Tracking Eastern Equine Encephalitis Virus Perpetuation in the Northeastern United States by Phylogenetic Analysis

Philip M. Armstrong,* Theodore G. Andreadis, John F. Anderson, Jason W. Stull, and Christopher N. Mores

The Connecticut Agricultural Experiment Station, New Haven, Connecticut; Department of Health and Human Services/Bureau of Disease Control, Concord, New Hampshire; Department of Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University, Skip Bertman Drive, Baton Rouge, Louisiana

Abstract. Epidemics and epizootics of eastern equine encephalitis virus (EEEV) occur sporadically in temperate regions where transmission is seasonal from late summer to early fall. These outbreaks may derive from viruses that overwinters locally or perhaps results from reintroduction of virus from other sites. To evaluate these possibilities, we compared the phylogenetic relationships of EEEV isolates obtained from mosquitoes collected during statewide arbovirus surveillance in Connecticut, in addition to isolates from concurrent outbreaks in southern New Hampshire and upstate New York. In Connecticut, viral isolates grouped into temporally discrete clades by year of isolation or over 2 years of sampling. Two or more clades arose in 2000, 2001, 2003, 2004, and 2006, possibly the result of separate introduction events into the state, whereas viruses from upstate New York and New Hampshire segregated into single clades that persisted for 2 or more years. New Hampshire viruses shared recent common ancestry to those isolated in Connecticut suggesting viral dispersal among these regions. These results provide additional evidence for independent episodes of EEEV overwintering in northern foci.

INTRODUCTION

Eastern equine encephalitis virus (EEEV; family Togaviridae, genus Alphavirus) is transmitted in discrete foci in eastern United States, primarily along the Gulf coast from Texas to Florida, the Atlantic coast from Florida to New Hampshire, and in inland sites in upstate New York and parts of the Midwest. The virus is maintained in an enzootic cycle involving Culiseta melanura mosquitoes and avian hosts and is closely associated with swamp habitats where vector mosquitoes breed.1,2 In temperate regions, transmission is seasonal and most prominent from late summer to early fall, whereas in Florida, the virus circulates throughout the year.3 Epidemics and epizootics occur sporadically when the virus overflows into human and domestic animal populations, resulting in high mortality and neurologic deficits in many survivors.4

The origin of EEEV outbreaks is not fully understood, particularly in temperate regions where transmission terminates each winter during periods of mosquito inactivity. The virus may overwinter locally by persisting in resident avian hosts or mosquito vectors and re-emerge when conditions favor viral amplification. Alternatively, EEEV could be reintroduced periodically perhaps by migrating birds from sub-tropical regions where transmission is continuous. Phylogenetic comparisons of EEEV strains suggest that local overwintering occurs in upstate New York where viral isolates were found to be genetically identical or nearly identical during 2- to 3-year intervals.5,6 However, broader phylogenetic trends indicate extensive movement of viruses throughout North America, supporting the possibility of annual importation of EEEV from other locations. EEEV isolates sampled over a 50-year period throughout the United States and Caribbean Basin were shown to form a single, highly conserved lineage with limited spatial structure and were distantly related to strains from South America.5,7,8

In this study, we describe the population dynamics of EEEV in Connecticut by conducting phylogenetic analyses on virus isolates obtained from field-collected mosquitoes during continuous statewide surveillance from 1996 to 2007. In particular, we test hypotheses concerning the relative importance of local overwintering versus annual importation of EEEV in this northeastern state. If the virus remains in the region for years until environmental conditions favor its amplification, the same variants of EEEV should reappear each year. Alternatively, if EEEV is re-imported annually from other sites, we expect to observe high rates of lineage turnover between transmission seasons with distinct variants appearing each year. In addition, we included viruses isolated from concurrent outbreaks in New York and New Hampshire in our analysis to compare patterns of virus persistence and movement in these regions.

MATERIALS AND METHODS

Virus isolation. EEEV was isolated from mosquitoes collected during the Connecticut Mosquito Surveillance Program from June through October of 1997–2007 and from September–October in 1996. Mosquitoes were trapped at up to 91 locations statewide and processed for virus isolation in Vero cell cultures as previously described.9,10 Representative viruses from each year were selected for sequencing except those isolated in 1997, which were not available. Viruses analyzed in this study were isolated mainly from Culiseta melanura mosquitoes representing different sampling dates and trapping locations throughout the state (Figure 1). Additional EEEV isolates were obtained from birds and mosquitoes captured in three counties during a recent outbreak involving human cases in New Hampshire from 2005 and 200611 and from blood-fed mosquitoes collected in Oswego County, NY, from 2004 to 2007.12

Reverse transcriptase-polymerase chain reaction and nucleotide sequencing. RNA was isolated from viruses passed once in Vero cell culture or directly from field specimens using the viral RNA kit (Qiagen, Valencia, CA). Reverse transcriptase-polymerase chain reaction (RT-PCR) was performed using the Titan One-Tube RT-PCR system (Roche Diagnostics, Indianapolis, IN) and primer pairs targeting a

* Address correspondence to Philip M. Armstrong, The Connecticut Agricultural Experiment Station, 123 Huntington St., P.O. Box 1106, New Haven, CT 06504. E-mail: philip.armstrong@po.state.ct.us
657-bp portion of the NSP3 gene \((\text{NSPfwd and NSPrev})\) and a 3,644-bp fragment spanning the capsid, \(E2\), and \(E1\) genes (EEE-7509 and E1rev) of EEEV (Table 1). For each RT-PCR reaction, 2 \(\mu\)L of extracted RNA was added to Master mix I containing 500 \(\mu\)mol/L dATP, 500 \(\mu\)mol/L dGTP, 500 \(\mu\)mol/L dCTP, 500 \(\mu\)mol/L dTTP, 12.5 \(\mu\)mol/L 1,4-dithiothreitol (DTT), and 1 \(\mu\)mol/L of each primer in a final volume of 20 \(\mu\)L. This mixture was heated to 85 \(^\circ\)C for 5 minutes and quickly chilled on ice. Master mix I was added to a second mix containing the Titan RT-PCR buffer and enzyme mix for a final volume of 50 \(\mu\)L. Amplification was performed as follows: 1 cycle of 50 \(^\circ\)C for 30 minutes and 94 \(^\circ\)C for 2 minutes, 10 cycles of 94 \(^\circ\)C for 15 seconds, 55 \(^\circ\)C for 30 seconds, and 68 \(^\circ\)C for 45 seconds or 3 minutes, followed by 25 cycles of 94 \(^\circ\)C for 15 seconds, 55 \(^\circ\)C for 30 seconds, and 68 \(^\circ\)C for 45 seconds or 3 minutes + 5 seconds per cycle, and 1 cycle of 68 \(^\circ\)C for 7 minutes. Amplification products of the appropriate size were purified using the QIAquick PCR purification kit (Qiagen) and sequenced in cycle-sequencing reactions using primers listed in Table 1 at the Keck Sequencing Facility (New Haven, CT).

**Genetic analyses.** Portions of the NSP3, capsid, E1, and E2 genes were sequenced to sample nucleotides from different coding regions of the genome. Overlapping sequence chromatograms were edited using ChromasPro version 1.33 (Technelysium, Tewantin, Australia) and edited sequences were deposited in Genbank (EU573359–EU573686). Multiple sequence alignments were generated by the ClustalW algorithm in Mega 4.0.\(^{13}\) Phylogenetic analyses were conducted by neighbor joining (NJ), maximum parsimony (MP), and maximum likelihood (ML) methods using PAUP 4.0 b10.\(^{14}\) NJ trees were constructed using the Jukes-Cantor model of sequence evolution. MP analyses were performed by the heuristic search method and treating characters as equally weighted and unordered. ML analyses were conducted by applying the heuristic search option and using the TrN+G model of sequence evolution selected by hierarchical likelihood ratio tests in Model Test 3.7.\(^{15}\) Support for individual nodes was evaluated by performing NJ analysis on 1,000 bootstrap replicates under the same substitution model selected for ML analysis. Phylogenetic trees were estimated from 88 concatenated NSP3, capsid, E2, and E1 sequences. Before joining molecular datasets, we tested whether sequences from these genes were phylogenetically incongruent using incongruence length difference (ILD) tests.\(^{16}\) ILD tests were performed in PAUP 4.0 b10 by parsimony analysis using 1,000 random partition replicates. Phylogenetic trees were rooted using the earliest isolate of EEEV (Ten Broeck; Virginia 1933) as previously described\(^{5}\) and displayed using TREEVIEW 1.6. Population subdivision was evaluated by estimating the average number of nucleotide substitution differences within and among temporally or spatially defined groups. Genetic variation among groups refers to fixed nucleotide differences between groups. Nucleotide distances were calculated using the Jukes-Cantor model, and SE was estimated by performing 1,000 bootstrap replicates in Mega 4.0. Nucleotide sequences were compared with the Genbank database using the BLASTN search algorithm.\(^{17}\)

**RESULTS**

To identify temporal-spatial patterns of EEEV transmission in Connecticut, we summarized relevant data from the statewide mosquito surveillance program. A total of 202 EEEV isolates were obtained from mosquitoes collected at 31 locales in seven Connecticut counties from 1996 to 2007 (Figure 1). The majority of these isolates were obtained from the southeastern corner of the state (New London County) where EEEV was detected repeatedly from many of the same sites over multiple years of sampling. Elsewhere, locations were widely scattered and virus detection was less consistent from

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer name</th>
<th>Sequence (5’-3’)</th>
<th>Genome position*</th>
<th>Used for</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSP3</td>
<td>NSPfwd</td>
<td>CAGACGACGATTTTACAGATTACG</td>
<td>4,833–4,854</td>
<td>Amplification sequencing</td>
</tr>
<tr>
<td></td>
<td>NSPrev</td>
<td>AAGCGCGAACGCAGTGGA</td>
<td>4,574–4,590</td>
<td>Sequencing</td>
</tr>
<tr>
<td>NSP4</td>
<td>EEE-7509</td>
<td>CCAAAAGCTACTACGGCTGA</td>
<td>7,509–7,528</td>
<td>Amplification</td>
</tr>
<tr>
<td>Capsid</td>
<td>CAPfwd</td>
<td>GGTTCCATCGTAAACGGGTGAC</td>
<td>7,727–7,746</td>
<td>Sequencing</td>
</tr>
<tr>
<td></td>
<td>CAPrev</td>
<td>GCGCGAAGCTTGGTGTCAG</td>
<td>8,126–8,143</td>
<td>Sequencing</td>
</tr>
<tr>
<td>E2</td>
<td>E2fwd</td>
<td>TCCACAGTGCCAAGGGTTTGA</td>
<td>9,102–9,122</td>
<td>Sequencing</td>
</tr>
<tr>
<td></td>
<td>E2rev</td>
<td>TCGTCGGCTTAATTCGCA</td>
<td>9,800–9,818</td>
<td>Sequencing</td>
</tr>
<tr>
<td>E1</td>
<td>E1fwd</td>
<td>GGTGAACCTCCCGGAAAAT</td>
<td>10,445–10,464</td>
<td>Sequencing</td>
</tr>
<tr>
<td></td>
<td>E1rev</td>
<td>GATCTTTCGTTGGCTTGCA</td>
<td>11,134–11,153</td>
<td>Amplification sequencing</td>
</tr>
</tbody>
</table>

* Position maps to sequence of EEEV isolate 82V2137.
year to year. EEEV was detected every year in Connecticut except during 1999, 2002, and 2005 (Figure 2). The number of virus isolations fluctuated annually with marked increases in 1996, 2003, and 2004 and relatively low levels of activity were detected during intervening years.

To determine the extent of viral genetic diversity, we sequenced portions of the NSP3, capsid, E2, and E1 genes from a subset of 65 EEEV isolates obtained during statewide surveillance in Connecticut 1996–2007, 10 viruses collected during an EEE epidemic in southern New Hampshire 2005–2006, and 7 viruses from well-established focus in upstate New York 2004–2007. Additional sequences from GenBank were also included, comprising six isolates from Virginia 1933, New Jersey 1960, Florida (1982, 1991, 1993), and Georgia 1997. Resulting alignments of EEEV sequences did not require the insertion of gaps except in the NSP3 alignment corresponding to a 12 amino acid deletion of EEEV strain M2006-10669 isolated from New Hampshire in 2006. Sequences were highly conserved across the different genes analyzed, resulting in relatively few variable and parsimony informative sites (Table 2). Mean nucleotide distances over all sequence pairs varied from 0.5% for the capsid gene to 1.2% for the NSP3 gene. To increase our sample of nucleotide sites for phylogenetic analysis, we combined data sets for the four genes for a total of 2,202 nucleotides sites, representing 19% of the genome.

Due to the limited polymorphism of EEEV, our study presents evidence of local overwintering, evolution, and extinction of EEEV strains in a number of northern sites, supporting the findings of previous studies conducted in upstate New York. Although EEEV sequences were highly conserved, there was sufficient genetic variation to differentiate strains and track their persistence and distribution in this region. Genetic variation arising during local evolution was evident by comparing strains from geographically isolated foci in southern New Hampshire and upstate New York. Viruses detected in New York from 2004 to 2007 formed a genetically homogenous clade, varying by zero to two nucleotide substitutions in pairwise sequence comparisons, and their NSP3 and E1 sequences were found to be identical or nearly identical virus strains were found to be dispersed in many counties throughout Connecticut, and viruses isolated from the same county fell into many distinct clades.

To determine whether viral genetic variation is shaped primarily by temporal or spatial dynamics in Connecticut, we estimated genetic divergence among viruses sampled in different years or different regions of the state (east or west of the Connecticut River). Most of the genetic variation (61.7%) occurred among viruses sampled in different years rather than within years (Table 3). In contrast, only 9.9% of the total genetic variation was attributed to divergence among viruses sampled in different regions of the state. Viral genetic variation seems to be temporally subdivided in Connecticut.

Analysis of isolates from New Hampshire and New York suggested that local overwintering of EEEV occurred in those locations as well (Figure 3). Viruses grouped into single clades from 2005–2006 in New Hampshire and 2004–2007 in New York. New Hampshire isolates shared recent common ancestry with those from Connecticut, including isolates from Fairfield County in 2003 (13073 and 13141) and from New Haven County in 2006 (12457). Viruses from upstate New York appeared as a genetically homogenous group that was most similar to a group collected in Connecticut from 2006 to 2007. BLAST searches against the GenBank database showed that NSP3 and E1 sequences of our New York viruses were identical or most similar to viral isolates from the same region during 2003–2005.

**DISCUSSION**

Our study presents evidence of local overwintering, evolution, and extinction of EEEV strains in a number of northern sites, supporting the findings of previous studies conducted in upstate New York. Although EEEV sequences were highly conserved, there was sufficient genetic variation to differentiate strains and track their persistence and distribution in this region. Genetic variation arising during local evolution was evident by comparing strains from geographically isolated foci in southern New Hampshire and upstate New York. Viruses detected in New York from 2004 to 2007 formed a genetically homogenous clade, varying by zero to two nucleotide substitutions in pairwise sequence comparisons, and their NSP3 and E1 sequences were found to be identical or nearly identical.
most similar to a larger sample of strains collected from the same region during 2003–2005 by Young and others.6 This suggests that EEEV perpetuated locally in this part of New York over 5 continuous years and was relatively isolated from viral immigration, as previously suggested.5,6,18 In contrast, viruses sampled from Connecticut were more genetically diverse and subject to high rates of population turnover, implying increased movement of new strains into this region.

EEEV was detected over many consecutive years during statewide surveillance in Connecticut for intervals lasting 2–3 years. This periodicity did not correspond to the persistence of a single strain over multiple years. Rather, our analysis suggested a more dynamic process with some strains proceeding to the next year, superimposed by the appearance of new variants. Two or more distinct variants of EEEV co-circulated in Connecticut during 2000, 2001, 2003, 2004, and 2006, but none was detected for > 2 years of sampling. The strongest evidence of local overwintering in this state/region was observed by detecting well-supported clades of identical or nearly identical strains from 2000–2001, 2003–2004, and 2006–2007. During these intervals, EEEV strains diverged over time, reflected by the stepwise chaining of phylogenetic tree branches.

**TABLE 3**
Partitioning of nucleotide diversity within EEE population from CT

<table>
<thead>
<tr>
<th>Variance Component</th>
<th>Mean Nucleotide Distance (%)</th>
<th>Differentiation Coefficient (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among years</td>
<td>0.47</td>
<td>61.7</td>
</tr>
<tr>
<td>Within years</td>
<td>0.29</td>
<td>0.4</td>
</tr>
<tr>
<td>Among regions</td>
<td>0.08</td>
<td>9.9</td>
</tr>
<tr>
<td>Within regions</td>
<td>0.68</td>
<td>0.8</td>
</tr>
<tr>
<td>Total</td>
<td>0.76</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Viruses isolated during an epidemic in New Hampshire were also included to evaluate its possible origins. In 2005, seven human cases of EEE were reported from southeastern New Hampshire, representing the first locally acquired, confirmed human cases from the state, in addition to four cases from Massachusetts.\textsuperscript{11} Interestingly, EEEV was not detected in Connecticut during the same year, indicating that virus amplification is highly focal and responsive to local conditions in individual swamp complexes. Despite this, the New Hampshire isolates were most similar to strains collected 2 years before in Connecticut in 2003 and again in 2006, suggesting both northward and southward movement of viruses among these regions. Such patterns of long-distance migration likely involve viremic birds that could transport the virus northward during spring migration, southward during fall migration, or in many directions during post-breeding wanderings as previously suggested.\textsuperscript{7,19} Viral migration from south to north was also inferred in a study by Young and others,\textsuperscript{6} whereby EEEV strains from southeastern United States were isolated years before genetically similar strains appeared in northern regions. Although we analyzed relatively few viral strains from the southern United States, a virus isolated from Georgia in 1997 was shown to be most similar to an isolate from Connecticut in 2003 (15176). These studies provide support for the periodic reintroduction of EEEV into northern sites from southern source populations; however, more extensive sampling of EEEV from southern regions is required to critically evaluate this hypothesis.

Taken together, our study found regional differences in the population dynamics of EEEV from the northeastern United States. The population from New York was relatively stable and homogenous from 2003 to 2007, whereas viruses from Connecticut were subject to continual change. We could not discern patterns in New Hampshire because viruses were sampled for only 2 years; nevertheless, there was strong support for overwintering of EEEV in this state/region as well. To date, the primary mechanism of viral overwintering remains unknown for EEEV in temperate climates. The mosquito vector, \textit{Culex melanura}, survives winter in the larval stage at this latitude,\textsuperscript{20–23} and therefore, EEEV could overwinter in the vector population by transovarial transmission of virus to the next generation. Despite early reports of transovarial transmission by isolation of EEEV from larval and male mosquitoes (cited as unpublished observations in Refs. 22 and 23), these results could not be confirmed in subsequent studies.\textsuperscript{23–26} Further attempts to recover EEEV from overwintering adults of several species of \textit{Culex} and \textit{Anopheles} as well as \textit{Cs. melanura} larvae in Connecticut were also unsuccessful.\textsuperscript{27} Alternatively, EEEV may survive through winter in chronically infected birds that could reinitiate transmission during relapsing viremias the following year. Field studies in New Jersey support this hypothesis where birds became viremic and developed antibodies against EEEV weeks before the virus was detected in mosquitoes.\textsuperscript{28} Early season sereconversions were also noted in birds captured from upstate New York,\textsuperscript{29} but this pattern was not observed during a subsequent EEE epizootic from the same locale.\textsuperscript{30} Despite the lack of a known mechanism for EEEV overwintering, our data support its occurrence in a number of different sites, suggesting that this phenomenon is broadly applicable to northern foci. Furthermore, differences seen between individual sites suggest a strong role of the local ecology in overwintering success. Identification of such ecological parameters could help determine mechanisms supporting EEEV survival, information that could be used to improve surveillance and intervention strategies against the virus.

Received March 17, 2008. Accepted for publication May 8, 2008.

Acknowledgments: The authors thank Andrew Main and Shirley Turrell for assistance in virus isolation and identification and Christine Bean, SueAnn MacRae, and Denise Bolton for laboratory testing of EEEV in New Hampshire. We also wish to acknowledge the technical assistance of our support staff: Shannon Finan, John Shepard, Michael Thomas, Nicholanna Halladay, Bonnie Hamid, Jodie Ortega, and Amanda Rahmann.

Financial support: This work was supported in part by grants from the Centers for Disease Control and Prevention (U50/CCU116806-01-I) and the US Department of Agriculture (58-6615-1-218, CONH00768, and CONH00773).

Authors’ addresses: Philip M. Armstrong, Theodore G. Andreadis, and John F. Anderson, The Connecticut Agricultural Experiment Station, 123 Huntington St., P.O. Box 1106, New Haven, CT 06504, Tel: 203-974-8461, Fax: 203-974-8502, E-mail: philip.armstrong@po.state.ct.us. Jason W. Stull, Department of Health and Human Services/Bureau of Disease Control, 29 Hazel Drive, Concord, NH 03301. Christopher N. Mores, Department of Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University, Skip Bertman Drive, Baton Rouge, LA 70803.

REFERENCES