PNAS RESEARCH ARTICLE | ECOLOGY

Phylogeographic reconstruction of the emergence and spread of Powassan virus in the northeastern United States

Chantal B. F. Vogels[a,1,2], Doug E. Brackney[b,1], Alan P. Dupuis II[c,d,1], Rebecca M. Robich[e,1], Joseph R. Fauver[a,f], Anderson F. Brito[a,g], Scott C. Williams[h], John F. Anderson[b], Charles B. Lubelczyk[e], Rachel E. Lange[c,d], Melissa A. Prusinski[i], Laura D. Kramer[c,d], Jody L. Gangloff-Kaufmann[j], Laura B. Goodman[k], Guy Baele[l], Robert P. Smith[e,1], Philip M. Armstrong[b,1], Alexander T. Ciota[c,d,1], Simon Dellicour[l,m,1,2], and Nathan D. Grubaugh[a,n,1,2]

Edited by Rasmus Nielsen, University of California Berkeley, Berkeley, CA; received October 21, 2022; accepted February 23, 2023

Powassan virus is an emerging tick-borne virus of concern for public health, but very little is known about its transmission patterns and ecology. Here, we expanded the genomic dataset by sequencing 279 Powassan viruses isolated from *Ixodes scapularis* ticks from the northeastern United States. Our phylogeographic reconstructions revealed that Powassan virus lineage II was likely introduced or emerged from a relict population in the Northeast between 1940 and 1975. Sequences strongly clustered by sampling location, suggesting a highly focal geographical distribution. Our analyses further indicated that Powassan virus lineage II emerged in the northeastern United States mostly following a south-to-north pattern, with a weighted lineage dispersal velocity of ~3 km/y. Since the emergence in the Northeast, we found an overall increase in the effective population size of Powassan virus lineage II, but with growth stagnating during recent years. The cascading effect of population expansion of white-tailed deer and *I. scapularis* populations likely facilitated the emergence of Powassan virus in the northeastern United States.

tick-borne flavivirus | deer tick virus | ticks | *Ixodes scapularis* | genomics

Significance

Our work provides important fundamental insights into the local transmission dynamics of an emerging tick-borne pathogen of public health concern. Without the availability of vaccines or specific treatments, prevention of Powassan virus infection is dependent on education and control. We identified that Powassan virus lineage II is maintained in highly localized transmission foci that have been maintained for several years, without introductions of new virus clades. This provides both opportunities for better education about high-risk areas and effective targeted control in Powassan virus foci with a long-lasting impact.

Reports of tick-borne diseases in the United States have been steadily rising, with more than 50,000 cases in 2019 (1). In the same year, a record number of 43 human cases of infection with an emerging tick-borne pathogen, Powassan virus (family: *Flaviviridae,* genus: *Flavivirus*), were reported (2–5). Powassan virus infection can cause severe neuroinvasive disease with long-lasting sequelae and high fatality rates in humans. Since its initial identification in 1958 (6), incidence rates of Powassan neuroinvasive disease in humans have dramatically risen in the United States, particularly during recent years in the Northeast (7). As Powassan virus infection is difficult to clinically diagnose (8) and most infections are asymptomatic (3), the reported cases are likely a vast underestimation of the true burden. The increasing number of human infections coupled with the lack of an effective vaccine or medicines, highlights the need to better understand local virus transmission patterns to guide targeted prevention and control measures.

Emergence of tick-borne diseases such as Lyme borreliosis, anaplasmosis, and babesiosis is associated with the spread of *Ixodes scapularis* ticks following the expansion of suitable habitats (9) and the reintroduction of their primary adult stage hosts, white-tailed deer (*Odocoileus virginianus*), into the northeastern United States (10, 11). Powassan virus consists of two genetically distinct lineages, of which lineage II, also thought to be primarily maintained by *I. scapularis* ticks and small mammals (12, 13), followed a similar path (3–5). However, very little is known about its ecology and transmission patterns. Genetic approaches, including phylogenetic and phylogeographic inference, are powerful tools to understand patterns of pathogen transmission and spread. However, until recently, there were only 23 near-complete Powassan virus genomes available from the United States. Thus, to inform future control efforts and mitigate public health risks, there is a critical need for new and innovative phylogeographic approaches to uncover how Powassan virus is being maintained in the Northeast and which factors facilitate or prevent its spread to other areas.

To investigate, we partnered with public health laboratories throughout the Northeast to sequence Powassan virus isolated from *I. scapularis* ticks in Connecticut, New York, and Maine. We sequenced samples collected in 2008 to 2019 from both historic endemic sites and from novel regions on the leading edge of expansion in the Northeast. With this expanded genomic dataset, we performed discrete and continuous phylogeographic analyses to answer important questions on the emergence and spread of Powassan virus lineage II, such as: 1) When did Powassan virus lineage II emerge in the Northeast? 2) How is Powassan virus lineage II locally maintained? 3) What are the patterns and velocity of

Author contributions: C.B.F.V., S.D., and N.D.G. designed research; C.B.F.V., D.E.B., A.P.D., R.M.R., J.R.F., S.C.W., J.F.A., C.B.L., R.E.L., M.A.P., L.D.K., J.L.G.-K., L.B.G., R.P.S., P.M.A., and A.T.C. performed research; C.B.F.V., A.F.B., G.B., S.D., and N.D.G. analyzed data; D.E.B., A.P.D., R.M.R., J.R.F., A.F.B., S.C.W., J.F.A., C.B.L., R.E.L., M.A.P., L.D.K., J.L.G.-K., L.B.G., R.P.S., P.M.A., and A.T.C. reviewed and approved the manuscript; and C.B.F.V., G.B., S.D., and N.D.G. wrote the paper.

Competing interest statement: N.D.G. is a consultant for Tempus Labs and the National Basketball Association for work related to COVID-19.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]C.B.F.V., D.E.B., A.P.D., R.M.R., R.P.S., P.M.A., A.T.C., S.D., and N.D.G. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: chantal.vogels@yale.edu, simon.dellicour@ulb.be, or nathan.grubaugh@yale.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2218012120/-/DCSupplemental.

Published April 11, 2023.

Downloaded from https://www.pnas.org by "KATHOLIEKE UNIVERSITEIT LEUVEN KU LEUVEN, 2BERGEN, BIOMEDISCHE BIBLIO" on April 27, 2023 from IP address 134.58.253.56.

spread? 4) Can increased transmission explain the recent increase in reported human cases? 5) What is the impact of environmental factors on the dispersal dynamics? Overall, our study provides important insights into the emergence and spread of Powassan virus in the northeastern United States. Our findings help to better identify potential high-risk areas for exposure, which will in turn help to direct future control efforts.

Results

Powassan Virus Phylogeny. Powassan virus consists of two genetically and ecologically distinct lineages (lineages I and II; Fig. 1). Prior to this study, there were 23 near-complete Powassan virus genome sequences publicly available from the United States. In this study, we sequenced an additional 279 Powassan viruses (2 belonging to lineage I, and 277 belonging to lineage II) from positive tick pools identified by public health laboratories in Connecticut, New York, and Maine from 2008 to 2019 (*SI Appendix*, Tables S1–S3). We created Nextstrain pages to visualize the Powassan genomic data with builds for all available genomes (14), and a more specific build for genomes available from the northeastern United States (15). Publicly available lineage I sequences were available from Russia, Canada (initial case in North America), and the United States, but Powassan virus lineage I had not been reported from the United States since the late 1970s (16, 17). As part of this study, we sequenced two lineage I genomes detected in *I. scapularis* ticks in 2019 from New York (18). Later, we sequenced two additional lineage I genomes detected in *I. cookei* (2020) and *Dermacentor variabilis* (2021) also from New York (19). Identification of Powassan virus lineage I in various tick species highlights how increased virus genomic surveillance can help to expand our knowledge of virus ecology.

Almost all of the more recent Powassan viruses detected from the United States, including what we sequenced for this study, belong to lineage II. Powassan virus lineage II, also referred to as "deer tick virus," consists of two geographically separated clades comprising viruses from the Midwest and the Northeast (Fig. 1). Powassan virus lineage II is the most prevalent in the Northeast, and we first carefully assessed the presence of temporal signal in these data. While the determination coefficient of the root-to-tip regression performed with TempEst (20) is relatively small (R^2 = 0.23; Fig. 1), we find very strong evidence in favor of temporal signal in the data set using a recently developed Bayesian method (21) [log Bayes factor = 41.8 (22)], enabling the use of molecular clocks to estimate time-calibrated phylogenies. As part of this analysis, we find that an uncorrelated relaxed clock with an underlying lognormal distribution provides a better model fit to the data compared to a strict clock model (Table 1). We estimate that the evolutionary rate of this clade is 8.25×10^{-5} substitutions/site/year [95% highest posterior density (HPD) interval: (8.23 to 10.46×10^{-5}); 95% HPD interval of the associated coefficient of variation: (0.48 to 0.75)]. Our estimate is higher than previous estimated evolutionary rates for all Powassan viruses [3.3×10^{-5} (23)] and NS5 coding sequences [3.9×10^{-5} to 5.4×10^{-5} (5, 24)], but lower than previous estimates based on envelope coding sequences [2.2×10^{-4} (5)]. This is likely a reflection of the recent emergence of lineage II in the region. Our work increased the number of publicly available Powassan virus lineage II sequences by more than ten-fold, enabling us to better understand the patterns of emergence and spread in the northeastern United States.

Emergence in the Northeast. The emergence of Powassan virus lineage II into the northeastern United States likely followed the reemergence of *I. scapularis* ticks, its primary vector, into the region. *Ixodes scapularis* originally colonized the Northeast thousands of years ago (25, 26); however, deforestation and restriction of white-tailed deer populations (primary reproductive host for adult *I. scapularis* ticks) during the 1800s greatly reduced *I. scapularis* populations in the Northeast (10, 27). Reforestation and increasing white-tailed deer populations during the mid-1900s then led to a reemergence of *I. scapularis* (10, 28, 29).

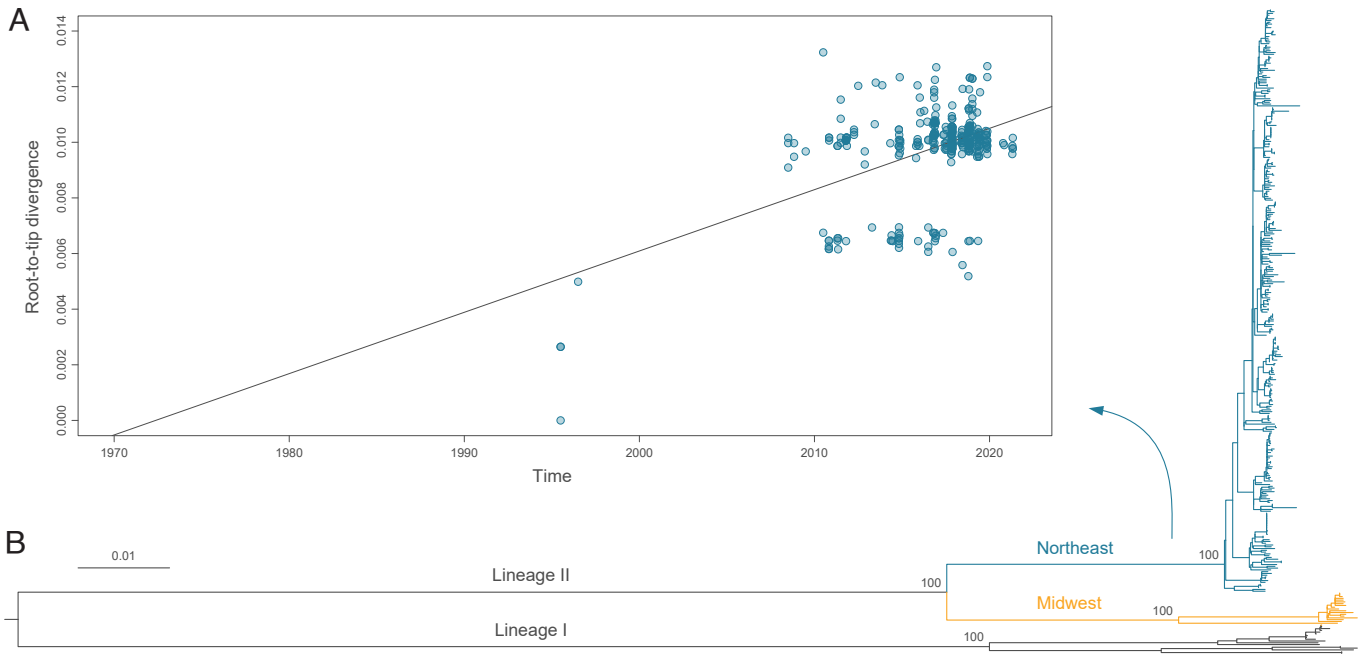

Fig. 1. Phylogenetic analysis of Powassan virus lineages. (*A*) Root-to-tip regression performed to assess the temporal signal within the Northeast clade (determination coefficient R^2 from the linear regression = 0.23). (*B*) Maximum likelihood tree was obtained from the phylogenetic analysis of publicly available Powassan virus genomes from the United States, Canada, and Russia. Powassan virus lineage II consists of two geographically separated clades in the Northeast and Midwest. Bootstraps support values (based on 1,000 replicates) are provided for the main internal nodes of the tree.

Downloaded from https://www.pnas.org by "KATHOLIEKE UNIVERSITEIT LEUVEN KU LEUVEN, 2BERGEN, BIOMEDISCHE BIBLIO" on April 27, 2023 from IP address 134.58.253.56.

Table 1. Bayesian evaluation of temporal signal analysis

Clock model	Model fit (gamma)	Model fit (lognormal)
Strict clock + dates	−36,575.48	−36,576.00
Strict clock − dates	−36,601.19	−36,606.66
Relaxed clock + dates	−36,512.57	−36,513.62
Relaxed clock − dates	−36,554.33	−36,556.68

Our expanded Powassan virus genomic data enabled us to reconstruct the dispersal history of Powassan virus lineage II in the northeastern United States. (Fig. 2*A*). Our discrete phylogeographic analysis estimates that the time to the most recent common ancestor (tMRCA) for the Northeast Powassan virus lineage II clade is between 1940.3 and 1974.7 (95% HPD interval; mean 1957.9). This means that lineage II emerged in the Northeast likely before the mid-1970s, corresponding to the reemergence of *I. scapularis* ticks.

Distinct Transmission Foci. Like tick-borne encephalitis virus in Europe (30, 31), Powassan virus is hypothesized to be primarily maintained within strict foci (32, 33), meaning that the virus does not routinely migrate between locations. To test this hypothesis, we examined our discrete phylogeographic reconstruction of Powassan virus lineage II in the Northeast (Fig. 2). We found that sequences strongly cluster by location (Fig. 2*A*) with relatively few transition events over the past 20 y (Fig. 2*B*).

We further explored the spatial structure using a continuous phylogeographic approach (Fig. 3). These findings again highlight the highly focal distribution, with rare long-distance dispersal events leading to the establishment of new foci (Fig. 3 *A*–*D*). These transmission foci are particularly clear in Connecticut, where, besides a single dispersal event between locations 1 (Westport) and 2 (Redding), we do not observe mixing between the five distinct locations (Fig. 3*D*, *Inset*). For example, we estimate that the lineage II viruses sequenced in location 2 (Redding) have been separated from location 3 (Bridgeport) for ~34 y (95% HPD interval: 26 to 42), despite being less than 20 km apart.

We found the same pattern of isolated transmission foci with limited mixing for New York and Maine, albeit with more short-distance dispersal events, particularly in New York (Fig. 2*B*, and 3*D*). These differences may be explained due to differences in sampling (e.g., Powassan virus sequenced from ticks in New York included adults collected from deer), environmental barriers to spread (e.g., separation of Connecticut locations by rivers), or other ecological factors.

Overall, our data suggest that after the initial emergence of Powassan virus lineage II in the Northeast, migration between nearby and long-distance locations was relatively rare. This supports the hypothesis that Powassan virus is primarily maintained in highly localized transmission foci.

Dispersal History. While we have shown that Powassan virus lineage II in the Northeast is primarily restricted to strict foci, we wanted to better understand the patterns and velocity of spread. Our spatially explicit continuous phylogeographic analysis indicates that Powassan virus lineage II emerged in the northeastern United States mostly following a south-to-north pattern (Fig. 3 *A*–*D*). We estimate that the virus first became established in southern New York and Connecticut by the late 1950s (1940.3 to 1974.7; Fig. 3*A*). This was followed by a few long-distance dispersal events to more northern regions, perhaps by infected ticks feeding on birds which can migrate over longer distances. We estimate that the virus finally became established in Maine by 1991 [95% HPD = (1971 to 2012)] through multiple introductions. Our estimates of the relatively recent dispersion of Powassan virus lineage II in the northern part of the Northeast

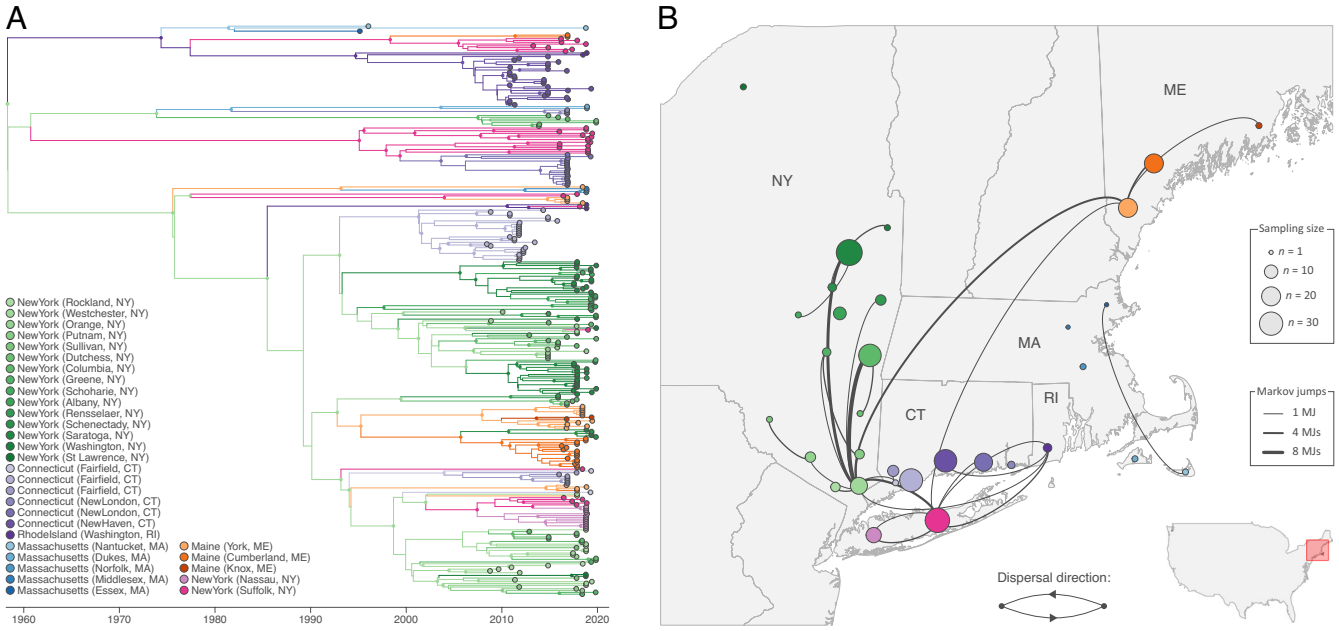

Fig. 2. Discrete phylogeographic analysis of the dispersal history of Powassan virus in the northeastern United States. (*A*) MCC tree with branches colored according to the locations inferred at the ancestral nodes. Tip nodes are colored according to their sampling location, and we only displayed internal nodes, using smaller dots and colored according to their inferred location, if they are associated with a posterior probability >0.95. (*B*) Sampling map and well-supported Markov jumps inferred by discrete phylogeographic inference. Sampling locations are displayed by dots, with the size being proportional to the number of Powassan virus genomic sequences sampled and included in our analyses. We only report Markov jumps associated with an adjusted Bayes factor support higher than three, which corresponds to positive support according to the scale of interpretation as previously defined (22).

Downloaded from https://www.pnas.org by "KATHOLIEKE UNIVERSITEIT LEUVEN KU LEUVEN, 2BERGEN, BIOMEDISCHE BIBLIO" on April 27, 2023 from IP address 134.58.253.56.

Downloaded from https://www.pnas.org by "KATHOLIEKE UNIVERSITEIT LEUVEN KU LEUVEN, 2BERGEN, BIOMEDISCHE BIBLIO" on April 27, 2023 from IP address 134.58.253.56.

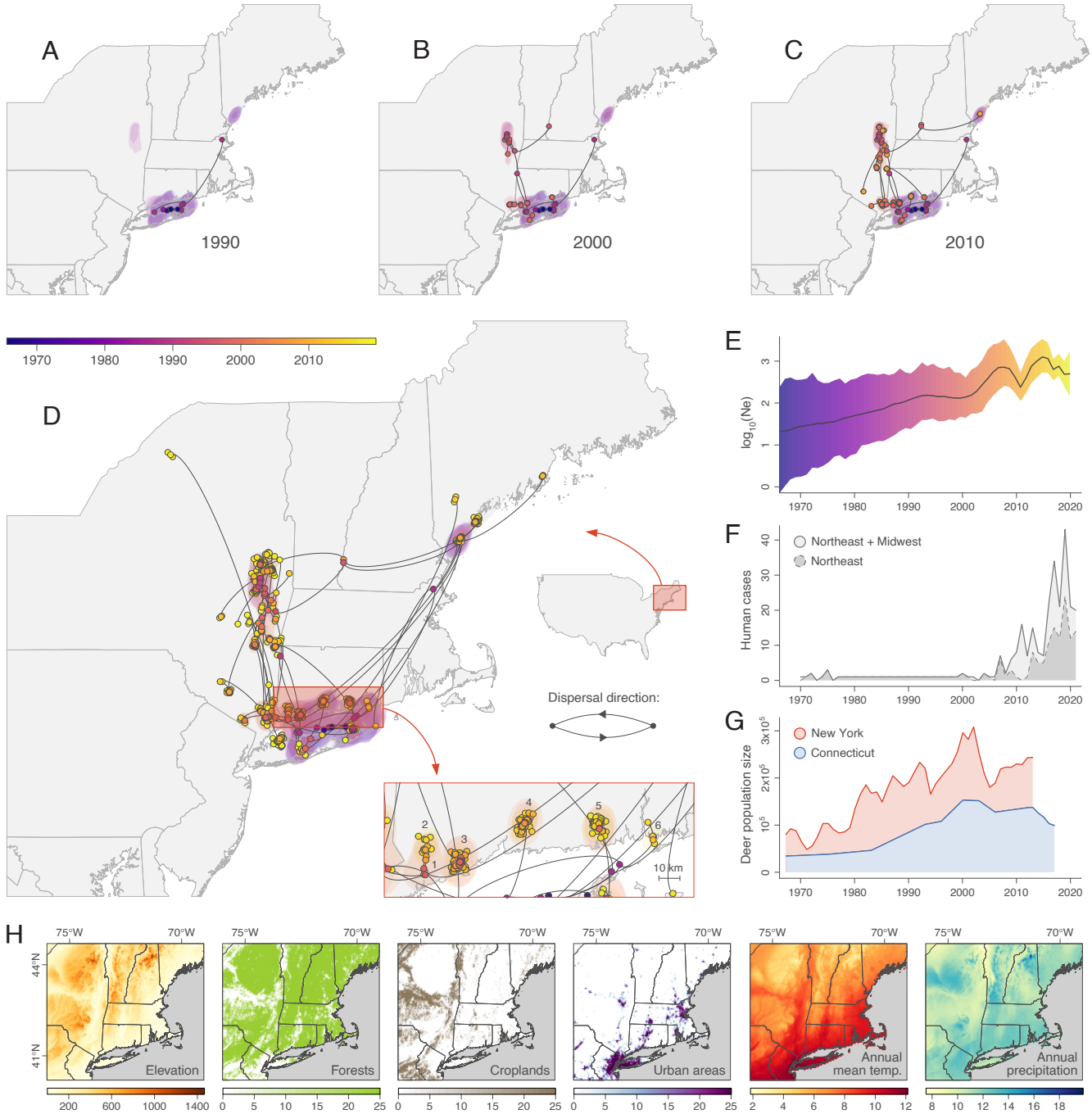

Fig. 3. Spatially explicit phylogeographic analysis of the dispersal history of Powassan virus in the northeastern United States. (*A–D*) Reconstruction of the dispersal history of Powassan virus lineages inferred by a spatially explicit phylogeographic analysis. We mapped branches of the MCC tree reported in Fig. 2 and whose nodes, as well as associated 80% highest posterior density (HPD) regions, are colored according to their time of occurrence. (*E*) Skygrid reconstruction of the evolution of the overall effective size of the viral population (Ne). (*F*) Evolution of the number of confirmed human cases. (*G*) Estimation of the white-tailed deer (*Odocoileus virginianus*) populations in the states of New York and Connecticut. (*H*) Environmental factors included in landscape phylogeographic analyses to test their impact on Powassan virus dispersal.

suggest that the virus is likely still emerging in parts of North America, following the northward expansion of *I. scapularis* (34).

We then used our continuous phylogeographic results to estimate the dispersal velocity of Powassan virus lineage II through the northeastern United States. We estimated a weighted lineage dispersal velocity of ~3 km/y [95% HPD = (2.6 to 3.8)])], which corresponds to a relatively slow dispersal capacity when compared to the estimates of the same metric obtained from the continuous phylogeographic analysis of other viruses (*SI Appendix*, Table S4). For instance, Powassan lineage II dispersed faster than the rodent-borne Lassa virus in western Africa [~1 km/y (35)], but considerably slower than the mosquito-borne West Nile virus when it invaded North America [~165 km/y (36)].

Population Size. We next investigated whether Powassan virus transmission has increased since its emergence in the Northeast, which could be a cause of the recent increase in reported human cases. We approached this by estimating the recent evolution of the virus-effective population size through a "skygrid" reconstruction, a nonparametric coalescent model assuming a panmictic population. Because the assumption of panmixia is never respected in practice, the effective population size (Ne) is sometimes interpreted as a

measure of the overall genetic diversity of the viral population (e.g., ref. 37); although it has also been argued that such interpretation could be a bit reductive (38). Since the emergence in the Northeast, we found an overall increase in the effective population size of Powassan virus lineage II, but with growth stagnating since ~2005 (Fig. 3*E*). The latter could be a reflection of the virus becoming established across all of our study sites, whereas the effective population size may still be increasing if we included additional sites in new emergence zones. Still, our data suggest that the increase in reported human cases in the region since 2010 (Fig. 3*F*) does not coincide with an increased virus-effective population size (Fig. 3*E*). Thus, our findings do not support the hypothesis that the recent uptick in human cases is due to a significant increase in Powassan virus transmission; rather, it may be caused by an increase of human exposure to infected ticks.

As hypothesized as a significant factor for the timing of Powassan virus lineage II emergence in the Northeast, our overall estimates for the virus-effective population size follow a similar trend as the population expansion of white-tailed deer in Connecticut and New York (Fig. 3*G*). Reforestation in the Northeast has led to dramatic increases in the white-tailed deer populations, followed by population expansion of *I. scapularis*, which in turn has facilitated the emergence of tick-borne pathogens such as *Borrelia burgdorferi* and *Babesia microti* (28). Our findings suggest that the cascading effect of population expansion of white-tailed deer and *I. scapularis* populations may also have facilitated the emergence of Powassan virus in the Northeast.

Impact of Environmental Factors on the Dispersal Dynamic. We exploited our spatially explicit phylogeographic reconstruction to investigate the impact of environmental factors on the dispersal dynamic of Powassan lineage II. As detailed in the Methods section, we tested the association between a series of environmental factors (Fig. 3*H*) and the dispersal location (39) as well as dispersal velocity of viral lineages (40). Our analyses revealed that inferred Powassan virus lineage II tended to avoid circulating in areas associated with relatively higher elevation [Bayes factor (BF) > 20; *SI Appendix*, Table S5]. However, outcomes of our analysis are strongly influenced by the sampling effort and pattern, and therefore we are able to describe environmental conditions related to dispersal locations of inferred viral lineages, but we cannot draw conclusions on the actual impact of those conditions on the dispersal (39). This observation could be related to higher abundance of *I. scapularis* at lower altitudes. Next, we investigated the impact of environmental factors on the dispersal velocity of Powassan virus lineage II. Our analyses did not highlight any environmental factor associated with the heterogeneity of Powassan lineage dispersal velocity across the study area (*SI Appendix*, Table S6). This means that none of the tested factors increased the correlation between dispersal duration and geographic distance, the latter thus remaining the main resistance factor to dispersal.

Discussion

Despite a rapid increase in the number of Powassan virus infections in humans over recent years, very little was known about the patterns of virus emergence and spread. By sequencing 279 Powassan virus genomes and using phylogeographic approaches, we have uncovered the patterns of virus emergence, transmission, and spread in the northeastern United States. Our analyses revealed that Powassan virus lineage II likely emerged in the Northeast around 1940 to 1975, following the population growth of white-tailed deer and expansion of *I. scapularis* tick populations (10, 28). Powassan virus lineage II is maintained in highly localized transmission foci, with few migration events between relatively nearby locations. Our continuous phylogeographic analysis revealed that Powassan virus lineage II likely emerged from southern Connecticut into more northern regions with a weighted lineage dispersal velocity of ~3 km/y. Although we found an overall upward trend in the virus effective population size over the last decades, the recent increase in the reported human cases of Powassan virus infection does not coincide with a higher effective population size. This suggests that the recent uptick in human cases is likely not due to a significant increase in Powassan virus enzootic transmission, but it may rather be due to other factors such as increased human exposure to infected ticks as well as an increase in case recognition. Our findings provide important insights into the local emergence patterns and transmission dynamics of Powassan virus in the northeastern United States. Insights into the highly localized transmission foci that sustain Powassan virus transmission across multiple years will help to identify areas with high risk of spillover to the human population, which can be targeted for prevention, education, or control efforts.

Our reconstruction of the emergence of Powassan virus in the Northeast follows similar patterns as the emergence reported for other tick-borne pathogens such as *Borrelia*, *Babesia*, and *Anaplasma* (41). Similar to Powassan virus, these pathogens are maintained in transmission cycles involving *I. scapularis* as the main vector and white-footed mice (*Peromyscus leucopus*) as the main host. Introduction of these tick-borne pathogens in the northeastern United States follows the reforestation in the 20th century leading to rapid population expansions of both white-tailed deer and *I. scapularis* (28, 34). Despite the similarities in emergence and ecology, other tick-borne pathogens seem to have more widespread distributions as compared to the highly focal distribution of Powassan virus. Previous studies on *Borrelia burgdorferi* reported high genetic diversity within local populations (42), lack of population genetic structure (25), and no strict genetic clustering by location within the Northeast (11). This suggests that despite the similarities in ecology, *B. burgdorferi* and Powassan virus are maintained by different mechanisms.

To explain the highly focal distribution, we have formulated two main hypotheses on how Powassan virus lineage II may be maintained in strict transmission foci (Fig. 4). Our first hypothesis is that vertical transmission from adult to the next larval stage plays a minor role in the Powassan virus transmission cycle. Adult *I. scapularis* ticks preferentially feed on larger mammals, such as white-tailed deer, and we would expect more mixing of Powassan virus clades if infected adults travel across larger distances when feeding on deer, particularly during early years of emergence when deer habitats were less fragmented. Inefficient vertical transmission from adult to larvae would explain why adult ticks and deer may move, while Powassan virus foci remain local. Moreover, low Powassan virus infection rates in ticks further reduce the probability of successful establishment in new areas (43). Previous studies have provided evidence for vertical transmission of Powassan virus (i.e., one of the six infected females transmitted Powassan virus to its progeny) (44), but it remains unclear what percentage of progeny within an egg batch becomes infected. Future studies can help to test this hypothesis by determining rates of vertical transmission in the laboratory, and determining infection rates of unfed larvae in the field.

Our second hypothesis is that backward transmission through tick life stages (e.g., from nymphs to larvae or adults to nymphs) is the main mechanism for Powassan virus maintenance (Fig. 4). Powassan virus clades better match the strong geographical structure of white-footed mice (45, 46) and other small mammals that share the same habitat rather than the weak structure of

Downloaded from https://www.pnas.org by "KATHOLIEKE UNIVERSITEIT LEUVEN KU LEUVEN, 2BERGEN, BIOMEDISCHE BIBLIO" on April 27, 2023 from IP address 134.58.253.56.

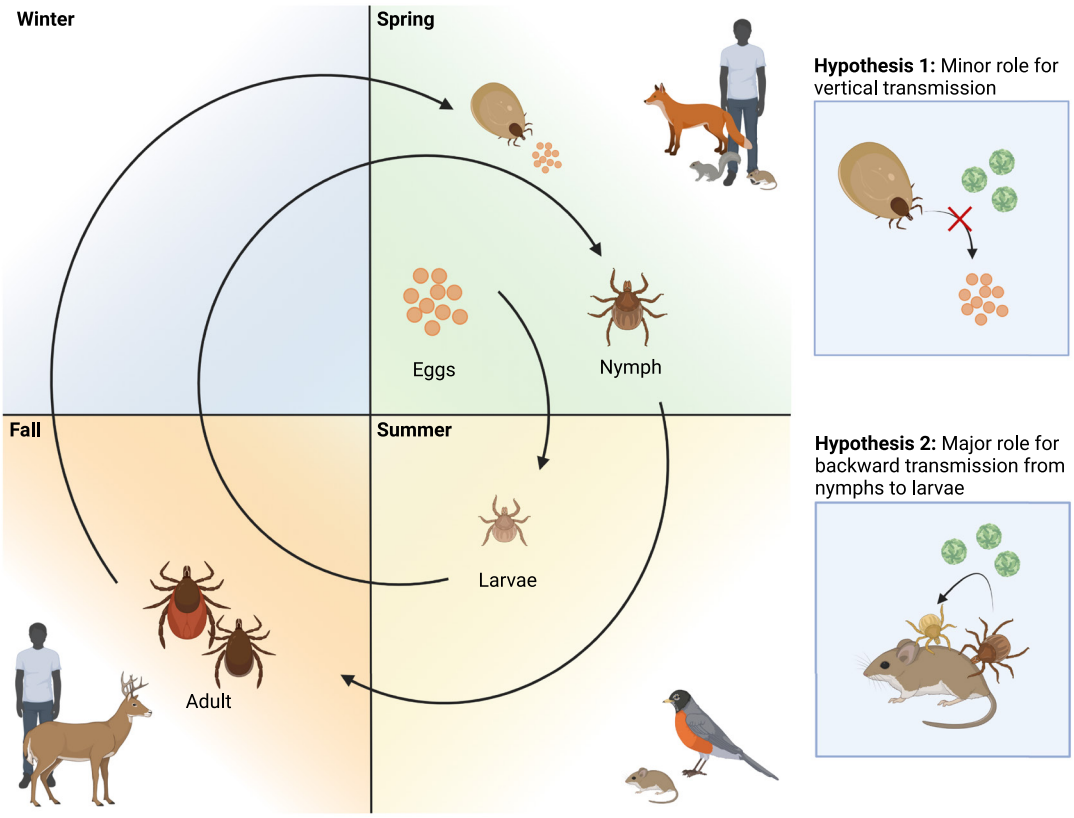

Fig. 4. *Ixodes scapularis* life cycle with proposed hypotheses that may explain the focal distribution of Powassan virus. Life cycle adapted from (41) and created with bioRender.com. Two hypotheses are proposed that may explain the focal distribution of Powassan virus.

I. scapularis ticks (26). This suggests that Powassan virus is mostly maintained by local transmission cycles that include larvae and nymphs feeding on small mammals, and adult ticks and other animals that can travel long distances are primarily dead-ends, although sporadic long-distance dispersal events may happen. Backward transmission from nymphs to larvae may either occur through direct cofeeding transmission in the absence of systemic infection of the rodent host, as has been reported for tick-borne encephalitis virus (47, 48), or through feeding on viremic hosts. To further investigate this hypothesis, we propose studies to determine the frequency of cofeeding between nymphs and larvae, to determine the size of transmission foci, and to perform comparative genetics between Powassan virus, *I. scapularis*, and small mammalian hosts, such as white-footed mice. Additionally, further research is needed to better understand the drivers and frequency of long-distance dispersal events that may result in the establishment of new transmission foci. Future ecological and genetic studies on host (e.g., bird and deer) and tick population movement are needed to better understand how movement of both host and vector may contribute to the occasional long-distance dissemination of Powassan virus.

Our study has several limitations. First, sampling strategy was not standardized across the northeastern United States nor was it standardized across time. Powassan virus surveillance activities are organized at the state level, and therefore there are differences in effort between states. For instance, samples sequenced from Connecticut and Maine were collected from the same sites during multiple years, whereas a much larger number of varying sites were sampled from year to year in New York. Sampling in New York also included tick collections from hunter-harvested deer, whereas sampling in Connecticut and Maine was exclusively done via drag sampling. Spatial heterogeneity in sampling effort will impact phylogeographic reconstructions, with unsampled locations being excluded from the reconstructed dispersal history, and undersampling of locations may lead to underestimation of the degree of connectivity between sites. Second, we attempted to include as many publicly available genomes in our analyses, but some genomes had to be excluded because they did not have complete metadata (e.g., missing collection date and location) or because they did not fit the molecular clock (e.g., Powassan sequences from human infections). Lastly, choice of molecular clock model may impact estimates for the evolutionary rate and the tMRCA. For example, when performing the analyses with a strict molecular clock model instead of a relaxed clock model, we get a lower estimated evolutionary rate of 3.76×10^{-5} substitutions/site/year (95% HPD = 2.59 to 5.45×10^{-5}) and an earlier tMRCA of 1881.3 (95% HPD = 1832.4 to 1934.8). This highlights the importance of molecular clock model choice based on best model fit and that we should be careful in our interpretation of the relaxed clock model estimates. Our estimates of emergence before the mid-1970s should be interpreted as emergence latest by this time period, and not as an exact estimate of the emergence time.

Our study has important implications for our understanding of Powassan virus transmission dynamics and future control. Currently, no vaccines or specific treatments are available for Powassan virus infection, which leaves prevention of disease highly dependent on education and control. Our estimates of the dispersal velocity can help to better inform the risk of regions in proximity of the current distribution of Powassan virus in the Northeast, which may be targeted for tick and virus surveillance as well as potential future control. Furthermore, the identification of highly localized transmission foci provides both opportunities for better education of the general public about high-risk areas and effective targeted control in Powassan virus hot spots. Eradication of

Downloaded from https://www.pnas.org by "KATHOLIEKE UNIVERSITEIT LEUVEN KU LEUVEN, 2BERGEN, BIOMEDISCHE BIBLIO" on April 27, 2023 from IP address 134.58.253.56.

Powassan virus in transmission foci that have been maintained for several years without introductions of new virus clades will likely result in highly effective and long-lasting control.

Methods

Sample Collection. Tick collections, nucleic acid extraction, and Powassan virus screening were done at the Connecticut Agricultural Experiment Station (CAES), MaineHealth Institute for Research, New York State Department of Health, and Cornell University (32, 43, 49–51). Briefly, the majority of ticks were collected by dragging a white cloth over the ground and low vegetation, and a smaller proportion were collected from vertebrate hosts (e.g., deer). All the collected ticks were sorted by species, life stage, collection site, and date before screening for pathogens. Individual or pooled ticks were homogenized, and nucleic acid was extracted according to manufacturer's instructions (*SI Appendix*, Table S1). Powassan virus was detected by RT-qPCR or nanochip assays (32, 43, 49–51). Samples collected from 2008 through 2016 in Connecticut were passaged once on BHK-21 cells before nucleic acid extraction.

Untargeted Metagenomic Illumina Sequencing. We initially used an untargeted metagenomic approach to sequence Powassan virus isolates which were passaged on cells at CAES. Our protocol is adapted from (52), and is openly available (53). In brief, 10 μL of the extracted nucleic acid was treated with DNase I (New England Biolabs), followed by a cleanup step using a ratio of 1.8:1 beads to sample. All the cleanup steps were done using Mag-Bind TotalPure NGS magnetic beads (Omega Bio-tek) with automated protocols for the Kingfisher flex purification system (Thermo Fisher Scientific). First-strand cDNA was synthesized using SuperScript IV VILO (Thermo Fisher Scientific) and second-strand cDNA using *Escherichia coli* DNA ligase and polymerase (New England Biolabs), followed by a cleanup step (1.8:1 beads-to-sample ratio). Libraries were prepared using the Nextera XT DNA library preparation kit for Illumina (Illumina, San Diego, CA), according to manufacturer's instructions but using less than the recommended reagent volumes (54). Individual and pooled libraries were quantified using the 1x dsDNA HS assay kit on the Qubit 4 (Thermo Fisher Scientific), and size distribution was determined using the high-sensitivity DNA kit on Bioanalyzer 2100 (Agilent, Santa Clara, CA). Pooled libraries were sequenced on the Illumina NovaSeq (paired-end 150) at the Yale Center for Genome Analysis. PCR duplicates were removed, reads were aligned to the reference genome using Bowtie2, and consensus genomes were called at a minimum frequency threshold of 0.75 and a minimum coverage of 10x using Geneious Prime 2020.0.4.

Targeted Amplicon-Based Sequencing. Although we were able to successfully sequence Powassan virus from cell culture-passage samples using untargeted metagenomics, we developed an amplicon-based sequencing approach to improve coverage when sequencing from tick homogenates. We used an adapted protocol (55), using Nextera XT for library prep as developed for SARS-CoV-2 amplicon-based sequencing (56). In brief, cDNA was synthesized from 10 μL RNA using SuperScript IV VILO (Thermo Fisher Scientific). Two separate primer pools were prepared by mixing equal volumes of each primer with a concentration of 10 μM (*SI Appendix*, Table S2). The two primer pools were used to generate tiled amplicons using Q5 high-fidelity 2X master mix (New England Biolabs), followed by a cleanup step and quantification using the Qubit. Amplicons were diluted to 1 ng/μL and combined for library prep as described above. Pooled libraries were quantified on Qubit and size distribution was determined on the bioanalyzer. Pooled libraries were sequenced on the Illumina NovaSeq (paired-end 150) at the Yale Center for Genome Analysis. Consensus genomes were generated at a minimum frequency threshold of 0.75 and a minimum coverage of 10x using iVar version 1.2.3. All sequencing data are publicly available under BioProject PRJNA889421 (57) and *SI Appendix*, Table S3.

Powassan Virus Phylogeny. We sequenced 279 Powassan virus genomes and estimated a maximum-likelihood tree using IQ-TREE version 1.6.12 with ultrafast bootstrap approximation (1,000 replicates) (58) to determine phylogenetic relationships between publicly available and newly sequenced Lineage I and II genomes.

Temporal Signal Assessment. To evaluate whether our Powassan virus dataset contains sufficient temporal signal and would permit time-calibrated analyses using molecular clock models (and hence constitutes a measurably evolving population), we performed a Bayesian Evaluation of Temporal Signal (BETS) analysis (21, 59). This analysis involves assessing the model fit to the data of both a strict clock and an uncorrelated relaxed clock with an underlying lognormal distribution, both with and without the sampling dates associated with the genomes in our dataset (Table 1). We employed generalized stepping-stone sampling (60) to accurately estimate the log marginal likelihood of each of these four models. For each log marginal likelihood estimation, we ran an initial Markov chain of 500 million iterations, followed by 500 power posteriors that are explored during 1 million iterations, logging every 1,000 iterations. We used the following priors: a lognormal prior distribution (mean = 1.0, SD = 1.25) on the transition–transversion ratio κ; a Dirichlet prior distribution (1.0; K = 4) on the nucleotide base frequencies; an exponential prior distribution (mean = 0.5) on the shape parameter describing the discretized gamma distribution used to model among-site rate heterogeneity; a CTMC reference prior (61) on the (constant) evolutionary rate under the strict clock model; a CTMC reference prior (61) on the mean evolutionary rate of the underlying lognormal distribution for the uncorrelated relaxed clock model; and an exponential prior distribution (mean = 1/3) on the SD of the underlying lognormal distribution for the uncorrelated relaxed clock model. Finally, we used both a gamma prior distribution (shape = 0.001; scale = 1,000.0) and a lognormal prior distribution (mean = 1; SD = 5; both on the log scale) on the constant population size, enabling us to test the sensitivity of BETS to different prior specifications. The results that we show in this manuscript are with a gamma prior distribution. BEAST ".xml" files for the BETS analyses are available on GitHub (https://github.com/grubaughlab/powassan-genomics/).

Discrete Phylogeographic Reconstruction. To investigate the dispersal history of Powassan virus lineages in the northeastern United States, we first conducted a discrete phylogeographic analysis using the Bayesian stochastic search variable selection (BSSVS) model (62) implemented in BEAST 1.10 (63). For this analysis, we considered each US county of origin as a distinct location, except for the Connecticut area where each sampling site was considered as a distinct discrete location. We modeled the branch-specific evolutionary rates according to a relaxed molecular clock with an underlying lognormal distribution (64) and the nucleotide substitution process according to a GTR+Γ parameterization(65), and we specified a skygrid coalescent model as the tree prior (66). Both the substitution and coalescent models were selected for their flexibility. Three independent Markov chain Monte Carlo (MCMC) algorithms were run for 5×10^8 iterations and sampled every 10^5 iterations. Resulting posterior distributions were eventually combined after having discarded 10% of sampled trees in each of them. We used the program Tracer 1.7 (67) for assessing the convergence and mixing properties, and that estimated sampling size (ESS) values associated with estimated parameters were all >200 after having combined the outputs of the three independent analyses. We then used the program TreeAnnotator 1.10 (63) to obtain a maximum clade credibility (MCC) tree. We reported Markov jumps between discrete locations as estimated by the BSSVS analyses and supported by an adjusted Bayes factor (BF) values >3, which correspond to at least "positive" statistical support following the scale of interpretation defined by Kass & Raftery (22). The adjusted BF support takes into account the relative abundance of samples by location (68) and is based on a tip-label swapping procedure similar to a tip-date randomization that can be performed to test for temporal signal (69). BEAST ".xml" files and concatenated log files for the discrete phylogeographic analysis are available on GitHub (https://github.com/grubaughlab/powassan-genomics/).

Continuous Phylogeographic Reconstruction. To reconstruct the dispersal history of Powassan virus lineages in a spatially explicit context, we performed a continuous phylogeographic analysis using the relaxed random walk (RRW) diffusion model (70, 71) implemented in the software package BEAST 1.10 (63), with a gamma distribution to model the among-branch heterogeneity in diffusion velocity. As for the discrete phylogeographic analysis, branch-specific evolutionary rates were modeled according to a relaxed molecular clock with an underlying lognormal distribution and the nucleotide substitution process according to a GTR+Γ parameterization (65), and we also specified a flexible skygrid model as the tree prior (66). The Markov chain Monte Carlo (MCMC) algorithm was run for 15×10^8 generations and sampled every 10^5 generations. We used the program Tracer 1.7 for assessing the convergence and mixing properties, and that estimated sampling size (ESS) values associated with estimated parameters were

Downloaded from https://www.pnas.org by "KATHOLIEKE UNIVERSITEIT LEUVEN KU LEUVEN, 2BERGEN, BIOMEDISCHE BIBLIO" on April 27, 2023 from IP address 134.58.253.56.

all >200, the program TreeAnnotator 1.10 to obtain a MCC tree, as well as the R package "seraphim" (72, 73) to extract the spatiotemporal information embedded within posterior trees and to estimate the weighted lineage dispersal velocity, the latter being defined as follows:

$$v = \frac{\sum_{i=1}^{n} d_i}{\sum_{i=1}^{n} t_i}$$

where d_i and t_i are the geographic distance traveled by the phylogeny branch and the duration of that branch, respectively. BEAST ".xml" files and concatenated log files for the continuous phylogeographic analysis are available on GitHub (https://github.com/grubaughlab/powassan-genomics/).

Landscape Phylogeographic Analyses. Landscape phylogeographic analyses aim at exploiting phylogeographic reconstructions to unravel the impact of environmental factors on the dispersal history and dynamic of viral lineages (74). Specifically, we implemented two previously introduced analytical procedures to investigate the impact of environmental factors on the dispersal location (39) and velocity (40) of Powassan virus lineages. Both analytical procedures here rely on the comparison between inferred and randomized spatially annotated trees, the latter sharing the time-scaled topology of the inferred trees but with phylogenetic branch positions that had been randomized across the study area. In practice, phylogenetic node positions were randomized within the study area, under the constraint that branch length (i.e., geographic distance connecting both branch nodes), branch duration, tree topology, and root position remained unchanged (72). The purpose of these randomizations is thus to obtain spatially annotated trees corresponding to the trees inferred by continuous phylogeography but along which we generated a diffusion process that has been impacted by any environmental factor.

We first investigated whether Powassan virus lineages tended to avoid or preferentially circulate within areas associated with particular environmental conditions. For this purpose, we extracted and subsequently averaged the environmental values at the tree node positions to obtain, for each environmental factor, a posterior distribution of mean environmental values at tree node positions. We then compared values obtained through inferred trees and their corresponding randomized trees by approximating a BF with the following formula (75): $BF = (p_e/(1-p_e))/(0.5/(1-0.5))$. To test whether a particular environmental factor e tended to attract viral lineages, p_e was defined as the frequency at which the environmental values from inferred trees were greater than values from randomized trees, and to test whether a particular environmental factor e tended to repulse viral lineages, p_e was defined as the frequency at which the environmental values from inferred trees were lower than values from randomized trees. We considered BF values > 20 and 3 < BF < 20 as strong and positive statistical supports, respectively (22).

Second, we investigated to what extent Powassan virus lineage dispersal velocity was impacted by environmental factors acting as conductance or resistance factors. For each branch in the inferred and randomized trees, we calculated an "environmental distance" using two path models: the least-cost path (76) and Circuitscape (77) algorithms, with the latter accommodating uncertainty in the travel route. An environmental distance is calculated first from the raster of the environmental variable, and second from a uniform "null" raster whose cell values are all set to "1." The environmental distance is a spatial distance that is weighted according to the values of the underlying environmental raster, and therefore constitutes a proxy for geographical distance when computed on the null raster. Each environmental variable was considered twice: once as a potential conductance factor that facilitates movement, and once as a potential resistance factor that impedes it. For each environmental variable, we also generated and tested several distinct rasters by transforming the original raster cell values with the following formula: $v_t = 1 + k(v_o/v_{max})$, where v_t and v_o are the transformed and original cell values and v_{max} the maximum cell value recorded in the raster. The rescaling parameter k here allows the definition and testing of different strengths of raster cell conductance or resistance, relative to the conductance/resistance of a cell with a minimum value set to "1," which corresponds to the "null" raster. For each environmental variable, we generated three distinct rasters using the following values for rescaling factor k: k = 10, 100, and 1,000. For these analyses, we estimated the statistic Q defined as the difference between the coefficient of determinations obtained i) when branch durations are regressed against the environmental distances computed on an environmental distance and ii) when branch durations are regressed against the environmental distances computed on the null raster. We estimated a Q statistic for each environmental raster and each of the 1,000 trees sampled from the posterior distribution. An environmental factor was only considered as potentially explanatory if both its distribution of regression coefficients and its associated distribution of Q values were positive (78), i.e., with at least 90% of positive values. In this case, the statistical support associated with the resulting Q distribution was compared with the corresponding null of distribution of Q values obtained when computing environmental distances for phylogenetic branches of randomized trees. Similar to the procedure used for the investigation of the impact of environmental factors on the dispersal locations of Powassan virus lineages, the comparisons between inferred and randomized distributions of Q values were formalized by approximating a Bayes factor support (40).

Data, Materials, and Software Availability. All data are included in this article, *SI Appendix*, GitHub (https://github.com/grubaughlab/powassan-genomics/) (79), and in BioProject PRJNA889421 (57).

ACKNOWLEDGMENTS. We would like to thank Anne Piantadosi, Erica Normandin, Pardis C. Sabeti, Rebekah McMinn, Greg Ebel, Heidi Goethert, Sam R. Telford III, Sebastian Lequime, Alexander A. Fisher, and Marc A. Suchard for their input in discussions on the methods and results of this study. This publication was made possible by CTSA Grant Number UL1 TR001863 from the National Center for Advancing Translational Science, a component of the NIH awarded to C.B.F.V., and the National Institute of Allergy and Infectious Diseases of the NIH under Award Number 1R56 AI149004-01A1 awarded to N.D.G. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH. G.B. acknowledges support from the Internal Funds KU Leuven under grant agreement C14/18/094 and the Research Foundation – Flanders ("Fonds voor Wetenschappelijk Onderzoek – Vlaanderen", G0E1420N). G.B. and S.D. acknowledge support from the Research Foundation – Flanders ("Fonds voor Wetenschappelijk Onderzoek – Vlaanderen", G098321N). S.D. also acknowledges support from the Fonds National de la Recherche Scientifique (F.R.S.-FNRS, Belgium; grant n°F.4515.22) and from the European Union Horizon 2020 project MOOD (grant agreement n°874850).

Author affiliations: [a]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT 06510; [b]Center for Vector Biology and Zoonotic Diseases, Department of Entomology, The Connecticut Agricultural Experiment Station, New Haven, CT 06511; [c]The Arbovirus Laboratory, New York State Department of Health, Wadsworth Center, Slingerlands, NY 12159; [d]Department of Biomedical Sciences, State University of New York at Albany School of Public Health, Albany, NY 12222; [e]Vector-borne Disease Laboratory, MaineHealth Institute for Research, Scarborough, ME 04074; [f]Department of Epidemiology, University of Nebraska Medical Center, Omaha, NE 68198; [g]Instituto Todos pela Saúde, São Paulo SP 01310-942, Brazil; [h]Department of Environmental Science and Forestry, The Connecticut Agricultural Experiment Station, New Haven, CT 06511; [i]New York State Department of Health, Bureau of Communicable Disease Control, Albany, NY 12237; [j]Department of Entomology, Cornell University, Ithaca, NY 14853; [k]Department of Public and Ecosystem Health, Cornell University, Ithaca, NY 14853; [l]Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven 3000, Belgium; [m]Spatial Epidemiology Lab, Université Libre de Bruxelles, Brussels 1050, Belgium; and [n]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06511

1. Centers for Disease Control and Prevention, National Center for Emerging and Zoonotic Infectious Diseases (NCEZID), Division of Vector-Borne Diseases (DVBD), Tickborne disease surveillance data summary (2019). https://www.cdc.gov/ticks/data-summary/index.html. Accessed 16 May 2019.
2. ArboNET, Arboviral Diseases Branch, Centers for Disease Control and Prevention, Powassan virus neuroinvasive disease cases reported by year, 2008–2017 (2018). https://www.cdc.gov/powassan/statistics.html. Accessed 22 April 2019.
3. M. E. Hermance, S. Thangamani, Powassan virus: An emerging arbovirus of public health concern in North America. *Vector Borne Zoonotic. Dis.* **17**, 453–462 (2017).
4. G. D. Ebel, Update on Powassan virus: Emergence of a North American tick-borne flavivirus. *Annu. Rev. Entomol.* **55**, 95–110 (2010).
5. K. N. Pesko, F. Torres-Perez, B. L. Hjelle, G. D. Ebel, Molecular epidemiology of Powassan virus in North America. *J. Gen. Virol.* **91**, 2698–2705 (2010).

Downloaded from https://www.pnas.org by "KATHOLIEKE UNIVERSITEIT LEUVEN KU LEUVEN, 2BERGEN, BIOMEDISCHE BIBLIO" on April 27, 2023 from IP address 134.58.253.56.

6. D. M. McLean, W. L. Donohue, Powassan virus: Isolation of virus from a fatal case of encephalitis. *Can. Med. Assoc. J.* **80**, 708–711 (1959).
7. Centers for Disease Control and Prevention, Powassan Virus Statistics & Maps (2021). https://www.cdc.gov/powassan/statistics.html. Accessed 8 October 2021.
8. A. Piantadosi *et al.*, Rapid Detection of Powassan Virus in a Patient With Encephalitis by Metagenomic Sequencing. *Clin. Infect. Dis.* **66**, 789–792 (2018).
9. D. E. Sonenshine, Range Expansion of Tick Disease Vectors in North America: Implications for Spread of Tick-Borne Disease. *Int. J. Environ. Res. Public Health* **15** (2018).
10. A. Spielman, M. L. Wilson, J. F. Levine, J. Piesman, Ecology of *Ixodes dammini*-borne human Babesiosis and Lyme disease. *Annual Review of Entomology* **30**, 439–460 (1985).
11. A. G. Hoen *et al.*, Phylogeography of *Borrelia burgdorferi* in the eastern United States reflects multiple independent Lyme disease emergence events. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15013–15018 (2009).
12. G. D. Ebel, E. N. Campbell, H. K. Goethert, A. Spielman, S. R. Telford, Enzootic transmission of deer tick virus in New England and Wisconsin sites. *The American Journal of Tropical Medicine and Hygiene* **63**, 36–42 (2000).
13. H. K. Goethert, T. N. Mather, R. W. Johnson, S. R. Telford 3rd., Incrimination of shrews as a reservoir for Powassan virus. *Commun Biol* **4**, 1319 (2021).
14. Grubaugh Lab, Nextstrain build for Powassan virus. https://nextstrain.org/community/grubaughlab/powassan-genomics/All. Accessed 2 February 2023.
15. Grubaugh Lab, Nextstrain build for Powassan virus in the northeastern United States. https://nextstrain.org/community/grubaughlab/powassan-genomics/NE. Accessed 2 February 2023.
16. CDC, Powassan virus isolated from a patient with Encephalitis. *Morb. Mortal. Weekly Rep.* **24**, 379 (1975).
17. A. J. Main, A. B. Carey, W. G. Downs, Powassan virus in *Ixodes cookei* and *Mustelidae* in New England. *J. Wildlife Dis.* **15**, 585–591 (1979).
18. R. E. Lange *et al.*, Identification and characterization of novel lineage 1 Powassan virus strains in New York State. *Emerg. Microbes Infect.* **12**, 2155585 (2023).
19. C. Hart *et al.*, Powassan virus lineage I in field-collected dermacentor variabilis ticks, New York, USA. *Emerg. Infect. Dis.* **29**, 415–417 (2023).
20. A. Rambaut, T. T. Lam, L. Max Carvalho, O. G. Pybus, Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2**, **vew007** (2016).
21. S. Duchene *et al.*, Bayesian evaluation of temporal signal in measurably evolving populations. *Mol. Biol. Evol.* **37**, 3363–3379 (2020).
22. R. E. Kass, A. E. Raftery, Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
23. A. N. Bondaryuk *et al.*, Phylogeography and re-evaluation of evolutionary rate of powassan virus using complete genome data. *Biology* **10**, 1282 (2021).
24. E. L. Subbotina, V. B. Loktev, Molecular evolution of the tick-borne encephalitis and Powassan viruses. *Mol. Biol.* **46**, 75–84 (2012).
25. P. T. Humphrey, D. A. Caporale, D. Brisson, Uncoordinated phylogeography of Borrelia burgdorferi and its tick vector. Ixodes scapularis. *Evolution* **64**, 2653–2663 (2010).
26. K. S. Walter, G. Carpi, A. Caccone, M. A. Diuk-Wasser, Genomic insights into the ancient spread of Lyme disease across North America. *Nat. Ecol. Evol.* **1**, 1569–1576 (2017).
27. X. Lee, K. Hardy, D. H. Johnson, S. M. Paskewitz, Hunter-killed deer surveillance to assess changes in the prevalence and distribution of Ixodes scapularis (Acari: Ixodidae) in Wisconsin. *J. Med. Entomol.* **50**, 632–639 (2013).
28. A. Spielman, The emergence of Lyme disease and human babesiosis in a changing environment. *Ann. N. Y. Acad. Sci.* **740**, 146–156 (1994).
29. R. J. Eisen, L. Eisen, C. B. Beard, County-scale distribution of Ixodes scapularis and Ixodes pacificus (acari: ixodidae) in the continental United States. *J. Med. Entomol.* **53**, 349–386 (2016).
30. S. E. Randolph, D. Miklisová, J. Lysy, D. J. Rogers, M. Labuda, Incidence from coincidence: patterns of tick infestations on rodents facilitate transmission of tick-borne encephalitis virus. *Parasitology* **118**, 177–186 (1999).
31. M. Labuda, S. E. Randolph, Survival strategy of tick-borne encephalitis virus: cellular basis and environmental determinants. *Zentralbl. Bakteriol.* **289**, 513–524 (1999).
32. J. F. Anderson, P. M. Armstrong, Prevalence and genetic characterization of Powassan virus strains infecting *Ixodes scapularis* in Connecticut. *Am. J. Trop. Med. Hygiene* **87**, 754–759 (2012).
33. D. E. Brackney, R. A. Nofchissey, K. A. Fitzpatrick, I. K. Brown, G. D. Ebel, Short report: Stable prevalence of Powassan virus in *Ixodes scapularis* in a northern Wisconsin Focus. *Am. J. Tropical Med. Hygiene* **79**, 971–973 (2008).
34. D. Fish, "Range expansion of Ixodes scapularis in the USA" in *Climate, Ticks and Disease*, P. Nuttall, Ed. (CABI, 2021), pp. 176–182.
35. R. Klitting *et al.*, Predicting the evolution of the Lassa virus endemic area and population at risk over the next decades. *Nat. Commun.* **13**, 1–15 (2022).
36. S. Dellicour *et al.*, Epidemiological hypothesis testing using a phylogeographic and phylodynamic framework. *Nat. Commun.* **11**, 5620 (2020).
37. A. Rambaut *et al.*, The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453**, 615–619 (2008).
38. S. D. W. Frost, E. M. Volz, Viral phylodynamics and the search for an "effective number of infections". *Lond. B Biol. Sci.* **365**, 1879–1890, Philos. Trans. R. Soc. (2010).
39. S. Dellicour *et al.*, Using phylogeographic approaches to analyse the dispersal history, velocity and direction of viral lineages - Application to rabies virus spread in Iran. *Mol. Ecol.* **28**, 4335–4350 (2019).
40. S. Dellicour *et al.*, Using viral gene sequences to compare and explain the heterogeneous spatial dynamics of virus epidemics. *Mol. Biol. Evol.* **34**, 2563–2571 (2017).
41. R. J. Eisen, L. Eisen, The blacklegged tick, ixodes scapularis: an increasing public health concern. *Trends Parasitol.* **34**, 295–309 (2018).
42. I. N. Wang *et al.*, Genetic diversity of ospC in a local population of Borrelia burgdorferi sensu stricto. *Genetics* **151**, 15–30 (1999).
43. R. M. Robich *et al.*, Prevalence and genetic characterization of deer tick virus (Powassan virus, lineage II) in *Ixodes scapularis* ticks collected in Maine. *Am. J. Trop. Med. Hyg.* **101**, 467–471 (2019).
44. A. Costero, M. A. Grayson, Experimental transmission of Powassan Virus (Flaviviridae) by *Ixodes scapularis* ticks (Acari: Ixodidae). *Am. J. Tropical Med. Hygiene* **55**, 536–546 (1996).
45. A. Rogic, N. Tessier, P. Legendre, F.-J. Lapointe, V. Millien, Genetic structure of the white-footed mouse in the context of the emergence of Lyme disease in southern Québec. *Ecol. Evol.* **3**, 2075–2088 (2013).
46. P. E. Howell, M. L. Delgado, K. T. Scribner, Landscape genetic analysis of co-distributed white-footed mice (Peromyscus leucopus) and prairie deer mice (Peromyscus maniculatus bairdii) in an agroecosystem. *J. Mammal.* **98**, 793–803 (2017).
47. M. Labuda, V. Danielova, L. D. Jones, P. A. Nuttall, Amplification of tick-borne encephalitis virus infection during co-feeding of ticks. *Med. Vet. Entomol.* **7**, 339–342 (1993).
48. S. E. Randolph, L. Gern, P. A. Nuttall, Co-feeding ticks: Epidemiological significance for tick-borne pathogen transmission. *Parasitol. Today* **12**, 472–479 (1996).
49. A. P. Dupuis *et al.*, Isolation of deer tick virus (Powassan virus, lineage II) from Ixodes scapularis and detection of antibody in vertebrate hosts sampled in the Hudson Valley, New York State. *Parasites Vectors* **6**, 185 (2013).
50. M. T. Aliota *et al.*, The prevalence of zoonotic tick-borne pathogens in *Ixodes scapularis* collected in the Hudson Valley, New York state. *Vector-borne and Zoonotic Diseases* **14**, 245–250 (2014).
51. Q. Yuan *et al.*, Active surveillance of pathogens from ticks collected in New York State suburban parks and schoolyards. *Zoonoses Public Health* **67**, 684–696 (2020).
52. C. B. Matranga *et al.*, Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. *Genome Biol.* **15**, 519 (2014).
53. Grubaugh Lab, Protocols in development. http://grubaughlab.com/open-science/protocols/. Accessed 8 October 2021.
54. N. D. Grubaugh *et al.*, An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* **20**, 8 (2019).
55. J. Quick *et al.*, Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017).
56. Andersen Lab, Protocols (2017). https://andersen-lab.com/secrets/protocols/. Accessed 8 October 2021.
57. Yale School of Public Health, Powassan virus genomics. *NCBI BioProject*. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA889421. Deposited 11 October 2022.
58. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
59. S. Duchene *et al.*, Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol.* **6**, veaa061 (2020).
60. G. Baele, P. Lemey, M. A. Suchard, Genealogical working distributions for Bayesian model testing with phylogenetic uncertainty. *Syst. Biol.* **65**, 250–264 (2016).
61. M. A. R. Ferreira, M. A. Suchard, Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat.* **36**, 355–368 (2008).
62. P. Lemey, A. Rambaut, A. J. Drummond, M. A. Suchard, Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* **5**, e1000520 (2009).
63. M. A. Suchard *et al.*, Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
64. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
65. S. Tavaré, Some probabilistic and statistical problems in the analysis of DNA sequenced. *Lect. Math. Life Sci.* **17**, 57–86 (1986).
66. M. S. Gill *et al.*, Improving Bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).
67. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
68. B. Vrancken *et al.*, Dynamics and dispersal of local human immunodeficiency virus epidemics within San Diego and across the San Diego-Tijuana border. *Clin. Infect. Dis.* **73**, e2018–e2025 (2021).
69. N. S. Trovão *et al.*, Host ecology determines the dispersal patterns of a plant virus. *Virus Evol* **1**, vev016 (2015).
70. P. Lemey, A. Rambaut, J. J. Welch, M. A. Suchard, Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol.* **27**, 1877–1885 (2010).
71. O. G. Pybus *et al.*, Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 15066–15071 (2012).
72. S. Dellicour, R. Rose, O. G. Pybus, Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data. *BMC Bioinformatics* **17**, 82 (2016).
73. S. Dellicour, R. Rose, N. R. Faria, P. Lemey, O. G. Pybus, SERAPHIM: Studying environmental rasters and phylogenetically informed movements. *Bioinformatics* **32**, 3204–3206 (2016).
74. S. Dellicour, B. Vrancken, N. S. Trovão, D. Fargette, P. Lemey, On the importance of negative controls in viral landscape phylogeography. *Virus Evol* **4**, vey023 (2018).
75. M. A. Suchard, R. E. Weiss, J. S. Sinsheimer, Models for estimating Bayes factors with applications to phylogeny and tests of monophyly. *Biometrics* **61**, 665–673 (2005).
76. E. W. Dijkstra, A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271 (1959).
77. B. H. McRae, Isolation by resistance. *Evolution* **60**, 1551–1561 (2006).
78. M. Jacquot, K. Nomikou, M. Palmarini, P. Mertens, R. Biek, Bluetongue virus spread in Europe is a consequence of climatic, landscape and vertebrate host factors as revealed by phylogeographic inference. *Proc. Biol. Sci.* **284**, 20170919 (2017).
79. Grubaugh lab, Powassan Genomics. *GitHub*. https://github.com/grubaughlab/powassan-genomics/. Deposited 11 January 2023.

Downloaded from https://www.pnas.org by "KATHOLIEKE UNIVERSITEIT LEUVEN KU LEUVEN, 2BERGEN, BIOMEDISCHE BIBLIO" on April 27, 2023 from IP address 134.58.253.56.