

# Dynamics of Eastern equine encephalitis virus during the 2019 outbreak in the Northeast United States

Verity Hill<sup>1,\*,@</sup>, Robert T. Koch<sup>1\*</sup>, Sean M. Bialosuknia<sup>2</sup>, Kiet Ngo<sup>2</sup>, Steven D. Zink<sup>2</sup>, Cheri A. Koetzner<sup>2</sup>, Joseph G. Maffei<sup>2</sup>, Alan P. Dupuis<sup>2</sup>, P. Bryon Backenson<sup>3</sup>, JoAnne Oliver<sup>4,5</sup>, Angela B. Bransfield<sup>6</sup>, Michael J. Misencik<sup>6</sup>, Tanya A. Petruff<sup>6</sup>, John J. Shepard<sup>6</sup>, Joshua L. Warren<sup>7,8</sup>, Mandev S. Gill<sup>9</sup>, Guy Baele<sup>10</sup>, Chantal B.F. Vogels<sup>1</sup>, Glen Gallagher<sup>11,12</sup>, Paul Burns<sup>11</sup>, Aaron Hentoff<sup>11</sup>, Sandra Smole<sup>11</sup>, Catherine Brown<sup>11</sup>, Matthew Osborne<sup>11</sup>, Laura D. Kramer<sup>2,13</sup>, Philip M. Armstrong<sup>1,5,#,@</sup>, Alexander T. Ciota<sup>2,13,#,@</sup>, Nathan D. Grubaugh<sup>1,7,14,#,@</sup>

<sup>1</sup> Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA

<sup>2</sup> The Arbovirus Laboratory, New York State Department of Health, Wadsworth Center, Slingerlands, NY, USA

<sup>3</sup> New York State Department of Health, Bureau of Communicable Disease Control, Albany, NY, USA

<sup>4</sup> New York State Department of Health, Bureau of Communicable Disease Control, Syracuse, NY, USA

<sup>5</sup> Division of Environmental and Renewable Resources, State University of New York at Morrisville - School of Agriculture, Business and Technology, Morrisville, NY, USA

<sup>6</sup> Center for Vector Biology and Zoonotic Diseases, Department of Entomology, The Connecticut Agricultural Experiment Station, New Haven, CT, USA

<sup>7</sup> Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

<sup>8</sup> Public Health Modeling Unit, Yale School of Public Health, New Haven, CT, USA

<sup>9</sup> Department of Statistics, University of Georgia, Athens, GA, USA

<sup>10</sup> Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium

<sup>11</sup> Massachusetts Department of Public Health, Boston, MA, USA

<sup>12</sup> Rhode Island State Health Laboratory, Rhode Island Department of Health, Providence, RI, USA

<sup>13</sup> Department of Biomedical Sciences, State University of New York at Albany School of Public Health, Albany, NY, USA

<sup>14</sup> Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

\* Co-first authors

# Co-senior authors

@ Co-corresponding authors: VH ([verity.hill@yale.edu](mailto:verity.hill@yale.edu)), PMA ([Philip.Armstrong@ct.gov](mailto:Philip.Armstrong@ct.gov)), ATC ([alexander.ciota@health.ny.gov](mailto:alexander.ciota@health.ny.gov)), NDG ([nathan.grubaugh@yale.edu](mailto:nathan.grubaugh@yale.edu))

## Abstract

Eastern equine encephalitis virus (EEEV) causes a rare but severe disease in horses and humans, and is maintained in an enzootic transmission cycle between songbirds and *Culiseta melanura* mosquitoes. In 2019, the largest EEEV outbreak in the United States for more than 50 years occurred, centered in the Northeast. To explore the dynamics of the outbreak, we sequenced 80 isolates of EEEV and combined them with existing genomic data. We found that, like previous years, cases were driven by frequent short-lived virus introductions into the Northeast from Florida. Once in the Northeast, we found that Massachusetts was important for regional spread. We found no evidence of any changes in viral, human, or bird factors which would explain the increase in cases in 2019. By using detailed mosquito surveillance data collected by Massachusetts and Connecticut, however, we found that the abundance of *Cs. melanura* was

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

44 exceptionally high in 2019, as was the EEEV infection rate. We employed these mosquito data  
45 to build a negative binomial regression model and applied it to estimate early season risks of  
46 human or horse cases. We found that the month of first detection of EEEV in mosquito  
47 surveillance data and vector index (abundance multiplied by infection rate) were predictive of  
48 cases later in the season. We therefore highlight the importance of mosquito surveillance  
49 programs as an integral part of public health and disease control.  
50

## 51 Introduction

52 Eastern equine encephalitis virus (EEEV) is a mosquito-borne virus that causes periodic outbreaks  
53 in humans and horses in the United States (US) since its discovery in 1933 (Giltner & Shahan,  
54 1933; TenBroeck & Merrill, 1933). The virus circulates in a bird-mosquito transmission cycle while  
55 infections of most mammals are considered “dead end hosts”. In humans, EEEV can cause  
56 severe disease, with an apparent case fatality rate of 30% and long-term neurological sequelae  
57 in more than half of those who survive (Lindsey et al., 2018). Still, diagnosed human cases are  
58 rare, with an average of 7 per year in the US. The largest human outbreak in more than 50 years  
59 was the 38 cases reported in 2019, including 12 deaths (Lindsey et al., 2020). The outbreak was  
60 not limited to the East Coast where cases are typically detected, as 10 of the human cases that  
61 year were reported from Michigan. The widespread EEEV cases in 2019 had significant impacts  
62 on the communities: many evening outdoor events were canceled to avoid mosquito exposure  
63 and aerial insecticide applications were the subject of public controversy (Shamus, 2019;  
64 Tunison, 2019). Thus, understanding the drivers of EEEV outbreaks and how to accurately  
65 communicate risk to the public is of high importance (Howard, 2019).

66 The key unanswered questions are (1) what factors facilitated the unprecedented EEEV activity  
67 in 2019, and (2) whether we can accurately estimate risk of human and horse infections? These  
68 are challenging questions to answer because the ecology of EEEV is complex, involving multiple  
69 species of bird and mosquito. *Culiseta melanura* serves as the main mosquito vector of EEEV in  
70 North America, lives in freshwater swamp habitats, and feeds primarily on passerine birds (Morris,  
71 1988). Historically, *Coquillettidia perturbans* was implicated in the spillover process to humans  
72 as bridge vectors (i.e., vectors which feed on both birds and mammals) (Armstrong & Andreadis,  
73 2022). This strict delineation between obligate avian and permissive feeders, however, is not so  
74 absolute, and *Cs. melanura* has also been found to occasionally feed on mammals in the  
75 Northeast (Molaei et al., 2006). In terms of viral dynamics, previous work demonstrated that EEEV  
76 circulates year round in Florida and is introduced into the Northeast through seasonal bird  
77 migration (Mundis et al., 2022; Tan et al., 2018), although this process does not happen  
78 predictably every year. Predicting the annual case dynamics is therefore difficult, having to take  
79 into account viral dynamics across multiple species and geographic scales.

80 In this study, we used a combination of phylodynamics, mosquito surveillance, and mathematical  
81 and statistical modeling to explore the dynamics of EEEV in the Northeast US, and specifically  
82 address factors behind the 2019 outbreak. We sequenced 80 EEEV isolates to add to the  
83 currently available genomic data, including 48 from 2019, and combined them with historical data  
84 to identify patterns influencing national and regional spread. We then explored which human,  
85 viral, mosquito, and ecological factors may contribute to years with many cases in humans and  
86 horses, with the aim of understanding if years like 2019 are predictable, or likely to be repeated.  
87 We confirmed that the 2019 outbreak was primarily driven by EEEV introductions from Florida  
88 rather than extended spread in the region. We also found that when there is regional spread in

89 the Northeast it mostly originates from Massachusetts. We found no viral, human, or avian factors  
90 which contributed to the 2019 outbreak, but found that mosquito surveillance data was able to  
91 explain much of the variation in human and horse cases, highlighting the importance of high-  
92 quality and routinely collected mosquito data.

93

## 94 Results

### 95 High EEEV mosquito infections rates within known transmission foci

96 The outbreak of EEEV in humans and horses in 2019 was primarily focused in the Northeast US,  
97 defined as New York, Connecticut, Massachusetts, New Hampshire, Vermont, Rhode Island,  
98 and Maine. Using routine surveillance data from Massachusetts, Connecticut, and New York, we  
99 found that (1) cases occurred within previously known EEEV transmission foci and (2) the high  
100 number of human and horse cases in 2019 corresponded with a high number of trapped *Cs.*  
101 *melanura* and a high EEEV mosquito infection rate.

102

103 During 2019 there were 19 human and 26 horse cases reported in the Northeast US, compared  
104 to 11 human and 20 horse cases in 2005, the second largest outbreak since surveillance began  
105 in 2003 (**Fig. 1A**). The earliest human cases in the region were reported in July in Massachusetts.  
106 This is slightly earlier than the earliest Northeastern horse cases, which were reported in August  
107 across all Northeastern states other than Vermont (**Fig. 1B**). The last cases were in September  
108 in Connecticut and Massachusetts (human) and October in New York (horse).

109

110 Historically, there have been two foci of transmission in the Northeast: an eastern, coastal focus  
111 which encompasses most of Massachusetts and Connecticut, and a western focus in central  
112 New York towards Lake Ontario. While cases were recorded in both foci in 2019, the vast majority  
113 were in the eastern focus in Massachusetts and Connecticut, with no human cases and only 8  
114 horse cases reported in upstate New York (**Fig. 1C**). Cases primarily occurred in counties which  
115 had reported EEEV-positive human, horses, or mosquitoes before 2019. Further, the two  
116 counties which had not previously detected EEEV (Cattaraugus county in the southwest and  
117 Ontario county further north, both in central New York) are adjacent to counties which have (**Fig.**  
118 **1C**). Therefore, there was not much geographical expansion in 2019, and cases fit into previously  
119 established transmission foci (**Fig. 1C**).

120

121 Using data from routine mosquito surveillance, we calculated the maximum-likelihood estimates  
122 (MLEs) of mosquito infection rates (see Methods). We found the number of tested *Cs. melanura*  
123 mosquitoes and EEEV infection rates in Massachusetts (MLE infection rate 2019 = 3.62/1000;  
124 2003-2018 average = 1.21/1000, 95% CI: 0.67-1.75/1000) and Connecticut (MLE infection rate  
125 2019 = 3.67/1000; 2003-2018 average = 0.94/1000, 95% CI: 0.36-1.52/1000), two states that  
126 had a high increase in cases compared to recent years (**Fig. 1A**), were high in 2019 (**Fig. 1D**).  
127 New York, which reported human and/or horse cases in 2014 (12 horse and 2 human cases),  
128 2015 (3 human cases), and 2018 (3 horse cases), had more of a normal EEEV year in 2019 (8  
129 horse cases). Subsequently, we did not find a high number of tested *Cs. melanura* mosquitoes  
130 and EEEV infection rates in New York in 2019 (MLE infection rate 2019 = 3.30/1000; 2003-2018

131 average = 2.49/1000, 95% CI: 1.53-3.45/1000). In general, the mosquito infection rate patterns  
132 in Massachusetts and Connecticut are more similar to each other compared to New York, which  
133 is likely indicative of the two distinct geographical foci of transmission described above.

134

135

136

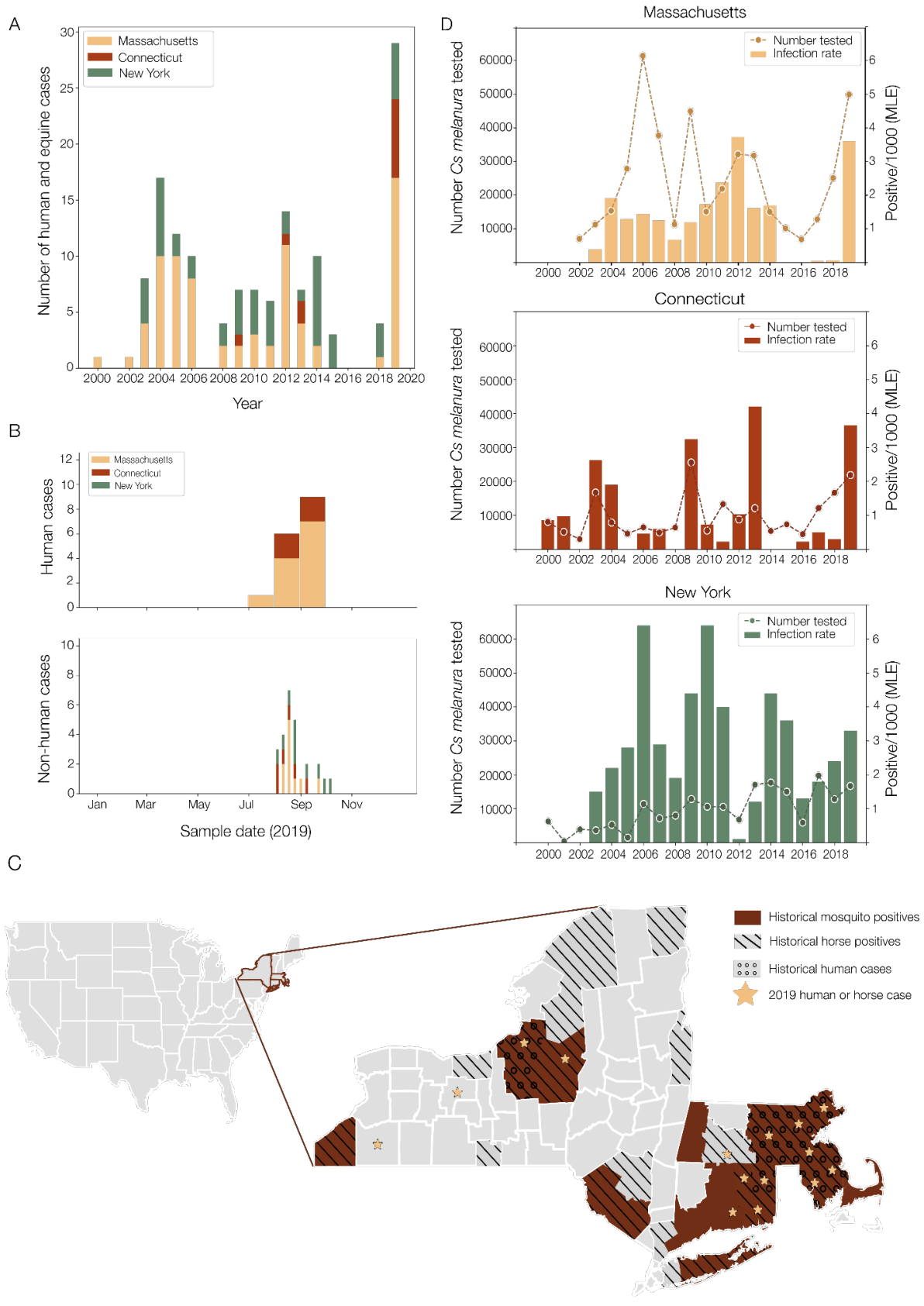
137

138

139

140

141



142  
143  
144  
145

146 **Figure 1 | Temporal and geographical characteristics of previous EEEV outbreaks after**  
147 **2000 and context of the 2019 outbreak in Massachusetts, Connecticut, and New York.**  
148 *A) Human and horse cases per year since 2000 in Massachusetts, Connecticut, and New York,*  
149 *colored by state. B) Human and veterinary cases in 2019 in these three states by sample date.*  
150 *Note that human cases are only available to the nearest month whereas veterinary cases have*  
151 *specific dates. C) Map showing geographical distribution of human and horse cases in 2019,*  
152 *shown as stars, relative to EEEV detections after 2000 in Massachusetts, Connecticut, and New*  
153 *York. Brown denotes counties where mosquito-positive pools have been sampled, hatched are*  
154 *non-human cases, and circles are human cases. D) EEEV infection rate of *Culiseta melanura**  
155 *mosquitoes and the number tested in Massachusetts, Connecticut, and New York by year from*  
156 *2000.*

157 The 2019 outbreak in the Northeast was caused by several recent virus  
158 introductions from Florida

159 To explore the underlying spatial dynamics of EEEV in the US, in particular to determine the timing  
160 and the source of viruses causing the 2019 outbreak, we sequenced 80 isolates of EEEV from  
161 Massachusetts, Connecticut, and New York. Adding our newly sequenced EEEV genomes to  
162 the existing publicly available whole genome sequences provided a sufficiently detailed dataset  
163 to perform a phylogeographic reconstruction on a state level. We found that EEEV transmission  
164 in Florida routinely seeds other locations across the eastern seaboard and was the source of  
165 multiple lineages causing the 2019 outbreak.

166  
167 Phylogeographic analysis to reconstruct virus spread requires sequence data from across spatial  
168 and temporal scales. Prior to the expansion of the nationwide arbovirus reporting system in 2003  
169 (Lindsey et al., 2012), genomic sequence data for EEEV were relatively sparse. There are,  
170 however, some sequences from across the full range of years, with the earliest sequence from  
171 the first recorded outbreak in 1933 (**Fig. 2A**). Within the Northeast, EEEV sequencing was most  
172 concentrated on samples from southeastern Massachusetts and upstate New York (**Fig. 2B**).  
173 Nationally, most sequences were from the Northeast and Florida, although only until 2014 for the  
174 latter. There was sporadic sequencing as far west as Texas, although there are no sequences  
175 from many of the intermediate states along the east coast (**Fig. 2C**).

176  
177 To provide additional geographical resolution within the Northeast and update the dataset to  
178 incorporate the 2019 outbreak, we sequenced an additional 80 isolates from across  
179 Massachusetts (n=17), Connecticut (n=38), and New York (n=25) (**Table S1**). They were  
180 primarily from *Cs. melanura* (n=63), although there were also isolates from other mosquito  
181 species: *Coquillettidia perturbans* (n=5), *Aedes vexans* (n=1), *Culex salinarius* (n=1), *Aedes*  
182 *canadensis* (n=1); as well as 9 sequences from horses and 1 from a turkey. These isolates were  
183 sampled from 2015-2019, with 48 from 2019, and were mostly from counties from which there  
184 were previously few or no sequences, particularly in Connecticut (**Fig. 2B**).

185  
186 Combining our new EEEV sequences with publicly available data (531 total sequences), we  
187 performed a joint phylogeographic and phylogenetic reconstruction (**Fig. 2D**). We estimate the  
188 time of origin of the US phylogeny to be 1923.2 (95% Highest Posterior Density (HPD): 1920.5-

189 1925.9), ten years before EEEV was first detected when it caused a large outbreak in horses in  
190 Virginia in 1933 (Giltner & Shahan, 1933). Further, we found a strong temporal signal for this  
191 dataset (**Fig. S1**) with an estimated evolutionary rate of  $1.86 \times 10^{-4}$  (95% HPD:  $1.77 \times 10^{-4}$ - $1.95 \times 10^{-4}$ )  
192 substitutions per site per year, in line with previous estimates (Tan et al., 2018).

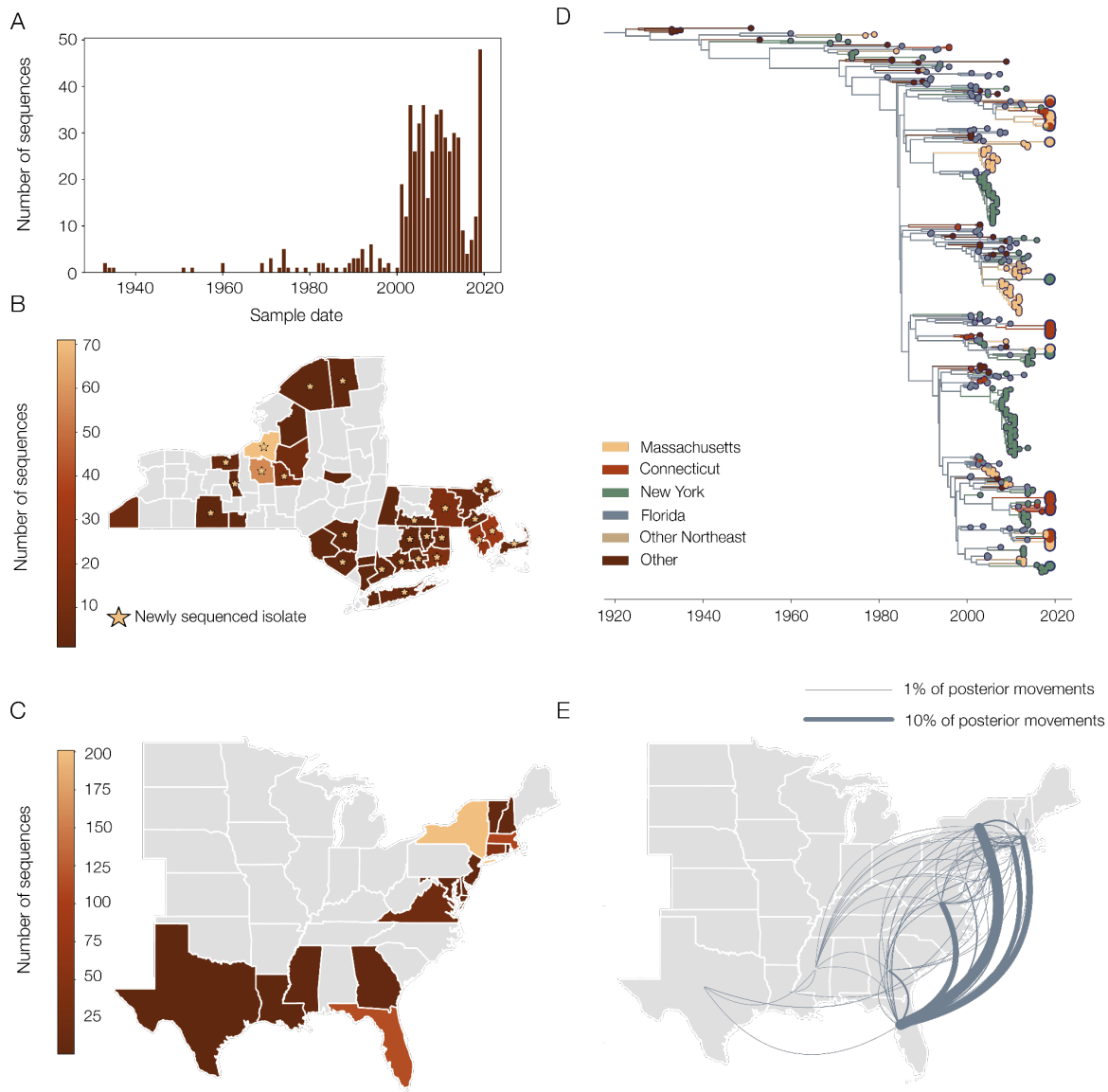
193

194 Supporting previous results of phylogeographic analysis (Tan et al., 2018), we found that Florida  
195 forms the backbone of the phylogeny. In other words, EEEV transmission in Florida acts as a  
196 source of virus introductions into other states (**Fig. 2D**). When investigating the movement  
197 between states across the whole posterior distribution of the phylogeographic reconstruction,  
198 we found that 86.7% of EEEV movements start in Florida and end in every other state in the  
199 dataset (**Fig. 2E**). It is worth noting that there are not many sequences from states outside of  
200 Florida and the Northeast, and so movements involving states such as Texas, Georgia, and  
201 Virginia should be interpreted with caution, and there are likely more regional dynamics which we  
202 could not uncover with this dataset.

203

204 Using our new EEEV sequences, we found that the 2019 outbreak in the Northeast involved  
205 several independent virus introductions (**Fig. 2D**). Following the national trends, we infer that  
206 each of these introductions originated in Florida and were not related to other previous EEEV  
207 clusters sequenced from the Northeast. Therefore, the 2019 outbreak in the Northeast consisted  
208 of multiple EEEV lineages most likely introduced from the reservoir population in Florida, as  
209 opposed to long-term regional persistence or a single introduction with explosive growth.

210



211  
 212 **Figure 2 | Phylogeographic reconstruction of EEEV from 2019 and prior outbreaks**  
 213 A) Number of EEEV sequences in the dataset over time by year of sampling. Note that the  
 214 nationwide reporting of surveillance data began in 2003. B) Location of EEEV sequences in the  
 215 dataset from Massachusetts, New York, and Connecticut to the county level. Stars indicate the  
 216 location of EEEV samples which were newly sequenced for this study. C) Location of all EEEV  
 217 sequences in the study to state level. D) Time-resolved phylogeny colored by location of nodes  
 218 from the discrete phylogeographic analysis. States in the Northeast are colored separately, but  
 219 non-Northeast and non-Florida states are grouped together. Larger tips represent EEEV  
 220 sequences from 2019. E) Movement of virus from the full posterior of the discrete  
 221 phylogeographic analysis. Direction is anti-clockwise, and width of lines corresponds to  
 222 frequency of movement across the posterior. Movements that make up fewer than 1% of the  
 223 total posterior have been filtered out.  
 224



225 EEEV lineages do not typically persist longer than a few years in the  
226 Northeast

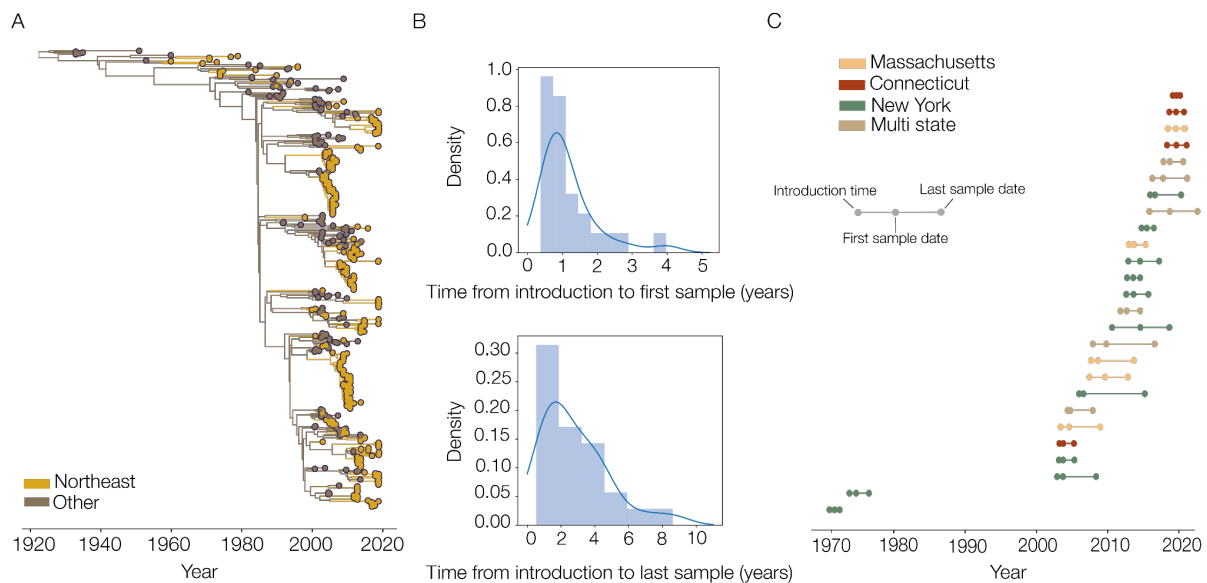
227 Having found that EEEV circulation in the Northeast, including the 2019 outbreak, is mostly driven  
228 by repeated introductions from Florida, we explored the maintenance of these lineages once  
229 they become established in the region. We found that viral lineages on average only persist for  
230 less than three years and are generally detected a year after introduction.

231  
232 Due to the lack of sequences from Florida in general, but especially after 2014, the clades  
233 estimated by our discrete phylogeographic analysis are likely not precise estimates of the timings  
234 of EEEV introductions into the Northeast (**Fig. 2D**). In particular, we expect that we have  
235 estimated larger and longer-lived clusters after 2014 than there are in reality, as they have not  
236 been broken up by EEEV sequences from Florida. To account for this, we calculated the average  
237 branch length in the maximum clade credibility (MCC) tree in Northeastern clusters of more than  
238 three sequences with more than half of their sequences from 2014 or earlier, which was 1.28  
239 years (standard deviation = 1.5 years). We then subdivided any clades that contained branches  
240 longer than the mean plus twice the standard deviation (4.3 years). This also included clusters  
241 from before 2014, indicating that sampling bias specifically involving Florida is an issue  
242 throughout the dataset. This procedure resulted in splitting the 61 EEEV introductions inferred  
243 from our discrete trait analysis into 75 separate introductions into the Northeast (**Fig. 3A**).

244  
245 Of these 75 introductions, we found 26 consisting of three or more sequences. On average,  
246 these took 1.17 (95% HPD: 0.46-3.27) years to be first detected in the genomic dataset, and  
247 circulated in the Northeast for an average of 2.8 (95% HPD: 0.65-7.5) years after being  
248 introduced (**Fig. 3B**). This supports previous analysis of the region showing limited multi-year  
249 maintenance in the region (Oliver et al., 2020; Young et al., 2008). The groups of lineages can be  
250 seen to roughly follow three wave patterns in the 2000s, between 2010 and 2014, and then 2014  
251 to 2019 (**Fig. 3C**), possibly corresponding to infection and immunity patterns within the bird  
252 population in the region.

253  
254 Of the 26 larger Northeastern introductions, 8 included samples from the 2019 outbreak, and  
255 half of these were solely composed of 2019 samples (**Fig. 3C**). We estimate that the average  
256 introduction time of these 2019 clusters was in mid 2017 (95% HPD: 2015-11-14 to 2019-01-  
257 15), and the earliest introduction was late 2015. Therefore, the EEEV lineages sequenced from  
258 2019 were introduced no more than 4 years before the outbreak and after the most recent  
259 Northeast EEEV outbreak in 2014.

260  
261  
262



263  
264  
265  
266  
267  
268  
269  
270  
271

### Figure 3 | Detection and persistence of EEEV lineages in the Northeast

A) Time-scaled phylogeny estimating EEEV introductions into the Northeast, taking sporadic sampling into account by splitting up clusters with very long internal branches (see Methods). B) Distributions of time from introduction to first (top) and last sample (bottom) for Northeastern clusters of more than three EEEV sequences. C) Time of first node in the Northeast, first sample, and last sample for each cluster with more than three EEEV sequences, colored by state

### 272 Multi-state spread of EEEV originating in Massachusetts

273 Considering that EEEV lineages have a short lifespan in the Northeast (**Fig. 3**), we subsequently  
274 investigated the extent to which they could spread within the region during that time. We found  
275 that most of the Northeast EEEV clusters were detected within a single state, suggesting that  
276 inter-state regional spread is rare. Where we did detect between-state spread, however,  
277 Massachusetts appeared to be an important regional source.

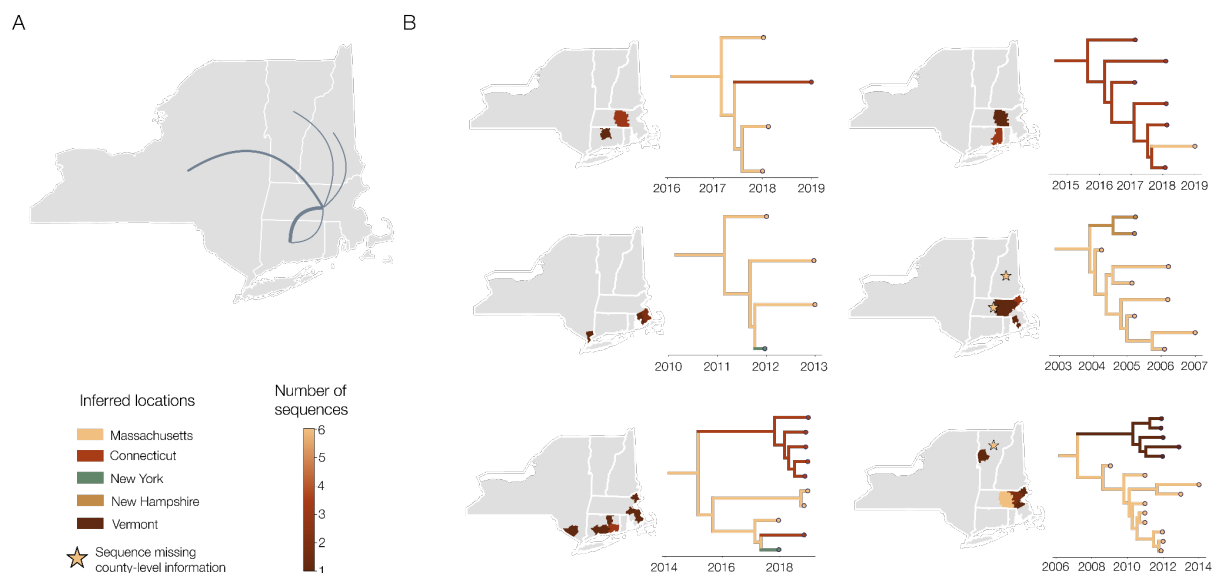
278  
279 Of the 26 Northeast clusters with three or more EEEV sequences (**Fig. 3C**), most (n=19) were  
280 from a single state, and only a single cluster was found in three states despite the density of  
281 sequencing in the region. Of the 7 clusters with sequences from multiple states, we found that  
282 most viral movements were inferred to be from Massachusetts to other states (7 out of 8  
283 movements in the MCC tree, **Fig. 4A**) and the origin of 5 out of 6 subtrees was in Massachusetts  
284 (**Fig. 4B**). The subtree that does not have an origin in Massachusetts circulated mostly in  
285 Connecticut but does contain a sequence from Massachusetts. Importantly, multi-state clusters  
286 were not necessarily sampled in neighboring counties (**Fig. 4B**), indicating true spread within the  
287 region, possibly by infected birds.

288  
289 It is worth noting that only two of these multi-state clusters contain sequences from New York,  
290 despite the majority of EEEV sequences being from that state, and its size and central location in  
291 the region. These New York sequences were from the southeast part of the state (**Fig. 4B**), closer  
292 to the prominent Massachusetts-Connecticut focus rather than the transmission clusters in

293 upstate New York. Therefore, while limited spread is happening in the region, it primarily only  
294 involves the eastern part of the region where smaller states are close together.

295  
296 While none of the multi-state clusters were solely sampled in 2019, we found that half of them  
297 contain EEEV sequences sampled in 2019 (**Fig. 4B**). Therefore, the 2019 outbreak appears to  
298 have a wider within-cluster geographical spread than previous ones (e.g., 2012-2014 sequences  
299 are only found in 2 subtrees; **Fig. 4B**). The spread in 2019 was possibly driven by the high  
300 numbers of infected mosquitoes (**Fig. 1D**) leading to more infected birds and more opportunities  
301 for cross-state movements. We emphasize, however, that cases were mostly reported in  
302 counties that had previous EEEV cases (**Fig. 1C**), and so this was strictly a within-cluster effect  
303 and the outbreak as a whole does not show geographical expansion.

304  
305



306  
307

#### 308 **Figure 4 | Phylogeographic spread of EEEV within the Northeast**

309 *A) Between-state movements from the maximum clade credibility (MCC) tree of the*  
310 *phylogeographic analysis within the six Northeastern EEEV clusters which have between-state*  
311 *movement. Width of lines relates to the number of movements, and direction is anti-clockwise.*  
312 *B) Maps showing sampling location to the county level for each EEEV cluster with multiple states*  
313 *where available. Stars indicate where there are sequences with information only at the state level*  
314 *( $n=2$  in New Hampshire,  $n=1$  in Massachusetts and  $n=3$  in Vermont). Corresponding subtrees*  
315 *are shown to the right of each map, colored by inferred location to the state level.*

316 Associations of virus, human, and bird factors with EEEV outbreaks were  
317 not found

318 The general epidemiology underlying the 2019 EEEV outbreak appears to be similar to previous  
319 outbreaks, with short-lived introductions from Florida driving transmission in the Northeast (**Fig.**  
320 **2-3**), and infections occurring in similar locations (**Fig. 1**). Therefore the change in the lead-up to  
321 the 2019 outbreak, which underlies the increase in cases, may be due to mosquito population,  
322 virus evolution, human behavior, or bird populations, and there are multiple possible factors that

323 are not mutually exclusive. Determining which factors were important for the 2019 outbreak may  
324 help us to understand what drives EEEV outbreaks in general. First, we ruled out that the 2019  
325 outbreak was driven by virus evolution. Then, we examined case demographics and aspects of  
326 human behavior, and we did not find any associations between the age, sex, or timing of human  
327 cases to outbreak years, nor were humans spending more time outside in 2019 compared to  
328 2018. Finally, we did not find strong evidence for the population sizes of key bird species as a  
329 correlate of cases or mosquito infection rates, though we are lacking important data on bird  
330 immunity to EEEV. Together, with the data available to us, virus genomics, case demographics,  
331 and bird population sizes fail to explain or predict EEEV outbreaks.  
332

### 333 Intrinsic viral factors and effective population size

334 A major concern when any pathogen leads to a large increase in case counts is that it has evolved  
335 to have a higher virulence or transmissibility. In the case of EEEV, increased transmissibility would  
336 refer to a higher infection success in either mosquitoes or birds. Increased virulence would lead  
337 to increased case ascertainment, as diagnosis occurs on hospitalized individuals and most  
338 human infections are probably asymptomatic (Morens et al., 2019). We would expect to see a  
339 phylogenetic signal for any intrinsic variability in transmissibility or virulence.

340  
341 We searched the alignment for any nucleotide substitutions, relative to a reference sequence  
342 from 2005, which are common to some or all of the 2019 EEEV sequences. There were 17  
343 substitutions across the genome shared by all of the 2019 sequences, a further two that more  
344 than 90% of them shared, and an additional two that 75% of them shared. These 21  
345 substitutions, however, were found across the whole phylogeny, in at least 93% of non-2019  
346 EEEV sequences. We found no shared substitutions unique to more than 75% the sequences  
347 sampled in 2019. This was expected given that these sequences are spread across the  
348 phylogeny (**Fig. 2D**), as there would have to be extremely strong positive selection leading to  
349 multiple instances of convergent evolution for unique shared substitutions to be possible.

350  
351 Past population size of viruses can be useful in inferring underlying transmission dynamics. In the  
352 case of EEEV, this can be done using a skygrid model (Gill et al., 2012), which we used to  
353 estimate changes in virus effective population size (in birds and mosquitoes) over time across the  
354 whole of the US. We found that there was a steady increase in effective population size in the  
355 1990s, which then plateaued and decreased more recently in the late 2010s (**Fig. S2A**).

356  
357 To explore if this could be formally connected to EEEV infections in humans and horses, we used  
358 an extension of the skygrid model to incorporate case data as covariates (Gill et al., 2016). We  
359 hypothesized that increased transmission in the mosquito-bird cycle could lead to more  
360 infections and reported cases in humans and horses. We tested 9 different covariates  
361 independently using data from 2003 onwards: all human and horse cases, only horse cases and  
362 only human cases each with no lag, a 1 year lag or a 2 year lag (**Fig. S2B**). We did not find any  
363 significant associations between the viral population size and any of these covariates. There is,  
364 therefore, no simple association between national circulation and reported cases, meaning that  
365 regional dynamics are likely more important for regional cases. More recent sequencing data in  
366 general, particularly from the Northeast to allow the inference of population dynamics around the

367 outbreak in 2019, could confirm this.  
368

### 369 Case demographics and human behavior

370 Human behavior could be one explanation for the increase in human EEEV cases in 2019, which  
371 could be revealed by analyzing the case demographics. Since 2003, we found that the cases  
372 have been predominantly male (including in 2019, where 62% of the cases were male), and this  
373 proportion was not found to vary across years (Fisher's exact test,  $p=0.838$ , **Fig. S3A**). In  
374 comparison, the proportion of cases in different age categories were found to vary across years  
375 ( $p=0.047$ , **Fig. S3B**). To determine if this variation was meaningful, we divided years with cases  
376 since 2003 into large outbreak years (5 or more human cases) or small outbreak years. When we  
377 compared these two groups, we did not find any evidence of a difference in age distribution  
378 (Fisher's exact test,  $p=0.322$ ). Finally, we compared the timing of human and horse cases in the  
379 Northeast. The median month of symptom onset in all years was either August or September,  
380 with earliest cases in July and latest in October. We tested whether the median month of infection  
381 correlates with the number of cases and found that it does not (Pearson's correlation,  $p=0.777$ ,  
382 **Fig. S3C-D**).

383  
384 To further explore whether human behavior may have changed in 2019, leading to a higher EEEV  
385 exposure rate, we compared the length of time individuals spent outside in relevant counties in  
386 Massachusetts, Connecticut, and New York (**Fig S3E**). We performed a paired t-test on the  
387 indoor activity seasonality metric for each of these counties in June, July, August, and September  
388 between 2018 and 2019 (data only available for these years) using data from (Susswein et al.,  
389 2022). We found no evidence of a difference in indoor-ness, and by extension the reciprocal  
390 outdoor-ness, between 2018 and 2019 in counties that had ever had a human case of EEEV  
391 ( $p=0.56$ ) or only those which had a case in 2019 ( $p=0.72$ ). Therefore, human behavior, the timing  
392 of cases, and case demographics do not appear to be the primary drivers of EEEV outbreaks  
393 based on our analysis of the available data. Although, understanding the behaviors and activities  
394 linked to human exposure is still an important area of research.  
395

### 396 Bird populations

397 Bird populations likely have a complicated role in EEEV outbreaks. While the presence of  
398 competent bird species is a necessity for EEEV transmission, abnormally large bird populations  
399 may dilute the virus, decreasing the likelihood of mosquito infection during blood feeding (Kramer  
400 & Ciota, 2015). Moreover, when sufficient opportunities exist for mosquitoes to feed on birds, it  
401 may limit the pursuit of mosquitoes to feed on alternate sources, such as humans and horses.  
402 Finally, bird population immunity is likely a key component of the potential for EEEV perpetuation,  
403 but studies that routinely collect bird serology data are difficult and not often conducted.  
404

405 While data on birds are limited, we examined the effect of the (1) absolute abundance and (2)  
406 proportion relative to all species of 8 bird species for which *Cs. melanura* has a high blood feeding  
407 preference (Armstrong & Andreadis, 2022; Molaei et al., 2016) on (A) human and horse cases  
408 and (B) mosquito EEEV infection rate using a Poisson regression. We also compared these  
409 metrics for 8 bird species for which a low feeding preference was estimated (Molaei et al., 2016)

410 as a control (**Fig. S4**). In New York, we found very few significant relationships between the  
411 abundance or proportion of any of the high-preference bird species to case counts or infection  
412 rates. There are, however, several significant relationships in the low-preference bird species,  
413 suggesting that this comparison is not valid for New York. Further, the comparisons to mosquito  
414 infection rate almost all have high p-values regardless of state, bird species, or metric. In  
415 Massachusetts and Connecticut, we found several positive relationships between high-  
416 preference bird species and case counts, notably the common grackle (abundance only;  
417  $p < 0.001$  for both states), chipping sparrow (proportion only;  $p < 0.01$  for both states), tufted  
418 titmouse (proportion only;  $p < 0.01$  for both states), and warbling vireo (abundance and proportion;  
419  $p < 0.01$  for both states). There are, however, some low-preference birds with significant  
420 relationships (although fewer), meaning that the relationship between specific bird presence and  
421 EEEV transmission/cases is complicated and not easily solved here.  
422

423 EEEV outbreaks are primarily driven by the infection dynamics within  
424 mosquito populations

425 After ruling out or not finding associations with other factors, we subsequently focused on  
426 mosquitoes where we have detailed temporal data. After the CDC expanded the arboviral  
427 surveillance reporting system (ArboNET) and federal funding to the states in the early 2000's  
428 (Hadler et al., 2015), many states began to routinely trap and test mosquitoes for multiple viruses,  
429 including EEEV and West Nile virus. This provides a glimpse into the transmission cycle of the  
430 virus, and allows us to explore if mosquito populations or infection rates are associated with  
431 outbreaks. We used mosquito surveillance data from all three states, and detailed data from  
432 Connecticut and Massachusetts, to develop a transmission model. Overall, we found that a high  
433 abundance of *Cs. melanura*, connected to environmental factors, is necessary, but not sufficient,  
434 for an outbreak of EEEV in humans and horses. When infection rates are included, however,  
435 these two metrics together are predictive of overall human and horse cases. We then applied this  
436 model to obtain predictions of total human and horse cases based on early season  
437 measurements.

438  
439 To begin with, we found when there were more EEEV-positive *Cs. melanura* pools (i.e.  
440 mosquitoes were tested in groups of up to 60 from the same trap and date), there were more  
441 human and horse cases in Massachusetts, Connecticut, and New York. (**Fig. 5A**). Further, in  
442 2019, there were more positive mosquito pools detected in Massachusetts and Connecticut than  
443 previously recorded. In comparison, the number of EEEV-positive mosquito pools detected in  
444 New York was more within the expected range of historic values, which aligns with the normal  
445 number of case counts in 2019 for that state. It must be noted that Massachusetts and  
446 Connecticut practice reactive sampling, and so more mosquito traps are set when EEEV-positive  
447 mosquitoes are found. To account for this, we also found the same strong correlation between  
448 the *Cs. melanura* EEEV infection rates (MLE per 1000 mosquitoes, see methods) and case  
449 counts (**Fig. S5A-B**). Therefore, our data suggest that human and horse cases are primarily  
450 driven by the abundance of infected mosquitoes more so than behaviors that increase exposure  
451 to mosquitoes.  
452

453 To further explore the within-season EEEV dynamics, we focused on Connecticut and  
454 Massachusetts, for which we have more detailed mosquito data. We plotted the monthly *Cs.*  
455 *melanura* abundance (mosquitoes/trap/day), EEEV infection rate, and vector index (abundance  
456 multiplied by the infection rate to estimate the relative number of infected mosquitoes (Fauver et  
457 al., 2016; Nasci et al., n.d.) for each state, highlighting 2019 (**Fig. 5B**). While there were  
458 considerable yearly variations in both states, in general *Cs. melanura* abundance rapidly  
459 increased in early summer and tended to peak by June to July. The EEEV infection rates lagged  
460 behind, peaking in August or September, with infected mosquitoes often being detected in  
461 Massachusetts earlier than Connecticut. Vector index, therefore, tended to rise earlier (July) and  
462 persist longer in Massachusetts than Connecticut, though it often peaks in August in both states.  
463 In 2019, we found that both the *Cs. melanura* abundance and EEEV infection rates were high in  
464 both states, leading to the highest vector index values in the 17-year dataset (**Fig. 5B**). Thus the  
465 2019 EEEV outbreak was driven by environmental conditions that supported larger *Cs. melanura*  
466 populations and transmission dynamics that led to high infection rates.

467  
468 We next investigated whether environmental data could help explain the high abundances or  
469 EEEV infection rates. To explore this, we used index P (Obolski et al., 2019), a modeled estimate  
470 of reproduction number of mosquito-borne viruses based on climate factors, including  
471 temperature and humidity, by including prior estimates for *Cs. melanura*-borne transmission of  
472 EEEV (**Table S2**). Our assumption was that index P would best model the vector index. While  
473 our index P estimates usually peaked in July to August, similar to the vector index, we found that  
474 it actually correlated most closely with the *Cs. melanura* abundance dynamics (**Fig. S5C**;  
475 correlation coefficient of 0.57, p-value < 0.0001 for Connecticut, 0.24 and p<0.05 for  
476 Massachusetts). Furthermore, our modeled index P in 2019 was relatively average across the 17  
477 years in both states (rank 9/17 for Connecticut, 7/17 for Massachusetts). Thus, for EEEV in the  
478 Northeast, index P is useful for describing the seasonal dynamics of transmission, but the  
479 temperature and humidity variables alone are not capable of describing which years will have a  
480 high number of EEEV-infected mosquitoes.

481  
482 To interrogate further how different factors combine to explain cases, we built a negative binomial  
483 regression model that uses index P, vector index, state, month of first EEEV detection, the year,  
484 and the month to explore counts of total human and horse cases (**Fig. 5C-D, Table S3**; results  
485 from different variations of the model are shown in **Fig. S6**). Because of the yearly variation in  
486 EEEV cases (**Fig. 5D**, see “Actual” cases), we found that year and being in the month of August  
487 or September lead to no significant changes in risk of cases (**Fig 5C**). Vector index (2.07, 95%  
488 CI: 1.5-3.0), the month of first detection compared to June (3.17, 95% CI: 1.34-7.92), index P  
489 with a 1 month lag (2.31, 95% CI: 1.19-4.65), and being in the month of October (4.18, 95% CI:  
490 1.39-13.55), however, all lead to a significant increase in the risk of human and horse cases (**Fig.**  
491 **5C**). Further, being in Massachusetts as compared to Connecticut had the largest increase in  
492 risk (11.95, 95% CI: 3.88-43.70). When used to fit cases each year, our modeled estimates  
493 tracked closely with observed cases, where cases with the model being off on average by 0.47  
494 cases per observation (Mean Absolute Error (MAE) = 0.47) and had strong explanatory power  
495 (Nagelkerke  $R^2 = 0.80$ ; **Fig. 5D**). Importantly, vector index was an important driver of the outbreak  
496 in 2019, where elevated levels correspond well with high case counts.

497

498 Next, we applied this model to examine if there were early season predictors which might indicate  
499 an increased risk of horse and human cases later in the season (**Fig. 5E**). While index P varies  
500 from year to year and is connected to mosquito abundance, it does not correlate well with vector  
501 index as it cannot predict when EEEV is present in the Northeast (**Fig. S5C**). We therefore held  
502 index P constant and simulated vector index across a range of values to predict the total human  
503 and horse cases for the year, and to explore its relationship with the month of first detection. We  
504 found that our model underpredicts cases when the vector index is low, as the real data points  
505 fall outside of the prediction intervals (**Fig. 5E**), indicating a potential lag in identification of EEEV  
506 cases. This could be due in part to stochastic chance in surveillance programs missing EEEV  
507 with low levels of virus circulation. As the vector index increases, however, our predicted values  
508 become closer to the observed data points.

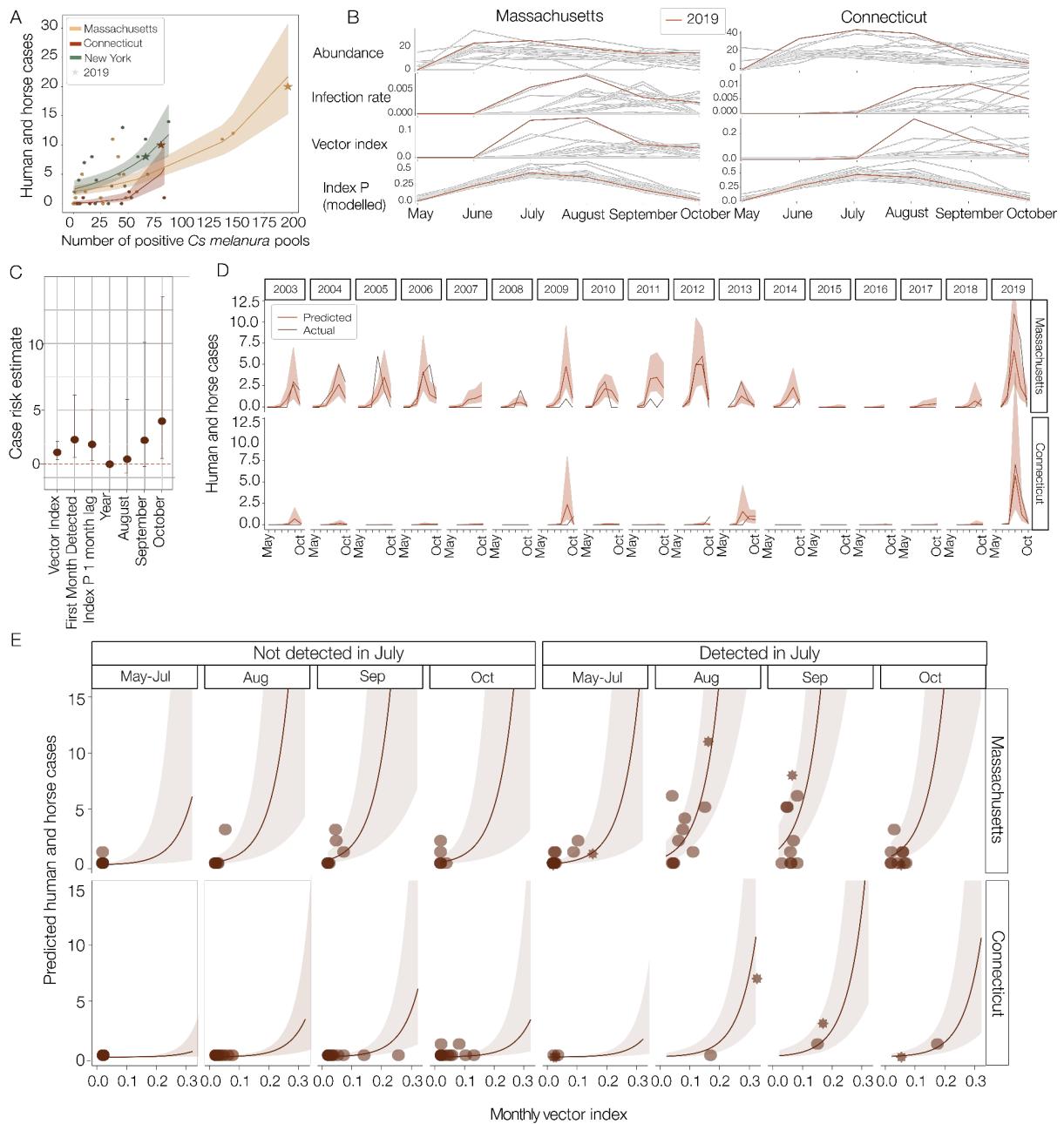
509  
510 The timing of the first detection of EEEV in mosquitoes is also important, and of particular note is  
511 that years with high case counts always first detected EEEV in July (**Fig. 5E**). This makes intuitive  
512 sense as the virus has more time to circulate to high levels in birds while there are sufficient  
513 mosquitoes present for it to transmit to humans. While the interaction between first EEEV  
514 detection in July and the vector index is not significant, this could be in part due to no cases  
515 being detected in years when the vector index was high but EEEV was first detected after July.  
516 The impact of the timing of the vector index is clear however, with a 220% increase in risk of  
517 cases when EEEV is detected in July (**Fig. 5E**). This is reflected in the empirical data, as the  
518 maximum number of cases in years when EEEV is detected after July is only 5.

519  
520 Our negative-binomial regression model could be used as a tool for mosquito surveillance  
521 programs to estimate the risk of cases, which typically are reported in August-October, based  
522 on the timing of first detection and the vector index. For example, if EEEV is detected in  
523 mosquitoes in July in Massachusetts, and the vector index in August is 0.15 (similar to the value  
524 measured in 2019), we predict 7.0 human and horse cases in total that year (95% CI: 2.9-16.6),  
525 compared to 2.2 cases (95% CI: 0.6-7.9) if EEEV is detected after July during a typical summer  
526 season. In Connecticut, the same situation would correspond to an expected 0.6 cases (95%  
527 CI: 0.2-1.8) for July detection and 0.2 cases (95% CI: 0.1 to 0.6) for a later detection. With a  
528 more extreme August vector index of 0.30 in Massachusetts (higher than has ever been  
529 measured), our model predicts 89.2 cases (95% CI: 17.0-468.5) if EEEV is detected in July versus  
530 28.2 cases (95% CI: 3.4-231.3) cases being detected later. While these predictions have wide  
531 confidence intervals, they provide a relative estimate of the possible extent of the cases in a given  
532 year, and therefore an indication of how much effort should be put into various mosquito control  
533 measures.

534  
535 Finally, while all of our previous results focused on *Cs. melanura*, *Coquillettidia perturbans* is also  
536 thought to be an important bridge vector for infections in horses and humans (Armstrong &  
537 Andreadis, 2022; Sherwood et al., 2020). This species is more abundant on average in  
538 Connecticut than *Cs. melanura* (**Fig. S5D**), with populations peaking in late June. However, only  
539 9 *Cq. perturbans* pools were positive for EEEV from 2003-2019 in the state, or 1.9% of all positive  
540 pools; and it was only the 8th most commonly positive mosquito species for EEEV. In 2019, there  
541 were six EEEV positive pools of *Cq. perturbans* sampled in Connecticut, making it the third most  
542 commonly sampled positive mosquito species for the year, although it still only made up 5% of



543 the positive mosquito pools. This does vary between states, however, as *Cq. perturbans* made  
 544 up a third of positive pools in Massachusetts in 2019 (Armstrong & Andreadis, 2022). Therefore,  
 545 *Cq. perturbans* may still have a role as a bridge vector, but the transmission dynamics of EEEV  
 546 in these states is primarily driven by *Cs. melanura*.  
 547  
 548



549  
 550

551 **Figure 5 | Model predictions of EEEV cases using mosquito infection estimates**

552 A) Poisson regression by state of the number of EEEV positive *Cs. melanura* pools sampled and  
 553 human and horse cases. B) Monthly trends across the year of mosquito abundance (*Cs.*  
 554 *melanura* collected per trap per day), infection rate (MLE per 1000 *Cs. melanura*), vector index  
 555 (abundance multiplied by the infection rate) and index P (modeled estimate of reproduction  
 556 number), for Connecticut and Massachusetts. 2019 values are highlighted in red. C) Results of a  
 557 negative binomial regression model for both Connecticut and Massachusetts combined of case

558 *risk predicted by vector index, first month of EEEV detection, index P with a 1 month lag, year,*  
559 *and risk compared to the month of May-July of cases in August, September, and October. The*  
560 *covariate for state was removed for scale, as its estimate is 11.95 (95% CI: 3.88 to 43.70)*  
561 *with Connecticut being the reference group. D) Actual and predicted case counts for each year*  
562 *based on the modeled results presented in C. Shaded areas denote error in the predicted values.*  
563 *E) Simulations using the previously described model. Predicted cases and intervals utilized a*  
564 *range of plausible vector index values and whether EEEV was detected in July or not, with index*  
565 *P held constant using the average value derived from the years studied for each month; and split*  
566 *by state. Shaded areas represent 95% confidence intervals and points are real data points from*  
567 *the EEEV mosquito surveillance program discussed elsewhere. Stars indicate values from 2019*  
568 *and darker points indicate the actual modeled value of Index P.*  
569

## 570 Discussion

571 The 2019 EEEV outbreak in the Northeast US was exceptional in that it had the highest number  
572 of human cases recorded in more than 50 years, along with dense *Cs. melanura* vector  
573 populations with high virus infection rates (**Fig. 1 & 5**). We found that other components of 2019  
574 were similar to outbreaks of years past. We showed that EEEV is not persistently maintained in  
575 the Northeast, with virus lineages typically going extinct within the region within a few years (**Fig.**  
576 **3**). Instead, the virus is frequently reintroduced into the region from Florida where EEEV is  
577 endemic (**Fig. 2**). Thus, EEEV outbreaks in the Northeast are limited by stochastic and  
578 deterministic factors that support new virus introductions, and to habitats that are suitable for the  
579 mosquito vectors. We identified two primary transmission foci in the Northeast: (1) an eastern,  
580 coastal focus which encompasses most of Massachusetts and Connecticut, and (2) a western  
581 focus in central New York towards Lake Ontario (**Fig. 1**). The 2019 outbreak primarily occurred  
582 in the former, especially in regards to human cases. We found that human and horse cases are  
583 associated with a high vector index (large number of EEEV infected *Cs. melanura* mosquitoes),  
584 and we constructed a model using environmental and mosquito surveillance data that could  
585 estimate cases (**Fig. 5**). Finally, we found that a high early season vector index can be used by  
586 surveillance systems to predict human EEEV risk and direct control efforts (**Fig. 5E**). It remains  
587 unclear as to what causes high EEEV infection rates in mosquitoes, and therefore what  
588 contributed to the exceptionally high rate in 2019.

589  
590 We conducted a phylodynamic analysis of our 80 newly-generated EEEV genome sequences  
591 combined with historical samples to explore the dynamics of EEEV in the Northeast of the US,  
592 with a focus on 2019 (**Fig. 2**). As both EEEV isolates and extracted RNA are designated as Select  
593 Agents in the US, it is extremely difficult to obtain clearance to store, transport, and in this case,  
594 sequence them. Thus our genomic dataset is critical to support future EEEV research. Using  
595 these additional sequences, we confirmed earlier studies which identified Florida as the source  
596 of EEEV introductions into the Northeast and other parts of the US (Tan et al., 2018) (**Fig. 2**). The  
597 documented endemic transmission in Florida and the limited number of EEEV cases reported  
598 outside of the US add further support to the hypothesis of Florida as a source for the Northeast.  
599 EEEV sequences from outside of the US, however, are limited in number (n=3) and we did not  
600 include these data in our analysis. Therefore, we cannot rule out a non-US external source of

601 introductions into the Northeast. Further sequencing from other countries, enabling a broader,  
602 regional analysis of EEEV would allow the investigation of the international dynamics of EEEV and  
603 the identification of at-risk regions.

604  
605

606 We found that while interstate spread of EEEV within the Northeast is not common,  
607 Massachusetts appears to be an important focus for when it does happen (**Fig. 4**). This may be  
608 in part due to bird migration routes, more consistent EEEV activity in Massachusetts compared  
609 to the surrounding states (**Fig. 1 & 5**), and/or that EEEV infection in mosquitoes in Massachusetts  
610 increases earlier in the year than it does in Connecticut (**Fig. 5**). Thus, controlling EEEV  
611 transmission in Massachusetts may help to alleviate some EEEV transmission in other states.

612

613 In discussing the movement of EEEV across the US, it is important to note that there is significant  
614 sampling bias in our dataset, in terms of both time and location of the sequences. Most of the  
615 sequences are from the Northeast US, and there are no sequences from Florida after 2014. While  
616 we have attempted to mitigate the latter point by splitting Northeastern clusters with long  
617 branches (see Methods), geographical heterogeneity can lead to overconfidence in the transition  
618 times (Layan et al., 2023). Further, the lack of EEEV sequences from other East Coast states may  
619 lead to an underestimate of the importance of those states in the spread of EEEV from Florida,  
620 and we are unable to examine the introduction or transmission dynamics in any other region of  
621 the country. After Florida, the Northeast has the most reported human and horse cases in the  
622 US. It therefore has an outsized importance for understanding the dynamics of EEEV. Obtaining  
623 data from other EEEV outbreaks (e.g. Michigan) would provide another opportunity to examine  
624 the outbreak dynamics. The sequencing of EEEV samples prior to 2003, a wider geographical  
625 range of samples in all time periods, and from Florida after 2014 would greatly add to the reliability  
626 and resolution of any phylodynamic study of EEEV.

627

628 While some characteristics of the EEEV outbreak in 2019, such as human case traits and climate  
629 factors, were within expected ranges, it is clear that many mosquito factors were unusually high  
630 (**Fig. 5**). For example, both Massachusetts and Connecticut had a very high abundance of *Cs.*  
631 *melanura* and the vector index for the latter state was the highest ever recorded. While EEEV has  
632 a complex ecology, by using data from detailed mosquito surveillance programs, we were able  
633 to find strong connections between mosquito infection rate, abundance (connected to climate  
634 factors), and human and horse cases. We then developed and applied a negative binomial  
635 regression model to utilize early season values of mosquito-specific predictors (most notably the  
636 month of first detection and vector index) to provide early estimates of overall case counts for  
637 that year. This important development will provide departments of health with estimates to help  
638 direct control strategies, and enable more effective communication of risk to the public.

639

640 Our models and estimates have several limitations. First, our negative binomial regression model  
641 to predict cases is likely more informative for Massachusetts than Connecticut, as there are more  
642 cases in Massachusetts to inform it. As such, the Connecticut predictions are heavily influenced  
643 by Massachusetts case counts, and given the highly localized nature of EEEV to forested wetland  
644 habitats, the model may not accurately represent conditions in Connecticut. In addition, our  
645 model was restricted to a monthly time scale to match monthly case data. Given the short

646 duration of the EEEV transmission season, certain dynamics may be missed on the weekly or  
647 even daily timescale. There was not enough case data within each month to explore impacts of  
648 detection on a month by month basis, making the month detection piece of the analysis less  
649 specific. While vector index is a direct risk estimate of EEEV cases, index P had a greater effect  
650 estimate in our model. This could be due to our model not fully accounting for the inherent  
651 seasonality of EEEV transmission that overlaps with high values of index P or that the true  
652 value of index P, which is modeled with a certain degree of uncertainty, lies a significant  
653 distance outside the determined estimate. Due to lack of information on EEEV in *Cs.*  
654 *melanura*, index P is estimated using the transmission probability, bird incubation period, bird  
655 infectious period and extrinsic incubation period derived from previous West Nile virus research  
656 of *Culex pipiens* species (Lourenço et al., 2020). This lack of data for EEEV priors may be  
657 confounding the predictive power of this parameter. Finally, veterinary cases are reported  
658 voluntarily and the true case counts include interventions undertaken in those years. These  
659 interventions include a highly recommended horse vaccination and mosquito spraying: the former  
660 will cloud the association between mosquito factors and case counts as exposures may not  
661 convert to cases. Both lead to an underestimate of true, unaltered case counts as the model is  
662 fitted to partially controlled epidemics.

663

664 Despite these limitations, on a short timescale, we have some predictive ability of human and  
665 horse cases of EEEV (**Fig. 5**). When considering longer timescales, we still cannot use climate-  
666 informed models, like index P, for annual predictions of outbreaks. This is because, while climatic  
667 factors are vital for the high abundance of *Cs. melanura* required for an intense EEEV year, they  
668 cannot predict when the virus will be present. Index P therefore may be a more effective way to  
669 predict EEEV case counts in Florida where the virus is continuously maintained (Bigler et al.,  
670 1976), but should be used with caution in scenarios where the virus must be introduced.

671

672 To provide longer-term predictions (i.e. on a time scale of years), we must therefore understand  
673 what drives introductions of EEEV from Florida into the Northeast. The missing piece here is  
674 large-scale studies on bird immunity, as waves of infection in the Northeastern bird population  
675 are likely driven by the renewal of the susceptible population, potentially through birth (Armstrong  
676 & Andreadis, 2022; Elias et al., 2017). Theoretically, the dynamics could be similar to those of  
677 Middle Eastern Respiratory Syndrome (MERS) (Dudas et al., 2018), wherein cycles of infection  
678 due to buildup of susceptibles in the reservoir population leads to spillover events into other  
679 species. Certainly, we found that EEEV appears to have waves of introductions into the Northeast  
680 which co-circulate before going extinct shortly afterwards, which would be expected from local  
681 depletion of susceptible birds.

682

683 In understanding EEEV outbreaks in humans and horses, we must look to a combination of  
684 dynamics of other arboviruses like West Nile and dengue in terms of the mosquito populations,  
685 as well as viruses that mostly exist in reservoir populations and spillover into humans, such as  
686 MERS. The complex interplay of these factors make long-term prediction with our current data  
687 sources difficult, as we do not have enough information on bird immunity and its interaction with  
688 EEEV transmission and spillover. We do know, however, that a large mosquito population,  
689 enabled by warm and wet conditions, is necessary, and an increase in years with warm and wet  
690 summers and mild winters may increase the frequency of outbreaks. Therefore, while the

691 interactions in EEEV with climatic factors are more complex than with some other arboviruses,  
692 climate change may still represent an increase in risk as more years will be permissive for  
693 outbreaks in mammals. In any case, all of these results rest on a timely and robust mosquito  
694 surveillance program, as currently exists in New York, Massachusetts, and Connecticut.  
695 Widespread and consistent trapping and rapid analysis provides the data required to calculate  
696 vector index, which is the strongest correlate of human and horse cases later in the season. It is  
697 imperative that programs like this, also of use for other mosquito-borne viruses such as West  
698 Nile virus, continue to be funded and expanded, even as competition for public health funding  
699 increases.  
700  
701

## 702 Materials and Methods

### 703 Ethics

704 No human samples or direct clinical data were used in this study. This study was determined to  
705 be Not Human Research by the Yale University Human Research Protection Program Institutional  
706 Review Board. All EEEV case data were aggregated and available from public surveillance  
707 databases as described below. Sequencing of remnant veterinarian samples by the Wadsworth  
708 Center were done following institutional protocols.

### 709 Case data and availability

710 EEE, caused by EEEV, is a notifiable human disease, therefore human case data is routinely  
711 reported to the federal government. Horse data reporting is voluntary and so is likely an  
712 underestimate. Data can be accessed on request from ArboNET  
713 (<https://www.cdc.gov/mosquitoes/mosquito-control/professionals/ArboNET.html>).  
714

715 Base layers for all map figures were taken from the Global Administrative Database ([gadm.org](https://gadm.org/)).

### 716 Mosquito surveillance

717 In New York State (NYS), mosquito surveillance was carried out in 13-43 counties throughout  
718 the state including an EEEV endemic area of Central NYS (Onondaga, Oswego, Oneida  
719 Counties), annually from May-October, as previously described (Oliver et al., 2018). Trapping was  
720 completed using a combination of Centers for Disease Control and Prevention (CDC) light traps,  
721 gravid traps and diurnal resting boxes. Resting boxes were primarily used in EEEV endemic areas  
722 to collect *Cs. melanura*. Mosquito specimens were sorted by species and pooled for testing.  
723 Pools of 10-60 mosquitoes were shipped on dry ice to the NYS Arbovirus Laboratory for EEEV  
724 testing by molecular and cellular methods. Specifically, pools containing a zinc-plated steel bead  
725 and 1ml mosquito diluent (20% heat-inactivated fetal bovine serum (FBS) in Dulbecco's  
726 phosphate-buffered saline plus 50 µg/mL penicillin/streptomycin, 50 µg/mL gentamicin, and 2.5  
727 µg/mL Fungizone) were homogenized using a mixer mill for 30 sec at 24Hz and centrifuged for  
728 5 min at 6000 rcf. Quantitative reverse transcriptase polymerase chain reactions (qRT-PCR) were  
729 performed using two distinct primer and probe sets, following RNA extraction and purification,

730 as previously described (Zink et al., 2013). In addition, 100  $\mu$ L of supernatant from *Cs. melanura*  
731 pools were inoculated on Vero cell culture and monitored for cytopathic effect. EEEV isolates  
732 obtained from a single round of amplification were used for sequencing.

733

734 In Massachusetts, mosquito surveillance was conducted from mid-May through to mid-October,  
735 as previously described (Kinsella et al., 2020; Molaei et al., 2013). Trapping was performed by  
736 the Massachusetts Department of Public Health (MDPH) in collaboration with 10 local mosquito  
737 control projects (MCP) at semi-variable frequencies visiting non-fixed trap sites spread across all  
738 14 counties. Site visitation frequency increased with high volume collections of vector species  
739 and narrowed to weekly-biweekly over time in correlation with increased site-specific target  
740 mosquito abundance. Targeted sites visitation frequency increased to weekly at minimum when  
741 EEEV activity was detected and persisted through the duration of the seasonal surveillance  
742 period. Weekly collections were performed at 10 fixed collection sites in Bristol and Plymouth  
743 counties known to be historically active EEEV sentinel sites. These site collections increased to  
744 twice a week after initial EEEV detection.

745

746 Mosquito collection methods varied depending on MCP, however nearly all successful *Cs.*  
747 *melanura* collections were performed using primarily CDC-Miniature Light traps with a CO<sub>2</sub>  
748 source (either dry ice or regulated tank flow ranging from 250-500cc), gravid traps baited with an  
749 infusion of lactalbumin-yeast-hay with oak leaves, or resting boxes placed primarily in locales with  
750 both deciduous and evergreen freshwater forested swamps. Light traps and gravid traps were  
751 placed in the early morning-late afternoon and retrieved 24 hours later, and resting boxes were  
752 visited once weekly. Mosquito trap canisters collected from the field were transported to the  
753 laboratory in an igloo cooler lined with dry ice, freeze-killed in an ultra-low -80°C freezer, identified  
754 by species using a dichotomous key to characterize morphological differences with a  
755 stereoscope, and pooled in sample vials of 5-50 female mosquitoes. Sample pools were grouped  
756 by species/trap site/date of collection before being submitted to the MDPH Molecular  
757 Diagnostics lab for arbovirus testing.

758

759 In Connecticut, mosquito trapping and arbovirus surveillance was conducted from the beginning  
760 of June through the end of October at 91 fixed collection sites, distributed among all 8 counties.  
761 Trapping locations, where *Cs. melanura* were likely to be collected or where there was historical  
762 record of EEEV activity, were established in sparsely populated rural settings that included  
763 permanent fresh-water swamps (red maple/white cedar) and bogs, coastal salt marshes, horse  
764 stables, and swamp-forest border environs. Additional trap sites are located in more densely  
765 populated urban or suburban locales, including parks, greenways, golf courses, undeveloped  
766 wood lots, sewage treatment plants, dumping stations, and temporary wetlands associated with  
767 waterways.

768 Mosquito trapping was conducted with CO<sub>2</sub> (dry ice)-baited CDC miniature light traps equipped  
769 with aluminum domes or gravid mosquito traps baited with a lactalbumin-yeast-hay infusion.  
770 Traps were placed in the field in the afternoon, operated overnight, and retrieved the following  
771 morning. Trapping frequency was minimally made once every ten days at each trap site over the  
772 course of the entire season. Mosquito trapping frequency was increased at EEEV-positive sites  
773 to twice per week after the virus was isolated from that site. Adult mosquitoes were transported  
774 alive to the laboratory each morning in an ice chest lined with cool packs. Mosquitoes were

775 immobilized with dry ice and transferred to chill tables where they were identified to species with  
776 the aid of a stereo microscope (90X) based on morphological characters. Female mosquitoes  
777 were pooled in groups of 50 or fewer by species, collection date, trap type, and collection site  
778 and stored at -80°C until processed for virus isolation. Processed mosquito pools were  
779 inoculated into Vero cell cultures and screened for cytopathic effect (CPE) as previously described  
780 (Armstrong et al., 2011). CPE positive virus cultures and the original mosquito pool were then  
781 tested for EEEV by TaqMan RT-PCR assay (Armstrong et al., 2012).

## 782 Maximum-likelihood estimation of infection rate, relative abundance, and 783 vector index

784 The maximum-likelihood estimation (MLE) of the infection rate is a pooled infection rate of positive  
785 EEEV pools, and was estimated using the CDC R software package PooledInfRate  
786 (<https://github.com/CDCgov/PooledInfRate>). This estimation procedure takes into account that  
787 there may be different numbers of positive mosquitoes in each positive pool, and so estimates  
788 the likely number of positive mosquitoes in each pool based on the overall number of positive  
789 pools. The relative abundance was calculated as the average number of mosquitoes captured  
790 per trap per night. Vector index is these two metrics multiplied together (Fauver et al., 2016;  
791 Nasci et al., n.d.).

## 792 RNA isolation and virus sequencing

793 RNA was extracted on the MagMax-96 Express robot (Applied Biosystems, Foster City, CA) with  
794 the Magmax Viral isolation kit (ThermoFisher Scientific, Waltham, MA), according to  
795 manufacturer's specifications. 50 µL of sample was added to 130 µL of lysis buffer containing  
796 20 µL of RNA binding beads diluted 1:1 with wash buffer. RNA was eluted in 90 µL of elution  
797 buffer. Primer pairs, ATAGGGTACGGTGTAGAGGCAACC, TGGTCCGGCATCCCCTTTCTTTAC,  
798 and CGTTAACGGAGGGGCACTGAAT, GCGTAGATGCCGGTAGATAACAAC, and  
799 AAAGCGCACCTCGTCAAGCATTCT, GCGGTGAGTCTTATCGGGTTTGTGTC, and  
800 CGAAACGGAATTGCAATGTCACTC, CTGATCATAGGCTCGGCTGTGCGTA, and  
801 CCAAAGGGGGTTACAGTCAAA, TCGGTGTGCGCAGAAGCAGTAGG, and  
802 CAAAAGTGCCGTCTCCAGTAGTGA, GAAATATTA AAAACAAAATAAAAACATAAAA, were used  
803 to generate 6 overlapping fragments of approximately 2.5 kb each using one-step superscript III  
804 RT-PCR with platinum Taq (Life Technologies, Carlsbad, CA). Each reaction utilized 5 µL of RNA,  
805 1 µL of enzyme, and a 0.4 µM final concentration of primer pairs in a total reaction volume of 25  
806 µL. Thermocycler amplification was completed using the following conditions: 55°C for 30 min;  
807 94°C for 2 min; 40 cycles of 94°C for 15 sec, 55°C for 30 sec, 68°C for 3.5 min; and a final  
808 extension of 68 °C for 10 min (Simpliamp by Applied Biosystems, Waltham, MA). Two µL of  
809 amplicons were visualized on a 1% agarose gel to confirm size and quality, and subsequently  
810 purified using Zymo DNA Clean and Concentrate (Zymo Research, Irvine, CA). Amplicons from  
811 individual isolates were pooled and sent to the Wadsworth Center Advanced Genomic  
812 Technologies Core for library preparation and indexing using the Nextera XT kit (Illumina, San  
813 Diego, CA) according to manufacturer's protocols.

814  
815 Sequencing was performed on the Illumina MiSeq platform (San Diego, CA). Paired-end reads  
816 were assembled to a 2014 Connecticut isolate from *Cs. melanura* (KX029260) deploying

817 Geneious Prime's reference mapping tool with high sensitivity and free end gaps using 10  
818 iterations of fine tuning and trimming paired read overhangs. Mean coverage/base ranged from  
819 703-2132x. Resultant consensus sequences were used for downstream analyses. We generated  
820 complete genome consensus sequences for all 80 sequenced isolates.

## 821 Nucleotide alignments and phylogenetic analysis

822 In total, 80 new samples of EEEV were sequenced from 2015-2019. These were deposited in  
823 Genbank with accession numbers OQ511733-OQ511812. 451 additional whole genome  
824 sequences from prior to this study were downloaded from Genbank. Those not from the US and  
825 without location data were removed, and the remaining combined with the new sequences to  
826 give a dataset of 523 sequences.

827  
828 We aligned the sequences using Mafft version 1.3.7 (Kato & Standley, 2013), and removed the  
829 non-coding regions at either end of the genome, giving a final multiple sequence alignment with  
830 a length of 11,277 bases. A maximum-likelihood phylogenetic tree was generated using IQ-TREE  
831 2.1.4 (Minh et al., 2020) and temporal signal was assessed in TempEst (Rambaut et al., 2016).  
832 A single molecular clock outlier (ID: AY722102) was removed, as described in (Hill & Baele,  
833 2019).

834  
835 We estimated introductions into the Northeast (defined as New York, Connecticut,  
836 Massachusetts, New Hampshire, Vermont, Rhode Island and Maine) by conducting a discrete  
837 trait phylogeographic analysis (DTA) at the state level in BEAST 1.10 (Suchard et al., 2018), with  
838 an asymmetric CTMC model. We also used a non-parametric skygrid coalescent model (Gill et  
839 al., 2012) estimated using Hamiltonian Monte Carlo sampling (Baele et al., 2020), an HKY  
840 substitution model (Hasegawa et al., 1985), and a strict clock model (Ferreira & Suchard, 2008).  
841 We used tip-date sampling for those sequences without exact sampling dates, with a starting  
842 date given as 0.6 of the way through the appropriate year (i.e. August) with a standard deviation  
843 of three months. We estimate Markov jump histories for the full posterior to obtain estimates of  
844 location transitions between states in the DTA, and summarize them using  
845 TaxaMarkovJumpHistoryAnalyzer (Lemey et al., 2020). We performed two independent  
846 replicates of this analysis, with each chain running for 100 million iterations, removing 10% for  
847 burn-in. Convergence and mixing were assessed in Tracer 1.7 (Rambaut et al., 2018).

848  
849 An introduction node was considered to be the first node which was inferred to be in one of the  
850 Northeastern states and downstream tips were counted as part of the cluster. One introduction  
851 left Massachusetts and returned to Florida. As this is unlikely given bird migration patterns, and  
852 the confidence in the location of the node was low (52% Massachusetts, 46% Florida), we instead  
853 used the child node which was found in Massachusetts, thereby excluding the Florida sequence  
854 from this cluster.

855  
856 As there are no sequences from Florida (which was estimated to be the backbone of the  
857 phylogeny) after 2014, and relatively few before, the DTA will underestimate the number of  
858 introductions in total as they will not be broken up by Florida sequences. The date of the  
859 introduction node (the first node in the cluster inferred to be in the Northeast) will therefore also  
860 be too far in the past, as the lack of Florida sequences in the cluster means that the location of



861 the node is inferred incorrectly. Therefore, we set out to split up some introductions which  
862 contained long branches and likely represented unsampled Floridian diversity. We identified all of  
863 the clusters which contained tips only from the Northeast with more than three sequences and  
864 more than 50% of the sequences sampled in 2014 or earlier (i.e. when the last Floridian sequence  
865 was sampled); and calculated their average branch length (1.28 years). We then traversed the  
866 tree, and for any branches where both the parent and child node were inferred to be in the  
867 Northeast, but were more than twice the standard deviation (1.52 years) above the average  
868 branch length (threshold = 4.33 years), we assigned the introduction node as the child of the pair  
869 instead of the parent. This therefore moved the introduction node closer to the present, and  
870 sometimes excluded Northeastern sequences on the sister branch, making separate  
871 introductions. The final number of introductions into the Northeast before this procedure was 42  
872 prior to 2014 and 19 after, and 49 and 26 respectively, adding 14 more introductions in total.

873

874 To test for nucleotide substitutions common to 2019 sequences, we compared the consensus  
875 sequences to the reference sequence used in (Yu et al., 2015), which has Genbank accession  
876 ID KJ469556.

## 877 Effective population size using a skygrid model

878 To identify any possible associations between effective population size and case counts, we  
879 applied the latter as covariates to the estimation of population size using the skygrid model (Gill  
880 et al., 2016). We began by comparing 12 possible covariates (all human and animal cases, just  
881 human cases, just horse cases, and all cases in Massachusetts, Connecticut, and New York with  
882 0, 1 and 2 year lags) to effective population estimated sizes using a skygrid model with no  
883 covariates (Gill et al., 2012). On the basis of this preliminary analysis, we ran formal analyses with  
884 all cases, human cases, and horse cases and the relevant lags. We also only used sequence and  
885 case data from 2003 onwards ( $n=423$ ), when the surveillance program began, in order to  
886 eliminate some noise from the data. BEAST runs were set up as above with tip-date sampling,  
887 HKY substitution models and a non-parametric skygrid coalescent model. Grid points were  
888 externally set to correspond to the start of each year.

889

890 Case counts for years between the inferred time of origin of the tree and 2003 were estimated  
891 independently using a normal prior, whose mean is the recorded case count. The standard  
892 deviation was calculated such that all national cases and horse case covariates were allowed to  
893 have  $\pm 5$  cases in the 95% confidence interval; and human covariates were allowed to have  $\pm 1$   
894 cases in the same interval, as the recording of the latter is more precise. This was also undertaken  
895 for 2021 data for both analyses with the 2 year lag covariates, as finalized data has not been  
896 released by ArboNET at time of writing.

897

898 Starting values for unobserved years were taken from news reports and other publicly available  
899 sources. Most of the researched outbreaks were reported in (Corrin et al., 2021), and data were  
900 supplemented by going to each of the references given in that paper. Data for outbreaks in Michigan  
901 were supplemented by information in (Stobierski et al., 2022), and Massachusetts data were  
902 supplemented using (Feemster, 1938; Grady et al., 1978; Massachusetts Department of Public  
903 Health, 2022). Data for all states were supplemented using CDC Morbidity and Mortality Weekly

904 reports in relevant years. Horse cases used are both confirmed and suspected, due to a lack of  
905 equine testing, especially in the earlier epidemics.  
906

## 907 Human outdoorness measurements

908 How much time people spent indoors or outdoors was taken from (Susswein et al., 2022), who  
909 used anonymized GPS data from mobile phones. This data was only available for 2018 and 2019.  
910 We took the average indoor/outdoorness for each county which had a case in 2019, or had ever  
911 had a case and performed a dependent t test for paired samples to compare the metric between  
912 years.

## 913 Bird abundance

914 Bird banding data to calculate abundance and proportion were obtained from the North  
915 American Bird Banding Program dataset (Celis-Murillo et al., 2022). All ages of birds were used.  
916 Bird species were selected using data from (Molaei et al., 2016), who calculated the relative  
917 feeding preference of *Cs. melanura* in Connecticut using a blood meal analysis and normalizing  
918 it by abundance of the bird species in question. High-preference birds in this analysis were the 8  
919 species identified in this paper, and low-preference birds were randomly selected from their list  
920 of birds for which *Cs. melanura* appeared to have no preference. We conducted Poisson  
921 regressions for each combination for the abundance or proportion of a bird species in a given  
922 year against the human and horse cases or mosquito infection rate (both defined above).

## 923 Mosquito transmission suitability model - index P

924 Index P is an estimate of the transmission potential by a single female mosquito using Bayesian  
925 methods and mosquito-virus specific priors (**Table S3**), including transmission probability as a  
926 function of temperature and relative humidity. Unlike the basic reproduction number,  $R_0$ , index P  
927 does not need to be greater than one to cause sustained transmission, as index P is multiplied  
928 by the ratio of human to mosquitoes to derive  $R_0$ . We estimated the variable index P using the  
929 Mosquito Virus Suitability Estimator (MVSE), an R package to download the functions for  
930 modeling the suitability of a given environment for mosquito-borne virus transmission (Obolski et  
931 al., 2019). Temperature and humidity data for each state were calculated by taking the average  
932 temperature for each day in the center of each county in the state. These county level data were  
933 then averaged to obtain a single temperature estimate for the entire state. All weather data were  
934 provided by [visualcrossing.com](https://visualcrossing.com). Index P was calculated from April to October, in line with the  
935 transmission season for EEEV. All negative temperatures were set to zero in the model to avoid  
936 erroneous results, as at zero or below, mosquitoes are in a hibernation state and so there is no  
937 mosquito activity that varies by temperature.

938  
939 We used a 1 month lag time from index P to cases to account for delays caused by requiring  
940 transmission into humans and horses, followed by incubation times and time to diagnosis.  
941 Therefore the relevant mosquito activity will be before human and horse cases are reported.  
942

943 While index P is seen as an important covariate in the negative binomial regression model, it is  
944 difficult to tease out the seasonality of the mosquito season with cases. Due to the sporadic  
945 nature of EEEV in the Northeast, index P is a necessary but not sufficient parameter to consider.  
946 To test whether temperature could be used as a proxy, an additional model was tested by  
947 replacing index P with temperature which resulted in a slightly higher Akaike information criterion  
948 (AIC) score (231.4 vs 232.3) and wider confidence interval. Given the widespread use of modeled  
949 mosquito viability parameters, we opted to keep index P. However, future research and  
950 surveillance programs may wish to utilize temperature, an easily obtained parameter with little  
951 uncertainty in its estimate.

## 952 Regression model fit to cases

953 We restricted the case data (described above) to Connecticut and Massachusetts for the model  
954 where we had sufficient mosquito data to calculate the Vector Index. Dispersion was detected  
955 with a dispersion parameter  $\alpha = 0.18$  ( $p$ -value = 0.019) using the dispersion test from the  
956 AER package in R (Cameron & Trivedi, 1990). This led us to fit a negative binomial regression  
957 model (estimated using ML) to predict human\_equine with vector\_index (formula: human\_equine  
958 ~ vector\_index + month\_detected\_july + indexP\_lag1 + st\_grp + year\_index + month\_f).

959  
960 Given the vast differences in cases between Massachusetts and Connecticut we attempted to  
961 run stratified models of each state separately. However the model failed to converge for  
962 Connecticut, likely due to low availability of cases. Thus a model combining both was used.  
963 Months were included as categorical variables with the months May through July as reference  
964 groups. These months were combined due to May and June having no cases, making parameter  
965 estimates fail to converge when separate. The parameter ‘first month detected July’ was a binary  
966 variable determined by identifying the month when the first non-zero value for the vector index  
967 occurred for each year. All continuous data were normalized so interpretation of estimates are  
968 for a 1 standard deviation increase in the term.

969  
970 Effect modification of vector index by month of detection was explored in addition to the original  
971 model without effect modification. AIC scores of the model with effect modification and without  
972 were nearly identical (231.4 vs 231.2 respectively) and an ANOVA test was conducted between  
973 the two models and found to have no difference ( $p$ -value = 0.14). While no difference was  
974 identified, this may be due in part to the small sample size of years where the month detected  
975 was after July.

976  
977 In addition to effect modification, given some of the complexity of index P, we explored the terms  
978 temperature 1-month lag and mosquito abundance 1-month lag in replacement of index P.  
979 Utilizing the two point rule of thumb for AIC, the abundance model performed worse (231.4 vs  
980 233.2) and its estimate was not significant (CI 95%: 0.57-1.62). The temperature performed  
981 slightly worse but similar to the index P model (AIC 231.4 vs 232.3) but had a larger standard  
982 error (CI 95%: 1.15-9.31). Given the desire to explore the utility of index P and its higher  
983 performance we opted to focus on this model. However, for simplicity sake future work may  
984 utilize temperature for its ease of use and similar performance (Figure S6 and Table S4).

985

986 Utilizing the DHARMA version 0.4.6 package in R ([https://CRAN.R-](https://CRAN.R-project.org/package=DHARMA)  
987 [project.org/package=DHARMA](https://CRAN.R-project.org/package=DHARMA)), we simulated the residuals 1,000 times to test for  
988 heteroskedasticity, zero inflation, and autocorrelation. Upon visual inspection no  
989 heteroskedasticity of the residuals was detected. The simulated ratio of expected to actual zeros  
990 was 1.01 (p-value = 0.776) The Durbin-Watson test for autocorrelation was conducted on a  
991 subset of the Massachusetts data to avoid duplicate time indexes. In addition to subsetting the  
992 data, index P with and without a 1-month lag was tested per the suggestion of the Durbin-Watson  
993 test to avoid lagging covariates. Autocorrelation was borderline but insignificant without the index  
994 P lag (DW = 1.59, p-value = 0.051).  
995

## 996 Acknowledgements

997 We thank the seasonal mosquito surveillance programs from Connecticut, Massachusetts, and  
998 each county in New York, Wadsworth Center Advanced Genomic Technologies Core for  
999 assistance with sequencing, D. Weinberger for help with the statistical model, J. Howard for his  
1000 thoughts, and P. Jack and S. Taylor for providing comments on an early draft. Research reported  
1001 in this publication was supported by the National Institute Of Allergy And Infectious Diseases of  
1002 the National Institutes of Health under Award Number DP2AI176740 (NDG), CTSA Grant Number  
1003 UL1 TR001863 from the National Center for Advancing Translational Science (NCATS), a  
1004 component of the National Institutes of Health (CBFV), the Cooperative Agreement Number  
1005 U01CK000509 funded by the Centers for Disease Control and Northeast Center of Excellence  
1006 (PMA, ATC), the KU Leuven Internal Funds Grant No. C14/18/094 (GB), and the Research  
1007 Foundation - Flanders award "Fonds voor Wetenschappelijk Onderzoek - Vlaanderen,"  
1008 G0E1420N, G098321N (GB). The content is solely the responsibility of the authors and does not  
1009 necessarily represent the official views of the National Institutes of Health or the Centers for  
1010 Disease Control and Prevention.  
1011

## 1012 Conflicts of interest

1013 The authors declare no conflicts of interest related to this work.

## 1014 Data availability

1015 New sequences were deposited in NCBI Genbank under accession numbers OQ511733-  
1016 OQ511812. Case data from the literature is available at [https://github.com/grubaughlab/eeev-](https://github.com/grubaughlab/eeev-genomics)  
1017 [genomics](https://github.com/grubaughlab/eeev-genomics), along with XMLs for the BEAST analyses. ArboNET case data is available on request  
1018 from the CDC [https://www.cdc.gov/arboNET/maps/ADB\\_Diseases\\_Map/index.html](https://www.cdc.gov/arboNET/maps/ADB_Diseases_Map/index.html). Original  
1019 shape files available from [gadm.org](https://gadm.org).

## 1020 References

1021

- 1022 Armstrong, P. M., & Andreadis, T. G. (2022). Ecology and Epidemiology of Eastern Equine  
1023 Encephalitis Virus in the Northeastern United States: An Historical Perspective. *Journal of*  
1024 *Medical Entomology*, 59(1). <https://doi.org/10.1093/jme/tjab077>
- 1025 Armstrong, P. M., Andreadis, T. G., Finan, S. L., Shepard, J. J., & Thomas, M. C. (2011).  
1026 Detection of infectious virus from field-collected mosquitoes by vero cell culture assay.  
1027 *Journal of Visualized Experiments: JoVE*, 52. <https://doi.org/10.3791/2889>
- 1028 Armstrong, P. M., Prince, N., & Andreadis, T. G. (2012). Development of a multi-target TaqMan  
1029 assay to detect eastern equine encephalitis virus variants in mosquitoes. *Vector Borne and*  
1030 *Zoonotic Diseases*, 12(10), 872–876.
- 1031 Baele, G., Gill, M. S., Lemey, P., & Suchard, M. A. (2020). Hamiltonian Monte Carlo sampling to  
1032 estimate past population dynamics using the skygrid coalescent model in a Bayesian  
1033 phylogenetics framework. *Wellcome Open Research*, 5.  
1034 <https://doi.org/10.12688/wellcomeopenres.15770.1>
- 1035 Bigler, W. J., Lassing, E. B., Buff, E. E., Prather, E. C., Beck, E. C., & Hoff, G. L. (1976).  
1036 Endemic Eastern Equine Encephalomyelitis in Florida: A Twenty-Year Analysis, 1955–  
1037 1974. *The American Journal of Tropical Medicine and Hygiene*, 25(6), 884–890.
- 1038 Cameron, A. C., & Trivedi, P. K. (1990). Regression-based tests for overdispersion in the  
1039 Poisson model. *Journal of Econometrics*, 46(3), 347–364.
- 1040 Celis-Murillo, A., Malorodova, M., & Nakash, E. (2022). *North American Bird Banding Program*  
1041 *Dataset 1960-2022 retrieved 2022-07-14* [Data set]. <https://doi.org/10.5066/P9BSM38F>
- 1042 Corrin, T., Ackford, R., Mascarenhas, M., Greig, J., & Waddell, L. A. (2021). Eastern Equine  
1043 Encephalitis Virus: A Scoping Review of the Global Evidence. *Vector Borne and Zoonotic*  
1044 *Diseases*, 21(5), 305–320.
- 1045 Dudas, G., Carvalho, L. M., Rambaut, A., & Bedford, T. (2018). MERS-CoV spillover at the  
1046 camel-human interface. *eLife*, 7. <https://doi.org/10.7554/eLife.31257>
- 1047 Elias, S. P., Keenan, P., Kenney, J. L., Morris, S. R., Covino, K. M., Robinson, S., Foss, K. A.,  
1048 Rand, P. W., Lubelczyk, C., Lacombe, E. H., Mutebi, J.-P., Evers, D., & Smith, R. P., Jr.  
1049 (2017). Seasonal Patterns in Eastern Equine Encephalitis Virus Antibody in Songbirds in  
1050 Southern Maine. *Vector Borne and Zoonotic Diseases*, 17(5), 325–330.
- 1051 Fauver, J. R., Pecher, L., Schurich, J. A., Bolling, B. G., Calhoun, M., Grubaugh, N. D.,  
1052 Burkhalter, K. L., Eisen, L., Andre, B. G., Nasci, R. S., LeBailly, A., Ebel, G. D., & Moore,  
1053 C. G. (2016). Temporal and Spatial Variability of Entomological Risk Indices for West Nile  
1054 Virus Infection in Northern Colorado: 2006-2013. *Journal of Medical Entomology*, 53(2),  
1055 425–434.
- 1056 Feemster, R. F. (1938). Outbreak of Encephalitis in Man Due to the Eastern Virus of Equine  
1057 Encephalomyelitis. *American Journal of Public Health and the Nation's Health*, 28(12),  
1058 1403–1410.
- 1059 Ferreira, M. A. R., & Suchard, M. A. (2008). Bayesian analysis of elapsed times in continuous-  
1060 time Markov chains. *The Canadian Journal of Statistics = Revue Canadienne de*  
1061 *Statistique*, 36(3), 355–368.
- 1062 Gill, M. S., Lemey, P., Bennett, S. N., Biek, R., & Suchard, M. A. (2016). Understanding Past  
1063 Population Dynamics: Bayesian Coalescent-Based Modeling with Covariates. *Systematic*  
1064 *Biology*, 65(6), 1041–1056.
- 1065 Gill, M. S., Lemey, P., Faria, N. R., Rambaut, A., Shapiro, B., & Suchard, M. A. (2012).  
1066 Improving Bayesian Population Dynamics Inference: A Coalescent-Based Model for  
1067 Multiple Loci. *Molecular Biology and Evolution*, 30(3), 713–724.
- 1068 Giltner, & Shahan. (1933). The 1933 outbreak of infectious equine encephalomyelitis in the  
1069 eastern states. *The North American Veterinarian*.
- 1070 Grady, G. F., Maxfield, H. K., Hildreth, S. W., Timperi, R. J., Jr, Gilfillan, R. F., Rosenau, B. J.,  
1071 Francy, D. B., Calisher, C. H., Marcus, L. C., & Madoff, M. A. (1978). Eastern equine  
1072 encephalitis in Massachusetts, 1957-1976. A prospective study centered upon analyses of  
1073 mosquitoes. *American Journal of Epidemiology*, 107(2), 170–178.

- 1074 Hadler, J. L., Patel, D., Nasci, R. S., Petersen, L. R., Hughes, J. M., Bradley, K., Etkind, P.,  
1075 Kan, L., & Engel, J. (2015). Assessment of Arbovirus Surveillance 13 Years after  
1076 Introduction of West Nile Virus, United States. *Emerging Infectious Diseases*, *21*(7), 1159–  
1077 1166.
- 1078 Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular  
1079 clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*(2), 160–174.
- 1080 Hill, V., & Baele, G. (2019). Bayesian estimation of past population dynamics in BEAST 1.10  
1081 using the Skygrid coalescent model. *Molecular Biology and Evolution*.  
1082 <https://doi.org/10.1093/molbev/msz172>
- 1083 Howard, J. (2019, October 3). *Why did EEE cases spike this year? It's complicated*. CNN.  
1084 <https://www.cnn.com/2019/10/03/health/eee-outbreak-2019-cause/index.html>
- 1085 Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7:  
1086 Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772.
- 1087 Kinsella, C. M., Paras, M. L., Smole, S., Mehta, S., Ganesh, V., Chen, L. H., McQuillen, D. P.,  
1088 Shah, R., Chan, J., Osborne, M., Hennigan, S., Halpern-Smith, F., Brown, C. M., Sabeti,  
1089 P., & Piantadosi, A. (2020). Jamestown Canyon virus in Massachusetts: clinical case series  
1090 and vector screening. *Emerging Microbes & Infections*, *9*(1), 903–912.
- 1091 Kramer, L. D., & Ciota, A. T. (2015). Dissecting vectorial capacity for mosquito-borne viruses.  
1092 *Current Opinion in Virology*, *15*, 112–118.
- 1093 Layan, M., Müller, N. F., Dellicour, S., De Maio, N., Bourhy, H., Cauchemez, S., & Baele, G.  
1094 (2023). Impact and mitigation of sampling bias to determine viral spread: Evaluating  
1095 discrete phylogeography through CTMC modeling and structured coalescent model  
1096 approximations. *Virus Evolution*, *9*(1), vead010.
- 1097 Lemey, P., Hong, S. L., Hill, V., Baele, G., Poletto, C., Colizza, V., O'Toole, Á., McCrone, J. T.,  
1098 Andersen, K. G., Worobey, M., Nelson, M. I., Rambaut, A., & Suchard, M. A. (2020).  
1099 Accommodating individual travel history and unsampled diversity in Bayesian  
1100 phylogeographic inference of SARS-CoV-2. *Nature Communications*, *11*(1), 1–14.
- 1101 Lindsey, N. P., Brown, J. A., Kightlinger, L., Rosenberg, L., Fischer, M., & ArboNET Evaluation  
1102 Working Group. (2012). State health department perceived utility of and satisfaction with  
1103 ArboNET, the U.S. National Arboviral Surveillance System. *Public Health Reports*, *127*(4),  
1104 383–390.
- 1105 Lindsey, N. P., Erin Staples, J., & Fischer, M. (2018). Eastern Equine Encephalitis Virus in the  
1106 United States, 2003–2016. *The American Journal of Tropical Medicine and Hygiene*, *98*(5),  
1107 1472.
- 1108 Lindsey, N. P., Martin, S. W., Staples, J. E., & Fischer, M. (2020). Notes from the Field:  
1109 Multistate Outbreak of Eastern Equine Encephalitis Virus - United States, 2019. *MMWR*.  
1110 *Morbidity and Mortality Weekly Report*, *69*(2), 50–51.
- 1111 Lourenço, J., Thompson, R. N., Thézé, J., & Obolski, U. (2020). Characterising West Nile virus  
1112 epidemiology in Israel using a transmission suitability index. *Euro Surveill: Bulletin*  
1113 *European Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*,  
1114 *25*(46). <https://doi.org/10.2807/1560-7917.ES.2020.25.46.1900629>
- 1115 Massachusetts Department of Public Health. (2022). *Arbovirus Surveillance Plan and historical*  
1116 *data* [Data set]. <https://www.mass.gov/lists/arbovirus-surveillance-plan-and-historical-data>
- 1117 Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler,  
1118 A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic  
1119 Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534.
- 1120 Molaei, G., Andreadis, T. G., Armstrong, P. M., Thomas, M. C., Deschamps, T., Cuebas-Incle,  
1121 E., Montgomery, W., Osborne, M., Smole, S., Matton, P., Andrews, W., Best, C., Cornine,  
1122 F., 3rd, Bidlack, E., & Teixeira, T. (2013). Vector-host interactions and epizootiology of  
1123 eastern equine encephalitis virus in Massachusetts. *Vector Borne and Zoonotic Diseases*,  
1124 *13*(5), 312–323.
- 1125 Molaei, G., Oliver, J., Andreadis, T. G., Armstrong, P. M., & Howard, J. J. (2006). Molecular

- 1126 identification of blood-meal sources in *Culiseta melanura* and *Culiseta morsitans* from an  
1127 endemic focus of eastern equine encephalitis virus in New York. *The American Journal of*  
1128 *Tropical Medicine and Hygiene*, 75(6), 1140–1147.
- 1129 Molaei, G., Thomas, M. C., Muller, T., Medlock, J., Shepard, J. J., Armstrong, P. M., &  
1130 Andreadis, T. G. (2016). Dynamics of Vector-Host Interactions in Avian Communities in  
1131 Four Eastern Equine Encephalitis Virus Foci in the Northeastern U.S. *PLoS Neglected*  
1132 *Tropical Diseases*, 10(1), e0004347.
- 1133 Morens, D. M., Folkers, G. K., & Fauci, A. S. (2019). Eastern Equine Encephalitis Virus -  
1134 Another Emergent Arbovirus in the United States. *The New England Journal of Medicine*,  
1135 381(21), 1989–1992.
- 1136 Morris, C. D. (1988). Eastern equine encephalomyelitis. In T. P. Monath (Ed.), *The Arboviruses:*  
1137 *epidemiology and ecology: Vol. III* (pp. 1–20). CRC Press.
- 1138 Mundis, S. J., Harrison, S., Pelley, D., Durand, S., & Ryan, S. J. (2022). Spatiotemporal  
1139 Environmental Drivers of Eastern Equine Encephalitis Virus in Central Florida: Towards a  
1140 Predictive Model for a Lethal Disease. *Journal of Medical Entomology*.  
1141 <https://doi.org/10.1093/jme/tjac113>
- 1142 Nasci, R. S., Doyle, M., Biggerstaff, B. J., & LeBailly, A. (n.d.). Calculation and application of a  
1143 vector index (VI) reflecting the number of WN virus infected mosquitoes in a population.  
1144 *71st Annual Meeting of the American*.
- 1145 Obolski, U., Perez, P. N., Villabona-Arenas, C. J., Thézé, J., Faria, N. R., & Lourenço, J. (2019).  
1146 MVSE: An R-package that estimates a climate-driven mosquito-borne viral suitability index.  
1147 *Methods in Ecology and Evolution / British Ecological Society*, 10(8), 1357–1370.
- 1148 Oliver, J., Lukacik, G., Kokas, J., Campbell, S. R., Kramer, L. D., Sherwood, J. A., & Howard,  
1149 J. J. (2018). Twenty years of surveillance for Eastern equine encephalitis virus in  
1150 mosquitoes in New York State from 1993 to 2012. *Parasites & Vectors*, 11(1).  
1151 <https://doi.org/10.1186/s13071-018-2950-1>
- 1152 Oliver, J., Tan, Y., Haight, J. D., Tober, K. J., Gall, W. K., Zink, S. D., Kramer, L. D., Campbell,  
1153 S. R., Howard, J. J., Das, S. R., & Sherwood, J. A. (2020). Spatial and temporal  
1154 expansions of Eastern equine encephalitis virus and phylogenetic groups isolated from  
1155 mosquitoes and mammalian cases in New York State from 2013 to 2019. *Emerging*  
1156 *Microbes & Infections*, 9(1), 1638–1650.
- 1157 Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior  
1158 Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, 67(5),  
1159 901–904.
- 1160 Rambaut, A., Lam, T. T., Max Carvalho, L., & Pybus, O. G. (2016). Exploring the temporal  
1161 structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus*  
1162 *Evolution*, 2(1), vew007.
- 1163 Shamus, K. J. (2019, September 18). Deaths from mosquito-borne EEE virus prompt calls to  
1164 cancel outdoor events in Michigan. *USA Today*.  
1165 [https://www.usatoday.com/story/news/nation/2019/09/17/mosquito-borne-eee-cancel-](https://www.usatoday.com/story/news/nation/2019/09/17/mosquito-borne-eee-cancel-outdoor-events-michigan-officials-urge/2358633001/)  
1166 [outdoor-events-michigan-officials-urge/2358633001/](https://www.usatoday.com/story/news/nation/2019/09/17/mosquito-borne-eee-cancel-outdoor-events-michigan-officials-urge/2358633001/)
- 1167 Sherwood, J. A., Stehman, S. V., Howard, J. J., & Oliver, J. (2020). Cases of Eastern equine  
1168 encephalitis in humans associated with *Aedes canadensis*, *Coquillettidia perturbans* and  
1169 *Culiseta melanura* mosquitoes with the virus in New York State from 1971 to 2012 by  
1170 analysis of aggregated published data. *Epidemiology and Infection*, 148, e72.
- 1171 Stobierski, M. G., Signs, K., Dinh, E., Cooley, T. M., Melotti, J., Schalow, M., Patterson, J. S.,  
1172 Bolin, S. R., & Walker, E. D. (2022). Eastern Equine Encephalomyelitis in Michigan:  
1173 Historical Review of Equine, Human, and Wildlife Involvement, Epidemiology, Vector  
1174 Associations, and Factors Contributing to Endemicity. *Journal of Medical Entomology*,  
1175 59(1), 27–40.
- 1176 Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018).  
1177 Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus*

- 1178 *Evolution*, 4(1). <https://doi.org/10.1093/ve/vey016>
- 1179 Susswein, Z., Rest, E. C., & Bansal, S. (2022). Disentangling the rhythms of human activity in  
1180 the built environment for airborne transmission risk. *medRxiv*, 2022.04.07.22273578.
- 1181 Tan, Y., Lam, T. T.-Y., Heberlein-Larson, L. A., Smole, S. C., Auguste, A. J., Hennigan, S.,  
1182 Halpin, R. A., Fedorova, N., Puri, V., Stockwell, T. B., Shilts, M. H., Andreadis, T.,  
1183 Armstrong, P. M., Tesh, R. B., Weaver, S. C., Unnasch, T. R., Ciota, A. T., Kramer, L. D.,  
1184 & Das, S. R. (2018). Large-Scale Complete-Genome Sequencing and Phylodynamic  
1185 Analysis of Eastern Equine Encephalitis Virus Reveals Source-Sink Transmission Dynamics  
1186 in the United States. *Journal of Virology*, 92(12). <https://doi.org/10.1128/JVI.00074-18>
- 1187 TenBroeck, C., & Merrill, M. H. (1933). A serological difference between eastern and western  
1188 equine encephalomyelitis virus. *Proceedings of the Society for Experimental Biology and*  
1189 *Medicine. Society for Experimental Biology and Medicine*, 31(2).
- 1190 Tunison, J. (2019, September 29). *Despite assurances, aerial mosquito spraying plan leaves*  
1191 *some concerned*. Mlive. [https://www.mlive.com/news/kalamazoo/2019/09/despite-](https://www.mlive.com/news/kalamazoo/2019/09/despite-assurances-aerial-mosquito-spraying-plan-leaves-some-concerned.html)  
1192 [assurances-aerial-mosquito-spraying-plan-leaves-some-concerned.html](https://www.mlive.com/news/kalamazoo/2019/09/despite-assurances-aerial-mosquito-spraying-plan-leaves-some-concerned.html)
- 1193 Young, D. S., Kramer, L. D., Maffei, J. G., Dusek, R. J., Backenson, P. B., Mores, C. N.,  
1194 Bernard, K. A., & Ebel, G. D. (2008). Molecular epidemiology of eastern equine encephalitis  
1195 virus, New York. *Emerging Infectious Diseases*, 14(3), 454–460.
- 1196 Yu, G.-Y., Wiley, M. R., Kugelman, J. R., Ladner, J. T., Beitzel, B. F., Eccleston, L. T.,  
1197 Morazzani, E. M., Glass, P. J., & Palacios, G. F. (2015). Complete Coding Sequences of  
1198 Eastern Equine Encephalitis Virus and Venezuelan Equine Encephalitis Virus Strains  
1199 Isolated from Human Cases. *Genome Announcements*, 3(2).  
1200 <https://doi.org/10.1128/genomeA.00243-15>
- 1201 Zink, S. D., Jones, S. A., Maffei, J. G., & Kramer, L. D. (2013). Quadruplex qRT-PCR assay for  
1202 the simultaneous detection of Eastern equine encephalitis virus and West Nile virus.  
1203 *Diagnostic Microbiology and Infectious Disease*, 77(2), 129–132.
- 1204