2005

Phylogenetic Analysis of North American West Nile Virus Isolates, 2001–2004: Evidence for the Emergence of a Dominant Genotype

C. Todd Davis
University of Texas Medical Branch, 301 University Blvd., Galveston, TX

Gregory D. Ebel
The University at Albany, State University of New York, 5668 State Farm Road, Slingerlands, NY

Robert S. Lanciotti
Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Public Health Services, US Department of Health and Human Services, PO Box 2087, Fort Collins, CO 80522, USA

Aaron C. Brault
University of California, Davis, CA

Hilda Guzman
University of Texas Medical Branch, 301 University Blvd., Galveston, TX

See next page for additional authors

Available online at www.sciencedirect.com

SCIENCE DIRECT

Virology 342 (2005) 252 – 265

VIROLOGY

www.elsevier.com/locate/yviro

Phylogenetic analysis of North American West Nile virus isolates, 2001–2004: Evidence for the emergence of a dominant genotype

C. Todd Davis[a], Gregory D. Ebel[b], Robert S. Lanciotti[c], Aaron C. Brault[d], Hilda Guzman[a], Marina Siirin[a], Amy Lambert[c], Ray E. Parsons[e], David W.C. Beasley[a], Robert J. Novak[f], Darwin Elizondo-Quiroga[g], Emily N. Green[d], David S. Young[b], Lillian M. Stark[h], Michael A. Drebot[i], Harvey Artsob[i], Robert B. Tesh[a], Laura D. Kramer[b], Alan D.T. Barrett[a,*]

[a]*Department of Pathology, Center for Biodefense and Emerging Infectious Diseases, and Sealy Center for Vaccine Development, University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77550-0609, USA*
[b]*The Arbovirus Laboratories, Wadsworth Center, New York State Department of Health, Department of Biomedical Sciences, School of Public Health, The University at Albany, State University of New York, 5668 State Farm Road, Slingerlands, NY 12159, USA*
[c]*Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Public Health Services, US Department of Health and Human Services, PO Box 2087, Fort Collins, CO 80522, USA*
[d]*Center for Vector-borne Diseases, Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, University of California, Davis, CA 95616, USA*
[e]*Mosquito Control Division, Harris County Public Health and Environmental Services, 3330 Old Spanish Trail, Bldg. D, Houston, TX 77021, USA*
[f]*Illinois Natural History Survey, Champagne, IL 61820, USA*
[g]*Universidad Autonoma de Nuevo Leon, San Nicolas de Los Garza, Nuevo Leon, Mexico*
[h]*Florida Department of Health, Bureau of Laboratories-Tampa, 3602 Spectrum Boulevard, Tampa, FL 33612, USA*
[i]*Zoonotic Diseases and Special Pathogens, National Microbiology Laboratory, Public Health Agency of Canada, 1015 Arlington Street, Winnipeg, Manitoba, Canada*

Received 12 May 2005; returned to author for revision 13 June 2005; accepted 7 July 2005
Available online 31 August 2005

Abstract

The distribution of West Nile virus has expanded in the past 6 years to include the 48 contiguous United States and seven Canadian provinces, as well as Mexico, the Caribbean islands, and Colombia. The suggestion of the emergence of a dominant genetic variant has led to an intensive analysis of isolates made across North America. We have sequenced the premembrane and envelope genes of 74 isolates and the complete genomes of 25 isolates in order to determine if a dominant genotype has arisen and to better understand how the virus has evolved as its distribution has expanded. Phylogenetic analyses revealed the continued presence of genetic variants that group in a temporally and geographically dependent manner and provide evidence that a dominant variant has emerged across much of North America. The implications of these findings are discussed as they relate to transmission and spread of the virus in the Western Hemisphere.
Published by Elsevier Inc.

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

Introduction

West Nile virus (WNV) (*Flaviviridae*:*Flavivirus*) has recently undergone a dramatic range expansion into much of North America. Until 1999, the geographical distribution of the virus was limited to Africa, the Middle East, India, Australasia, and western and central Asia with occasional epizootics and epidemics in Europe (Murgue et al., 2002; Hall et al., 2002). But, since the summer of 1999, the distribution of WNV has expanded to include the contiguous U.S. states and seven Canadian provinces, as

* Corresponding author. Fax: +1 409 772 2500.
E-mail address: abarrett@utmb.edu (A.D.T. Barrett).

0042-6822/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.virol.2005.07.022

well as Mexico and the Caribbean islands (Estrada-Franco et al., 2003; Komar et al., 2003; Blitvich et al., 2003; Dupuis et al., 2003; Quirin et al., 2004) and Colombia (Komar, personal communication). Because of its relatively recent introduction, studies concerning the evolution of WNV are important to understand the extent to which the virus has mutated as its temporal and geographic distribution have expanded and to recognize the accumulation of mutations in the genome that may be important to transmission and perpetuation of this virus in nature. Nucleic acid sequencing of WNV isolates collected across the U.S. since 1999 has identified mutations to the genome when compared to the prototype New York strain, WN-NY99 (GenBank accession no. AF196835). These mutations reveal the presence of distinct genetic variants that group in a temporally and geographically dependent manner (Lanciotti et al., 2002; Beasley et al., 2003; Davis et al., 2003; Ebel et al., 2004) and suggest the emergence of a dominant genetic variant that is now distributed across much of North America. Additional genetic and phenotypic studies suggest that the increased isolation frequency of the dominant genetic variant during 2002 and 2003 may be due to the possibility of enhanced mosquito transmission efficiency of this variant in *Culex pipiens*, one of the main vectors in the northeastern United States (Ebel et al., 2004). In order to assess the genetic divergence of WNV as its spatial and temporal distribution have expanded and to identify mutations in the WNV genome that may be conserved in the North American WNV population, we have characterized both partial and complete genome sequences of WNV isolates collected from across North America. The prM and E genes of 74 WNV isolates collected during 2001–2004 from the U.S., Canada, and Mexico were sequenced and compared to other North American WNV sequences available in GenBank. The complete genomes of 25 isolates were sequenced in order to attain a more robust comparison of the genetic relationships of WNV isolates because of the relatively few phylogenetically informative sites in the prM and E genes of isolates studied. Sequence alignments were constructed in order to identify nucleotide and amino acid mutations in the WNV genome that may support evidence that a dominant genetic variant has emerged throughout North America. The individual genes of the WNV were compared to test whether certain genes may be more prone to the accumulation of mutations and thus may act as surrogates to complete genome sequencing when performing phylogenetic analyses. Phylogenetic trees were generated using both individual viral gene and complete genomic sequences in order to compare the relationships among WNV isolates generated using each method and to identify those genes of WNV that are the most parsimony informative in the phylogenetic analysis of North American isolates. Collectively, the data suggest that there has been continued divergence of WNV as the temporal and spatial distribution of the virus has expanded.

Results

Nucleotide sequencing of North American WNV isolates

The prM and E genes of 74 North American WNV isolates collected from 2001 to 2004 were sequenced and new sequences were deposited in GenBank with information pertaining to source, collection date, and location of the isolate (Accession No. DQ158207–DQ158250). The complete genomes of four isolates made from dead birds and mosquitoes in Texas in 2003 had previously been sequenced and deposited in GenBank (Accession No. AY712945–AY712948; Davis et al., 2004). Complete genomes of 21 additional isolates were sequenced and also submitted to GenBank (Accession no. DQ164186–DQ164206) (Table 1). Table 2 shows the percentage nucleotide and deduced amino acid sequence divergence among the complete genomes and ORFs of all newly sequenced WNV isolates and NY99. All North American West Nile viruses maintain a high degree of both nucleotide and amino acid sequence conservation over the complete genome with an average degree of nucleotide and amino acid divergence from NY99 of 0.24% and 0.09%, respectively. In comparison to NY99, both the nucleotide and amino acid divergence tended to increase as the year of isolation increased with the exception of a few isolates (i.e., TX 2004 Harris 4). The highest degree of nucleotide divergence from NY99 was 0.39% for AZ 2004, and the degree of amino acid sequence divergence was highest for NY 2003 Rockland, which had seven amino acid substitutions from NY99. Overall, the highest degree of both nucleotide and amino acid divergence was between AZ 2004 and NY 2003 Rockland at 0.58% and 0.38%, respectively. Interestingly, the degree of nucleotide and amino acid divergence did not always correlate for a given isolate because the majority of the nucleotide mutations were synonymous (data not shown). Each of the 25 isolates that were completely sequenced had at least one amino acid substitution relative to the consensus sequence derived from an alignment of all completely sequenced isolates (Table 3). In general, isolates made during 2003 and 2004 had accumulated a larger number of amino acid substitutions than isolates made in 2001 and 2002. Table 3 shows that the majority of the isolates share a conserved amino acid substitution at position E159 from a V → A with the exception of some isolates from New York, 2001–2003, and an isolate from Ohio, 2002. Several amino acid substitutions were also shared by isolates made in the same state, though not necessarily the same year. Deduced amino acid substitutions were identified in each of the 25 isolates with the largest number of amino acid substitutions occurring in NS5.

Phylogenetic analysis of WNV isolates

Several phylogenetic trees were generated from prM and E genes and untranslated regions, as well as complete

Table 1
West Nile virus isolates sequenced for this study

Abbreviation	Isolate	Year of isolation	Location	Source	Passage history	Genbank accession no.
FL 2001 Palm Beach	Florida 2001	2001	Palm Beach, Florida	Catbird	V2	DQ080072
NY 2001 Suffolk	32010157	2001	Suffolk Co., NY	*Culex pipiens/restuans*	V1	DQ164194
FL 2002 Sumter	Florida 2002	2002	Sumter, Florida	Horse	V2	DQ080071
NY 2002 Queens	02003011	2002	Queens Co., NY	American crow	P	DQ164186
NY 2002 Broome	02003557	2002	Broome Co., NY	American crow	P	DQ164187
NY 2002 Clinton	02002758	2002	Clinton Co., NY	American crow	P	DQ164193
NY 2002 Nassau	34020055	2002	Nassau Co., NY	*Culex pipiens/restuans*	V2	DQ164195
Ontario 2002	Ontario	2002	Ontario, Canada	American crow	V2	DQ158245
Saskatchewan 2002	Saskatchewan	2002	Saskatchewan, Canada	Black-billed magpie	V2	DQ158246
Manitoba 2002	Manitoba	2002	Manitoba, Canada	Amercican crow	V2	DQ158244
GA 2002 1	68955	2002	Georgia	Human—plasma	P	DQ164196
GA 2002 2	68960	2002	Georgia	Human—brain	P	DQ164197
OH 2002	81948	2002	Ohio	Human—plasma	P	DQ164202
IN 2002	81931	2002	Indiana	Human—plasma	P	DQ164200
TX 2002 1	80025	2002	Texas	Human—plasma	P	DQ164198
TX 2002 2	80022	2002	Texas	Human—plasma	P	DQ164205
NY 2003 Westchester	03000360	2003	Westchester Co., NY	Amercican crow	P	DQ164188
NY 2003 Alabany	03001986	2003	Albany Co., NY	American crow	P	DQ164189
NY 2003 Suffolk	03002018	2003	Suffolk Co., NY	American crow	P	DQ164190
NY 2003 Chautauqua	03002086	2003	Chautauqua Co., NY	American crow	P	DQ164191
NY 2003 Rockland	03002094	2003	Rockland Co., NY	American crow	P	DQ164192
AZ 2003 1	03-az-mp-1623	2003	Arizona	Magpie	V1	DQ158207
AZ 2003 2	03-az-mp-1681	2003	Arizona	Magpie	V1	DQ158208
AZ 2003 3	03-az-mp-1799	2003	Arizona	Magpie	V1	DQ158209
CA 2003 Los Angeles 1	03-ca-crow-s0331532	2003	Los Angeles Co., CA	American crow	V1	DQ158210
CA 2003 Los Angeles 2	03-ca-crow-s0334814	2003	Los Angeles Co., CA	American crow	V1	DQ158211
CA 2003 Los Angeles 3	Mosq. grla1131	2003	Los Angeles Co., CA	*Culex tarsalis*	V2	DQ158212
CA 2003 Los Angeles 4	03-ca-mp-grla-1260	2003	Los Angeles Co., CA	Magpie	V1	DQ158214
CA 2003 Los Angeles 5	Bird 9173	2003	Los Angeles Co., CA	American crow	V2	DQ158217
CA 2003 San Bernadino	Bird 9172	2003	San Bernadino Co., CA	American crow	V2	DQ158218
CA 2003 1	03-ca-mp-impr-102	2003	Imperial Valley, CA	Magpie	V1	DQ158215
CA 2003 2	03-ca-mp-impr-1075	2003	Imperial Valley, CA	Magpie	V1	DQ158216
CA 2003 3	Mosq. impr1143	2003	Imperial Valley, CA	*Culex tarsalis*	V2	DQ158213
MX 2003 Nuevo Leon	Mosq. 9488	2003	Nuevo Leon, Mexico	*Culex quinquefasciatus*	V2	AY963775
CO 2003 Larimer	Bird 9185	2003	Larimer Co., CO	*Culex pipiens*	V2	DQ158219
CO 2003 1	Colorado 3068	2003	Colorado	Red-tailed hawk	P	DQ164204
CO 2003 2	Colorado 3258	2003	Colorado	Magpie	P	DQ164203
NB 2003 Brown	Bird 9239	2003	Brown Co., Nebraska	Bluejay	V2	DQ158236
NB 2003 Dakota	Bird 9241	2003	Dakota Co., Nebraska	Bluejay	V2	DQ158237
TX 2003 Harris 1	Bird 9045	2003	Harris Co., TX	Bluejay	V1	DQ158223
TX 2003 Harris 2	Bird 9114	2003	Harris Co., TX	Bluejay	V1	DQ158222
TX 2003 Harris 3	Mosq. V4095	2003	Harris Co., TX	*Culex quinquefasciatus*	V1	DQ158224
TX 2003 Harris 4	Mosq. V4369	2003	Harris Co., TX	*Culex quinquefasciatus*	V1	AY712948
TX 2003 Harris 5	Mosq. V4096	2003	Harris Co., TX	*Culex quinquefasciatus*	V1	DQ158226
TX 2003 Harris 6	Bird 1153	2003	Harris Co., TX	Mourning dove	V1	AY712945
TX 2003 Harris 7	Bird 1171	2003	Harris Co., TX	Great-tailed grackle	V1	AY712946
TX 2003 Harris 8	Bird 1175	2003	Harris Co., TX	Bluejay	V1	DQ158220
TX 2003 Harris 9	Bird 1240	2003	Harris Co., TX	Bluejay	V1	DQ158221
TX 2003 Harris 10	Bird 1461	2003	Harris Co., TX	Bluejay	V1	AY712947
TX 2003 Harris 11	Mosq. V4370	2003	Harris Co., TX	*Culex quinquefasciatus*	V1	DQ158225
TX 2003 Montgomery 1	Bird 1519	2003	Montgomery Co., TX	Bluejay	V1	DQ158227
TX 2003 Montgomery 2	Bird 1574	2003	Montgomery Co., TX	Bluejay	V1	DQ158228
TX 2003 Montgomery 3	Bird 1576	2003	Montgomery Co., TX	Bluejay	V1	DQ158229
TX 2003 Jefferson 1	Bird 1881	2003	Jefferson Co., TX	Mourning Dove	V1	DQ158230
TX 2003 Jefferson 2	Bird 2073	2003	Jefferson Co., TX	Bluejay	V1	DQ158231
TX 2003 Wharton	Bird 2071	2003	Wharton Co., TX	Bluejay	V1	DQ158232
TX 2003	Texas 82229	2003	Texas	Human—plasma	P	DQ164199
LA 2004 New Iberia	Bird 2409	2004	New Iberia Parish, Louisiana	Northern cardinal	V2	DQ080061
IL 2004 Madison	Illinois—9515	2004	Madison Co., Illinois	American crow	V2	DQ158250
IL 2004 Schuyler	Illinois—9517	2004	Schuyler Co., Illinois	American crow	V2	DQ158249
IL 2004 Knox	Illinois—9519	2004	Knox Co., Illinois	Bluejay	V2	DQ158248

Table 1 (*continued*)

Abbreviation	Isolate	Year of isolation	Location	Source	Passage history	Genbank accession no.
IL 2004 New Marion	Illinois—9520	2004	New Marion Co., Illinois	Bluejay	V2	DQ158247
AZ 2004	Arizona 2004	2004	Arizona	Human—plasma	P	DQ164201
MX 2004 Sonora	Mexico 2004	2004	Sonora, Mexico	Human—plasma	V2	AY963774
TX 2004 Harris 1	Bird 2419	2004	Harris Co., TX	Bluejay	V1	DQ158233
TX 2004 Harris 2	Bird 2541	2004	Harris Co., TX	Mourning dove	V1	DQ158234
TX 2004 Harris 3	Bird 3218	2004	Harris Co., TX	Bluejay	V1	DQ158235
TX 2004 Harris 4	Bird 3588	2004	Harris Co., TX	Bluejay	V1	DQ164206
TX 2004 Parmer	Bird 9460	2004	Parmer Co., TX	House sparrow	V1	DQ158238
TX 2004 Brazoria 1	Bird 9461	2004	Brazoria Co., TX	Bluejay	V1	DQ158239
TX 2004 Brazoria 2	Bird 9473	2004	Brazoria Co., TX	Bluejay	V1	DQ158241
TX 2004 Swisher	Bird 9472	2004	Swisher Co., TX	House sparrow	V1	DQ158240
TX 2004 Randall	Bird 9477	2004	Randall Co., TX	House sparrow	V1	DQ158242
TX 2004 Galveston	Bird 9483	2004	Galveston Co., TX	Bluejay	V1	DQ158243

Isolates in bold have been completely sequenced. P = RNA extracted from infected tissue. V1 = RNA extracted from original isolation in Vero cells. V2 = RNA extracted after single Vero cell passage.

genomes in order to ascertain the phylogenetic relationships of the isolates studied. RT-PCR was used to sequence a 2004 nucleotide region of the prM and E genes of 74 WNV isolates to broaden the scope of our phylogenetic analysis because there are a larger number of partial sequences (primarily the prM and E genes) of WNV isolates available in GenBank. A Bayesian analysis was used to generate a consensus tree based on the prM and E alignment of these 74 North American WNV isolates in comparison to all other North American WNV isolates for which the prM and E sequence was available in GenBank (Fig. 1). The tree was rooted with the most closely related Old World WNV, Israel-1998 (Accession no. AY033389) in order to generate a tree illustrating more parsimony informative sites. This analysis revealed three distinct clades comprised of North American isolates. The apical clade was comprised of isolates collected from all over North America which were collected from 2002 to 2004 and has been termed the "North America 2002–2004" clade. Within this clade were many subclades, which consisted of isolates that were highly conserved at the nucleotide sequence level and some which were identical to one another. Most of the isolates within these subclades were made in the same state and during the same year, although there was evidence that some isolates made in close geographic proximity tended to cluster into subclades. Basal to the North American 2002–2004 clade was a sister clade that represents an intermediate grouping of isolates that share some degree of sequence identity with the North American 2002–2004 clade. This sister clade was made up primarily of isolates from eastern regions of the U.S. including New York, Alabama, Indiana, Illinois, Louisiana, and eastern Texas. An isolate from Tabasco State, Mexico made up a second sister clade because of a conserved nucleotide substitution at position 2466 in the envelope protein gene that was also found in all apical isolates. Two additional Mexican isolates from Sonora State and Nuevo Leon State did not group with the isolate from Tabasco State, which again suggests the occurrence of multiple introductions of WNV into Mexico as has been noted by others (Estrada-Franco et al., 2003; Blitvich et al., 2004). Interestingly, both of the Mexican isolates fall into the most apical clade, which also contains isolates collected from Arizona in 2003 and 2004, several isolates collected in Texas in 2004, and an isolate from Colorado in 2003. A second clade, termed the "Eastern U.S." clade, was comprised of isolates collected from the Eastern U.S. states. The majority of these isolates were made prior to 2003, with the exception of two isolates from New York. This clade also contained isolates from the Northeastern U.S. that were made during the early stages of the WNV epidemic including isolates from New York in 1999. Although genetically more distantly related to Israel-1998 than the Eastern U.S. isolates, the most basally positioned clade was comprised of "Southeast coastal Texas" isolates from 2002. The basal position of this clade occurred as a result of several non-parsimony informative nucleotide mutations shared by isolates in the Eastern U.S. clade and the North American 2002–2004 clade, none of which were found in the Southeast coastal Texas 2002 clade. Interestingly, isolates made from the southeast coast of Texas in 2003 and 2004 belonged to the North American 2002–2004 clade, suggesting that this genotype has become extinct or displaced in this particular region.

To further define the phylogenetic relationships among North American WNV isolates and to confirm the relationships as delineated by the prM and E sequences, the complete genomic sequencing of 25 isolates was undertaken to provide a more robust comparison. A phylogenetic tree was generated by maximum likelihood analysis to show the relationship between the newly sequenced isolates and those North American isolates available in GenBank. A total of 36 North American WNV isolates from this and previous studies group in a monophyletic clade of the Lineage I WNV strains, which sits just apical to the closest related Old

Table 2
Percentage nucleotide/amino acid divergence over the complete genome/ORF among WNV isolates

	NY99AF196835	NY 2001 Suffolk	NY 2002 Nassau	NY 2002 Clinton	NY 2002 Queens	NY 2002 Broome	GA 2002 1	GA 2002 2	TX 2002 1	TX 2002 2	IN 2002	OH 2002	NY 2003 Westchester	NY 2003 Albany	NY 2003 Suffolk	NY 2003 Chautauqua	NY 2003 Rockland	CO 2003 1	CO 2003 2	TX 2003	TX 2003 Harris 4	TX 2003 Harris 6	TX 2003 Harris 7	TX 2003 Harris 10	AZ 2004	TX 2004 Harris 4
NY99AF196835		**0.15**	**0.18**	**0.18**	**0.17**	**0.24**	**0.24**	**0.23**	**0.23**	**0.23**	**0.16**	**0.30**	**0.18**	**0.25**	**0.28**	**0.22**	**0.26**	**0.32**	**0.34**	**0.25**	**0.24**	**0.33**	**0.35**	**0.25**	**0.39**	**0.21**
NY 2001 Suffolk	0.06		**0.28**	**0.28**	**0.27**	**0.34**	**0.34**	**0.33**	**0.33**	**0.33**	**0.26**	**0.40**	**0.26**	**0.35**	**0.38**	**0.32**	**0.36**	**0.42**	**0.41**	**0.35**	**0.34**	**0.43**	**0.45**	**0.35**	**0.49**	**0.34**
NY 2002 Nassau	0.06	0.12		**0.15**	**0.12**	**0.33**	**0.21**	**0.19**	**0.17**	**0.17**	**0.13**	**0.34**	**0.31**	**0.22**	**0.23**	**0.18**	**0.35**	**0.28**	**0.31**	**0.20**	**0.21**	**0.29**	**0.32**	**0.20**	**0.35**	**0.26**
NY 2002 Clinton	0.06	0.12	0.06		**0.14**	**0.34**	**0.17**	**0.15**	**0.15**	**0.15**	**0.13**	**0.35**	**0.31**	**0.22**	**0.23**	**0.15**	**0.37**	**0.26**	**0.29**	**0.18**	**0.17**	**0.27**	**0.30**	**0.20**	**0.32**	**0.28**
NY 2002 Queens	0.03	0.09	0.03	0.03		**0.32**	**0.20**	**0.18**	**0.16**	**0.16**	**0.10**	**0.33**	**0.30**	**0.21**	**0.20**	**0.17**	**0.34**	**0.27**	**0.30**	**0.19**	**0.20**	**0.28**	**0.31**	**0.19**	**0.34**	**0.25**
NY 2002 Broome	0.15	0.20	0.20	0.20	0.17		**0.41**	**0.39**	**0.37**	**0.37**	**0.33**	**0.44**	**0.36**	**0.42**	**0.43**	**0.38**	**0.19**	**0.48**	**0.49**	**0.40**	**0.41**	**0.45**	**0.48**	**0.40**	**0.55**	**0.41**
GA 2002 1	0.06	0.12	0.06	0.06	0.03	0.20		**0.02**	**0.22**	**0.22**	**0.19**	**0.42**	**0.37**	**0.28**	**0.31**	**0.21**	**0.44**	**0.33**	**0.35**	**0.23**	**0.22**	**0.34**	**0.36**	**0.26**	**0.38**	**0.34**
GA 2002 2	0.03	0.09	0.03	0.03	0.00	0.17	0.03		**0.20**	**0.20**	**0.17**	**0.40**	**0.35**	**0.26**	**0.29**	**0.19**	**0.42**	**0.31**	**0.34**	**0.21**	**0.20**	**0.32**	**0.34**	**0.24**	**0.36**	**0.33**
TX 2002 1	0.09	0.15	0.09	0.09	0.06	0.23	0.09	0.06		**0.02**	**0.17**	**0.38**	**0.35**	**0.26**	**0.27**	**0.19**	**0.40**	**0.29**	**0.32**	**0.21**	**0.22**	**0.30**	**0.33**	**0.23**	**0.36**	**0.31**
TX 2002 2	0.12	0.17	0.12	0.12	0.09	0.26	0.12	0.09	0.03		**0.17**	**0.38**	**0.35**	**0.26**	**0.27**	**0.19**	**0.40**	**0.29**	**0.32**	**0.21**	**0.22**	**0.30**	**0.33**	**0.23**	**0.36**	**0.31**
IN 2002	0.03	0.09	0.03	0.03	0.00	0.17	0.03	0.00	0.06	0.09		**0.34**	**0.29**	**0.20**	**0.23**	**0.16**	**0.35**	**0.26**	**0.31**	**0.20**	**0.19**	**0.27**	**0.30**	**0.20**	**0.34**	**0.26**
OH 2002	0.06	0.12	0.12	0.12	0.09	0.20	0.12	0.09	0.15	0.17	0.09		**0.43**	**0.43**	**0.44**	**0.37**	**0.47**	**0.45**	**0.50**	**0.41**	**0.40**	**0.48**	**0.51**	**0.41**	**0.56**	**0.45**
NY 2003 Westchester	0.09	0.15	0.15	0.15	0.12	0.23	0.15	0.12	0.17	0.20	0.12	0.15		**0.38**	**0.41**	**0.34**	**0.39**	**0.44**	**0.44**	**0.38**	**0.37**	**0.45**	**0.48**	**0.38**	**0.50**	**0.37**
NY 2003 Albany	0.09	0.15	0.09	0.09	0.06	0.23	0.09	0.06	0.12	0.15	0.06	0.15	0.17		**0.32**	**0.25**	**0.44**	**0.35**	**0.40**	**0.29**	**0.28**	**0.36**	**0.39**	**0.29**	**0.43**	**0.35**
NY 2003 Suffolk	0.09	0.15	0.09	0.09	0.06	0.23	0.09	0.06	0.12	0.15	0.06	0.15	0.17	0.12		**0.28**	**0.45**	**0.38**	**0.41**	**0.30**	**0.31**	**0.39**	**0.42**	**0.30**	**0.44**	**0.36**
NY 2003 Chautauqua	0.03	0.09	0.03	0.03	0.00	0.17	0.03	0.00	0.06	0.09	0.00	0.09	0.12	0.06	0.06		**0.41**	**0.26**	**0.33**	**0.22**	**0.19**	**0.31**	**0.34**	**0.24**	**0.35**	**0.32**
NY 2003 Rockland	0.20	0.26	0.26	0.26	0.23	0.12	0.26	0.23	0.29	0.32	0.23	0.26	0.29	0.29	0.29	0.23		**0.51**	**0.52**	**0.43**	**0.44**	**0.50**	**0.53**	**0.43**	**0.58**	**0.44**
CO 2003 1	0.03	0.09	0.03	0.03	0.00	0.17	0.03	0.00	0.06	0.09	0.00	0.09	0.12	0.06	0.06	0.00	0.23		**0.44**	**0.34**	**0.31**	**0.32**	**0.34**	**0.34**	**0.44**	**0.42**
CO 2003 2	0.15	0.20	0.15	0.15	0.12	0.29	0.15	0.12	0.17	0.20	0.12	0.20	0.23	0.17	0.17	0.12	0.35	0.12		**0.34**	**0.35**	**0.45**	**0.48**	**0.36**	**0.44**	**0.43**
TX 2003	0.06	0.12	0.06	0.06	0.03	0.20	0.06	0.03	0.09	0.12	0.03	0.12	0.15	0.09	0.09	0.03	0.26	0.03	0.15		**0.24**	**0.34**	**0.37**	**0.25**	**0.39**	**0.34**
TX 2003 Harris 4	0.12	0.17	0.12	0.12	0.09	0.26	0.12	0.09	0.15	0.17	0.09	0.17	0.20	0.15	0.15	0.09	0.32	0.09	0.20	0.12		**0.34**	**0.36**	**0.26**	**0.38**	**0.34**
TX 2003 Harris 6	0.12	0.17	0.12	0.12	0.09	0.26	0.12	0.09	0.15	0.17	0.09	0.17	0.20	0.15	0.15	0.09	0.32	0.09	0.20	0.12	0.17		**0.03**	**0.34**	**0.48**	**0.43**
TX 2003 Harris 7	0.17	0.23	0.17	0.17	0.15	0.32	0.17	0.15	0.20	0.23	0.15	0.23	0.26	0.20	0.20	0.15	0.38	0.15	0.26	0.17	0.23	0.06		**0.37**	**0.51**	**0.45**
TX 2003 Harris 10	0.15	0.20	0.15	0.15	0.12	0.29	0.15	0.12	0.17	0.20	0.12	0.20	0.23	0.17	0.17	0.12	0.35	0.12	0.23	0.15	0.20	0.20	0.26		**0.41**	**0.32**
AZ 2004	0.17	0.23	0.17	0.17	0.15	0.32	0.17	0.15	0.20	0.23	0.15	0.23	0.26	0.20	0.20	0.15	0.38	0.15	0.20	0.17	0.23	0.23	0.29	0.26		**0.49**
TX 2004 Harris 4	0.09	0.15	0.09	0.09	0.06	0.23	0.09	0.06	0.12	0.15	0.06	0.15	0.17	0.12	0.12	0.06	0.29	0.06	0.17	0.09	0.15	0.15	0.20	0.12	0.20	

Percentage nucleotide divergence displayed in boldface; amino acid divergence displayed in regular type.

Table 3
Deduced amino acid substitutions from WN-NY99 found in each completely sequenced isolate

Residue	WN-NY99	NY 2001 Suffolk	NY 2002 Nassau	NY 2002 Clinton	NY 2002 Queens	NY 2002 Broome	GA 2002 1	GA 2002 2	TX 2002 1	TX 2002 2	IN 2002	OH 2002	NY 2003 Westchester	NY 2003 Albany	NY 2003 Suffolk	NY 2003 Chautauqua	NY 2003 Rockland	CO 2003 1	CO 2003 2	TX 2003	TX 2003 Harris 4	TX 2003 Harris 6	TX 2003 Harris 7	TX 2003 Harris 10	AZ 2004	TX 2004 Harris 4
CAP-59	Ile			Val																						
CAP-67	Asp	Asn																								
CAP-90	Thr																								Ala	
prM-4	Asn																				Asp					
prM-156	Val																					Ile	Ile			
E-159	Val		Ala	Ala	Ala		Ala	Ala	Ala	Ala	Ala			Ala	Ala	Ala		Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
E-169	Ile																								Val	
E-176	Tyr					His											His									
E-178	Leu												Gln													
E-188	Thr																									Ala
E-208	Thr																									
E-263	His																Gln									
NS1-111	Leu																Phe									
NS1-135	Val					Ile											Ile									
NS1-144	Pro	Ser																								
NS1-254	His																		Tyr							
NS2A-34	Met					Leu											Leu									
NS2A-102	Gln								His	His																
NS2A-224	Ala												Thr													
NS2B-84	Asn											Asp														
NS3-106	Val																Ala									
NS3-328	Glu																							Lys		
NS3-336	Leu																									
NS3-486	Phe													Leu											Ser	
NS3-539	Arg						Lys																			
NS3-603	Ser																									
NS4A-3	Ile		Met										Leu													
NS4A-85	Ala																		Thr						Thr	
NS4A-135	Val																							Met		
NS4B-83	Ala											Ser														
NS4B-119	Phe					Leu											Leu									
NS4B-202	Ile																		Thr							
NS4B-240	Ile								Met	Met																
NS4B-241	Thr																				Ala					
NS4B-245	Ile																		Val							
NS4B-249	Glu																					Gly	Gly			
NS5-200	Arg																						Leu			
NS5-277	Ser														Arg											
NS5-287	Arg	Lys																								
NS5-296	His																				Tyr					
NS5-314	Lys																								Arg	
NS5-531	Lys													Arg												
NS5-560	Ala																									Ser
NS5-619	Ala																							Ser		
NS5-688	Ala																						Asp			
NS5-804	Ala																					Val	Val			
Total		3	2	2	1	4	2	1	3	3	1	2	3	3	2	1	7	1	5	1	4	4	6	4	6	3

Amino acid positions determined from the sequence alignment of the 25 completely sequenced isolates in comparison to WN-NY99 (Accession no. AF196835).

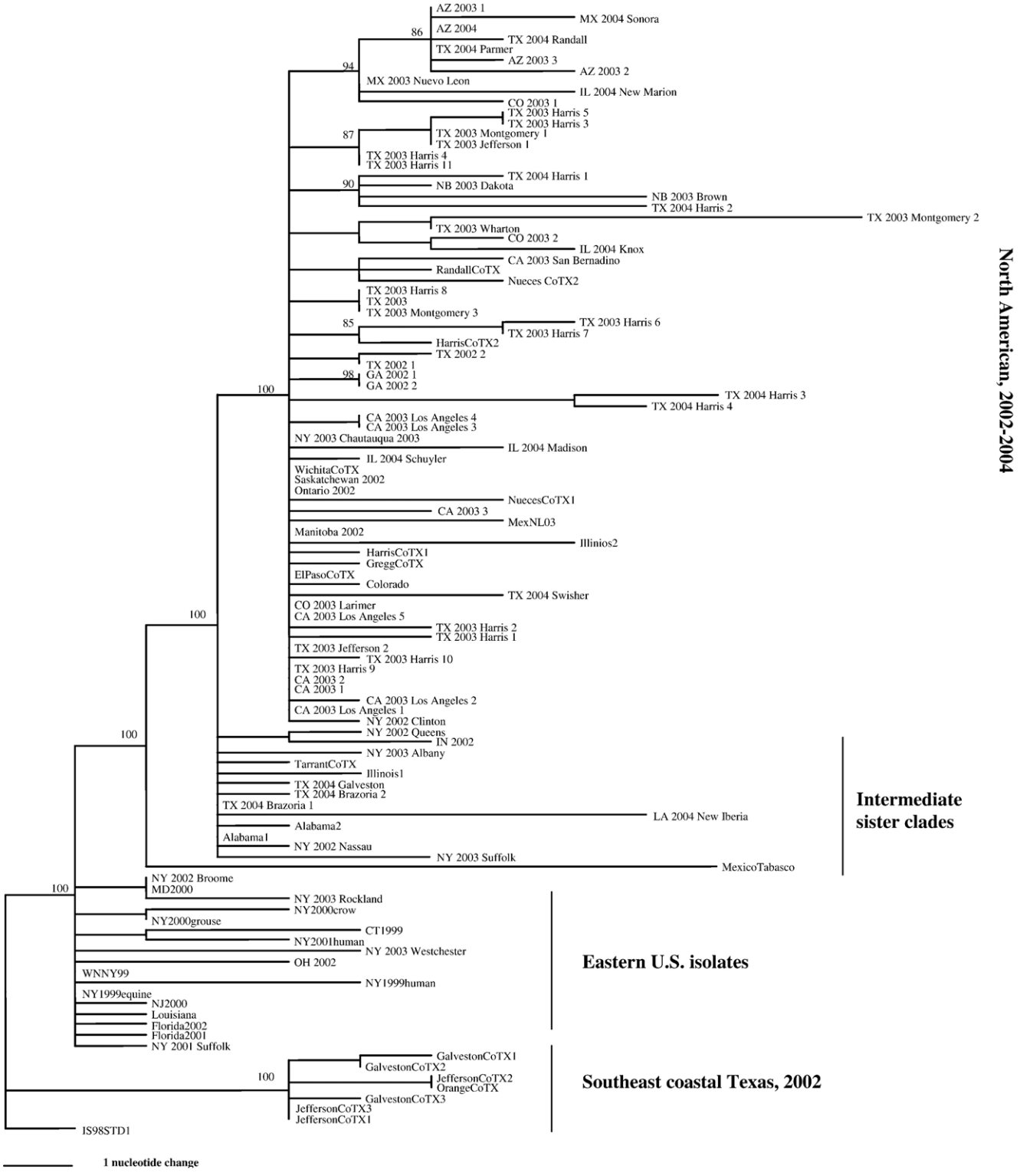

Fig. 1. Phylogenetic tree constructed by Bayesian analysis of prM and E genes (2004 ntds.) of 108 North American WNV isolates.

World strain, Israel-1998. The complete genome analysis generated a tree that also illustrates the presence of the North American 2002–2004 clade, relative to the Eastern U.S. clade and illustrates the presence of more highly resolved subclades within each of the larger clades. As was noted in the prM and E gene tree, the majority of fully resolved subclades were comprised of isolates that were made in the same geographic region and during the same year. It is interesting to note that the complete genome analysis placed an isolate from the southeast coast of Texas in the same clade as other isolates from the Eastern U.S. Because of the lack of nucleotide mutations in isolates

belonging to the Eastern U.S. clade relative to the North American 2002–2004 clade, sub-groupings were less resolved and the clade consisted of a large polytomy. Interestingly, the resolved sub-groupings were comprised of northeastern isolates that were made primarily in 2002 and 2003. Also of interest was an intermediate clade found between the two larger clades that is comprised of an isolate from Tabasco State, Mexico, an isolate from Ohio, 2002, and an isolate from New York, 2000, indicating the likelihood that additional clades within the North American monophyletic clade will continue to be identified as additional genome sequences become available (Fig. 2).

Fig. 3 shows two additional phylogenetic trees that represent the overall tree topology generated from analyses of each of the individual genes and untranslated regions of the WNV genome. Bootstrap values are shown only at clades and subclades that were resolved in each of the trees generated by analysis of individual genes or regions. The "group B" topologies are similar to those seen in the complete genome analysis in that the two large clades can

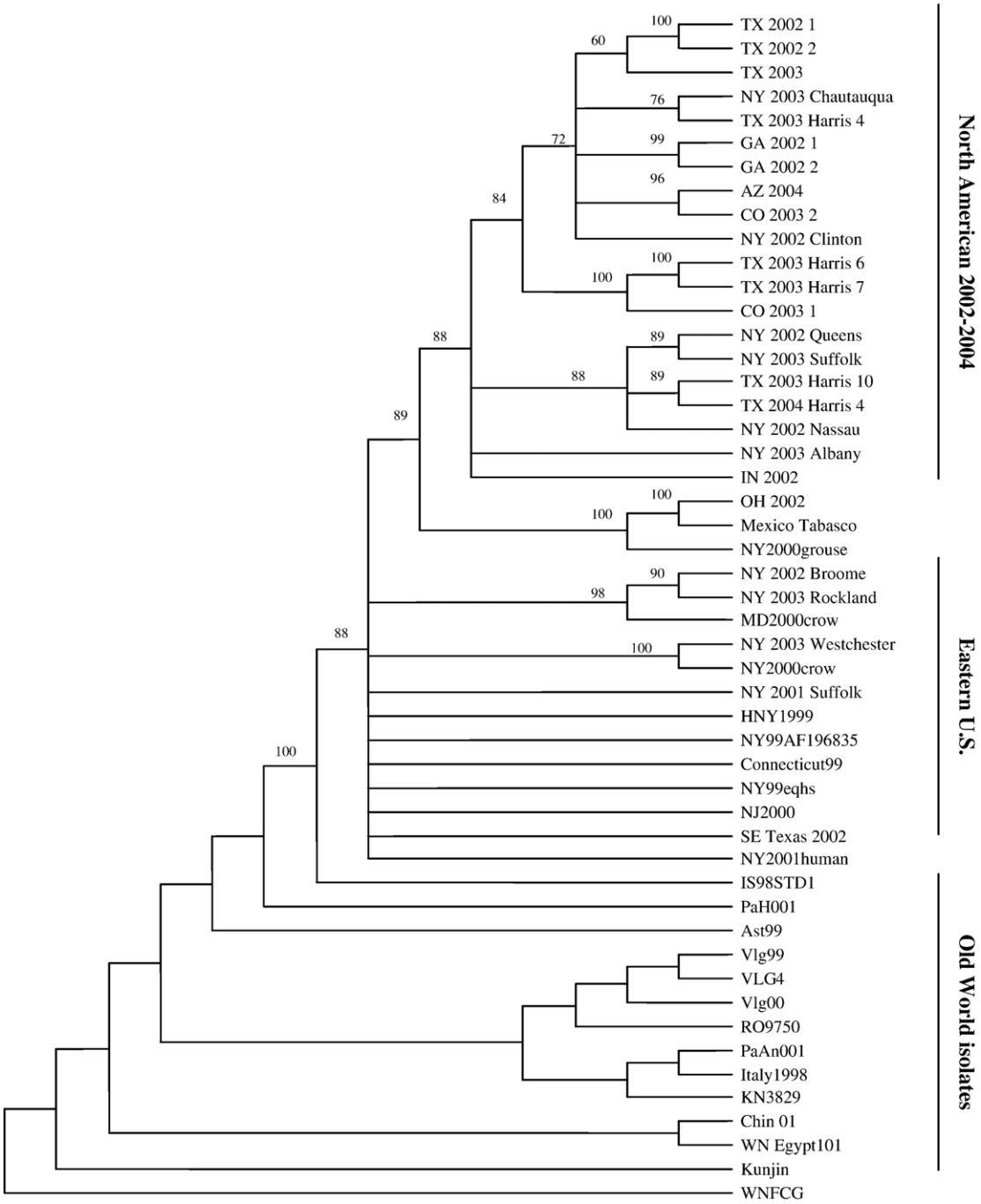

Fig. 2. Phylogenetic tree generated by maximum likelihood analysis of an alignment of the complete genomes of 25 newly sequenced North American WNV isolates with other U.S. and Old World isolates.

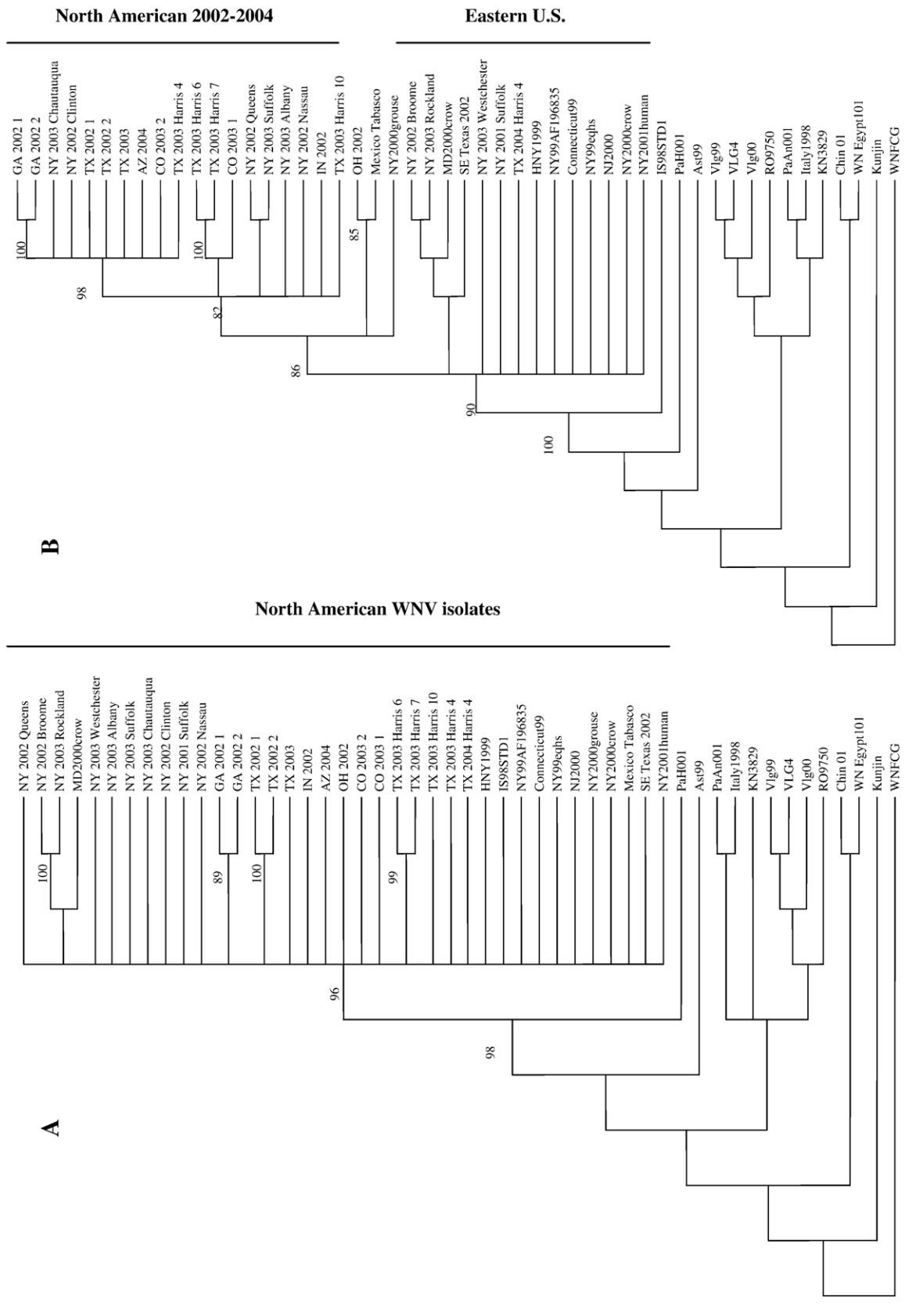

Fig. 3. Representative phylogenetic trees generated by maximum likelihood analysis of each of the individual WNV genes, and untranslated regions showing presence or absence of clade topology with North American, 2002–2004 clade separate from Eastern U.S. clade. Group "A" represents tree topology generated from analyses of 5′UTR, capsid, NS1, NS2B, and NS4A. Group "B" represents tree topology generated from analyses of prM, E, NS2A, NS3, NS4B, NS5, and 3′UTR.

Table 4
Nucleotide mutations conserved in all isolates of North American clade 2002–2004 relative to WN-NY99

5′UTR	Capsid	prM	E	NS1	NS2A	NS2B	NS3	NS4A	NS4B	NS5	3′UTR
None	None	660 (C to U)	1442 (U to C)[a] 2466 (C to U)	None	3774 (U to C) 4146 (A to G)	None	4803 (C to U) 6138 (C to U) 6238 (C to U) 6426 (C to U)	None	6996 (C to U)	7938 (U to C) 9352 (C to U)	10851 (A to G)

Nucleotide positions correspond to WN-NY99.
[a] Encodes E159 (Val to Ala) amino acid substitution.

be readily distinguished from one another with average bootstrap values >85. The group B topology was generated from analyses of the prM, E, NS2A, NS3, NS4B, NS5, and 3′UTR. In contrast, "group A" trees (generated from analyses of the 5′UTR, capsid, NS1, NS2B, and NS4A) do not delineate between the two clades and illustrate only the monophyletic North American clade relative to Old World isolates. Furthermore, group A trees do not place Israel-1998 as a distinct outlier to the North American clade. A lack of parsimony informative mutations within certain genes or untranslated regions is likely to relate to the lack of topology seen in the group A trees. Accordingly, the complete genome alignment of the completely sequenced WNV isolates was used to identify the nucleotide and deduced amino acid mutations that are conserved in isolates belonging to the North American 2002–2004 clade. Table 4 shows that no conserved mutations have accumulated in the genes (capsid, NS1, NS2B, and NS4A) or untranslated region (5′UTR) that produce trees with no resolution of clade topology. In contrast, at least one nucleotide mutation is conserved in individual genes or the untranslated region that results in trees that show the presence of multiple clades. The largest number of conserved mutations was found in the NS3 gene. Despite the occurrence of 13 nucleotide mutations in each of the isolates belonging to the North American 2002–2004 clade, there has been only a single amino acid substitution conserved in these isolates (E-V159A).

Discussion

Since the detection of WNV in the U.S. in 1999, studies concerning the molecular epidemiology of the virus have proven useful for tracing the geographic and temporal spread and interhost transmission of the virus (Davis et al., 2003; Ebel et al., 2004; Solomon et al., 2003; Estrada-Franco et al., 2003; Blitvich et al., 2004) and in detecting phenotypic variants (Beasley et al., 2004; Davis et al., 2004; Ebel et al., 2004). Because the location and year of the first virus isolations have been well documented (Lanciotti et al., 1999, 2002), it has been possible to make genetic and phenotypic comparisons between isolates made at the onset of the North American WNV epizootic and isolates made from all subsequent years and over a broad geographic distribution (Anderson et al., 2001; Beasley et al., 2003; Davis et al., 2003, 2004; Ebel et al., 2001, 2004). Consequently, genome sequences of isolates made in 1999 and 2000 provide a genetic baseline allowing for the identification of novel mutations to the genomes of more recently isolated strains, which can then be used to infer phylogenetic relationships among isolates. Previous phylogenetic studies have focused on sequencing complete genomes of virus isolates from the Northeastern U.S. and only partial sequences (primarily the prM and/or E genes) of isolates from other regions of North America. Accordingly, this study has relied on both partial and complete genome sequences of isolates in order to further characterize the phylogenetic relationships of North American WNV isolates and to better understand how this virus has evolved since its introduction 6 years ago.

Our laboratories sequenced the prM and E genes of the genome for 74 North American isolates because previous studies of WNV molecular epidemiology have relied on sequencing of this region. Furthermore, because of the relatively few complete nucleotide sequences of North American WNV isolates available for analysis prior to this study, we have sequenced complete genomes of isolates representing a wide range of host species, collection years, and locations. Although the prM and E gene sequences demonstrate an accurate and reliable phylogenetic representation, analysis of complete genomes of North American WNV isolates provides a more robust comparison of the evolutionary relationships between isolates and reveals additional mutations to the WNV genome that may have implications for the phenotypic properties that these isolates express (Davis et al., 2004; Beasley et al., 2004; Ebel et al., 2004). Moreover, complete genome analysis allows one to compare isolates at the individual gene level in order to identify those genes which may be more reliable for the comparison of parsimony informative sites among isolates. Therefore, this study has used a variety of methods to infer phylogenetic relationships of North American WNV isolates and to further define the evolution of WNV.

Data from complete nucleotide and amino acid sequence analysis supports previous findings suggesting that WNV has continued to diverge from progenitor isolates (those made in the Northeastern U.S. at the early stages of WNV spread) as the temporal and geographic distribution of the virus has expanded. It was also evident from this analysis that isolates displaying the highest degree of nucleotide and amino acid sequence divergence, in general, were collected

after 2002. This may be significant as 2002 was the year of the largest recorded epidemic of arboviral encephalitis in North America and saw a vast geographic expansion by the virus (CDC, 2002). The nucleotide divergence may reflect the nature of the error-prone polymerase of WNV in that as the virus continues to replicate over time and geographic location, additional mutations to the genomes of viruses will continue to arise. Although the majority of these mutations are silent, it is apparent from this and other studies that amino substitutions are not uncommon and have resulted in phenotypic variation within the North American WNV population (Beasley et al., 2004; Davis et al., 2004; Ebel et al., 2004). Despite the continuing divergence of the virus from progenitors, many of the accumulated nucleotide mutations to the genome have become fixed in currently circulating viruses. Many of these fixed mutations were found in isolates from Georgia, Texas, and Indiana in 2002 indicating their presence since at least 2002 and were most likely found in those viruses that were circulating at the forefront of the western expansion of the virus. Consequently, all isolates sequenced from the western U.S. contain the majority of those fixed mutations shown in Table 4, a result which may reflect the rapid westward progression of the virus from 2002 to 2004. In contrast, there has been only a single conserved amino acid substitution (E-V159A) shared in the majority of isolates sequenced for this study, perhaps reflecting the inability of the virus population to sustain substitutions in viral proteins which may be disadvantageous to viral fitness. Incidentally, this particular substitution is found in many Old World WNV strains (data not shown) and may represent a non-critical residue that has reverted back to a residue common in ancestral strains. Undoubtedly, as the virus continues to evolve, additional amino acid substitutions will become fixed in the population at large, the consequences of which remain speculative at this point.

Phylogenetic analysis of prM and E gene sequences revealed three distinct clades within the North American WNV monophyletic group. As noted by previous studies, the majority of isolates collected since 2002 belong to a single clade that has been termed the North American 2002–2004 clade. The emergence of this clade appears to correspond with the increased intensity of transmission and western progression of WNV across the U.S. during 2002 and indicates that, as the distribution of the virus expanded, mutations to the genomes of circulating viruses began to accumulate and became fixed in the overall virus population. The finding that the majority of isolates made from 2002 to 2004 belong to this clade suggests the emergence of a dominant genetic variant whose distribution is now uniform across all of North America. It is also apparent that this variant has began to replace other genetic variants in certain regions (i.e., Eastern U.S. and Southeast coastal Texas variants). Findings by Ebel et al. (2004) suggest that this displacement and rapid spread may be correlated with increased transmission efficiency of the dominant genetic variant in *Culex* spp. mosquitoes when compared to other variants. Additional phenotypic studies will be required to characterize any fitness advantages that this variant may have over others.

Analysis of the prM and E genes of North American WNV isolates also separated many isolates of the North American 2002–2004 clade into subclades that represent isolates from certain regions of North America. Of note was a subclade consisting of isolates from Mexico, Arizona, and Texas, suggesting that transmission of viruses in those regions may be the result of localized virus spread by resident birds rather than by migratory birds that may introduce more distantly related viruses into a particular area. Although not always the case, several subclades were made up of isolates from the same state during the same year, illustrating that genetic distance among WNV isolates often reflects geographic and temporal distances. In contrast, some intermediate clades, as illustrated in the complete genome tree, may also represent the long-distance spread of WNV by migratory birds, in contrast to the more localized spread by non-migratory species. Evidence that migratory birds are capable of the dispersal of discrete genetic variants over long distances is supported by the close genetic relatedness of the Tabasco isolate to a 2002 Ohio isolate.

As was noted by Blitvich et al. (2004), Bayesian analysis of WNV isolates that differ by only a few informative sites can be used as a more efficient method for the generation of equally or more robust phylogenetic trees in comparison to more commonly used phylogenetic methods (i.e., parsimony, likelihood analysis). Similar topologies were created by maximum parsimony, neighbor-joining, and Bayesian methods, and the Bayesian analysis generated equally or more significant confidence values at internal nodes. Bayesian analysis was preferred to other methods because the analysis was also less computationally exhaustive due to the large number of isolates compared in the prM and E sequence alignment. While analysis using complete genome sequences is clearly the most informative method of describing evolutionary relationships among viral isolates, a lack of complete sequences in public databases makes this technique less useful. Thus, Bayesian analysis may provide a more efficient method for describing the relationships between WNV isolates that share high nucleotide identities with few informative sites.

Complete nucleotide sequence analysis revealed similar tree topology as that found in the analysis of prM and E genes. The North American monophyletic group in comparison to Old World isolates remained divided into two separate clades, the North American 2002–2004 and Eastern U.S. clade. It is likely that the Southeast coastal Texas clade was not resolved by the complete genome analysis because only a single isolate from this clade has been completely sequenced. The results demonstrate that all isolates collected after 2002 with the exception of a few

isolates from New York in 2003 are contained in the North American 2002–2004 clade. Thus, it is evident from the complete genome analysis that a dominant WNV variant has emerged in all of the regions sampled in this study. This analysis also reveals that isolates collected in close spatial and temporal proximity cluster in discrete subclades. Despite the existence of a large polytomy in the Eastern U.S. clade, the more highly evolved North American clade contains more fully resolved subclades, again suggesting that isolates collected outside of the Northeastern U.S. and after 2002 depict the continuing divergence of isolates from progenitors and that their phylogenetic relationships can be more resolved because of a larger number of conserved mutations to their genomes. The existence of these groupings continues to suggest that the rate of evolution (albeit on a microevolutionary scale) occurring between transmission seasons and as WNV spreads from region to region is rapid enough to drive the emergence of genetic variants and is likely to continue to do so as the virus distribution increases. Even though the overall tree topology was conserved between the prM and E gene analysis and the complete genome analysis, when using alignments of individual genes or untranslated regions, the phylogenetic relationships of North American West Nile viruses become less clear. While analyses of certain genes or untranslated regions were able to generate the topologies illustrated by complete genome or prM and E analyses, many did not. Furthermore, in those analyses that did result in tree topology demonstrating the separation of clades, the resolution of subclades was far less pronounced. The differences in tree topology based on individual genes and untranslated regions may also reflect the ability of some genes or untranslated regions to withstand mutations to the genome, while others are unable to withstand mutations because of structural or functional constraints. Alternatively, the rates of evolution of individual genes or untranslated regions may differ because of selective pressures that drive the evolution of certain genes/regions. Thus, while certain individual genes or untranslated regions may be able to act as surrogates to complete genomes when conducting phylogenetic analyses, combinations of genes appear to provide a more accurate depiction of genetic relatedness among isolates in lieu of complete genome sequencing. Despite this finding, it is likely that, as more sequence data become available for isolates, the topologies derived from the analysis of each gene or untranslated region will become more consistent. As additional sequence information becomes available from geographic regions underrepresented by this study and in regions where WNV isolates have yet to be obtained for sequencing (i.e., Caribbean islands and Central America), the molecular epidemiology of this emerging virus will continue to be defined. This and previous studies have provided a database of sequence information which can be utilized in the future to trace the spread of WNV as its distribution expands and to further define the modes of transmission and spread that are important to the maintenance and evolution of this virus in nature.

Materials and methods

Strains

All newly sequenced WNV isolates came from a variety of sources including birds, mosquitoes, horses, and humans as described previously (Lillibridge et al., 2004; Davis et al., 2003). A total of 74 isolates were collected from several locations including Texas, Nebraska, Ohio, Indiana, Illinois, Georgia, Florida, Louisiana, Arizona, California, Colorado, Mexico, and Canada (Table 1).

RT-PCR

The prM and E genes of each isolate and complete genomes of 25 isolates were sequenced by RT-PCR from RNA extracted from infected tissues, cell culture supernatants of original isolations, or supernatants after a single passage in African green monkey (Vero) cell culture. RT-PCR protocols, primer sequences, and sequencing methods have been described elsewhere and are available upon request (Davis et al., 2003, 2004).

Phylogenetic analysis

Following completion of either partial or complete genome sequencing, nucleotide and deduced amino acid sequences were aligned with sequences from other WNV isolates found in GenBank which represented homologous sequence regions. The AlignX program of the Vector NTI Suite software package (Informax, Frederick, MD) was used to generate all alignments. The GenBank accession numbers for isolates used in the generation of phylogenetic trees are found in each figure. Bayesian analysis was used to generate a phylogenetic tree based on the prM–E alignment of the 74 newly sequenced North American WNV isolates with other North American WN viruses available in GenBank and rooted with the most closely related Old World WNV, Israel-1998 (Accession no. AY033389) in order to create a tree illustrating more parsimony informative sites. Phylogenies were generated by the program MRBAYES (Version 2.0) (Huelsenbeck and Ronquist, 2001) using the Metropolis coupled Markov chain Monte Carlo algorithm run with four chains over 150,000 generations under a general time-reversible model with a burn-in time of 50,000 generations. Rate heterogeneity was estimated using a γ distribution for the variable sites. The consensus phylogram among the 108 WNV isolates analyzed was then generated from the MRBAYES output file using PAUP (Version 4.0b11, Sinauer Associates, Sunderland, MA) with clade credibility values at relevant nodes to demonstrate statistical support

for each clade. The Bayesian consensus tree was compared to trees generated by neighbor-joining and maximum parsimony analyses using PAUP in order to determine if similar tree topologies were generated using different methods. Additional phylogenetic trees were generated from alignments of complete genomes or individual genesntranslated regions (5′UTR, Capsid, prM, E, NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5, and 3′UTR) comparing the 25 completely sequenced isolates to all other WNV complete genome sequences available in GenBank. Maximum likelihood analysis was used to generate all additional trees using PAUP under the general time-reversible model with a γ distribution of substitution rates. Confirmation of tree topology was done using maximum parsimony and neighbor-joining methods with statistical support for relevant clades provided by 500 bootstrap replicates using the neighbor-joining analysis.

Acknowledgments

CTD is supported by the James W. McLaughlin Fellowship Fund and would like to thank the UTMB Protein Chemistry Laboratory for nucleic acid sequencing and Dr. Scott Weaver for helpful discussions. GDE and LDK would like to thank Kristen A. Bernard for helpful discussions, the New York State Department of Health, the New York State Wildlife Pathology Unit, the Wadsworth Center Arbovirus Laboratories, and the Wadsworth Center Molecular Genetics Core facility and is funded in part by NIH contract N01-AI25490 and by CDC contract U50/CCU223671-02. ACB is funded in part by the California Mosquito Research Program and wishes to thank the California Department of Health Services, the California Animal Health and Food Safety Laboratory, as well as the California Mosquito and Vector Control Districts for providing the materials from which viral isolates were made. WNV isolates from Arizona were graciously provided by Ronald C. Cheshier of the Arizona Department of Health Services. DEQ is supported by the Tropical Vector-borne Viral and Rickettsial Infections grant (5D43TW006590) sponsored by the Fogarty International Center. ADTB is supported in part by Centers for Disease Control and Prevention (CDC) cooperative grant U90/CCU620916. RBT is supported in part by NIH contracts N01-AI25489 and N01-AI30027. We also thank Dr. Wayne Kramer from the Nebraska Department of Health and Human Services for providing additional virus isolates.

References

Anderson, J.F., Vossbrinck, C.R., Andreadis, T.G., Beckwith, 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.C.W., Davis, C.T., Guzman, H., Vanlandingham, D.L., Travassos da Rosa, A.P.A., Parsons, R.E., Higgs, 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.C., Davis, C.T., Estrada-Franco, J., Navarro-Lopez, R., Campomanes-Cortes, A., Tesh, R.B., Weaver, S.C., Barrett, A.D.T., 2004. Complete genome sequence of a West Nile virus isolate from Mexico and identification of an attenuating mutation. Emerging Infect. Dis. 10 (12), 2221–2224.

Blitvich, B.J., Fernandez-Salas, I., Contreras-Cordero, J.F., Marlenee, N.L., Gonzalez-Rojas, J.I., Komar, N., Gubler, D.J., Calisher, C.H., Beaty, B.J., 2003. Serologic evidence of West Nile virus infection in horses, Coahuila State, Mexico. Emerging Infect. Dis. 9, 853–856.

Blitvich, B.J., Fernandez-Salas, I., Contreras-Cordero, J.F., Marlenee, N.L., Gonzalez-Rojas, J.I., et al., 2004. Phylogenetic analysis of West Nile virus, Nuevo Leon State, Mexico. Emerging Infect. Dis. 10, 1314–1317.

CDC, 2002. Provisional surveillance summary of the West Nile virus epidemic—United States, January–November 2002. Morb. Mort. Wkly. Rep. 51, 1129–1133.

Davis, C.T., Beasley, D.C.W., Guzman, H., Raj, P., D'Anton, M., Novak, R.J., Unasch, T.R., Tesh, R.B., Barrett, A.D.T., 2003. Genetic variation among temporally and geographically distinct West Nile virus isolates collected in the Unites States, 2001 and 2002. Emerging Infect. Dis. 9, 1423–1429.

Davis, C.T., Beasley, D.C.W., Siirin, M., Guzman, H., Tesh, R.B., Barrett, A.D.T., 2004. Emergence of attenuated West Nile virus isolates in Texas, 2003. Virology 330, 342–350.

Dupuis II, A.P., Marra, P., Kramer, L.D., 2003. Serological evidence of West Nile virus transmission, Jamaica, West Indies. Emerging Infect. Dis. 9, 860–863.

Ebel, G.D., Dupuis, A.P., Ngo, K., Nicholas, D., Kauffman, E., Jones, S.A., et al., 2001. Partial genetic characterization of West Nile virus strains, New York State, 2000. Emerging Infect. Dis. 7, 650–653.

Ebel, G.D., Carricaburu, J.E., Young, D.S., 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.

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

Hall, R.A., Broom, A.K., Smith, D.W., Mackenzie, J.S., 2002. The ecology and epidemiology of Kunjin virus. Curr. Top. Microbiol. Immunol. 267, 253–269.

Huelsenbeck, J.P., Ronquist, F.R., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Komar, O., Robbins, M.B., Klenk, K., Blitvich, B.J., Marlenee, N.L., Burkhalter, K.L., Gubler, D.J., Gonzalvez, G., Pena, C.J., Peterson, A.T., Komar, N., 2003. West Nile virus transmission in resident birds, Dominican Republic. Emerging Infect. Dis. 9, 1299–1302.

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 US. Science 286, 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, 96–105.

Lillibridge, K.M., Parsons, R., Randle, Y., Travassos da Rosa, A.P.A., Guzman, H., Siirin, M., Wuithiranyagool, T., Hailey, C., Beasley, D.W., Higgs, S., Pascual, R., Meyer, T., Barrett, A.D.T., Tesh, R.B., 2004.

The 2002 introduction of West Nile virus into Harris County, TX, an area historically endemic for St. Louis encephalitis. Am. J. Trop. Med. Hyg. 70, 676–681.

Murgue, B., Zeller, H., Duebel, V., 2002. The ecology and epidemiology of West Nile virus in Africa, Europe, and Asia. Curr. Top. Microbiol. Immunol. 267, 195–221.

Quirin, R., Salas, M., Zientara, S., Zeller, H., Labie, J., Murri, S., Lefrancois, T., Petitclerc, M., Martinez, D., 2004. West Nile virus, Guadeloupe. Emerging Infect. Dis. 10, 706–708.

Solomon, T., Fisher, A.F., Beasley, D.W., Mandava, P., Granwehr, B.P., Langsjoen, H., Travassos Da Rosa, A.P., Barrett, A.D., Tesh, R.B., 2003. Natural and nosocomial infection in a patient with West Nile encephalitis and extrapyramidal movement disorders. Clin. Infect. Dis. 36 (11), 140–145.