Diversity and evolution of West Nile virus In Illinois and The United States, 2002–2005

Available online at www.sciencedirect.com
ScienceDirect
Virology 360 (2007) 143–149
VIROLOGY
www.elsevier.com/locate/yviro

Diversity and evolution of West Nile virus in Illinois and the United States, 2002–2005

Luigi Bertolotti[1], Uriel Kitron, Tony L. Goldberg *

University of Illinois, Department of Pathobiology, 2001 South Lincoln Avenue, Urbana, IL 61802, USA

Received 8 August 2006; returned to author for revision 23 August 2006; accepted 14 October 2006
Available online 20 November 2006

Abstract

Evolutionary analyses of West Nile virus (WNV) have been limited by uneven sampling across geographic regions and over time. In this study, an expanded data set of 68 WNV envelope gene sequences from the Midwest (Illinois) was created and combined with published sequences to investigate spatial and temporal structuring in the United States viral population. Results indicate an overall lack of geographic structure to WNV in the United States, supporting the notion of WNV as a rapidly expanding pathogen not significantly restricted in its spread by geographic distance. However, analyses of viral genetic diversity show a steady increase in WNV nucleotide-level diversity over time. Additionally, evolutionary rate calculations indicate that WNV has evolved at approximately 0.85×10^{-3} substitutions/site/year, largely through neutral substitution and purifying selection. Overall, these results show WNV across the United States to be a panmictic viral population that is diversifying and evolving.

Keywords: West Nile virus; Flavivirus; Molecular epidemiology; Viral evolution; Phylogenetics

Introduction

West Nile virus (WNV; Flaviviridae; *Flavivirus*) is a single stranded, positive-sense RNA virus member of the Japanese encephalitis (JE) serocomplex that is transmitted primarily through the bite of infected mosquitoes. WNV was first isolated and identified in Uganda in 1937 (Smithburn et al., 1940) and was introduced into the USA approximately 60 years later (Lanciotti et al., 1999). After its initial appearance in New York in 1999 (Lanciotti et al., 1999), WNV spread rapidly throughout the United States, Canada (Gancz et al., 2004), and Mexico (Estrada-Franco et al., 2003). The virus is now endemic in North America.

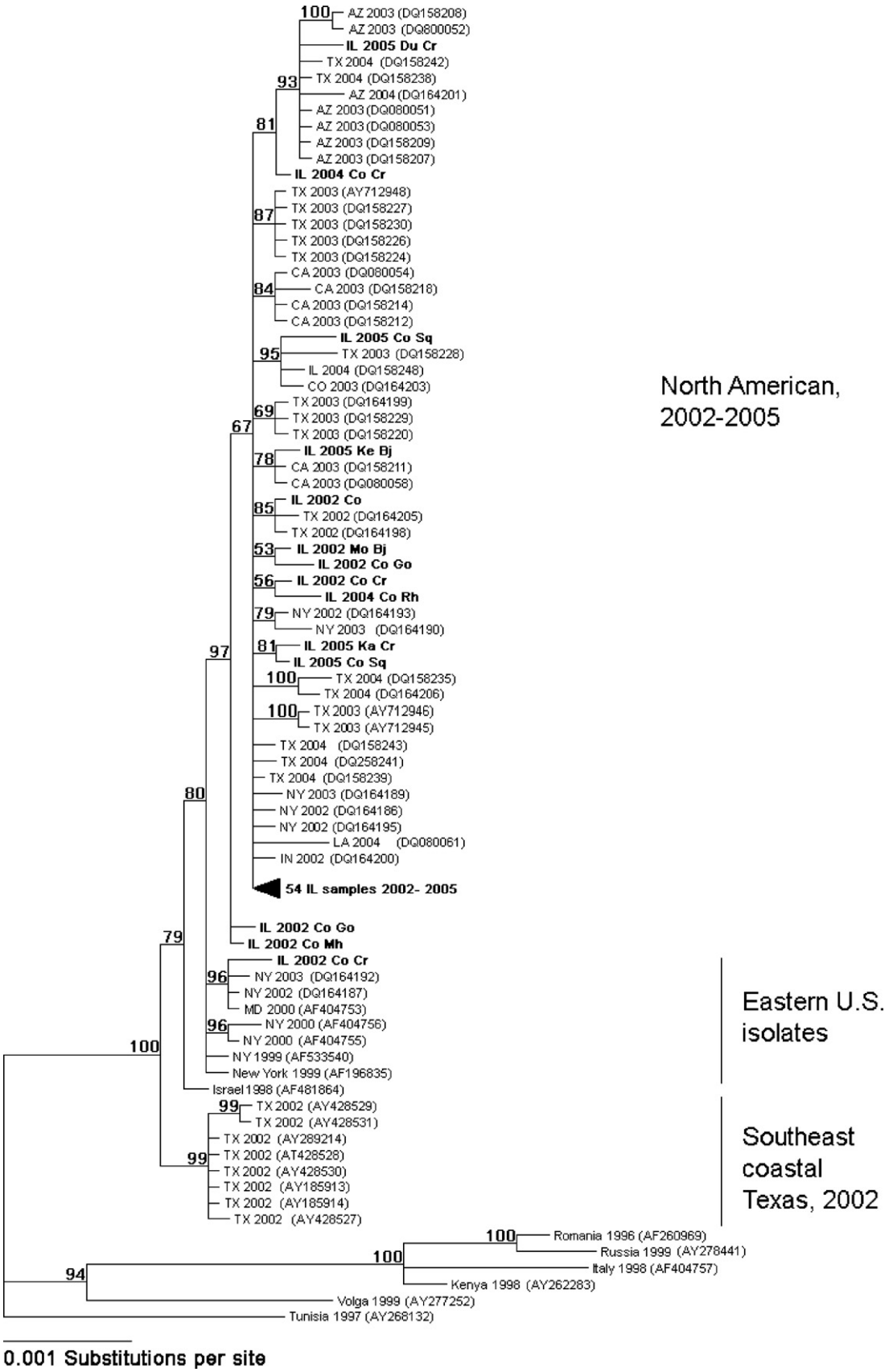
Initial investigations of WNV indicated that the virus appeared to have low genetic variability (Anderson et al., 2001; Lanciotti et al., 2002; Beasley et al., 2003). The requirement for the virus to maintain infectivity in both vertebrate hosts and insect vectors has been cited as a constraint accounting for this property (Jerzak et al., 2005). Nevertheless, more recent research has suggested that a dominant strain may have begun to replace earlier North American WNV variants (Davis et al., 2005). The rate at which WNV in the United States may be evolving or adapting to local conditions is currently unclear.

To date, studies of the evolution of WNV in the USA have been hampered by lack of representative data from multiple geographic regions over multiple years. Conspicuously lacking have been adequate viral sequence data from the Midwestern United States, where the virus has had a profound impact on avian and human health. In 2002, Illinois led the nation in WNV-related human cases and deaths (Huhn et al., 2005). Despite the importance of states like Illinois to the epidemiology of WNV, only six WNV genetic sequences from this region were available in GenBank as of June 2006.

To add to available WNV sequence data, and to enhance representation of the Midwestern U.S. region, we generated a large number of WNV envelope gene sequences from the state of Illinois, selected to span the widest possible range of geographic locations and times of collection. By combining our

* Corresponding author. Fax: +1 217 244 7421.
E-mail address: tlgoldbe@uiuc.edu (T.L. Goldberg).
[1] Present address: Università degli Studi di Torino, Facoltà di Medicina Veterinaria, Via L. da Vinci 44, Grugliasco (TO), Italy.

0042-6822/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2006.10.030


AZ 2003 (DQ158208)
AZ 2003 (DQ800052)
IL 2005 Du Cr
TX 2004 (DQ158242)
TX 2004 (DQ158238)
AZ 2004 (DQ164201)
AZ 2003 (DQ080051)
AZ 2003 (DQ080053)
AZ 2003 (DQ158209)
AZ 2003 (DQ158207)
IL 2004 Co Cr
TX 2003 (AY712948)
TX 2003 (DQ158227)
TX 2003 (DQ158230)
TX 2003 (DQ158226)
TX 2003 (DQ158224)
CA 2003 (DQ080054)
CA 2003 (DQ158218)
CA 2003 (DQ158214)
CA 2003 (DQ158212)
IL 2005 Co Sq
TX 2003 (DQ158228)
IL 2004 (DQ158248)
CO 2003 (DQ164203)
TX 2003 (DQ164199)
TX 2003 (DQ158229)
TX 2003 (DQ158220)
IL 2005 Ke Bj
CA 2003 (DQ158211)
CA 2003 (DQ080058)
IL 2002 Co
TX 2002 (DQ164205)
TX 2002 (DQ164198)
IL 2002 Mo Bj
IL 2002 Co Go
IL 2002 Co Cr
IL 2004 Co Rh
NY 2002 (DQ164193)
NY 2003 (DQ164190)
IL 2005 Ka Cr
IL 2005 Co Sq
TX 2004 (DQ158235)
TX 2004 (DQ164206)
TX 2003 (AY712946)
TX 2003 (AY712945)
TX 2004 (DQ158243)
TX 2004 (DQ258241)
TX 2004 (DQ158239)
NY 2003 (DQ164189)
NY 2002 (DQ164186)
NY 2002 (DQ164195)
LA 2004 (DQ080061)
IN 2002 (DQ164200)
54 IL samples 2002- 2005
IL 2002 Co Go
IL 2002 Co Mh
IL 2002 Co Cr
NY 2003 (DQ164192)
NY 2002 (DQ164187)
MD 2000 (AF404753)
NY 2000 (AF404756)
NY 2000 (AF404755)
NY 1999 (AF533540)
New York 1999 (AF196835)
Israel 1998 (AF481864)
TX 2002 (AY428529)
TX 2002 (AY428531)
TX 2002 (AY289214)
TX 2002 (AT428528)
TX 2002 (AY428530)
TX 2002 (AY185913)
TX 2002 (AY185914)
TX 2002 (AY428527)
Romania 1996 (AF260969)
Russia 1999 (AY278441)
Italy 1998 (AF404757)
Kenya 1998 (AY262283)
Volga 1999 (AY277252)
Tunisia 1997 (AY268132)
North American, 2002-2005
Eastern U.S. isolates
Southeast coastal Texas, 2002
0.001 Substitutions per site

new data with published sequences from other regions of the country, we were able to conduct a relatively thorough examination of the geographic pattern of genetic diversity and rate of evolution of WNV within the USA.

Results

We generated 68 new WNV ENV sequences of 1575 nucleotides each, representing 28 Illinois counties, 9 host species, and four years (2002–2005). All new sequences, including information on source, collection date, and geographic origin, were deposited in GenBank (Accession No. DQ874402–DQ874469). A phylogenetic tree representing the relationships among these sequences and selected sequences from the literature is shown in Fig. 1. Little geographic substructuring is evident in the tree, in that major sub-clades (similar to those identified by Davis et al., 2005) contain sequences from different regions of the United States. The position of the Israel 1998 sequence (AF841864) within the North American clade is probably spurious, resulting from small numbers of phylogenetically informative nucleotide positions; Davis et al. (2005) have previously noted this phenomenon in trees built from partial genome sequences.

Table 1 shows the results of hierarchical analyses of molecular variance conducted to examine geographic and temporal subdivisioning within WNV (184 sequences included). Most WNV genetic diversity is contained within years or within States (91% and 89% of total variance, respectively, $P<0.001$ for both values). Temporal subdivisioning among years and geographical subdivisioning among regions/within states are low but statistically significant (8.7% and 11.29%, respectively, $P<0.001$ for both values). Results did not differ substantially when the analysis was repeated on 33 complete WNV genomes from the published literature (93% and 91% of total variance within years and states, respectively, 7.21% of total variance among years, and 19.7% of total variance among regions/within states).

Based on results from our analyses of molecular variance (AMOVA), we investigated changes in the genetic diversity of WNV within the Midwest over time. We considered only the Midwest because this was the only region where we had adequate sample sizes from 2002 to 2005. We expressed genetic diversity as nucleotide diversity (Nei, 1987), or mean proportion of nucleotide differences among sequences within each year, corrected for multiple substitutions using the same model of molecular evolution used during phylogenetic analyses. Fig. 2 shows a significant trend of increasing genetic diversity over time ($R^2=0.36$, $t=21.7$, $P<0.0001$). The slope of the line is 0.73×10^{-3}, indicating that the WNV ENV gene in the Midwest has diversified at a rate of approximately 0.07% per year between 2002 and 2005.

Table 1
Hierarchical analyses of molecular variance (AMOVA) for WNV envelope gene sequence data from the United States, 1999–2005[a]

Variance component	*df*	Variance	% Total	P[b]	Φ Statistic
Among years[c]					
Among years σ^{2a}	6	0.19288	8.70	<0.0001	$\Phi_{ST}=0.08705$
Within years σ^{2c}	177	2.02296	91.30		
Among regions[d]					
Among regions σ^{2a}	3	−0.00812	−0.37	0.22847	$\Phi_{CT}=-0.00366$
Among states, within regions σ^{2b}	13	0.25069	11.29	<0.0001	$\Phi_{SC}=0.11248$
Within states σ^{2c}	167	1.97797	89.08	<0.0001	$\Phi_{ST}=0.10924$

[a] Viral sequences were divided into either years (1999–2005) or geographic regions (Northeast, South, Midwest, and West).
[b] Probability of obtaining a more extreme variance component and Φ statistic than the observed values by chance alone, calculated from 16,000 random permutations of the data, implemented with the computer program ARLEQUIN, version 3.01 (Excoffier et al., 2005).
[c] Sample sizes: 1999 ($n=5$), 2000 ($n=4$), 2001 ($n=2$), 2002 ($n=67$), 2003 ($n=57$), 2004 ($n=27$), 2005 ($n=22$).
[d] Sample sizes: Northeast ($n=19$), South ($n=61$), Midwest ($n=78$), West ($n=26$).

Results of our analysis of WNV evolutionary rate (Fig. 3) show a positive and significant association between evolutionary distance from the United States common ancestor and time ($R^2=0.29$, $t=7.0$, $P<0.0001$). The rate estimate generated using a Bayesian method incorporating relaxed molecular clock assumptions implemented with the computer program BEAST (Drummond and Rambaut, 2003) was 0.85×10^{-3}, with 95% confidence limits of 0.66×10^{-3} to 1.06×10^{-3}. Synonymous substitutions account almost entirely for this rate of evolution ($R^2=0.20$; $t=5.38$; $P<0.0001$); inferred numbers of non-synonymous substitutions were not significantly related to time ($R^2=0.018$; $t=1.49$; $P=0.14$).

Discussion

Our results demonstrate that WNV within the United States, as represented by the ENV gene, shows very little geographic subdivisioning at regional scale. The overall topology of our tree (Fig. 1) is consistent with previous studies showing broad geographic admixture among isolates and the presence of a basal coastal Texas clade (Davis et al., 2005). However, we have documented significant temporal subdivisioning in the virus, manifest both as an increase in genetic diversity over time

Fig. 1. Phylogenetic tree constructed by Bayesian analysis of 132 WNV envelope gene sequences (length 1575 nucleotides). Names of groups of sequences follow Davis et al. (2005). Newly sequenced viruses are indicated in bold. Illinois sequence names are followed by the year of collection, county name (Bo: Bond, Ch: Champaign, Cl: Clinton, Co: Cook, Cr: Crawford, Dk: Dekalb, Do: Douglas, Du: Dupage, Fa: Fayette, Gr: Grundy, He: Henry, Ir: Iroquois, Ka: Kane, Ke: Kendall, La: Lake, Le: Lee, Ma: Macoupin, Mh: Mchenry, Ml: Mclean, Mo: Moultrie, Pe: Perry, Ri: Rock Island, Rf: Rockford, Sc: Saintclair, Sp: Springfield, Ve: Vermilion, Wl: Will, Wi: Winnebago), and species of origin (Go: great horned owl, *Bubo virginianus*; Rh: red tailed hawk, *Buteo jamaicensis*; Fi: finch, *Carpodacus mexicanus*; Cr: crow, *Corvus brachyrhynchos*; Bj: blue jay, *Cyanocitta cristata*; Mh: merlin hawk, *Falco columbarius*; Ke: kestrel, *Falco punctatus*; Ma: magpie, *Pica pica*; Sq: squirrel, *Sciurus carolinensis*). Other United States sequence names include state abbreviation, collection year, and GenBank accession number (in parentheses). Posterior probabilities of clades are indicated above branches.

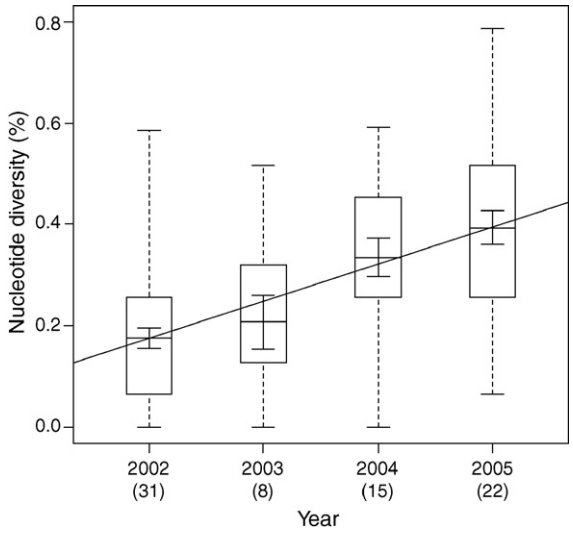


Fig. 2. Genetic diversity in WNV from the Midwestern USA, 2002–2005. Genetic diversity is expressed as nucleotide diversity (Nei, 1987), or mean proportion of nucleotide differences among sequences within each year, corrected for multiple substitutions using a model of molecular evolution derived from a hierarchical likelihood ratio test approach implemented using the computer program Modeltest, version 3.7 (Posada and Crandall, 2001). Lines and small error bars indicate means and standard errors, respectively. Boxes indicate third quartiles, and large error bars indicate ranges. The solid line is a least squares regression line indicating the rate of increase of viral genetic diversity over time ($R^2=0.3626$, $t=21.69$, $P<0.0001$). Sample sizes for each year are indicated in parentheses.

within the Midwest and as a significant rate of neutral evolution for the viral population across the USA. Our evolutionary rate data are consistent with the introduction of WNV into the USA in 1999, in that our analysis could not exclude 1999 as the year of the last common ancestor of all United States WNV sequences.

The low degree of geographic subdivisioning evident in our analyses indicates that ample viral gene flow has occurred among regions of the USA since the introduction of WNV. This may reflect rapid spread of the virus from the Northeast to the rest of the country and the occurrence of multiple long-distance transmission events. Patterns of bird migration within the USA might have predicted a different result; migratory pathways are used to explain, at least in part, the movement and spread of avian infections, such that Eastern, Midwestern, and Western regional differences in WNV might have been expected (Reed et al., 2003). Moreover arthropod populations maintain WNV infection locally and over the winter (Tesh et al., 2004), which might also be expected to foster the evolution of geographically localized viral subtypes. Nevertheless, our data showed no such geographic substructure. This conclusion is unlikely to be an artifact of the gene region selected (ENV), since an analysis of 33 whole WNV genome sequences from the literature yielded essentially the same results.

Other studies have claimed that WNV in the USA is a relatively static virus (Lanciotti et al., 2002; Beasley et al., 2003). Our findings, however, demonstrate that WNV is evolving at a steady rate. Our estimate of the WNV ENV gene evolutionary rate (0.85×10^{-3} substitutions per site per year) is similar to that of the immunodominant genes of other RNA viruses such as human influenza B (HA gene rate $=1.1\times10^{-3}$; Yamashita et al., 1988), Dengue-4 (E gene rate $=0.79\times10^{-3}$; Rambaut, 2000), hepatitis C (E1 gene rate $=0.74\times10^{-3}$; Smith et al., 1997), South American eastern equine encephalitis viruses (E2 gene rate $=0.43\times10^{-3}$; Weaver et al., 1994), and measles virus (H gene rate $=0.4\times10^{-3}$; Rima et al., 1997). We caution that the evolutionary rate of WNV could differ in different parts of the world, especially if more temperate climates lead to decreased seasonality in viral transmission (Tesh et al., 2004) and/or to increased numbers of viral genome replication cycles due to direct temperature effects (Weaver et al., 1999). Although the evolution of the WNV ENV gene shows an overall pattern of drift and purifying selection, the continued accumulation of mutational changes in WNV raises concerns, since variants of the virus that display divergent biological properties could evolve by chance. Even minor amino acid substitutions can account for such phenotypic changes in WNV, as has been shown by the presence of single-residue attenuating substitutions in Mexican isolates of the virus (Beasley et al., 2004). Because we have also documented a steady increase of WNV genetic diversity with time, the likelihood of such events may be increasing.

The overall pattern of lack of geographic substructure, drift, and purifying selection that we have documented for WNV is similar to patterns described for other arboviruses that are also transmitted through avian enzootic hosts (e.g. Western equine

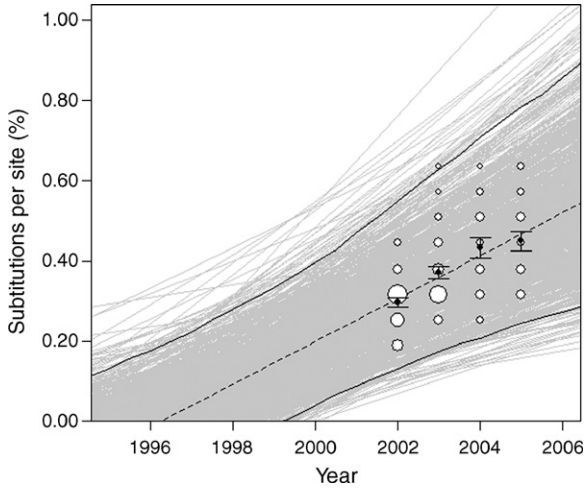


Fig. 3. Evolution of the WNV envelope gene in the United States. Black circles indicate mean evolutionary distances for all US sequences ($n=120$) from their last common ancestor, calculated along the tree shown in Fig. 1; bars indicate standard errors. The dashed line is a least-squares regression line indicating that the WNV envelope gene has evolved significantly between 2002 and 2004 ($R^2=0.2929$, $t=6.992$, $P<0.0001$). White circles indicate observed evolutionary distances from the common ancestor for all sequences included in the analysis and are proportional in diameter to the number of sequences analyzed. Grey lines are least-squares regression lines derived from 1000 bootstrap replicates of the data; solid lines indicate 95% confidence intervals around the evolutionary rate estimate. The evolutionary rate of the WNV ENV gene was estimated from these data to be 0.85×10^{-3}, using the computer program BEAST (Drummond and Rambaut, 2003; see text for full explanation), which is marginally higher than the slope of the line shown (0.54×10^{-3}).

encephalitis virus, North American Eastern equine encephalitis virus; Weaver et al., 1999; Weaver, 2006). We caution, however, that this rate and pattern of WNV evolution may not be generalizable to other locations to which the virus has spread (e.g. Mexico, the Caribbean). Eastern equine encephalitis viruses, for example, have evolved at different rates in North America and South America, possibly reflecting local differences in temperature and/or enzootic transmission cycles (Weaver et al., 1994). As WNV continues to evolve in North America, its rate and pattern of evolution may also begin to vary in response to local conditions.

Overall, our results show that WNV in the USA is a geographically panmictic viral population that is nevertheless evolving and diversifying at a rate comparable to that of other positive sense RNA viral pathogens, and according to a pattern of drift and purifying selection characteristic of other arboviruses. Future monitoring of the rate and pattern of viral evolution in the USA will be critical for understanding how WNV may or may not begin to adapt to local conditions and/or host species.

Materials and methods

Sample collection

We collected tissues from infected animals that had been submitted to the Veterinary Diagnostic Laboratory at University of Illinois at Urbana-Champaign and the Centralia Animal Disease Laboratory of the Illinois Department of Agriculture between 2002 and 2005. The samples were all WNV positive as determined by immunohistochemistry. Samples were chosen to span the widest possible geographic range (28 different Illinois Counties, range of geographic distances between sample collection locations = 0–502 km), dates of isolation, and species of origin (Table 2). All samples were stored at −80 °C until RNA extraction.

We selected for sequence analysis the WNV Envelope (ENV) gene. This gene codes for a 53 kDa protein (Petersen and Roehrig, 2001) that represent the primary target for virus-neutralizing antibodies (Rey et al., 1995; Beasley and Barrett, 2002). In addition to being among the most variable of WNV genes, the ENV gene has also been shown to contain phylogenetic information that compares favorably with whole viral genome sequencing (Davis et al., 2005). We added to our newly generated sequences all homologous United States WNV ENV gene sequences available in GenBank as of June, 2006.

Table 3
Sequences of primers used for WNV envelope gene reverse transcription and PCR amplification

	Primer name[a]	Sequence (5′–3′)	Direction
RT-PCR[b]	WNV-450	CGCCAGCGTAGGAGCAGTTAC	Forward
	WNV-2928c	AGAAGTGGAGGATTTTGGATTTGG	Reverse
PCR-E1[c]	WNV-871	CTGGTGGCAGCCGTCATTGGTTGG	Forward
	WNV-1700c	CACACGCCACGAAGCAGTCTG	Reverse
PCR-E2[c]	WNV-1595	CGTTCTTGGTCCATCGTGAGTG	Forward
	WNV-2666c	AAATGTGGGAAGCAGTGAAGGACG	Reverse

[a] Numbers refer to nucleotide positions in the viral genome to which the 5′ base of each primer anneals.
[b] Primer pair used for initial long reverse-transcription PCR.
[c] Primer pairs used for nested amplifications of ENV gene segments 1 and 2.

RNA extraction, ENV gene amplification, and sequencing

We extracted total RNA from tissue samples using the Qiagen® RNeasy® Fibrous Tissue Mini kit (Qiagen, Valencia, CA), according to the manufacturer's instructions, performing the final elution in a 50 µl volume. To amplify the ENV gene, we used a long reverse transcription PCR followed by a subsequent amplification of two overlapping ENV gene fragments. Primer sequences are shown in Table 3 and were from Davis et al. (2005). For the reverse transcription and initial PCR step, we used the Qiagen® OneStep RT-PCR kit with 2.5 µl RNA template. We followed the manufacturer's instructions and performed reactions in 25 µl volumes; reactions included 24 pM of each primer, 50 U of Invitrogen RNaseOUT (Invitrogen Life Technologies, Carlsbad, CA.) and 1 µg of bovine serum albumin. After the reverse transcription step (1 h at 48 °C, 15 min at 95 °C), we subjected reactions to 35 cycles of PCR as follows: 94 °C for 30 s, 48 °C for 30 s, 68°C for 2.5 min, followed by a terminal 10 min extension at 68 °C and an indefinite soak at 4 °C.

We used 1 µl of RT-PCR product as template for two subsequent amplifications of overlapping ENV gene segments with primers WNV-871 and WNV-1700c, and WNV-1595 and WNV-2666c, respectively (Table 3). Reactions were performed in 25 µl volumes containing 1× Failsafe™ Premixed buffer (Epicentre Biotechnologies, Madison, WI), 12 pM of each primers and 0.625 U of Enzyme mix. We used Failsafe™ Premix Buffer F for the PCR of the first ENV gene segment (E1) and Failsafe™ Premix Buffer B for the PCR of the second (E2) ENV gene segment. After an initial incubation at 94 °C for 4 min, we subjected reactions to 35 cycles of PCR as follows: 94 °C for 30 s, annealing temperature for 30 s (55 °C for PCR of

Table 2
Sources and years of collection of WNV positive vertebrate tissues from Illinois used for viral sequencing

Year	*Cyanocitta cristata*	*Corvus brachyrhynchos*	*Carpodacus mexicanus*	*Bubo virginianus*	*Falco punctatus*	*Pica pica*	*Falco columbarius*	*Buteo jamaicensis*	*Sciurus carolinensis*	Total
2002	8	8		6	1	1	1		2	27
2003	3	5								8
2004	2	7	1					1		11
2005	2	17							3	22
Total	15	37	1	6	1	1	1	1	5	68

the E1 segment and 59 °C for PCR of the E2 segment), 72 °C for 45 s, followed by 7 min terminal extension at 72 °C and an indefinite soak at 4 °C.

We electrophoresed second-round amplicons in 1% agarose gels, which we stained with ethidium bromide and visualized under ultraviolet light. We purified amplicons from gels using the ZymoClean Gel DNA Recovery Kit (Zymo Research, Orange, CA) and sequenced them directly and in both directions with the same primers used for second-round PCR. Sequencing was performed at the University of Illinois Roy J. Carver Biotechnology Center. We hand edited all sequences and resolved ambiguous bases by repeated sequencing of both DNA strands.

Analyses

We aligned sequences using the computer program ClustalX (Thompson et al., 1997). We conducted population genetic analyses with the computer programs Arlequin ver. 3.01 (Excoffier et al., 2005) and MEGA ver. 3.1 (Kumar et al., 2004). To investigate spatiotemporal subdivisioning within the United States WNV population, we conducted analyses of molecular variance (AMOVA; Excoffier et al., 1992). We tested the significance of spatiotemporal subdivisions using 16,000 permutations of the data, performed with Arlequin. To assess whether more complete sequencing of the WNV genome might show a different spatiotemporal pattern, we repeated our AMOVA analyses using 33 full length WNV genomes available from the literature (GenBank accession numbers AF196835, AF260976, AF404753–AF404756, AY289214, AY712945–AY712948, AY795965, DQ164186–DQ164206, and DQ211652).

For phylogenetic analyses, we first created a preliminary tree of all published United States sequences and selected from among these a subset of sequences representing all major viral clades, including those named in Davis et al. (2005). To this set of sequences we added our newly generated unique sequences from Illinois. In both cases, we selected a model of molecular evolution using a hierarchical likelihood ratio test approach and the Akaike information criterion (Akaike, 1973) implemented with the computer program ModelTest ver. 3.7 (Posada and Crandall, 2001). We used Bayesian methods implemented with the computer program MrBayes ver. 3.1.1 (Ronquist and Huelsenbeck, 2003) to create a final phylogenetic tree and to assess statistical support for clades. Specifically, we executed a Markov chain Monte Carlo search for 1,000,000 generations using two runs with four chains (temperature = 0.05) and represented the results of the search as a 50% majority rule consensus tree. We calculated tree statistics and performed phylogenetic manipulations with the computer program PAUP* ver. 4.0b10 (Swofford, 2000).

To examine the rate of WNV evolution within the United States, we used our final Bayesian tree to calculate root-to-tip distances between all United States sequences and their most recent common ancestor. We then conducted linear regression analyses on these distances using year of collection/isolation as a predictor variable. To test the significance of the regression, and to obtain 95% confidence intervals, we used a bootstrap approach in which we replicated the regression analysis using root-to-tip distances derived from 1000 bootstrap data sets generated using the Seqboot module of the computer program PHYLIP (version 3.65, Felsenstein, 2004). To avoid potential problems of phylogenetic non-independence of sequences, we generated point estimates of the rate of evolution of the ENV gene, as well as 95% confidence intervals, using a Bayesian method that relaxes standard molecular clock assumptions (Drummond et al., 2006), implemented with the computer program BEAST, version 1.3 (Drummond and Rambaut, 2003). To examine selection in WNV ENV sequences, we estimated synonymous and non-synonymous substitution rates using the computer program DnaSP version 4.10.7 (Rozas et al., 2003). We performed all statistical analyses and procedures with the computer program R (R Development Core Team, 2006).

Acknowledgments

We thank J. Andrews and D. Webb for providing diagnostic samples, G. Ebel for providing primer sequences and advice on PCR methods, K. Vashisht for helping with laboratory analyses, and S. Loss and G. Hamer for logistical support. E. Walker, M. Ruiz, J. Brawn, and two anonymous reviewers provided helpful comments on the manuscript. This material is based upon work supported by the NSF/NIH Ecology of Infectious Diseases program under Award No #0429124.

References

Akaike, H., 1973. Information theory and an extension of an extension of the maximum likelihood principle. Proc. 2 Inter. Symposium on Information Theory, Budapest, pp. 267–281.

Anderson, J.F., Vossobrinck, 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. 98, 12885–12889.

Beasley, D.W.C., Barrett, A.D.T., 2002. Identification of neutralizing epitopes within structural domain III of the West Nile Virus envelope protein. J. Virol. 76, 13097–13100.

Beasley, D.W.C., Davis, T.C., Guzman, H., Vanlandingham, D.L., Travassos de Rosa, A.P.A., Parsons, R.E., Higghs, S., Tesh, R.B., Barrett, A.D.T., 2003. Limited evolution of West Nile virus has occurred during its southwesterly spread in the United States. Virology 309, 190–195.

Beasley, D.W., Davis, C.T., Estrada-Franco, J., Navarro-Lopez, R., Campomanes-Cortes, A., Tesh, R.B., Weaver, S.C., Barrett, A.D., 2004. Genome sequence and attenuating mutations in West Nile virus isolate from Mexico. Emerging Infect. Dis. 10, 2221–2224.

Davis, T.C., Ebel, G.D., Lanciotti, R.S., Brault, A.C., Guzman, H., Siirin, M., Lambert, A., Parsone, R.E., Beasley, D.W.C., 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.T., 2005. Phylogenetic analysis of West Nile virus isolates 2001–2004: evidence for the emergence of dominant genotype. Virology 342, 252–265.

Drummond, A.J., Rambaut, A., 2003. BEAST v1.0, Available from http://evolve.zoo.ox.ac.uk/beast/.

Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.

Estrada-Franco, J.G., Navarro-Lopez, R., Beasley, D.W.C., Coffey, L., Carrara, A.S., Travassos da Rosa, A., Clemens, T., Wang, E., Ludwig, G.V., Campomanes Cortes, A., Paz Ramirez, P., Tesh, R.B., Barrett, A.D.T., Weaver, S.C., 2003. West Nile virus in Mexico: evidence of widespread circulation since July 2002. Emerging Infect. Dis. 9, 1604–1607.

Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131, 479–491.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online, vol. 1, pp. 47–50.

Felsenstein, J., 2004. PHYLIP (Phylogeny Inference Package) Version 3.6. Department of Genome Sciences, University of Washington, Seattle.

Gancz, A.Y., Barker, I.K., Lindsay, R., Dibernardo, A., McKeever, K., Hunter, B., 2004. West Nile virus outbreak in north American owls, Ontario, 2002. Emerging Infect. Dis. 10, 2135–2142.

Huhn, G.D., Austin, C., Langkop, C., Kelly, K., Lucht, R., Lampman, R., Novak, R., Haramis, L., Boker, R., Smith, S., Chudoba, M., Gerber, S., Conover, C., Dworkin, M., 2005. The emergence of West Nile virus during a large outbreak in Illinois in 2002. Am. J. Trop. Med. Hyg. 72, 768–776.

Jerzak, G., Bernard, K.A., Kramer, L.D., Ebel, G.D., 2005. Genetic variation in West Nile virus from naturally infected mosquitoes and birds suggests quasispecies structure and strong purifying selection. J. Gen. Virol. 86, 2175–2183.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

Lanciotti, R.S., Roehrig, J.T., Deubel, V., Smith, J., Parker, M., Steele, K., 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 U.S. Science 286, 2333–2337.

Lanciotti, R.S., Ebel, G.D., Deubel, V., Kerst, A.J., Murri, S., Meyer, R., Bowen, M., McKynney, N., Morrill, W.E., Crabtree, M.B., Kramer, L.D., Roehring, 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, 96–105.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Petersen, L.R., Roehrig, J.T., 2001. West Nile virus: a reemerging global pathogen. Emerging Infect. Dis. 7, 611–614.

Posada, D., Crandall, K.A., 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. 50, 580–601.

Rambaut, A., 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. Bioinformatics 16, 395–399.

R Development Core Team, 2006. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Reed, K.D., Meece, J.K., Hekel, J.S., Shukla, S.K., 2003. Birds, migration and emerging zoonoses: west nile virus, lyme disease, influenza a and enteropathogens. Clin. Med. Res. 1, 5–12.

Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., Harrison, S.C., 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. Nature 375, 291–298.

Rima, B.K., Earle, J.A.P., Baczko, K., ter Meulen, V., Liebert, U.G., Carstens, C., Carabana, J., Caballero, M., Celma, M.L., Fernandez-Munoz, R., 1997. Sequence divergence of measles virus haemagglutinin during natural evolution and adaptation to cell culture. J. Gen. Virol. 78, 97–106.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496–2497.

Smith, D.B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Yap, P.L., Simmonds, P., 1997. The origin of hepatitis C virus genotypes. J. Gen. Virol. 78, 321–328.

Smithburn, K.C., Hughes, T.P., Burke, A.W., Paul, J.H., 1940. A neurotropic virus isolated from the blood of a native of Uganda. Am. J. Trop. Med. 20, 471–492.

Swofford, D.L., 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA.

Tesh, R.B., Parsons, R., Siirin, M., Randle, Y., Sargent, C., Guzman, H., Wuithiranyagool, T., Higgs, S., Vanlandingham, D.L., Bala, A.A., Haas, K., Zerinque, B., 2004. Year-round West Nile Virus activity, Gulf coast region, Texas and Louisiana. Emerging Infect. Dis. 10, 1649–1652.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Weaver, S.C., 2006. Evolutionary influences in arboviral disease. Curr. Top. Microbiol. Immunol. 299, 285–314.

Weaver, S.C., Hagenbaugh, A., Bellew, L.A., Gousset, L., Mallampalli, V., Holland, J.J., Scott, T.W., 1994. Evolution of alphaviruses in the eastern equine encephalomyelitis complex. J. Virol. 68, 158–169.

Weaver, S.C., Powers, A.M., Brault, A.C., Barrett, A.D., 1999. Molecular epidemiological studies of veterinary arboviral encephalitides. Vet. J. 157, 123–138.

Yamashita, M., Krystal, M., Fitch, W.M., Palese, P., 1988. Influenza B virus evolution: co-circulating lineages and comparison of evolutionary pattern with those of influenza A and C viruses. Virology 163, 112–122.