Available online at www.sciencedirect.com

ScienceDirect

Virology 363 (2007) 220–228

VIROLOGY

www.elsevier.com/locate/yviro

Phylogenetic analysis of WNV in North American blood donors during the 2003–2004 epidemic seasons

Belinda L. Herring [a,b], Flavien Bernardin [a,b], Sally Caglioti [c], Susan Stramer [d], Leslie Tobler [a], William Andrews [e], Lawrence Cheng [f], Sarah Rampersad [g], Cherie Cameron [g], John Saldanha [g], Michael P. Busch [a,h], Eric Delwart [a,b,*]

[a] *Blood Systems Research Institute, San Francisco, CA 94118, USA*
[b] *Department of Medicine, University of California, San Francisco, CA 94118, USA*
[c] *Blood Systems Laboratory, Tempe AZ, USA*
[d] *American Red Cross, Gaithersburg, MD, USA*
[e] *Chiron, Emeryville, CA, USA*
[f] *Roche Molecular Systems, Alameda, CA, USA*
[g] *Canadian Blood Services, Ottawa, ON, Canada*
[h] *Department of Laboratory Medicine, University of California, San Francisco CA 94143, USA*

Received 4 December 2006; returned to author for revision 9 January 2007; accepted 18 January 2007
Available online 26 February 2007

Abstract

West Nile Virus (WNV) collected from 179 human blood donors in 25 US states and three Canadian provinces during the 2003 and 2004 epidemic seasons were genetically analyzed. The evolution of WNV during its Western spread was examined by envelope (E) gene sequencing of all 179 cases and full open reading frame sequencing of a subset of 20 WNV to determine if geographic and temporal segregation of distinct viral variants had occurred. Median joining network analysis was used to examine the genetic relationship between E gene variants and identified four large genetic clusters showing the gradual accumulation of mutations during the virus' western expansion. Two related WNV variants and their descendents, undetected in prior years, expanded in frequency. Apparent founder effects were observed in some regional outbreaks possibly due to local WNV colonization by a limited number of viruses. Amino acid mutations associated with newly expanding genetic variants reflect either selectively neutral mutational drift and/or mutations providing replicative advantages over the previously dominant forms of WNV.

Keywords: West Nile Virus; Blood donors; North America; Evolution; Phylogenetics

Introduction

In 1999, West Nile Virus (WNV) (*Flaviviridae*, *Flavivirus*) was introduced in the western hemisphere causing an outbreak characterized by an equine and avian epizootic and 62 cases of humans encephalitis in New York City resulting in seven deaths (Murgue et al., 2002; Nash et al., 2001). Full genome sequences of early North American WNV variants demonstrated the close genetic relationship between US WNV variants and those obtained in 1998 and 1999 from Israel (Jia et al., 1999; Lanciotti et al., 1999). Since then, WNV distribution has expanded to most parts of North America including all contiguous US states, Canada, the Caribbean islands and Mexico (Blitvich et al., 2003; Estrada-Franco et al., 2003; Komar et al., 2005; Lefrancois et al., 2005). WNV has also been a significant public health issue since 1999 as thousands have been symptomatically infected (http://www.cdc.gov/ncidod/dvbid/westnile/).

RNA viruses are notable for their high mutations rates and therefore their potential for rapid evolution (Holland et al., 1982) although for WNV the need to replicate in both insects and birds may impose special constraints on their genomic plasticity. Focusing largely on sequencing of the pre-membrane (prM) and the envelope (E) genes, as well as full open reading frames (ORF), prior analyses have revealed mutations relative

* Corresponding author. Blood Systems Research Institute, 270 Masonic Ave, San Francisco, CA 94118, USA. Fax: +1 415 2762311.
E-mail address: delwarte@medicine.ucsf.edu (E. Delwart).

0042-6822/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2007.01.019

to the prototype 1999 North American WNV virus during the expansion of the virus across North America (NY99, GenBank Accession no. AF196835). In some cases viral mutations could be grouped temporally or geographically (Beasley et al., 2003; Davis et al., 2003, 2005; Ebel et al., 2004; Lanciotti et al., 2002). These studies have highlighted the emergence of a new dominant variant lineage, named the North American 2002–2004 clade, increasingly prevalent in numerous North American areas in recent years (Davis et al., 2005; Ebel et al., 2004).

Because of the transient viremia in humans which resolves prior to the onset of symptoms, very few WNV variant sequences have been derived from infected humans and most WNV have been derived from infected mosquito pools, directly or after culture in Vero cells, as well as various avian species and horses. Large-scale investigation of the WNV actually transmitted to humans has therefore not been previously possible. In 2006 less than 15 WNV sequences available in Genbank were labeled as derived from humans.

In 2002 it was determined that WNV could be transmitted through blood transfusions (CDC, 2002; Harrington et al., 2003; Hollinger and Kleinman, 2003; Pealer et al., 2003). As a consequence screening of the blood supply for WNV RNA was initiated in 2003, removing a significant number of potentially infectious donations (Busch et al., 2005, 2006; Kleinman et al., 2005; Stramer et al., 2005; Montgomery et al., 2006). We report here on the E gene and full ORF sequences of WNV variants from voluntary blood donors collected across the US and Canada in 2003 and 2004.

Results

Amplification of WNV samples and sequence analysis

The E gene was amplified by nested PCR from 179 plasma samples from WNV RNA positive blood donors and directly sequenced (see Materials and methods). Plasma samples were obtained from 46 blood centers in 25 US states and 3 Canadian provinces during the 2003 and 2004 WNV epidemic seasons (Busch et al., 2006; Montgomery et al., 2006; Stramer et al., 2005) (Table 1). Nucleotide sequences of the E gene were compared to the prototype strain from 1999 NY (AF196835). Overall, nucleotide differences were observed at a total of 52 positions (3.6%) and the majority of changes ($n=45$) were synonymous. Only two out of 179 WNV sequences were identical to NY99, both collected in 2003 in central Florida. An amino acid change at E-V159A resulting from a U>C mutation at position 1442 was observed in all but 3 (all from Central Florida) of our 179 WNV sequences collected in 2003 and 2004. While the large majority of mutations were sporadic (observed in only one virus) others were more common necessitating the use of network analysis to facilitate visualization of the relationships between these closely related variants.

Ambiguous or mixed bases were also observed at numerous positions following direct sequencing of the WNV RT-PCR products, indicating the presence of genetically mixed viral quasi-species in these recently infected, pre-seroconversion, blood donors. Such viral quasi-species diversity may have very rapidly evolved following a clonal human infection or result from inoculations with multiple WNV variants from insect bite(s) (Jerzak et al., 2005).

Table 1
Geographical distribution of plasma samples

State	City	Year	
		2003	2004
Arizona	Tempe		10
Arkansas	Little Rock		1
California	San Francisco		1
	Los Angeles		3
	La Habra		1
	San Diego		1
	Ontario		1
Colorado	Penrose	2	
	Grand Junction		8
Florida	Miami	2	1
	Orlando	1	
	Tampa	9	
	Gainsville	3	
Illinois	Glenview	2	
Indiana	Indianapolis	3	
	South Bend	3	
Iowa	Davenport	1	
	Sioux City	2	
Louisiana	New Orleans	6	
	Shreveport		2
Minnesota	Minneapolis	3	
	Rochester	1	
	Sanborn		1
Missouri	Kansas City	3	
Nebraska	Lincoln	3	
Nevada	Las Vegas		2
New Mexico	Albuquerque	3	8
New York	New York	3	
North Dakota	Bismark	12	
Ohio	Elyria	1	
Oklahoma	Oklahoma City	8	
Pennsylvania	Bethlehem	1	
South Carolina	Goose Creek		1
South Dakota	Rapid City	13	1
	Sioux Falls	3	
Tennessee	Memphis	1	
	Jackson	2	
Texas	Lubbock	4	
	Amarillo	7	
	Dallas	2	
	San Antonio	5	
	El Paso	4	
	Houston	1	
	Tyler	5	
Wisconsin	Milwaukee	4	
Wyoming	Cheyenne	2	
Canada			
Alberta		3	
Manitoba		1	
Sasketechuan		8	
Total		137	42

Network analysis of E gene variants

Median joining network analysis was used here to facilitate visualization of the genetic relationship between the WNV

(Bandelt et al., 1999). Network analysis was performed using sequences from this study and other US and Canadian derived WNV sequences available in Genbank at the time of analysis. Networks are labeled by years of sample collection (1999–2004) (Fig. 1a) and by geographic origin (eastern, central, mountain, and pacific states time zones) (Fig. 1b).

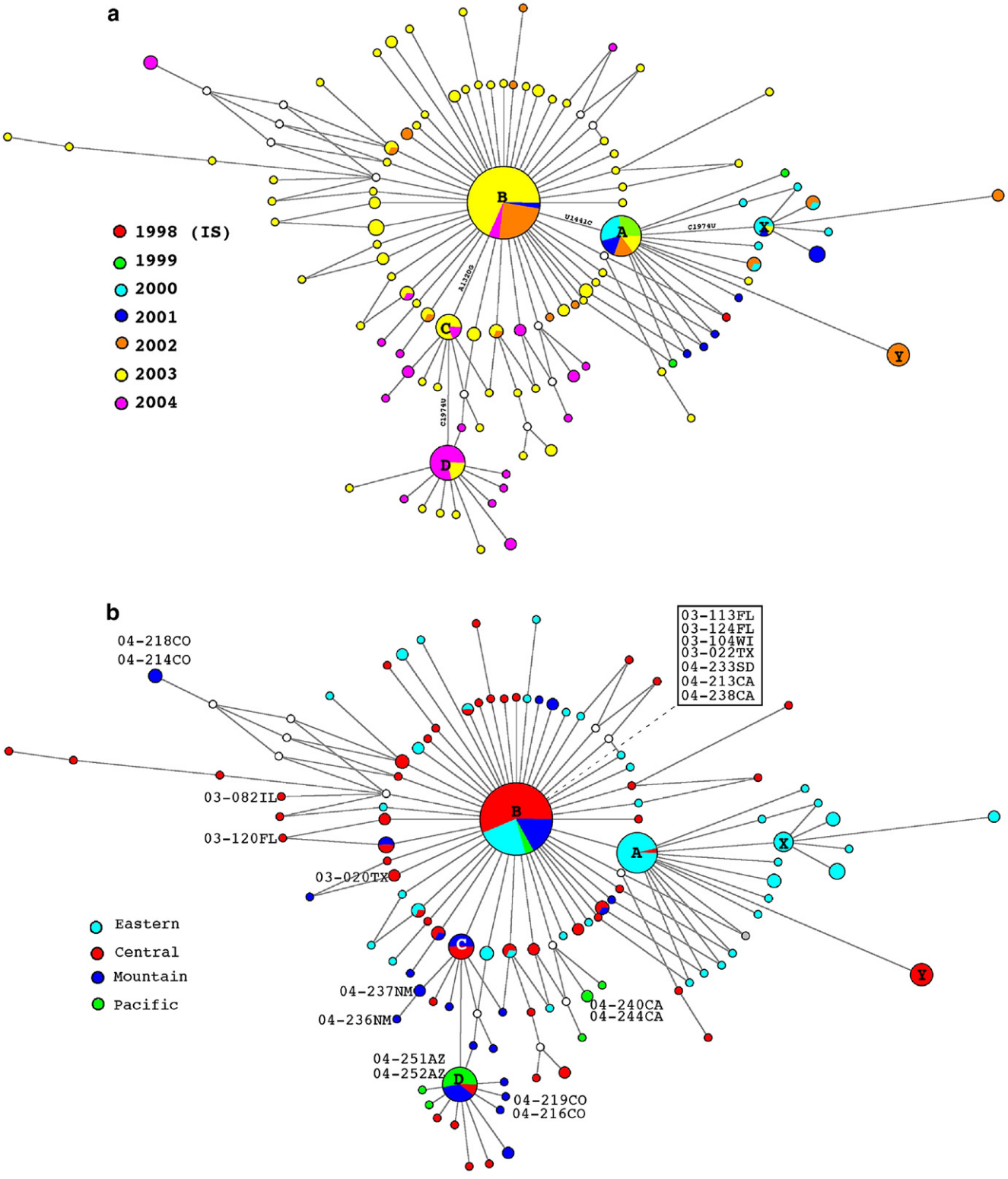

Fig. 1. Genetic network analysis of E gene region of North American WNV. (a) Labeled by year of collection from 1998 (Israel) to 2004. Location of E gene mutations linking nodes X, A, B, C, and D is included. (b) Labeled by time zone. Network position of blood donors WNV variants whose full genomes are analyzed in Fig. 2 corresponds to colored circle closest to variant name. Nodes labeled A–D, X, and Y are defined in text. White circles correspond to hypothetical intermediate variants. Multiple lines converging on a single peripheral variant indicate alternative mutational pathways.

The length of network connections reflects the number of mutations.

The 1998 sequence from Israel (Fig. 1a in red) is the pre-existing WNV variant most closely related to those of the 1999 N. American outbreak. In the E gene region analyzed, one to two mutations separated IS-98 from all US variants collected in 1999 (Fig. 1a in green). All 1999 US variants fell within node A or directly radiated from it (Fig. 1a in green). Also connected to node A and carrying the ancestral North American nucleotide U at position 1442 were all the variants collected in 1999 and 2000 and the large majority of 2001 WNV. The 26 WNV variants within node A were exclusively from the Eastern US states with a single exception from Louisiana (Fig. 1b). All variants radiating from node A were also exclusively from Eastern US states except for a genetically homogenous but highly divergent group of 8 WNV variants detected in 2002 from birds and insects in Southern Texas (Davis et al., 2003) (Fig. 1b node Y). No member nor descendent of this divergent group of Southern Texas WNV variants have since been reported (Davis et al., 2005).

One or more node A variant(s) carrying the non-synonymous U1442C mutation (resulting in E-V159A) founded E gene node B which contains the largest number of WNV in this analysis and corresponds to the North American, 2002–2004 cluster defined by Davis et al. (2005) (Figs. 1a and b node B). The majority of WNV in blood donors in central US states and the majority of 2003 variants overall belonged to node B and its radiating variants.

One or more node B variant(s) carrying a synonymous A1320G mutation then founded node C consisting of 10 WNV from the central and mountain states collected in 2003 and 2004 (Figs. 1a and b).

One or more node C variant(s) carrying a synonymous C1974U mutation then gave rise to the D node (Figs. 1a and b). The majority of node D and radiating variants in viremic donors were collected in the mountain states in 2004, mainly from Arizona.

Using network analysis, a stepwise accumulation of mutations in the E gene loci was therefore associated with the westward progression of WNV (node A>U1442C>node B>A1320G>node C>C1974U>node D). Interestingly the synonymous C1974U mutation was also seen linking node A to the small X node (Figs. 1a and b node X) consisting of 6 WNV and 11 radiating variants all collected in NY state in 2000 to 2003 (Fig. 1a). The synonymous C1974U mutations therefore appeared in two genetic backgrounds possibly reflecting its independent emergence on at least two occasions.

The geographic distribution of variants did not appear random as in some cases clusters of closely related variants were collected during the same season by the same blood banks. Such uneven WNV variant distribution indicated that some of the local WNV outbreaks may have been initiated by only a few or by multiple, but genetically closely related, founder viruses. Nine out of ten of the 2004 WNV from Tempe AZ were members of node D and one radiated from node D. Clusters of WNV variants carrying the same E gene signature mutations seen in no other WNV were also detected in the following areas. Four out of five 2004 WNV from the Los Angeles area contained a unique mutation at position 2257. Three out of eight variants collected in Grand Junction, CO in 2004 contained a four-mutation genetic signature (1065, 1089, 1620, 2070). Three out of five 2003 variants from San Antonio, TX contained a distinct mutation at position 2278 while the other two San Antonio variants both contained a three-mutations signature (1221, 1239, 1995). Three out of eight variants from 2003 Oklahoma City had distinct mutations at positions 1134, 1983, and 1995. Both variants from 2004 Shreveport, LA had the same 2304 mutation. A large fraction of WNV in each of the above blood collection centers therefore contained viruses that were genetically related and distinguishable from those in other areas during the same year.

Amplification and sequencing of the nearly complete WNV open reading frame

In order to confirm and extend the results of the E gene analysis and to further examine the possibility of founder effects in local outbreaks, we selected 20 samples from different geographical areas for full genome sequencing (Grand Junction, CO 2004 ($n=4$), Tampa, FL 2003 ($n=3$), Amarillo, TX 2003 ($n=2$), Los Angeles, CA 2004 (n=2), Tempe, AZ 2004 ($n=2$), Albuquerque, NM ($n=2$), San Francisco, CA 2004 ($n=1$), San Diego, CA 2004 ($n=1$), Rapid City, SD 2004 ($n=1$), Glenview, IL 2003 ($n=1$), Milwaukee, WI 2003 ($n=1$)). RT-PCR and sequencing were performed directly from human plasma (Table 2 and see Materials and methods).

Table 2
Primers used for the amplification and sequencing of the WNV genome

			NY99 pos.
nPCR primers			
F1	+	CTGACAAACTTAGTAGTGTTTGTGAGG	20–46
nF1[a]	+	TGCGAGCTGTTTCTTAGCAC	67–86
F2	+	CACCTTTGTGGTTGATGGTC	2862–2881
nF2[a]	+	TCGCGCTTGGAATAGCTTAG	2910–2929
nR1[a]	−	CCTTTCAAGCTTCCACGTATC	3091–3111
R1	−	GCCCTGGTTTTGTGTCTTG	3249–3267
F3	+	ATTGCCCTTTGCCTACAACG	5725–5744
nF3[a]	+	GCCTACAACGTGCTGGAAAG	5735–5754
nR2[a]	−	GGCATGTTGATGTTGTCCAG	6097–6116
R2	−	TCTCTCTTCTCCTCTGAGCCG	6184–6204
F4a	+	TGGTGTTACTATATGGAACC	7939–7959
nF4a[a]	+	GGAACATTGTCACCATGAAGAG	8042–8063
nR3[a]	−	TGCCACGTCGAACTGTACTC	8545–8564
R3	−	TGAGAGGAGCCTGACCACTC	8654–8673
nR4a[a]	−	TGAGGGCTTACATGGATCAC	10,597–10,616
R4a	−	TATCGCAGACTGCACTCTCC	10,696–10,715
Sequencing			
F2b	+	TGTCGTGCTATTGCTTTTGG	930–949
S1	+	CTCCACACAGGTTGGAGCCACTCAGGCAGG	1431–1460
R2b	−	TCTATGGCACACCCAGTGTC	2470–2489
SEG1	+	ACTTACACTGATGTGTTACG	3673–3692
SEG2	+	AGTGATGCAGAAATTACAGG	4408–4427
SEG3	+	GATGGATGAGCCAATCCCAG	5121–5140
SEG4	+	GAAAGATGCCTGAGCACTTC	6497–6516
SEG5	+	AAGTCACCCTCACCGTTACG	7229–7248
SEG6	+	GCGAGTGTTCAAAGAGAAGG	8742–8761
SEG7	+	GTGGACAAGTTGTCACCTAC	9491–9510
SEG8	+	GAGGACATCTGGTGTGGCAG	10,222–10,241

[a] PCR primers also used for sequencing. +, sense primers; − antisense primers.

When compared with the ancestral US strain (NY 1999 Genbank AF196835), we detected mutations in 275 positions out of the 10508 nucleotide sequenced. An average of 0.29% of the nucleotides differed between the 2003–2004 sequences and the NY99 reference strain (range of 0.2–0.35%). Relative to the Eastern US clade consensus sequence nine synonymous mutations were identified in most of the newly sequenced 2003–2004 genomes (C660U (C), C2466U (E), A4146G (NS2A), C4803U (NS3), C6138U (NS3), C6238U, C6426U (NS3), C6996U (NS4B), U7938C (NS5) and C9352U (NS5)). Except for C660U, present in only 15/20 of sequenced genomes and C6238U present in only 11/20 genomes, these nine synonymous mutations were found in 19/20 of the 2003–2004 sequenced genomes and were therefore associated with the replacement of the early (1999–2003) Eastern US clade by the North American 2003–2004 clade and its derived sub-clusters including nodes C and D. One WNV (Fig. 2 03-113 Tampa FL), phylogenetically basal to the North American 2003–2004 clade, carried only 4/9 of these mutations (A4146G, C6138U, C6426U, and U7938C) presumably representing an intermediate form. A phylogenetic analysis of the complete genomes was performed including representative genomes sampled by Davis et al. (2005) (Fig. 2). The 03-113 Tampa FL ORF, whose E gene fell within node B, clustered with 2 other database sequences (Fig. 2 OH 2002 and Mexico, Tabasco) and was phylogenetically located betwen the two major genome clades, the Eastern US and the North American 2002-2004 clades (Davis et al., 2005).

The overall topology of that tree reflected the lines of descent seen in the E gene network analysis. The 2004 NM variants (04-236NM and 04-327NM) branched off the same lineage that led to the 2004 AZ and CO variants (04-251AZ, 04-252AZ, 04-216CO, 04-219CO) just as these NM variants radiated from node C and the AZ and CO variants radiated from node D (Figs. 1b and 2). The four node D variants were closely related at the genome level with another Arizona 2004 genome sequence derived by Davis et al. (2005) (AZ 2004-DQ164201). The other two 2004 CO variants (04-214CO, 04-218CO) branched off together from the major group of sequences corresponding to the North American 2002–2004 cluster much as they radiated

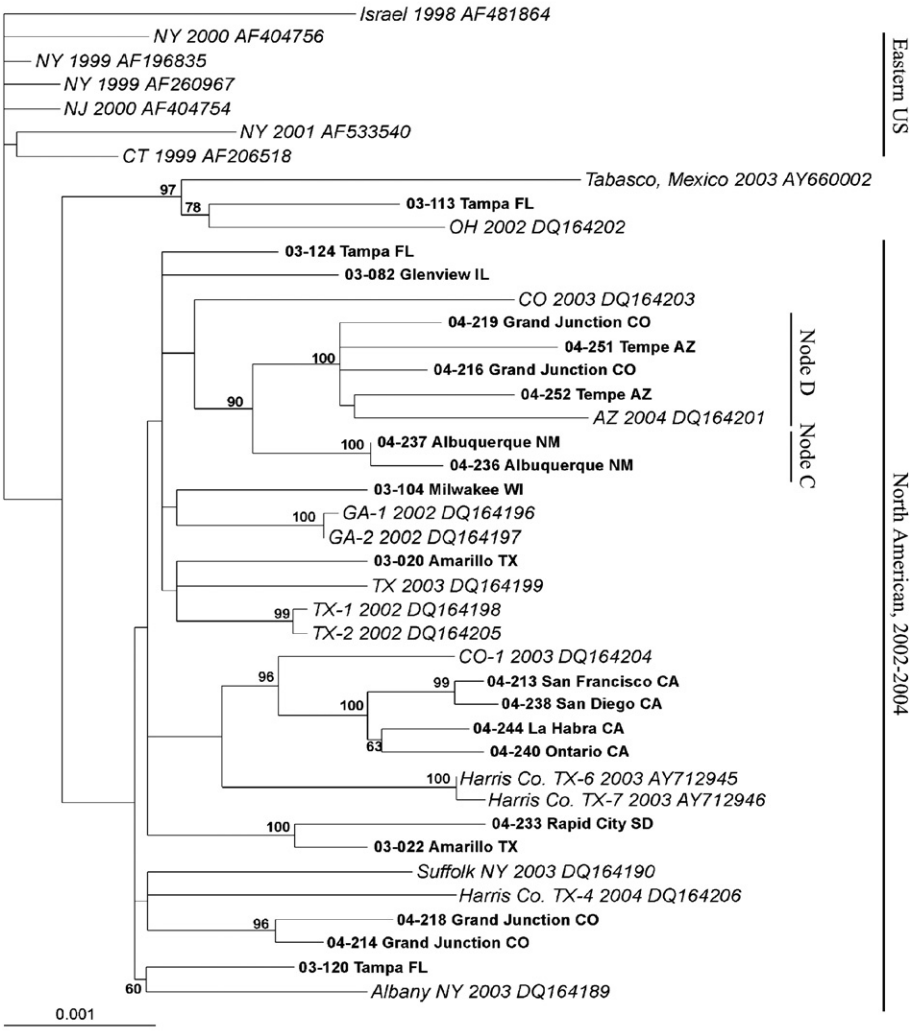

Fig. 2. Phylogenetic analysis of WNV genomes. Blood donors WNV are shown in thick font while the reference US WNV variant (NY 1999), 1998 variant from Israel (Israel 1998), and selected variants from Davis et al. (2005) are in italic font. Variants belonging to genetic network defined nodes or previously defined clusters are located left of the vertical bars.

from node B in the network tree (Figs. 1b and 2). Whole genome phylogenetic analysis also uncovered relationships not detected using only E gene sequence. Of the four 2004 CA variants genome sequenced, only the two from the Los Angeles area clustered using E gene alone in the network analysis (04-240CA, 04-244CA) while the other two variants (04-213CA, 04-238CA) from the San Francisco and San Diego area fell into node B (Fig. 1b). Using full genome sequences these four viruses clustered together with bootstrap values of 100/100 together with another avian variant reported by Davis et al. (2005) (CO 2003 1-DQ164204) reflecting an apparent common ancestry for all four sampled Californian WNV that was not detected using E gene alone (Fig. 2). A 2003 variant from TX (03-022TX) and another from 2004 in SD (04-233SD) both in node B by E gene network analysis (Fig. 1b) showed a close relationship by whole genome analysis (100/100 bootstrap) (Fig. 2). The other 5 WNV genomes sequences were part of the major North American 2002–2004 cluster (Davis et al., 2005) but did not specifically cluster with other database variants (Fig. 2). In the network analysis, these 5 variants belonged to or radiated from node B (Fig. 1b). Full genome sequencing therefore confirmed, with increased confidence, clustering detected using E gene alone in the network analysis and in some cases revealed previously undetected relationships.

Amino acid signatures were also seen in the genome-based clusters. The Mountain state cluster consisting of two Albuquerque, NM variants plus two Tempe, AZ and two Grand Junctions, CO variants (Fig. 2) (radiating from node C and D in E gene respectively) (Fig. 1b), all carried the non-synonymous mutations NS4A A85T and NS5 K314R. The other two Grand junctions CO variants clustering separately (radiating from node B in env) carried the unique NS5, R422K mutation. The four Pacific cluster genomes from California (Fig. 2) were characterized by E V431I, NS4A F92L, and NS4A V135M non-synonymous mutations.

Discussion

We report here on a genetic analysis of WNV from various North American locations during the 2003 and 2004 epidemic seasons and compare them to pre-existing WNV variants mostly derived from mosquitoes and bird surveillance. This study included 179 infected human blood donors and used direct amplification and sequencing of WNV from plasma without prior tissue culture viral amplifications. Both E gene and nearly complete genome sequences were generated to analyze the evolution of this epidemic, focusing on WNV variants that reached their human hosts as a result of insect bites.

Our sampling revealed the presence of 3 main clusters of WNV circulating in the continental US and Western Canadian provinces in 2003–2004 (nodes B–D in Fig. 1) with particularly notable expansion of node D in 2004. A recent study suggested the presence of 4 major WNV clades; the "Southeast Costal Texas" (node Y in Fig. 1) (Davis et al., 2003), the "Eastern US isolates" clade (node A in Fig. 1), the "Intermediate sister" clades, and the "North America 2002–2004" clade (node B in Fig. 1) (Davis et al., 2005). We did not identify any sequences belonging to the Southeast Costal Texas node Y clade in blood donors in 2003–2004. As others have suggested, this variant may now be extinct (Davis et al., 2005).

The Eastern US clade (node A) was largely comprised of pre-2002 sequences and has significantly decreased in frequency since 2002 (Beasley et al., 2003; Davis et al., 2005; Ebel et al., 2004). This study identified only three 2003 sequences from Florida as belonging to this cluster, suggesting that this variant may also have become extinct as no Eastern US clade (node A) virus has been reported in 2004–2005 (Bertolotti et al., in press; Davis et al., 2005; Snapinn et al., in press). Only one member of the Intermediate sister clades cluster was identified in 2003 in Central Florida (03-120FL clustering weakly with NY 2003 Albany DQ164189) (Fig. 2) (Davis et al., 2005). The rest of our sequences belonged to the North America 2002–2004 group (node B). The identification of frequent WNV variants belonging to or directly radiating from nodes C and D, the growing frequency of node C and D diagnostic E gene mutations in 2003 and 2004, and the high bootstrap values for these two clusters using full genome data led us to define them as significant genetic groups rather than simply minor variants radiating from node B (i.e. expanding sub-clusters of the North America 2002–2004 group) (Davis et al., 2005). Further WNV sampling and analysis in mountain and pacific states in 2005 and 2006 will be required to determine if node C and D or further descendents remain dominant.

In 2003 the node C and radiating variants appeared largely restricted to the Northern half of the central and mountain states (Alberta, North Dakota, South Dakota, Wyoming, Colorado and Iowa). In 2004 this lineage also appeared further South in New Mexico, Arkansas, and again Colorado. The transportation of WNV by migratory and dispersing resident birds is well documented and is the most likely source for the spread of WNV throughout North America. It is interesting to speculate on a role for migrating birds on the continental central flyway (http://www.birdnature.com/flyways.html) in bringing node C variant south in late 2003. The node D and radiating variants were initially detected in 2003 thinly distributed among node B variants throughout the central states from Canada to Texas but became the dominant variants in 2004 among the sampled viruses in Colorado and Arizona. Spread of this lineage westward to New Mexico, Arizona, Colorado, and Nevada may have been facilitated by bird hosts of WNV that have a more limited range. The California WNV genetic cluster detectable in the Los Angeles area in 2004 (and also including San Francisco and San Diego variants when analyzing full genome sequence data) descends from the main North America 2002–2004 cluster (node B) rather than the node D variants seen the same year in Arizona or Nevada. The 2004 California WNV therefore appeared to have leaped from the central states while the 2004 mountain states (AZ, CO, NV) viruses were mostly derived from 2003 minority variant(s) from the central states.

Also of interest was the observation that in a subset of cases a high frequency of WNV variants collected the same year by the same blood bank contained genetic signatures distinct from all other sampled WNV. The majority of WNV in the Los Angeles, CA area, San Antonio, TX, Shreveport, LA, and significant

minority in Grand Junction, CO and Oklahoma City, OK, carried viral polymorphisms seen in no other regions. Nine out of ten variants collected in Tempe, AZ were identical in their E gene all belonging to node D. Geographic concentrations of closely related variants may be explained by viral founder effects whereby only one or a few WNV initially colonize a local area. Descendents of one or a few introduced viruses may rapidly infect the pool of local birds through insect vectors thereby amplifying any genetic polymorphisms initially carried by the founding virus. The majority of human infections in the same locality (i.e. collected by the same blood bank) would therefore be derived from one or a very few local WNV colonization events. The complexity of an arbovirus expansion into geographically diverse and complex environments, followed in some areas by yearly re-emergences, is doubtlessly subject to many influences. We propose that the high level of viral genetic homogeneity seen during some outbreaks reflects the stochastic introduction of one or a very few infected vectors into a susceptible area followed by rapid local amplification. The highly focalized distribution of WNV outbreaks, as seen by a zip code study of WNV prevalence in blood donors in central plains states, is consistent with such stochastic WNV introductions precipitating local outbreaks (Kleinman et al., 2005). A recent study using available Genbank sequence from North America plus 68 animal WNV E sequences from different counties in Illinois showed low but significant structure in their geographic distribution together with temporally increasing genetic diversity and divergence from their last common ancestor (Bertolotti et al., in press). Our results confirm the growing genetic distance of North American WNV from their presumably ancestral NY99 sequence and, at least in some locations, a detectable degree of geographic clustering of related variants (Beasley et al., 2003; Davis et al., 2003, 2004, 2005).

WNV in human hosts was also sampled in two blood centers in consecutive years. In Albuquerque, NM the three 2003 strains sampled belonged to node B (and radiating variants) while all eight tested in 2004 belonged to node C or D (and radiating variants) suggesting a re-introduction of WNV rather than over wintering and re-emergence of the local 2003 variants. Twelve WNV variants were sampled from Rapid City, SD in 2003 and two in 2004. Both time points showing the presence of node B and D variants so that viral diversity remained high in this region of repeated WNV epidemics due to either multiple re-introductions and/or over-wintering of 2003 variants.

WNV variants from 2003 and 2004 therefore confirmed the ongoing expansion of the North American 2002–2004 clade into viremic blood donors (Davis et al., 2005) (node B) and also identified expanding variants (nodes C and D), all sharing the V159A mutation in the E gene relative to their progenitor in node A. WNV containing the E-V159A mutations have been reported as early as 2001 in Alabama and in 2002 in Texas, Georgia, Illinois, Colorado, and Western Canada (Davis et al., 2005). In NY state this mutation was found in 55% of 2002 and 85% of 2003 strains (Beasley et al., 2003; Davis et al., 2005; Ebel et al., 2004). Viruses carrying the V159A mutation display higher transmission efficiency in Culex spp (Ebel et al., 2004), possibly accounting for these successful expansion of these variants whose initial explosive growth is also reflected in the rapid generation of novel E-V159A derived variants (Snapinn et al., in press). It is interesting to speculate whether the amino acid mutations associated with node D expansion, the 2004 Californian cluster or the Grand Junction, CO cluster represent neutral polymorphisms or selected mutations conferring selective advantages to these successful WNV variants. Replicative fitness advantages might be of a general nature or restricted to the regional conditions encountered where these variants became dominant (e.g. replication in local species of birds and insect vectors). Surprisingly the same synonymous E gene mutation (C1974U) found in node D was also detected in a different genetic background leading to node X found in 2000 in New York state indicating that this mutation arose and was amplified twice. Whether other mutations characterizing expanding node C and D variants reflect ongoing selection and adaptation or simple genetic drift will require detailed phenotypic comparisons of these expanding WNV variants (Ebel et al., 2004).

Materials and methods

Human samples

Infected human plasma samples were obtained as part of normal blood donor screening for WNV RNA at blood centers across the USA and Canada (Busch et al., 2005; Kleinman et al., 2005). Samples were obtained from 46 blood centers in 25 states and from three Canadian provinces during the epidemic seasons of 2003 and 2004 (Table 1). A total of 179 samples were sequenced. All studies were approved by the UCSF CHR.

E gene RT-nPCR and sequencing

Total viral RNA was extracted from plasma aliquots using the QIAamp viral RNA minikit (Qiagen, Valencia, CA) as per the manufacturer's instructions and collected in 50 μl of elution buffer containing 40 U of Protector RNAse inhibitor (Roche Diagnostic Corporation, Indianapolis, IN). First strand cDNA synthesis was initiated using 10 μl of RNA, 200 ng random oligonucleotide hexamers, and 200 U of Superscript II RNAse H^- reverse transcriptase (Invitrogen, Carlsbad, CA) in a final volume of 20 μl. A 1564 bp fragment of the E gene was then amplified by nested PCR. First round PCR reactions (50 μl) contained 5 μl cDNA and 10 pmol each of sense WNV-E-F1 CTGGTGGCAGCCGTCATTGGTTGGATGCTTGGG and antisense WNV-E-R1 ACACTCCACTTCCACATCTCAGCTCTTGCCG primers (first round with 1.25 U of Taq DNA polymerase (Promega, Madison, WI), 1.5 mmol/L $MgCl_2$, and 200 mmol/L dNTPs). Second round reactions were initiated with 2 μl from first round reaction and included 10 pmol each of sense WNV-E-F2 CCATGCAGAGAGTTGTGTTTGTCGTGC and antisense WNV-E-R2 GCCGGCTGATGTCTATGGCACACCC primers. PCR cycling conditions for the amplification of WNV E gene were denaturation at 94 °C for 30 s, annealing at 54 °C for first PCR round and 56 °C for second round for 30 s, extension at 72 °C for 70 s for 35 cycles. Amplified products were directly sequenced following purification using the QIAquick PCR

Purification Kit (Qiagen). Sequencing was preformed using both second round forward and reverse primers and an additional internal primer (S1 CTCCACACAGGTTGGAGCCACTCAGGCAGG) and the BigDye Version 3.0 Dye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) on an ABI 3700 automated capillary sequencer.

Full genome RT-nPCR and sequencing

WNV full genome was amplified and sequenced in four fragments. Viral RNA was extracted as described above using 280 µl of plasma. First strand cDNA synthesis was initiated using 14 µl of RNA and 0.5 µg of either primer R1, R2, R3, or R4a (Table 2) and 400 U of murine leukemia virus reverse transcriptase (Promega) in a final volume of 25 µl. For amplification of each portion of the genome, a nested PCR was performed using 5 µl of each cDNA and TaKaRa Ex Taq DNA polymerase (TaKaRa Shuzo Co., LTD, Shiga, Japan). The first round primers used for each of the four genomic fragments were F1 and R1, F2 and R2, F3 and R3, and F4a and R4a. The second round primers were respectively nF1 and nR1, nF2 and nR2, nF3 and nR3, and nF4a and nR4a (Table 2). The PCR cycles were run as follows: 3 min at 94 °C, followed by 10 cycles of 15 s at 94 °C, 30 s at 54 °C (fragment 4) or 55 °C (fragment 1) or 56 °C (fragments 2 and 3), 3 min at 68 °C, and 20 cycles of 15 s at 94 °C, 30 s at 54 °C, 55 °C or 56 °C, 3 min plus 10 extra seconds added at each new cycle at 68 °C. A single band of 2.8–3.2 kb, depending on the amplicon, was detected on 0.8% agarose gel. PCR products were purified and sequenced as described above. Sequencing primers are listed in Table 2. Sequences were edited manually using EditView and were assembled into a single contig and then aligned using the Seqman and Megalign programs, respectively (DNAstar, Inc., Madison, WI).

Network analysis

A median joining network was constructed using NETWORK version 4.200 free software obtained from http://www.fluxus-engineering.com (Bandelt et al., 1999). Program default settings were used.

Phylogenetic analysis

Nearly complete E gene sequences from genomic position 1055–2413 (E gene ORF is 967–2469) and nearly full WNV ORF from genomic position 108–10616 (WNV ORF is 97–10395) were aligned using ClustalX (Thompson et al., 1997).

Phylogenetic analysis was performed using PAUP* 4.0 (Swofford, 2002). Maximum Likelihood (ML) phylogenetic trees were constructed using a heuristic search employing the general time reversible model of substitution as determined by hierarchical likelihood ratio test scores criteria in Modeltest 3.06 (Posada and Crandall, 1998). A starting tree was obtained by stepwise addition and branch swapping was performed using sub-pruning regrafting. Base frequency, gamma distribution, and transition transversion ratios were determined by Modeltest 3.06. Bootstrap analysis (100 replicates) was performed on each ML tree. Pair-wise genetic distances were calculated using nucleotide sequences and ML settings identical to those used for phylogenetic analysis.

WNV sequences generated from this study have been submitted to GenBank under accession numbers DQ431693-DQ431712 for full genomes and DQ911843-DQ912021 for E region.

Acknowledgments

We acknowledge support from the Blood Systems Foundation, participating blood centers, as well as the CDC/NCID (R01-CI-000214).

References

Bandelt, H.J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16 (1), 37–48.

Beasley, D.W., Davis, C.T., Guzman, H., Vanlandingham, D.L., Travassos da Rosa, A.P., Parsons, R.E., Higgs, S., Tesh, R.B., Barrett, A.D., 2003. Limited evolution of West Nile virus has occurred during its southwesterly spread in the United States. Virology 309 (2), 190–195.

Bertolotti, L., Kitron, U., Goldberg, T.L., in press. Diversity and evolution of West Nile virus in Illinois and The United States, 2002–2005. Virology (Available online since Nov 20, 2006)

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 (7), 853–856.

Busch, M.P., Caglioti, S., Robertson, E.F., McAuley, J.D., Tobler, L.H., Kamel, H., Linnen, J.M., Shyamala, V., Tomasulo, P., Kleinman, S.H., 2005. Screening the blood supply for West Nile virus RNA by nucleic acid amplification testing. N. Engl. J. Med. 353 (5), 460–467.

Busch, M.P., Wright, D.J., Custer, B., Tobler, L.H., Stramer, S.L., Kleinman, S.H., Prince, H.E., Bianco, C., Foster, G., Petersen, L.R., Nemo, G., Glynn, S.A., 2006. West Nile virus infections projected from blood donor screening data, United States, 2003. Emerg. Infect. Dis. 12 (3), 395–402.

CDC, 2002. West Nile Virus Activity—United States, October 10–16, 2002, and update on West Nile Virus infections in recipients of blood transfusions. Morb. Mortal Wkly. Rep. 51 (41), 929–931.

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

Davis, C.T., Beasley, D.W., Guzman, H., Siirin, M., Parsons, R.E., Tesh, R.B., Barrett, A.D., 2004. Emergence of attenuated West Nile virus variants in Texas, 2003. Virology 330 (1), 342–350.

Davis, C.T., Ebel, G.D., Lanciotti, R.S., Brault, A.C., Guzman, H., Siirin, M., Lambert, A., Parsons, R.E., Beasley, D.W., Novak, R.J., Elizondo-Quiroga, D., Green, E.N., Young, D.S., Stark, L.M., Drebot, M.A., Artsob, H., Tesh, R.B., Kramer, L.D., Barrett, A.D., 2005. Phylogenetic analysis of North American West Nile virus isolates, 2001–2004: evidence for the emergence of a dominant genotype. Virology 342 (2), 252–265.

Ebel, G.D., Carricaburu, J., Young, D., Bernard, K.A., Kramer, L.D., 2004. Genetic and phenotypic variation of West Nile virus in New York, 2000–2003. Am. J. Trop. Med. Hyg. 71 (4), 493–500.

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 (12), 1604–1607.

Harrington, T., Kuehnert, M.J., Kamel, H., Lanciotti, R.S., Hand, S., Currier, M., Chamberland, M.E., Petersen, L.R., Marfin, A.A., 2003. West Nile virus infection transmitted by blood transfusion. Transfusion 43 (8), 1018–1022.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., VandePol, S., 1982. Rapid evolution of RNA genomes. Science 215 (4540), 1577–1585.

Hollinger, F.B., Kleinman, S., 2003. Transfusion transmission of West Nile virus: a merging of historical and contemporary perspectives. Transfusion 43 (8), 992–997.

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 (Pt. 8), 2175–2183.

Jia, X.Y., Briese, T., Jordan, I., Rambaut, A., Chi, H.C., Mackenzie, J.S., Hall, R.A., Scherret, J., Lipkin, W.I., 1999. Genetic analysis of West Nile New York 1999 encephalitis virus. Lancet 354 (9194), 1971–1972.

Kleinman, S., Glynn, S.A., Busch, M., Todd, D., Powell, L., Pietrelli, L., Nemo, G., Schreiber, G., Bianco, C., Katz, L., 2005. The 2003 West Nile virus United States epidemic: the America's Blood Centers experience. Transfusion 45 (4), 469–479.

Komar, O., Robbins, M.B., Contreras, G.G., Benz, B.W., Klenk, K., Blitvich, B.J., Marlenee, N.L., Burkhalter, K.L., Beckett, S., Gonzalvez, G., Pena, C.J., Peterson, A.T., Komar, N., 2005. West Nile virus survey of birds and mosquitoes in the Dominican Republic. Vector Borne Zoonotic Dis. 5 (2), 120–126.

Lanciotti, R.S., Roehrig, J.T., Deubel, V., Smith, J., Parker, M., Steele, K., Crise, B., Volpe, K.E., Crabtree, M.B., Scherret, J.H., Hall, R.A., MacKenzie, J.S., Cropp, C.B., Panigrahy, B., Ostlund, E., Schmitt, B., Malkinson, M., Banet, C., Weissman, J., Komar, N., Savage, H.M., Stone, W., McNamara, T., Gubler, D.J., 1999. Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. Science 286 (5448), 2333–2337.

Lanciotti, R.S., Ebel, G.D., Deubel, V., Kerst, A.J., Murri, S., Meyer, R., Bowen, M., McKinney, N., Morrill, W.E., Crabtree, M.B., Kramer, L.D., Roehrig, J.T., 2002. Complete genome sequences and phylogenetic analysis of West Nile virus strains isolated from the United States, Europe, and the Middle East. Virology 298 (1), 96–105.

Lefrancois, T., Blitvich, B.J., Pradel, J., Molia, S., Vachiery, N., Pallavicini, G., Marlenee, N.L., Zientara, S., Petitclerc, M., Martinez, D., 2005. West Nile virus surveillance, Guadeloupe, 2003–2004. Emerging Infect. Dis. 11 (7), 1100–1103.

Montgomery, S.P., Brown, J.A., Kuehnert, M., Smith, T.L., Crall, N., Lanciotti, R.S., Macedo de Oliveira, A., Boo, T., Marfin, A.A., 2006. Transfusion-associated transmission of West Nile virus, United States 2003 through 2005. Transfusion 46 (12), 2038–2046.

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

Nash, D., Mostashari, F., Fine, A., Miller, J., O'Leary, D., Murray, K., Huang, A., Rosenberg, A., Greenberg, A., Sherman, M., Wong, S., Layton, M., 2001. The outbreak of West Nile virus infection in the New York City area in 1999. N. Engl. J. Med. 344 (24), 1807–1814.

Pealer, L.N., Marfin, A.A., Petersen, L.R., Lanciotti, R.S., Page, P.L., Stramer, S.L., Stobierski, M.G., Signs, K., Newman, B., Kapoor, H., Goodman, J.L., Chamberland, M.E., 2003. Transmission of West Nile virus through blood transfusion in the United States in 2002. N. Engl. J. Med. 349 (13), 1236–1245.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14 (9), 817–818.

Swofford, D.L., 2002. PAUP: Phylogenetic Analysis Using Parsimony (and Other Methods) 4.0 Beta. Florida State University.

Snapinn, K.W., Holmes, E.C., Young, D.S., Bernard, K.A., Kramer, L.D., Ebel, G.D., in press. Declining Growth Rate of West Nile Virus in North America. J. Virol. (2006 Dec 20; Epub ahead of print).

Stramer, S.L., Fang, C.T., Foster, G.A., Wagner, A.G., Brodsky, J.P., Dodd, R.Y., 2005. West Nile virus among blood donors in the United States, 2003 and 2004. N. Engl. J. Med. 353 (5), 451–459.

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 (24), 4876–4882.