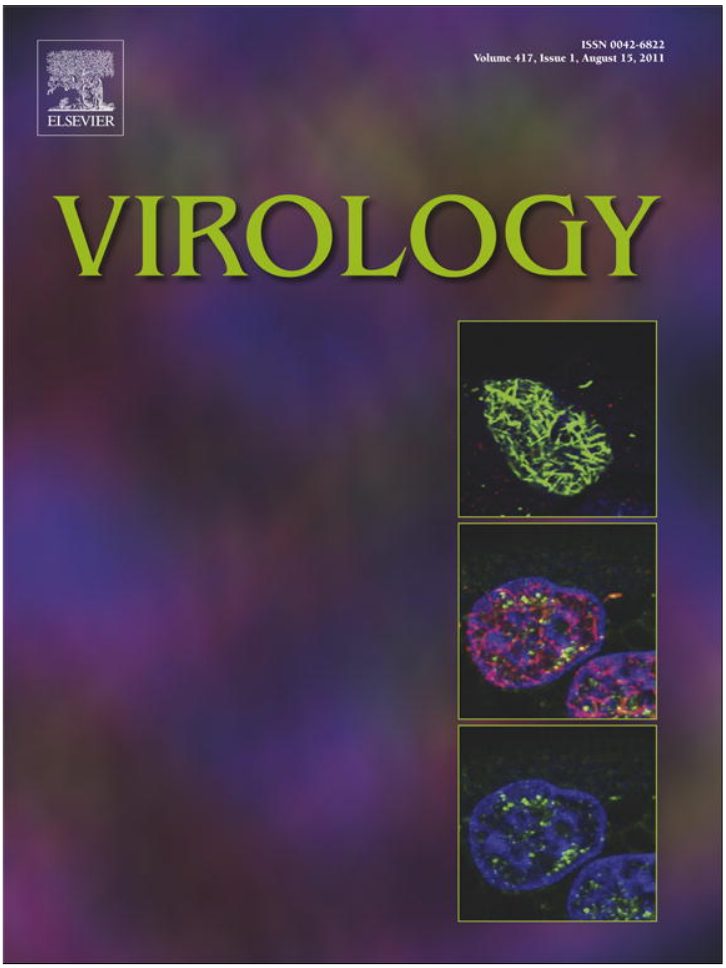

Virology 417 (2011) 203–210

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

# Molecular evolution of West Nile virus in a northern temperate region: Connecticut, USA 1999–2008

Philip M. Armstrong [a,*], Charles R. Vossbrinck [a], Theodore G. Andreadis [a], John F. Anderson [a], Kendra N. Pesko [b], Ruchi M. Newman [c], Niall J. Lennon [c], Bruce W. Birren [c], Gregory D. Ebel [b], Mathew R. Henn [c]

[a] Center for Vector Biology and Zoonotic Diseases, The Connecticut Agricultural Experiment Station, New Haven, CT 06504, USA
[b] Department of Pathology, University of New Mexico School of Medicine, Albuquerque, NM 87131, USA
[c] Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA

ARTICLE INFO

*Article history:*
Received 12 May 2011
Accepted 8 June 2011
Available online 1 July 2011

*Keywords:*
Phylogeny
*Flaviviridae*
*Flavivirus*
West Nile virus
Molecular epidemiology
Viral evolution

ABSTRACT

West Nile virus (WNV) has become firmly established in northeastern US, reemerging every summer since its introduction into North America in 1999. To determine whether WNV overwinters locally or is reseeded annually, we examined the patterns of viral lineage persistence and replacement in Connecticut over 10 consecutive transmission seasons by phylogenetic analysis. In addition, we compared the full protein coding sequence among WNV isolates to search for evidence of convergent and adaptive evolution. Viruses sampled from Connecticut segregated into a number of well-supported subclades by year of isolation with few clades persisting ≥2 years. Similar viral strains were dispersed in different locations across the state and divergent strains appeared within a single location during a single transmission season, implying widespread movement and rapid colonization of virus. Numerous amino acid substitutions arose in the population but only one change, V→A at position 159 of the envelope protein, became permanently fixed. Several instances of parallel evolution were identified in independent lineages, including one amino acid change in the NS4A protein that appears to be positively selected. Our results suggest that annual reemergence of WNV is driven by both reintroduction and local-overwintering of virus. Despite ongoing evolution of WNV, most amino acid variants occurred at low frequencies and were transient in the virus population.



## Introduction

Invasive pathogens threaten the health of immunologically-naïve human and wildlife populations as illustrated by the introduction of West Nile virus (WNV; *Flaviviridae*, *Flavivirus*) into North America in 1999. Since that time, this virus has spread throughout the Western Hemisphere where it has caused more than 30,000 confirmed human cases and 1200 deaths in the US, and imposed substantial mortality on native bird populations. WNV has become firmly established across the continental US by perpetuating in an enzootic cycle involving ornithophilic mosquitoes (mainly *Culex* species) and passerine bird hosts (Komar, 2003; Kramer et al., 2008). Humans and other mammals are dead-end hosts in the transmission cycle, becoming infected when mosquito vectors feed opportunistically on both viremic birds and mammalian hosts (Molaei et al., 2006; Weaver and Barrett, 2004).

The introduction of WNV as a point source into the New York City area, and its continued perpetuation for over a decade in this region, provide an opportunity to evaluate the evolutionary processes acting on an invading virus when it enters a new environment. WNV is a single-stranded, positive-sense RNA virus that exhibits higher mutation rates than DNA-based organisms (May et al., 2011). The viral genome is relatively small, approximately 11 kb in length, making genomic sequencing and analyses from a large number of samples feasible. The acquisition and sequencing of virus isolates during the onset of the outbreak gives us access to the ancestral genotype (Anderson et al., 1999; Lanciotti et al., 1999). The first isolates of WNV (designated as NY99) were shown to be genetically similar to a strain isolated from Israel in 1998 (Lanciotti et al., 1999, 2002). Initial analysis of WNV from Connecticut revealed a homoplasy free phylogeny with low genetic variability during the first two years of the outbreak (Anderson et al., 2001). In 2002, another variant (designated as WN02) arose, rapidly displaced the NY99 strain, and spread throughout North America (Davis et al., 2003, 2005; Ebel et al., 2004; Grinev et al., 2008; Herring et al., 2007). The mechanistic basis for this genotype replacement is related to viral fitness differences. WN02 variants were shown to replicate and disseminate more rapidly in colonies of *Culex pipiens* collected from New York and Pennsylvania,

* Corresponding author at: Center for Vector Biology and Zoonotic Diseases, The Connecticut Agricultural Experiment Station, 123 Huntington St., P.O. Box 1106, New Haven, CT 06504, USA. Fax: +1 203 974 8502.
*E-mail address:* Philip.Armstrong@ct.gov (P.M. Armstrong).

0042-6822/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2011.06.006

and *Culex tarsalis* from California (Ebel et al., 2004; Kilpatrick et al., 2008; Moudy et al., 2007), perhaps due to the fixation of a single amino acid substitution in the envelope protein (Davis et al., 2003, 2005; Ebel et al., 2004).

Phylogenetic comparisons of WNV also indicate an overall lack of geographic structure in North America (Bertolotti et al., 2007; Davis et al., 2005; Grinev et al., 2008), implying extensive movement of viral strains throughout this region. Birds could serve as an effective vehicle for dispersing viruses over long distances, thereby mixing strains from different geographic regions. Nevertheless, regional variants of WNV have arisen in South Texas (Davis et al., 2003, 2005), southwestern US (Herring et al., 2007; McMullen et al., 2011), and on the Pacific coast (Herring et al., 2007). These findings suggest that virus may perpetuate and evolve in relative isolation under certain circumstances. Viruses sampled from Chicago, Illinois, in contrast, were shown to contain a mixture of both locally-derived and exogenous virus strains over a three-year period (Amore et al., 2010).

In the northeastern US, WNV transmission is highly seasonal, re-emerging every summer and continuing into fall until mosquito feeding ceases. The primary mechanism(s) for reinitiating and sustaining transmission in this and other temperate regions is not well understood. WNV may persist and evolve locally, perhaps surviving through winter in vertically-infected, hibernating mosquitoes (Anderson and Main, 2006; Andreadis et al., 2010; Bugbee and Forte, 2004; Farajollahi et al., 2005; Nasci et al., 2001). Alternatively, virus transmission may be reseeded annually by the reintroduction of new viral strains from other geographic regions. By intensively sampling virus on both local and statewide scales, we examined the molecular evolution of WNV in Connecticut over 10 consecutive years by phylogenetic analysis. The full coding region of the WNV genome was sequenced and analyzed to differentiate WNV strains, track their distribution and persistence, and monitor evolutionary divergence. In addition, we analyzed the patterns of amino acid substitution to search for evidence of convergent and adaptive evolution within this geographic region.

## Results

### *Nucleotide sequence analysis*

Our analysis included the entire coding sequence and flanking portions of the 5′ and 3′ un-translated regions from 100 WNV isolates from Connecticut, 33 from other US states, one isolate from Mexico and one from Israel (Supplementary Table). Within Connecticut, 53 WNV sequences originated from the town of Stratford during 1999 and 2001–2008, 42 viral sequences came from 21 other towns during 2003, and the remaining sequences were from Greenwich in 1999 (n = 1), Milford in 2000 (n = 3), and Shelton in 2000 (n = 1) (Fig. 1). Thus, our sample represented a large number of sequences from the same location in Connecticut (Stratford) over many consecutive years, as well as sequences from a large number of towns during a single year (2003). WNV was not detected in Stratford in 2000 so we included strains from the nearby towns of Milford and Shelton during that year. The resulting alignment comprised a total of 10,393 nucleotide positions representing 94.2% of the genome, 977 variable sites, and 407 parsimony informative sites. Mean nucleotide distance over all sequence pairs was 0.3%. The majority of virus sequences were genetically unique, except for three sequence pairs and one group of four viruses that were identical to each other. The mean nucleotide substitution rate was $5.83 \times 10^{-4}$ substitutions/site/year which is consistent with previous estimates for WNV (Amore et al., 2010; Bertolotti et al., 2007; May et al., 2011).

### *Phylogenetic analysis*

Fig. 2 depicts the phylogenetic relationships among WNV isolates based on maximum likelihood analysis of nucleotide sequences.

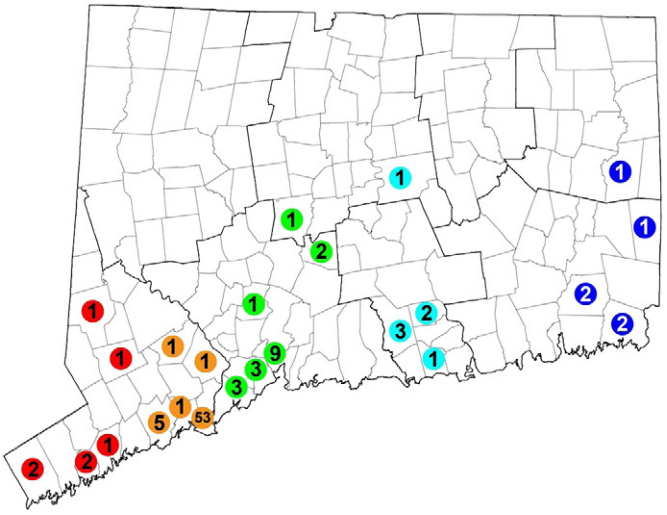


**Fig. 1.** Map of Connecticut showing the geographic location and number of WNV isolates analyzed in this study. Collection sites are color coded from west (warm colors) to east (cool colors) along a longitudinal gradient.

Viruses segregated into three major groups as previously defined: NY99, intermediate (INT), and WN02 genotypes (Davis et al., 2005; Ebel et al., 2004). Earlier WNV isolates sampled from northeastern US (1999–2003) and one isolate from Texas during 2002 formed the ancestral NY99 genotype. The INT genotype contained six isolates from Connecticut, Florida, Ohio, New York, and Mexico (2000–2003). The remaining viruses clustered together to form the WN02 genotype and had originated from sites throughout the US (2002–2008). The WN02 genotype appears to have completely supplanted the NY99 genotype in support of previous findings (Davis et al., 2005; Ebel et al., 2004; Grinev et al., 2008; Herring et al., 2007). WNV isolates from Stratford, Connecticut are highlighted with a black dot in Fig. 2. These viruses were genetically diverse with many strains grouping into well-supported subclades. Most of these subclades were defined by year of isolation with the exception of two clades that were detected from 2002 to 2003 and a larger group sampled from 2003 to 2006. Viruses sampled from Stratford, other Connecticut towns and US states were distributed throughout the phylogeny suggesting virus dispersal among these geographic regions.

To evaluate the spatial distribution of WNV in Connecticut, we restricted our phylogenetic analysis to 46 isolates obtained statewide during 2003 (Fig. 3). We focused on this year because virus activity was more widely distributed throughout the state in comparison to other years (Andreadis et al., 2004). Taxa were color coded according to their geographic location in the state and did not appear to be structured by region. However, viruses often segregated into subclades on a finer geographic scale that corresponded to a particular trapping location. These clades were generally detected transiently within a single location followed by the appearance of new variants in the same location, as seen in Darien, Fairfield, New Haven, Stratford, and West Haven, Connecticut. In addition, similar WNV strains were sometimes dispersed in different regions of Connecticut, indicating widespread migration across the state.

### *Detection of recombination*

To evaluate the potential contribution of recombination to WNV evolution, we analyzed our dataset by the SBP and GARD recombination detection methods. These analyses search for evidence of phylogenetic incongruence among fragments in the alignment to identify potential recombination breakpoints and then compare goodness of fit scores for recombination versus non-recombination

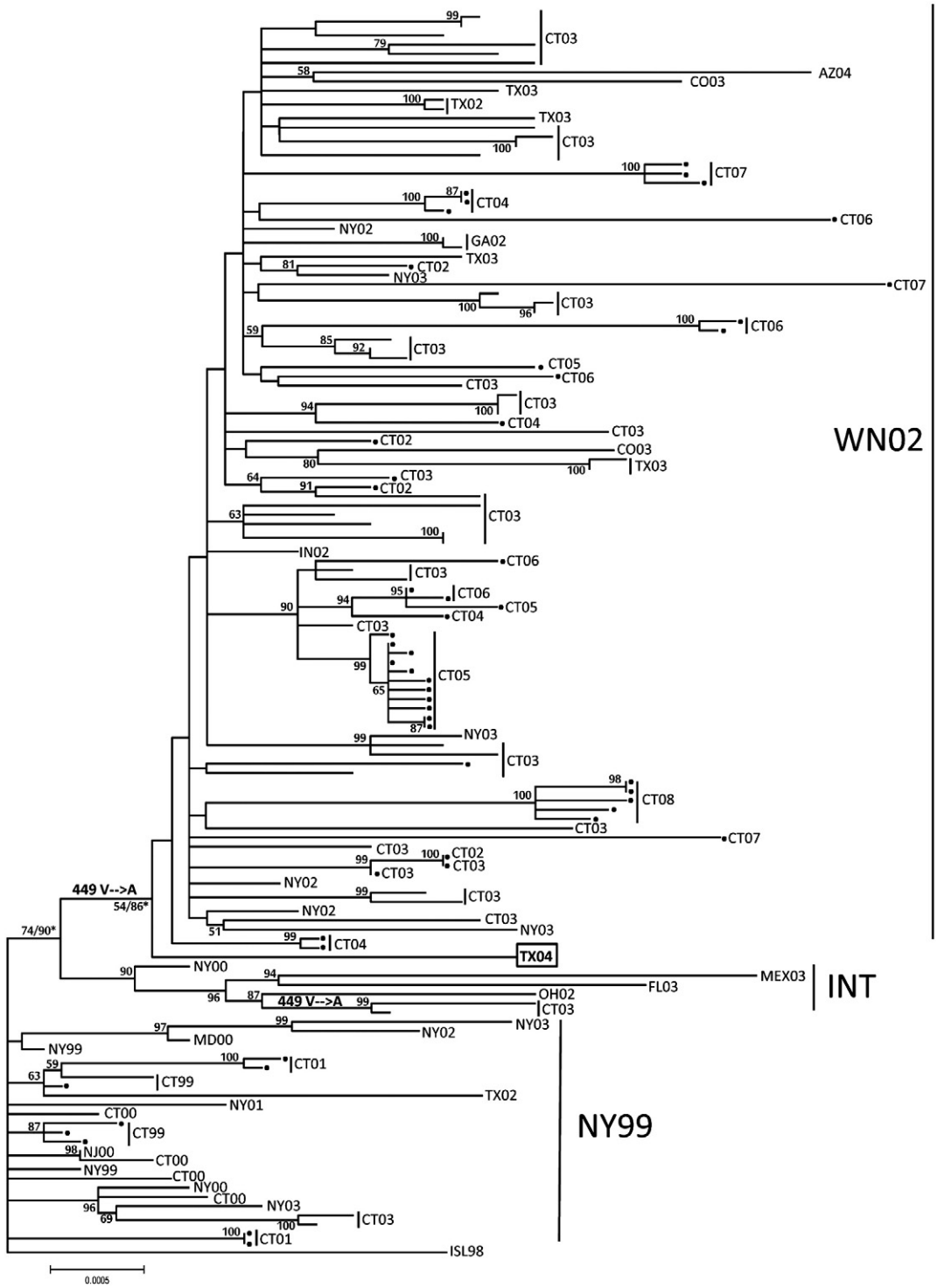


**Fig. 2.** Phylogenetic tree depicting relationships of 135 WNV isolates based on maximum likelihood analysis of 10,393 nucleotides sites. Viruses are labeled by the state or country of origin followed by the year of isolation. Numbers at nodes indicate bootstrap values >50%. Branch lengths are proportional to the number of nucleotide substitutions. Stratford taxa are highlighted with a black dot. Asterisk indicates bootstrap values with and without recombinant taxon TX04.

models. The GARD method detected evidence for recombination at a single breakpoint predicted at nucleotide position 5979 (cAIC = 15.5, $p = 0.0002$). These results were supported by the SBP method with evidence of recombination at breakpoint position 5985 and a cAIC = 15.4. Phylogenetic trees were generated from each side of the proposed recombination breakpoint and compared to identify inconsistent relationships among taxa (data not shown). One virus isolated from Harris County Texas in 2004 (TX04; Genbank no. DQ164206) displayed the genetic signature of hybridization by recombination among parental strains in the NY99 and WN02 genotypes (Table 1). A total of 9 nucleotide differences defined the NY99 and WN02 genotypes. The TX04 strain shared the WN02 sequence at five of these positions: 1442, 2466, 4146, 4803, and 9352, whereas it contained the NY99 sequence at positions 6138, 6426, 6996, and 7938. No other recombinant sequences were identified using the GARD and SBP methods when the TX04 isolate was removed from our dataset. Removal of this sequence also affected the bootstrap values on the full-length maximum likelihood tree. Bootstrap support

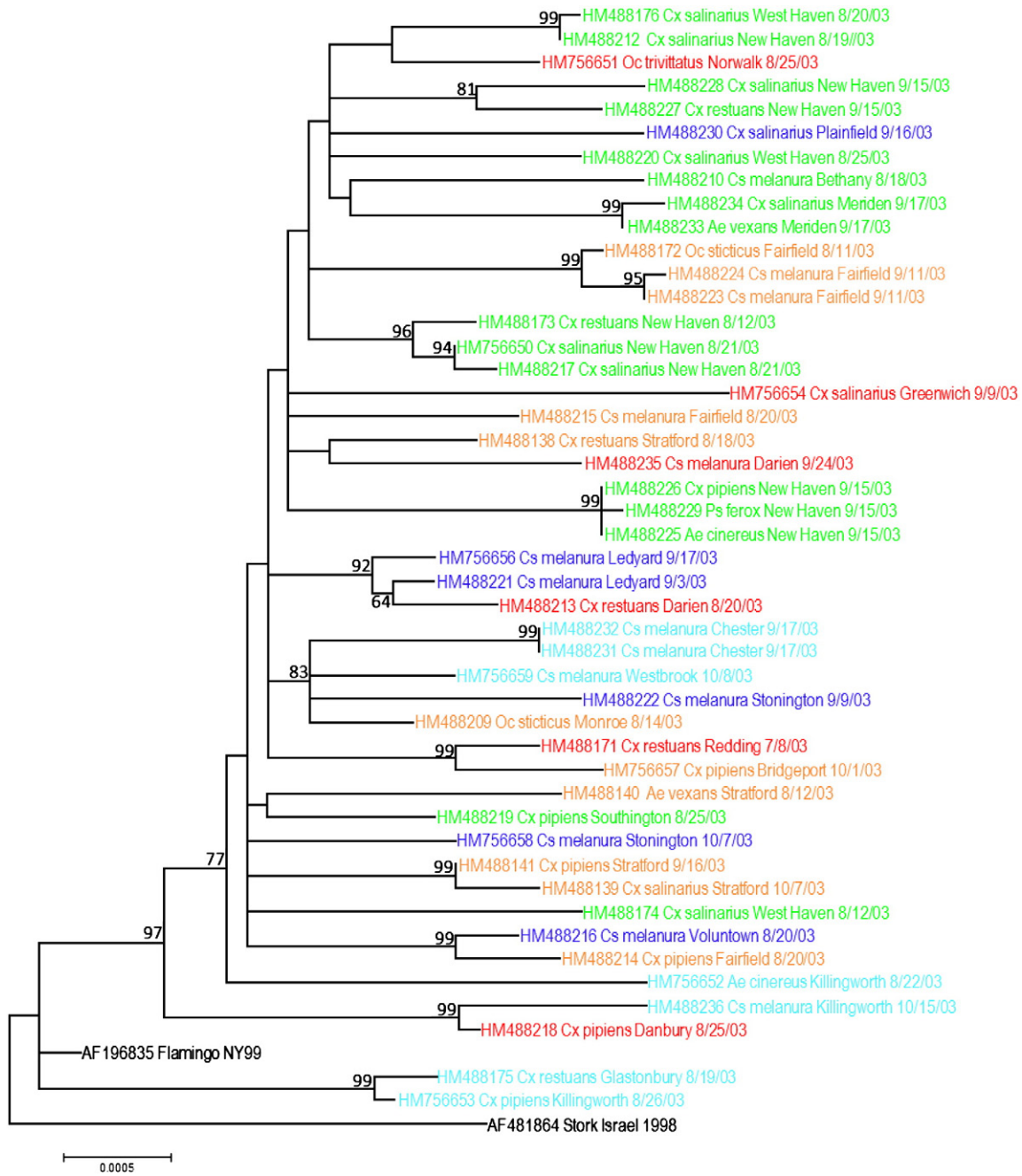


**Fig. 3.** Phylogenetic tree generated by maximum likelihood analysis of 10,393 nucleotide sites from WNV isolates collected in Connecticut during 2003. Taxon names specify the Genbank number, mosquito species, town, and date of collection and are color-coded according to their location shown in Fig. 1. Numbers at nodes indicate bootstrap values >50%. Branch lengths are proportional to the number of nucleotide substitutions.

**Table 1**
Nucleotide differences among the NY99 and WN02 genotypes, and recombinant sequence — TX04.

| Position | NY99 | WN02 | TX04 |
|---|---|---|---|
| 1442 | U | C | C |
| 2466 | C | U | U |
| 4146 | A | G | G |
| 4803 | C | U | U |
| 6138 | C | U | C |
| 6426 | C | U | C |
| 6996 | C | U | C |
| 7938 | U | C | U |
| 9352 | C | U | U |

increased at nodes defining the INT and WN02 genotypes from 74 to 90% and 54 to 86% after removing the TX04 isolate (Fig. 2).

*Amino acid analysis*

We then analyzed the patterns of amino acid diversity among WNV isolates in our sample. The open reading frame translated into a full-length polyprotein alignment of 3433 amino acids which varied at 182 positions. Most amino acid variants occurred at low frequencies in this dataset. A total of 118 amino acid variants occurred only once in the sample and of the remaining 67 amino acid polymorphisms, most were shared by only two (n = 45) or three taxa (n = 12). Only one substitution, V → A at position 159 of the envelope protein, became

fixed after it first appeared in 2002. This change was mapped on to our phylogenetic tree (Fig. 2) and appears to have occurred in two separate lineages: once to give rise to the WN02 genotype and again within the INT genotype.

To search for evidence of convergent evolution, we mapped all the amino acid substitutions onto our phylogenetic tree. For this and all subsequent analyses, the aforementioned recombinant sequence (TX04) was excluded from the dataset. A total of 19 parallel amino substitutions were identified in unrelated lineages (Table 2). Five of these changes arose in both NY99 and WN02 genotypes, two in the intermediate and WN02 genotypes, and one change in all three genotypes. The remaining parallel substitutions were mapped along different branches within the WN02 genotype. In addition, one change was identified as a possible reversion back to the sequence of the 1998 Israel strain, S→P→S at position 54 of the NS5 protein. Only one of these changes Y→H at position 355 of the NS3 gene resulted in a charge difference.

*Selection analysis*

To determine whether amino acid positions are subject to negative or positive selection pressures, we used maximum likelihood methods to estimate and compare rates of synonymous (dS) versus non-synonymous (dN) substitution. Significant departures in these rates provide strong evidence for either negative selection against amino acid change (when dS>dN) or positive selection to change the protein sequence (when dN>dS). The mean dN/dS ratio for the entire coding sequence was 0.07, indicating that the vast majority of nucleotide substitutions are silent changes and that, overall, the WNV genome is subject to strong purifying (negative) selection. There was strong support for negative selection in 407 codons by FEL analysis and 204 codons by SLAC analysis. Positive selection was identified by FEL ($p=0.02$) and SLAC ($p=0.09$) analysis at position 135 of the NS4A gene, whereby a V→M substitution was inferred in 6 separate instances (Table 2).

## Discussion

In this study, we examined the molecular evolution of WNV in Connecticut by intensively sampling virus over 10 successive years. Our sampling was stratified by location during 2003 and by year in the town of Stratford to discern patterns of lineage turnover in a stable WNV focus. Viruses from Stratford were genetically diverse as evidenced by their positions throughout the phylogram. These isolates tended to group by year of isolation with 2 or more distinct viral strains or subclades appearing in a given year. Viruses sampled from a range of locations across Connecticut grouped into subclades that were detected transiently within a single location and sometimes contained viruses from different regions of the state. This indicates high rates of WNV dispersal in the environment and supports epidemiologic observations of rapid viral spread across North America. The population from Connecticut appears to be constantly turning over with an influx of new strains between years and within a given transmission season.

Our analysis also provides support for the occurrence of local overwintering of WNV in this region. One metric for the contribution of virus overwintering versus reintroduction is indicated by the degree of viral clade localization. We observed evidence of year-to-year continuity with two viral subclades persisting in Connecticut from 2002 to 2003 and one from 2003 to 2006. The virus could survive through winter in resident birds (Garmendia et al., 2000) or mosquitoes, as previously shown for *C. pipiens* (Andreadis et al., 2010; Bugbee and Forte, 2004; Farajollahi et al., 2005; Nasci et al., 2001). WNV has been detected in hibernating *C. pipiens* from nearby New York City (Andreadis et al., 2010; Nasci et al., 2001) and was shown to persist in unfed vertically-infected mosquitoes from Connecticut (Anderson et al., 2006; Anderson and Main, 2006). This could provide a plausible mechanism for local overwintering of virus as indicated in our analysis. Alternatively, the occasional instances of viral lineage persistence could represent the annual return of the same WNV strains to Connecticut from another geographic region. Given this scenario, the virus would still require a mechanism to overwinter during periods of mosquito inactivity, unless it was derived from a distant southern source where transmission could be continuous. This seems unlikely given the lack of evidence for stable year-around WNV transmission in the southeastern US, Caribbean Basin, or Central America. Moreover, migratory birds infected by WNV in any of these locations would likely clear the infection before arriving in the Northeast.

Our results indicate that there are multiple opportunities for WNV to reestablish transmission within a given locale by either re-introduction or local overwintering of virus. The population from Stratford appears to derive from a mixture of both local and exogenous virus strains in support of previous findings in Chicago, Illinois (Amore et al., 2010). However, none of the local clades appeared to be permanently established. Given the constant influx of new virus strains by an avian host, WNV transmission will become quickly reinstated once favorable environmental condition return.

**Table 2**
Predicted parallel amino acid changes.

| Protein | Site | Root | Inferred substitution | No. viruses w/ substitution | Location–year of amino acid variants | Genotype of amino acid variants |
|---|---|---|---|---|---|---|
| Capsid | 44 | I | $I\rightarrow_2T$ | 2 | CT03, CT04 | WN02 |
| Envelope | 159 | V | $V\rightarrow_2A$ | 104 | Widespread after 2002 | INT, WN02 |
| NS2A | 34 | M | $M\rightarrow_2L$ | 4 | MD00, NY02-03, CT03 | NY99, WN02 |
| | 43 | V | $V\rightarrow_2A$ | 2 | CT03, CT07 | WN02 |
| | 90 | M | $M\rightarrow_2V$ | 3 | CT01, CT06 | NY99, WN02 |
| | 119 | H | $H\rightarrow_2Y$ | 2 | FL03, CT03 | INT, WN02 |
| NS2B | 119 | V | $V\rightarrow_2I$ | 3 | CT01, CT03 | NY99, WN02 |
| NS3 | 106 | V | $V\rightarrow_2A$ | 2 | CT03, NY03 | NY99, WN02 |
| | 355 | Y | $Y\rightarrow_3H$ | 3 | CT03, CT06, CT08 | WN02 |
| | 466 | P | $P\rightarrow_2S$ | 2 | CT03, CT07 | WN02 |
| NS4A | 135 | V | $V\rightarrow_6M$ | 10 | CT02–03, CT05–06, TX03 | WN02 |
| NS4B | 202 | I | $I\rightarrow_2T$ | 2 | CO03, CT06 | WN02 |
| | 240 | I | $I\rightarrow_3M$ | 11 | CT00, CT04, CT08, NY03, TX02 | NY99, WN02 |
| | 241 | T | $T\rightarrow_2A$ | 4 | TX03, CT05–06 | WN02 |
| | 245 | I | $I\rightarrow_3V$ | 4 | CO03, CT00, FL03, MEX03 | INT, NY99, WN02 |
| NS5 | 54 | S | S→P→S | 1→132→1 | ISL98→N. America→CT03 | Reversion in WN02 |
| | 258 | V | $V\rightarrow_2A$ | 2 | CT03, CT06 | WN02 |
| | 312 | D | $D\rightarrow_2E$ | 2 | CT07, TX03 | WN02 |
| | 314 | K | $K\rightarrow_2R$ | 2 | AZ04, CT03 | WN02 |

Attempts to extinguish WNV transmission by vector control efforts will be temporary if applied focally or during a limited time period even if the measures are completely effective. Together, our findings suggest that WNV transmission is resilient to control attempts or unfavorable weather conditions because sites will quickly become recolonized by virus.

Phylogenetic patterns described in this study are strikingly similar that of other avian, mosquito-borne viruses found in this region of the US Eastern equine encephalitis virus (EEEV; *Togaviridae*; *Alphavirus*) and Highlands J virus (*Togaviridae*: *Alphavirus*) exhibit limited spatial structure in eastern North America, tending to group by year of isolation (Cilnis et al., 1996; Weaver et al., 1991). EEEV strains from northeastern US were shown to segregate into distinct clades that were detected regionally from 1 to 5 years (Armstrong et al., 2008; Young et al., 2008). These clades eventually disappeared to be superseded by new virus strains. These findings contrast sharply to the phylogeography of Jamestown canyon virus (JCV; *Bunyaviridae*; *Orthobunyavirus*) that perpetuates in a deer–mosquito cycle within this region (Andreadis et al., 2008; Armstrong and Andreadis, 2007). JCV was found to be geographically structured within Connecticut rather than temporally, in contrast to EEEV and WNV (Armstrong and Andreadis, 2007). JCV variants were stably maintained in this region and included one lineage detected over a 40 year period. The genetic structure of these mosquito-borne viruses appears to be influenced by the mobility of the main vertebrate host.

WNV isolates sequenced for this study were obtained mainly from mosquitoes collected during the statewide surveillance program and from an ecological study in the town of Stratford, but also included four viruses from American Crows in 1999. Given the limited flight range of mosquitoes relative to birds, infected mosquitoes are more likely to contain viruses circulating in proximity to their respective collection sites. Thus, our mosquito-based sample should provide accurate information about the spatial distribution of WNV variants within the state. Our conclusions about viral population change were based largely on our sample from Stratford. We chose this site because it's an active focus with recurrent WNV transmission each year, allowing us to track patterns of viral clade replacement. WNV turnover is likely to be more pronounced in sites with less stable transmission but such sites are not represented over multiple years in this study.

In this paper, we tracked the frequency and distribution of amino acid substitutions that arose during 10 years of WNV evolution. Most of these changes were detected at low frequencies which would be predicted for selectively neutral or nearly-neutral mutations in a large, expanding population. Under these conditions, most amino acid changes would be lost due to genetic drift, consistent with observations in this study. One notable exception was the E159 substitution that was rapidly driven to fixation within two years of its appearance in 2002, as noted in other US regions (Amore et al., 2010; Chisenhall and Mores, 2009; Davis et al., 2003, 2005; Ebel et al., 2004; Grinev et al., 2008; Herring et al., 2007; McMullen et al., 2011). This substitution may confer a selective advantage for the virus, allowing it to rapidly displace the NY99 genotype (Ebel et al., 2004). If this substitution is important, then we expect to observe evidence of strong negative selection acting on the codon position after the change. Negative selection is indicated by an excess of silent or synonymous changes to preserve the amino acid sequence. One such silent substitution was observed at this codon position in three viruses from Connecticut during 2003 (Genbank nos. HM756651, HM488176, and HM488227).

Evidence for convergent or parallel evolution across viral lineages was found by mapping amino acid substitutions on to the WNV phylogeny. There were a number of instances of the same amino acid substitution arising independently in different lineages. These results are consistent with an earlier analysis that included WNV isolates sampled from throughout the world and thus, with a much deeper evolutionary history (May et al., 2011). This suggests that only a limited number of amino acid changes are permitted due to functional constraints. The adaptive benefit of these convergent changes is not clear; however, positive selection was identified at one position (135 V→M of the NS4A protein). This amino acid substitution was found in a number of lineages circulating in Connecticut and in Texas. The selection pressure forcing this change might be related to the putative functions of the NS4A protein. This includes involvement in the viral RNA replication complex (Mackenzie et al., 1998; Shiryaev et al., 2009) and immune evasion by interfering with interferon signaling pathways (Liu et al., 2004, 2006). Similar examples of positive selection were identified in WNV from North America, including another amino acid substitution at position 85 A→T of the NS4A protein (McMullen et al., 2011). Adaptive convergent evolution was also identified at position 249 T→P of the NS3 helicase, occurring prior to WNV introduction into North America (Brault et al., 2007). This amino acid substitution was shown to affect viral growth properties and virulence in American Crows, and was associated with outbreaks of avian disease. Future monitoring and functional analyses are needed to assess the significance of the changes identified in this study.

During the course of our analysis, we observed a possible instance of intermolecular recombination among WNV strains circulating in North America. Out of 135 viral genomes examined, a single WNV sequence from Texas displayed the genetic signature of viral recombination. Infrequent episodes of genetic recombination have been inferred for a number of flaviviruses (Twiddy and Holmes, 2003), including among WNV strains circulating in Africa (Pickett and Lefkowitz, 2009). This process could result in rapid genetic change; however, these observations should be interpreted with caution. Natural recombinants should be verified, ideally by re-sequencing plaque-purified virus to ensure against a possible mixed infection and sequencing artifacts (Rico-Hesse, 2003). The recombinant sequence identified in this study was unexpected but clearly shows the imprint of hybridization among strains in the NY99 and WN02 genotypes. This is based on the analysis of a previously submitted Genbank sequence and therefore, requires further confirmation.

In conclusion, our analysis describes the patterns of viral lineage turnover and protein evolution within a region supporting stable WNV transmission. We observed evidence of local overwintering of virus but without permanent establishment of local populations. Moreover, we documented the monthly and yearly appearance of distinct variants, implying rapid re-colonization of virus in a given locale. Numerous nucleotide changes have arisen since its introduction into North America, but negative selection appeared to constrain changes at the protein level. Finally, we identified several instances of convergent evolution, including one amino acid change that appears to be positively selected.

## Materials and methods

### *Virus strains*

WNV isolates sequenced in this study were recovered from mosquitoes (n=95) or crows (n=4) collected during the statewide surveillance program in Connecticut (Andreadis et al., 2004) or during targeted mosquito trapping efforts in Stratford, Connecticut (Anderson et al., 2006) (Supplementary Table). Of these, 57 viruses were obtained each year from an active WNV focus in Stratford during 1999 and 2001–2008 or in the adjacent towns of Milford and Shelton during 2000. An additional 42 WNV isolates were derived from mosquitoes collected in 2003 from 21 towns throughout Connecticut. Bird tissues and whole mosquitoes were processed and screened for virus infection in Vero cell cultures as previously described (Andreadis et al., 2004).

### *RNA isolation and tiling PCR*

Viral RNA was isolated from primary WNV cultures (QIAmp viral RNA mini kit, Qiagen) and the RNA genome reverse transcribed to

cDNA with Superscript III reverse transcriptase (Invitrogen), random hexamers (Roche) and a specific oligonucleotide targeting the 3′ end of the target genome sequences. Four overlapping PCR products, each of size ~3 kb, were designed to capture the WNV coding region. Four primer pairs (1F: AGTAGTTCGCCTGTGTGAGCTGAC; 1R: ATGGGCCCTGGTTTTGTGTCTTGT; 2F: CGGCAAGAGCTGAGATGTG-GAAGT; 2R: CCTCAGTCCAATGGGCGAAGTT; 3F: CGCCGGTAAAACAAG-GAGGATTCT; 3R: GCAGCCAGTCCTCAACCATTTCAA ; 4F: KACGGTRACAGCGGCAACAC; 4R: CGGTTCTGAGGGCTTACATG ) were synthesized with a 5′ amino modifier C6 (Integrated DNA Technologies) to prevent ligation in the 454 library construction and allow for even coverage across the genome (ref: Harismendy O, Frazer K. Biotechniques. 2009 Mar;46(3):229-31. PMID: 19317667). PCR was performed using the high fidelity polymerase PfuUltra II Fusion HS DNA polymerase (Stratagene) with 40 cycles of amplification. For post-PCR quality control purposes the products were run on pre-cast 1% agarose E-Gels (Invitrogen). Each reaction was quantified using the Quant-iT PicoGreen dsDNA assay (Invitrogen). Based on these concentrations, 50 ngs of each reaction was pooled for a total of 200 ngs and the volume was brought up to 100 uL with TE for shearing and library construction.

*Library construction, sequencing, and assembly*

Whole WNV genomes were sequenced using the Broad Institute's viral genome sequencing and assembly pipeline (http://www.broadinstitute.org/annotation/viral/WNV/). Pooled PCR products were prepared for sequencing on the 454 Genome Sequencer FLX Titanium (Roche) using standard protocols with the following modifications. Each sample received a 454 library adapter that had been synthesized with an in-house designed 5–8 base molecular tag or barcode (Lennon et al. (2010) Genome Biology 11(2):R15). Post adapter ligation, sample batches of up to 48 were pooled by volume to create sequence-ready libraries. Emulsion PCR and sequencing were performed according to manufacturers' protocols. The library was loaded in to a picotiter plate (PTP) yielding ~50× coverage for each sample. Sequence reads were binned by molecular barcode and sent to their respective project directories for assembly and analysis. Resulting sequence reads were trimmed of primer sequences, filtered for high quality, assembled *de novo* and annotated using the Broad Institute's in-house viral assembly and annotation algorithms. All genome sequences newly determined here have been deposited in GenBank and assigned accession numbers (Supplementary Table).

*Genetic analysis*

WNV sequences were combined with 36 previously published sequences available on Genbank (Supplementary Table) for a total of 135 sequences and aligned by the ClustalW algorithm. Phylogenetic relationships were evaluated by maximum-likelihood (ML) analysis in Mega 5.2 (Tamura et al., in press). The analysis employed the GTR + G + I substitution model and nearest neighbor interchange heuristic search method. The optimal nucleotide substitution model was identified and implemented after performing ML fits of 24 different models in Mega. Support for individual nodes was obtained by performing 500 bootstrap replicates. The nucleotide substitution rate was estimated by dividing the number of base substitutions from the NY99 strain (Genbank no. AF196835) by the number of years of divergence for all sequences from 2000 to 2008. Estimates of evolutionary divergence were conducted using the maximum composite likelihood model in Mega.

A second dataset of 94 WNV sequences was created by eliminating identical or nearly identical sequences, and contained the first 9999 bps of the open reading frame, in order to comply with alignment size restrictions for recombination and selection detection programs. Evidence for recombination in the alignment was tested by performing single break point (SBP) and genetic algorithm recombination detection (GARD) methods using the Datamonkey web server (http://www.datamonkey.org) (Kosakovsky Pond et al., 2006). Tests for positive and negative selection were performed by the fixed effects likelihood (FEL) and single-likelihood ancestor counting (SLAC) methods on the Datamonkey web server (Kosakovsky Pond and Frost, 2005; Pond and Frost, 2005).

Supplementary materials related to this article can be found online at doi:10.1016/j.virol.2011.06.006.

## Acknowledgments

We thank the following members of our support staff for their technical assistance: Angela Bransfield, Shannon Finan, Bonnie Hamid, John Shepard, and Michael Thomas. This work was supported in whole or in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract HHSN272200900018C, Centers for Disease Control and Prevention (U50/CCU116806-01-1), and the US Department of Agriculture (58-6615-1-218, CONH00768, and CONH00773). KP was supported by National Institutes of Health, National Research Service Award Institutional Training Grant 5T32-AI07538-13.

## References

Amore, G., Bertolotti, L., Hamer, G.L., Kitron, U.D., Walker, E.D., Ruiz, M.O., Brawn, J.D., Goldberg, T.L., 2010. Multi-year evolutionary dynamics of West Nile virus in suburban Chicago, USA, 2005–2007. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 365 (1548), 1871–1878.

Anderson, J.F., Main, A.J., 2006. Importance of vertical and horizontal transmission of West Nile virus by *Culex pipiens* in the Northeastern United States. J. Infect. Dis. 194 (11), 1577–1579.

Anderson, J.F., Andreadis, T.G., Vossbrinck, C.R., Tirrell, S., Wakem, E.M., French, R.A., Garmendia, A.E., Van Kruiningen, H.J., 1999. Isolation of West Nile virus from mosquitoes, crows, and a Cooper's hawk in Connecticut. Science 286 (5448), 2331–2333.

Anderson, J.F., Vossbrinck, C.R., Andreadis, T.G., Iton, A., Beckwith III, W.H., Mayo, D.R., 2001. A phylogenetic approach to following West Nile virus in Connecticut. Proc. Natl. Acad. Sci. U S A 98 (23), 12885–12889.

Anderson, J.F., Andreadis, T.G., Main, A.J., Ferrandino, F.J., Vossbrinck, C.R., 2006. West Nile virus from female and male mosquitoes (Diptera: Culicidae) in subterranean, ground, and canopy habitats in Connecticut. J. Med. Entomol. 43 (5), 1010–1019.

Andreadis, T.G., Anderson, J.F., Vossbrinck, C.R., Main, A.J., 2004. Epidemiology of West Nile virus in Connecticut: a five-year analysis of mosquito data 1999–2003. Vector Borne Zoonotic Dis. 4 (4), 360–378.

Andreadis, T.G., Anderson, J.F., Armstrong, P.M., Main, A.J., 2008. Isolations of Jamestown Canyon virus (Bunyaviridae: Orthobunyavirus) from field-collected mosquitoes (Diptera: Culicidae) in Connecticut, USA: a ten-year analysis, 1997–2006. Vector Borne Zoonotic Dis. 8 (2), 175–188.

Andreadis, T.G., Armstrong, P.M., Bajwa, W.I., 2010. Studies on hibernating populations of *Culex pipiens* from a West Nile virus endemic focus in New York City: parity rates and isolation of West Nile virus. J. Am. Mosq. Control. Assoc. 26 (3), 257–264.

Armstrong, P.M., Andreadis, T.G., 2007. Genetic relationships of Jamestown Canyon virus strains infecting mosquitoes collected in Connecticut. Am. J. Trop. Med. Hyg. 77 (6), 1157–1162.

Armstrong, P.M., Andreadis, T.G., Anderson, J.F., Stull, J.W., Mores, C.N., 2008. Tracking eastern equine encephalitis virus perpetuation in the northeastern United States by phylogenetic analysis. Am. J. Trop. Med. Hyg. 79 (2), 291–296.

Bertolotti, L., Kitron, U., Goldberg, T.L., 2007. Diversity and evolution of West Nile virus in Illinois and the United States, 2002–2005. Virology 360 (1), 143–149.

Brault, A.C., Huang, C.Y., Langevin, S.A., Kinney, R.M., Bowen, R.A., Ramey, W.N., Panella, N.A., Holmes, E.C., Powers, A.M., Miller, B.R., 2007. A single positively selected West Nile viral mutation confers increased virogenesis in American crows. Nat. Genet. 39 (9), 1162–1166.

Bugbee, L.M., Forte, L.R., 2004. The discovery of West Nile virus in overwintering *Culex pipiens* (Diptera: Culicidae) mosquitoes in Lehigh County, Pennsylvania. J. Am. Mosq. Control. Assoc. 20 (3), 326–327.

Chisenhall, D.M., Mores, C.N., 2009. Diversification of West Nile virus in a subtropical region. Virol. J. 6, 106.

Cilnis, M.J., Kang, W., Weaver, S.C., 1996. Genetic conservation of Highlands J viruses. Virology 218 (2), 343–351.

Davis, C.T., Beasley, D.W., Guzman, H., Raj, R., D'Anton, M., Novak, R.J., Unnasch, T.R., Tesh, R.B., Barrett, A.D., 2003. Genetic variation among temporally and geographically distinct West Nile virus isolates, United States, 2001, 2002. Emerg. Infect. Dis. 9 (11), 1423–1429.

Davis, C.T., Ebel, G.D., Lanciotti, R.S., Brault, A.C., Guzman, H., Siirin, M., Lambert, A., Parsons, R.E., Beasley, D.W., Novak, R.J., Elizondo-Quiroga, D., Green, E.N., Young, D.S., Stark, L.M., Drebot, M.A., Artsob, H., Tesh, R.B., Kramer, L.D., Barrett, A.D., 2005.

Phylogenetic analysis of North American West Nile virus isolates, 2001–2004: evidence for the emergence of a dominant genotype. Virology 342 (2), 252–265.
Ebel, G.D., Carricaburu, J., Young, D., Bernard, K.A., Kramer, L.D., 2004. Genetic and phenotypic variation of West Nile virus in New York, 2000–2003. Am. J. Trop. Med. Hyg. 71 (4), 493–500.
Farajollahi, A., Crans, W.J., Bryant, P., Wolf, B., Burkhalter, K.L., Godsey, M.S., Aspen, S.E., Nasci, R.S., 2005. Detection of West Nile viral RNA from an overwintering pool of *Culex pipens pipiens* (Diptera: Culicidae) in New Jersey, 2003. J. Med. Entomol. 42 (3), 490–494.
Garmendia, A.E., Van Kruiningen, H.J., French, R.A., Anderson, J.F., Andreadis, T.G., Kumar, A., West, A.B., 2000. Recovery and identification of West Nile virus from a hawk in winter. J. Clin. Microbiol. 38 (8), 3110–3111.
Grinev, A., Daniel, S., Stramer, S., Rossmann, S., Caglioti, S., Rios, M., 2008. Genetic variability of West Nile virus in US blood donors, 2002–2005. Emerg. Infect. Dis. 14 (3), 436–444.
Herring, B.L., Bernardin, F., Caglioti, S., Stramer, S., Tobler, L., Andrews, W., Cheng, L., Rampersad, S., Cameron, C., Saldanha, J., Busch, M.P., Delwart, E., 2007. Phylogenetic analysis of WNV in North American blood donors during the 2003–2004 epidemic seasons. Virology 363 (1), 220–228.
Kilpatrick, A.M., Meola, M.A., Moudy, R.M., Kramer, L.D., 2008. Temperature, viral genetics, and the transmission of West Nile virus by *Culex pipiens* mosquitoes. PLoS Pathog. 4 (6), e1000092.
Komar, N., 2003. West Nile virus: epidemiology and ecology in North America. Adv. Virus Res. 61, 185–234.
Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22 (5), 1208–1222.
Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D., 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. 23 (10), 1891–1901.
Kramer, L.D., Styer, L.M., Ebel, G.D., 2008. A global perspective on the epidemiology of West Nile virus. Annu. Rev. Entomol. 53, 61–81.
Lanciotti, R.S., Roehrig, J.T., Deubel, V., Smith, J., Parker, M., Steele, K., Crise, B., Volpe, K.E., Crabtree, M.B., Scherret, J.H., Hall, R.A., MacKenzie, J.S., Cropp, C.B., Panigrahy, B., Ostlund, E., Schmitt, B., Malkinson, M., Banet, C., Weissman, J., Komar, N., Savage, H.M., Stone, W., McNamara, T., Gubler, D.J., 1999. Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. Science 286 (5448), 2333–2337.
Lanciotti, R.S., Ebel, G.D., Deubel, V., Kerst, A.J., Murri, S., Meyer, R., Bowen, M., McKinney, N., Morrill, W.E., Crabtree, M.B., Kramer, L.D., Roehrig, J.T., 2002. Complete genome sequences and phylogenetic analysis of West Nile virus strains isolated from the United States, Europe, and the Middle East. Virology 298 (1), 96–105.
Lennon, N.J., Lintner, R.E., Anderson, S., Alvarez, P., Barry, A., Brockman, W., Daza, R., Erlich, R.L., Giannoukos, G., Green, L., Hollinger, A., Hoover, C.A., Jaffe, D.B., Juhn, F., McCarthy, D., Perrin, D., Ponchner, K., Powers, T.L., Rizzolo, K., Robbins, D., Ryan, E., Russ, C., Sparrow, T., Stalker, J., Steelman, S., Weiand, M., Zimmer, A., Henn, M.R., Nusbaum, C., Nicol, R., 2010. A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. Genome Biol. 11 (2), R15.
Liu, W.J., Chen, H.B., Wang, X.J., Huang, H., Khromykh, A.A., 2004. Analysis of adaptive mutations in Kunjin virus replicon RNA reveals a novel role for the flavivirus nonstructural protein NS2A in inhibition of beta interferon promoter-driven transcription. J. Virol. 78 (22), 12225–12235.
Liu, W.J., Wang, X.J., Clark, D.C., Lobigs, M., Hall, R.A., Khromykh, A.A., 2006. A single amino acid substitution in the West Nile virus nonstructural protein NS2A disables its ability to inhibit alpha/beta interferon induction and attenuates virus virulence in mice. J. Virol. 80 (5), 2396–2404.
Mackenzie, J.M., Khromykh, A.A., Jones, M.K., Westaway, E.G., 1998. Subcellular localization and some biochemical properties of the flavivirus Kunjin nonstructural proteins NS2A and NS4A. Virology 245 (2), 203–215.
May, F.J., Davis, C.T., Tesh, R.B., Barrett, A.D., 2011. Phylogeography of West Nile virus: from the cradle of evolution in Africa to Eurasia, Australia, and the Americas. J. Virol. 85 (6), 2964–2974.
McMullen, A.R., May, F.J., Guzman, H., Bueno, R., Dennett, J.A., Tesh, R.B., Barrett, A.D.T., 2011. Evolution of new genotype of West Nile Virus in North America. Emerg. Infect. Dis. 17 (5), 785–793.
Molaei, G., Andreadis, T.G., Armstrong, P.M., Anderson, J.F., Vossbrinck, C.R., 2006. Host feeding patterns of *Culex* mosquitoes and West Nile virus transmission, northeastern United States. Emerg. Infect. Dis. 12 (3), 468–474.
Moudy, R.M., Meola, M.A., Morin, L.L., Ebel, G.D., Kramer, L.D., 2007. A newly emergent genotype of West Nile virus is transmitted earlier and more efficiently by *Culex* mosquitoes. Am. J. Trop. Med. Hyg. 77 (2), 365–370.
Nasci, R.S., Savage, H.M., White, D.J., Miller, J.R., Cropp, B.C., Godsey, M.S., Kerst, A.J., Bennett, P., Gottfried, K., Lanciotti, R.S., 2001. West Nile virus in overwintering *Culex* mosquitoes, New York City, 2000. Emerg. Infect. Dis. 7 (4), 742–744.
Pickett, B.E., Lefkowitz, E.J., 2009. Recombination in West Nile Virus: minimal contribution to genomic diversity. Virol. J. 6, 165.
Pond, S.L., Frost, S.D., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21 (10), 2531–2533.
Rico-Hesse, R., 2003. Microevolution and virulence of dengue viruses. Adv. Virus Res. 59, 315–341.
Shiryaev, S.A., Chernov, A.V., Aleshin, A.E., Shiryaeva, T.N., Strongin, A.Y., 2009. NS4A regulates the ATPase activity of the NS3 helicase: a novel cofactor role of the non-structural protein NS4A from West Nile virus. J. Gen. Virol. 90 (Pt 9), 2081–2085.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., in press. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. [Epub ahead of print]
Twiddy, S.S., Holmes, E.C., 2003. The extent of homologous recombination in members of the genus *Flavivirus*. J. Gen. Virol. 84 (Pt 2), 429–440.
Weaver, S.C., Barrett, A.D., 2004. Transmission cycles, host range, evolution and emergence of arboviral disease. Nat. Rev. Microbiol. 2 (10), 789–801.
Weaver, S.C., Scott, T.W., Rico-Hesse, R., 1991. Molecular evolution of eastern equine encephalomyelitis virus in North America. Virology 182 (2), 774–784.
Young, D.S., Kramer, L.D., Maffei, J.G., Dusek, R.J., Backenson, P.B., Mores, C.N., Benard, K.A., Ebel, G.D., 2008. Molecular epidemiology of eastern equine encephalitis virus, New York. Emerg. Infect. Dis. 14 (3), 454–460.