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 5.83×10^{-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.

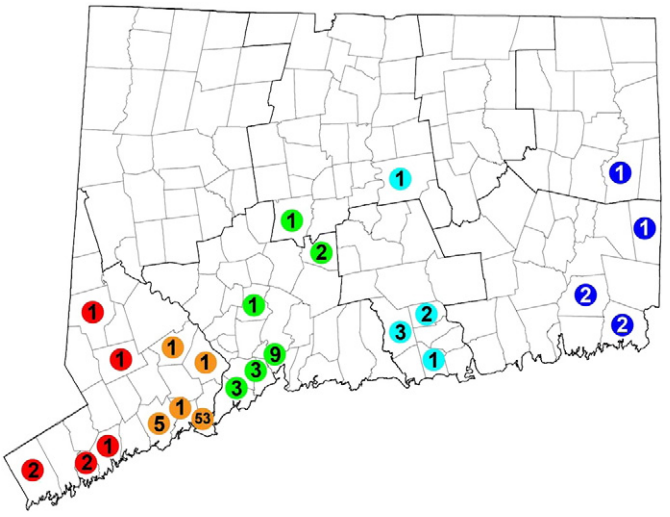

Fig. 1. Map of Connecticut showing the geographic location and number of WNV isolates analyzed in this study. Collection sites are color coded from west (warm colors) to east (cool colors) along a longitudinal gradient.

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