. Dis.* **2015**, *9*, e0003569. [[CrossRef](#)]
47. Carnieli, P.; Oliveira, R.D.N.; Macedo, C.I.; Castilho, J.G. Phylogeography of rabies virus isolated from dogs in Brazil between 1985 and 2006. *Arch. Virol.* **2011**, *156*, 1007–1012. [[CrossRef](#)] [[PubMed](#)]
48. Brunner, K.; Marston, D.; Horton, D.L.; Cleaveland, S.; Fooks, A.R.; Kazwala, R.; Ngeleja, C.; Lembo, T.; Sambo, M.; Mtema, Z.J.; et al. Elucidating the phylodynamics of endemic rabies virus in eastern Africa using whole-genome sequencing. *Virus Evol.* **2015**, *1*, vev011. [[CrossRef](#)]

-
49. Dibia, I.N.; Sumiarto, B.; Susetya, H.; Putra, A.A.G.; Scott-Orr, H.; Mahardika, G.N. Phylogeography of the current rabies viruses in Indonesia. *J. Veter-Sci.* **2015**, *16*, 459–466. [[CrossRef](#)] [[PubMed](#)]
 50. Vakulenko, Y.; Deviatkin, A.; Lukashev, A. The Effect of Sample Bias and Experimental Artefacts on the Statistical Phylogenetic Analysis of Picornaviruses. *Viruses* **2019**, *11*, 1032. [[CrossRef](#)]