Journal of NeuroVirology, 11: 544–556, 2005

ISSN: 1355-0284 print / 1538-2443 online
DOI: 10.1080/13550280500385229

Taylor & Francis
Taylor & Francis Group

Expression and evolutionary analysis of West Nile virus (Merion Strain)

Mathura P Ramanathan,[1] Jerome A Chambers,[1] Jesse Taylor,[2] Bette T Korber,[3] Mark D Lee,[1] Aysegul Nalca,[4] Kesan Dang,[1] Panyupa Pankhong,[5] Watcharee Attatippaholkun,[5] and David B Weiner[1]

[1]Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; [2]Institute of cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9JT, UK; [3]Los Alamos National Laboratory, Los Alamos, New Mexico, USA; [4]Homeland Security and Infectious Disease Research Division, Southern Research Institute, Frederick, Maryland, USA; [5]Department of Clinical Chemistry, Faculty of Medical Technology, Mahidol University Bangkok, Thailand

The authors report a new strain of West Nile virus (WNV) with the expression analysis of its individual open reading frames. Since its sudden appearance in the summer of 1999 in New York City, the virus has spread rapidly across the continental United States into Canada and Mexico. Besides, its rapid transmission by various vectors, the spread of this virus through organ transplantation, blood transfusion, and mother-child transmission through breast milk is of concern. In order to understand molecular variations of WNV in North America and to generate new tools for understanding WNV biology, a complete clone of WNV has been constructed. Investigations so far have focused only on half of its genes products and a detailed molecular and cell biological aspects on all of WNV gene have yet to be clearly established. The open reading frames of WNV were recovered through an reverse transcriptase–polymerase chain reaction (RT-PCR)-PCR using brain tissue from a dead crow collected in Merion, PA, and cloned into a mammalian expression vector. The deduced amino acid sequences of individual open reading frames were analyzed to determine various structural motifs and functional domains. Expression analysis shows that in neuronal cells, C, NS1, and NS5 proteins are nuclear localized whereas the rest of the antigens are confined to the cytoplasm when they are expressed in the absence of other viral antigens. This is the first report that provides an expression analysis as well as intracellular distribution pattern for all of WNV gene products, cloned from an infected bird. Evolutionary analysis of Merion strain sequences indicates that this strain is distinct phylogenetically from the previously reported WNV strains. *Journal of NeuroVirology* (2005) **11,** 544–556.

Keywords: phylogenetic analysis; Merion strain; West Nile virus; WNV antigens; WNV pathogenesis

Introduction

West Nile virus (WNV) is a mosquito-transmitted flavivirus that produces potentially fatal disease in humans and some mammals and birds. Since its appearance in 1999 in New York, WNV has spread across the continental United States into Canada and Mexico. The genus, Flavivirus covers 70 viruses classified on the basis of their serological and genetic relatedness (Chambers *et al*, 1990). WNV is grouped under Japanese encephalitis virus (JEV) complex, which includes St. Louis encephalitis (SLE), Murray Valley encephalitis (MVE), and Kunjin viruses. WNV is a single-stranded RNA virus with a positive polarity RNA genome of approximately 11 kb in length. The genomic RNA carries a type 1 cap at its 5′ end but does not contain a polyadenylate tract at the 3′ end of the genome (Brinton, 2002). The genome has a single open reading frame encoding for a polyprotein that is

Address correspondence to David B. Weiner, PhD, Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, 422 Curie Boulevard, 505 Stellar-Chance Laboratories, Philadelphia, PA 19104, USA. E-mail: dbweiner@mail.med.upenn.edu

This study was supported in part by the funding from NIH to D.B.W. The technical help from Xinyu Zhao, Center for Bioimaging, Department of Pathology and Laboratory Medicine, University of Pennsylvania Medical School, is sincerely acknowledged.

Received 20 April 2005; revised 8 June 2005; accepted 22 September 2005.

cleaved cotranslationally and post-translationally at specific sites by host and viral proteases to produce the virion and replicase components. The three structural proteins of the virus such as C, preM(M), E are encoded at the 5′ end of the genome while the remaining 3′ portion encodes seven nonstructural proteins, *viz.,* NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 (Brinton, 2002).

WNV is a potential human, equine, and avian neuropathogen (Nash *et al*, 2001) (Campbell *et al*, 2002). Apparent increases in the frequency of neuralogic infections, human case-fatality rates, and horse and bird deaths in the Northern Hemisphere raised the question of whether emergence of WNV strains with increased pathogenesis occurred in North America or whether the virulence of the virus had previously been underestimated (Campbell *et al*, 2002). Several reports established that WNV is transmissible through blood transfusion (Couzin, 2003; Lefrere *et al*, 2003), organ transplantation (DeSalvo *et al*, 2004; Iwamoto *et al*, 2003; Kiberd and Forward, 2004; Kusne and Smilack, 2005), breast feeding (Williams, 2004), and in uterointraplacental transmission of WNV from an infected mother (Alpert *et al*, 2003; Williams, 2004). The development of new strategies for WNV infection will benefit from a more detailed knowledge of the biology of this infectious agent.

Though studies on the molecular aspects on flaviviruses are not new, investigations exclusively on the molecular biology of WNV are scarce. Information on the biology of WNV genes are restricted to the description of half of the WNV antigens including capsid (Yamshchikov and Compans, 1994; Yang *et al*, 2001, 2002), envelope, NS2B-NS3 complex (Yamshchikov and Compans, 1994), and NS5 (Grun and Brinton, 1986, 1987). We have cloned and expressed all of the WNV open reading frames from the carcass of a dead crow collected from suburban Philadelphia, Merion Station, Pennsylvania, in September 2002. In this paper, we report the expression analysis and intracellular distribution pattern of individual open reading frames of this strain. To gain more insights into the potential variation in the genome sequence of WNV since its outbreak, we analyzed the sequence of Merion strain and compared this sequence to other isolates in the database. Phylogenic analysis shows that WNV undergoes significant mutations over time. Interestingly, the Merion strain appears to be distinct from rest of the reported strains collected in the northern hemisphere.

Results

Recovery and cloning of WNV genes from crow carcasses

A fresh dead crow was identified on the street at the intersection of Hazelhurst Street and Highland Avenue in Lower Merion, PA. This dead bird was dissected and the brain tissues were used for cloning

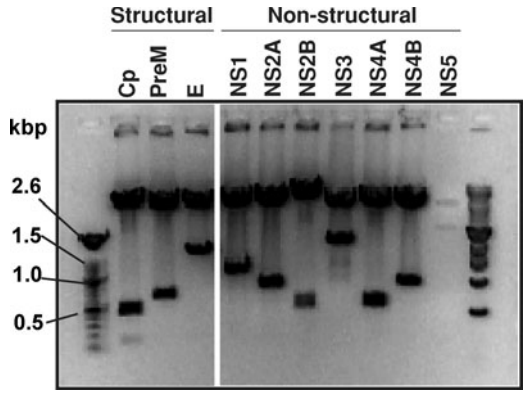

Figure 1 Cloning of WNV Merion strain open reading frames. Total RNA content was extracted from the tissue samples of dead crow and used to prepare total cDNA by random priming. Merion strain WNV open reading frames were amplified and cloned into pcDNA3.1 His A vector. The open reading frames encoding three structural genes, *viz.*, capsid, membrane and envelope, and seven nonstructural genes were amplified through RT-PCR and ligated into a mammalian expression vector. Figure 1 reveals the restriction analysis of WNV expression plasmids encoding three structural and seven nonstructural genes. The length of individual open reading frames corresponds to the existing entries from the database. An aliquot of 100-bp ladder DNA is run in the same gel to reveal the lengths of individual genes specifically.

WNV sequences by reverse transcriptase–polymerase chain reactin (RT-PCR). PCR analysis of brain tissues reveals positive signals for WNV gene sequences. Figure 1 reveals the length of individual WNV open reading frames, as judged from their restriction analysis. The length of amplified open reading frames of structural genes, C, PreM, and E were 369, 498, and 1503 bp, respectively, as determined by sequencing. Among the nonstructural genes, the size of the first open reading frame, NS1, is 1056 bp in length. Following this, the NS2A, NS2B, and NS3 open reading frames are of 693, 393, and 1857 bp, repectively. NS4A and NS4B genes were 447 and 765 bp in length, respectively. NS5 represents the longest open reading frame among all of WNV genes (2718 bp, Figure 1). The resulting sequences of these constructs were verified by comparing them with WNV genome sequences available on the public database (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) at the National Center for Bioinformatics (NCBI) website.

Presence of structural motifs and functional domains in WNV proteins

The deduced amino acid sequence from the cloned WNV genes were analyzed using modular servers such as SMART (Simple Modular Architecture Research Tool) program at (http://www.embl-heidelberg.de/predictprotein/predictprotein.html) EMBL, and PORT II server (http://psort.nibb.ac.jp) to identify structural motifs and functional domains present in the cloned WNV proteins. Figure 2 shows the deduced amino acid sequences of Merion strain open reading frames with an emphasis on structural motifs. Among the structural gene products, the

protein C appeared to include various motifs. The capsid protein, besides being rich in basic residues, includes frequent leucine repeats, a transmembrane domain between the residues 44 and 66, and a low complexity region spanning residues 97 to 108. Interestingly, it also includes a HALZ region (homeobox-associated leucine zipper) and a putative bipartite nuclear localization signal. PreM protein possesses two low-complexity regions spanning residues 85 through 93 as well as 153 through 162. Envelope harbors a low complexity region that spans from residues 316 through 327 of the protein, and two transmembrane domains in the regions of amino acids 340 to 362 and 369 to 388.

WNV- Merion Strain (3433 aa)

Capsid

MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPIRFVLALLAFFRFTAI
APTRAVLDRWRGVN**KQTAMKHLLSFKKELGTLTSAINRRSSKQKKRGGKTGIAVMIGLI**
ASVGA (Shaded- Transmembrane domain; Boxed-Low complexity domain; Bolded-HALZ domain; Underlined- Bipartite NLS)

PreM

VTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVRAMDVGYMCDDTITYECPVLSAGN
DPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLTVQTHGESTLANKKGAWMDSTKATR
YLVKTESWILRNPGYALVAAVIGWMLGSNTMQRVVFVVLLLLVAPAY
(Boxed- Low complexity domain)

Envelope

FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYC
YLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRDWGNGCGLFGKGSIDTCAKFA
CSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVGATQAGRFSI**A**PAAPSYTLKL
GEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWFMDLNLPWSSAGSTVWRNRETLM
EFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSSNTVKLTSGHLKCRVKMEKLQLKGT
TYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDGPCKVPISSVASLNDLTPVGRLVTVNP
FVSVATANAKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTLKGAQRLAA
LGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLLWMGINARDR
SIALTFLAVGGVLLFLSVNVHA
(Shaded- Transmembrane domain; Boxed- Low complexity domain)

NS1

DTGCAIDISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSV
SRLEHQMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGK
SILFAPELANDTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECD
SKIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDG**T**
LESDLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGP
ATRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNA

NS2A

YNADMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVIL
VGAAFAE**P**NSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYHD
ARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRILLLM
VGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKR
(Shaded- Transmembrane domain)

NS2B

GWPATEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADI
SWESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKR (Shaded- Transmembrane domain)

Figure 2 Structural and functional motifs from WNV gene products. The deduced amino acid sequences of WNV open reading frames are given as per the order of their occurrence from the polyprotein. The residues unique to the Merion strain are in bold. The highlighted motifs and domains were determined with high confidence using online servers listed under Materials and Methods. *(Continued)*

NS3

GGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFHTLWHTTKGAALMS
GEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKNVKNVQTKPGVFKTPE
GEIGAVTLD**S**PTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYISAIVQGERMDEPIPAGFE
PEMLRKKQICVMDLHPGAGKTRRILPQIIKEAINRRLRTAVLAPTRVVAAEMAEALRGLP
IRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNYNLFVMDEAHFTDPASIAARGY
ISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQTEIPDRAWNSGYEWITEYTGKTV
WFVPSVKMG**NEIALCLQRAGKKVVQLNRKSYETEYPKCKNDDWDFVITTDISEMGANFKA**
SRVIDSRKSVKPTIITEGEGRVILGEPSAVTAASAAQRRGRIGRNPSQVGDEYCYGGHTN
EDDSNFAHWTEARIMLDNINMPNGLIAQFYQPEREKVYTMDGEYRLRGEERKNFLELLRT
ADLPVWLAYKVAAAGVSYHDRRWCFDGPRTNTILEDNNEVEVITKLGERKILRPRWIDAR
VYSDHQALKAFKDFASGKR

(Shaded- DEXDc domain; Bolded- HELICc domain; Underlined- Bipartite NLS)

NS4A

SQIGLIEVLGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALL
SVMTMGVFFLLM**R**RKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIP
EPEKQRSQTDNQ**P**AVFLICVMTLVSAVAA

(Shaded- Transmembrane domain)

NS4B

NEMGWLDKTKSDISSLFGQRIEVKENFSMGEFLLDLRPATAWSLYAVTTAVLTPLLKHLI
TSDYINTSLTSINVQASALFTLARGFPFVDVGVSALLLAAGCWGQVTLTVTVTAATLLF
CHYAYMVPGWQAEAMRSAQRRTAAGIMKNAVVDGIVATDVPELERTTPIMQKKVGQIMLI
LVSLAAVVVNPSVKTVREAGILITAAAVTLWENGASSVWNATTAIGLCHIMRGGWLS**R**LS
ITWTLIKNMEKPGLKR (Boxed- Low complexity domain)

NS5

GGAKGRTLGEVWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTA
KLRWLVERRFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYG
WNIVTMKSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFC
VKVLCPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVL
LGRMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFKE
KVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFEEQN
QWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSRAIWFM
WLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKIYADDTAG
WDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGRTVMDVISR**G**
DQRGSGQVVTYALNT**I**TNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKVRTWLFENGEER
LSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWYDWQQVPFCSNHFT
ELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYAQMWLLLYFHRRDLRL
MANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVWIEENEWMEDKTPVEKWS
DVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIGDEKYVDYMSSLKRYEDTTL
VEDTVL

(Boxed- Low complexity domain)

Figure 2 (Continued)

The NS1 protein appears to have no significant domains, repeats, motifs, or features that could be predicted with confidence by SMART or by PORT II servers. Both NS2A and NS2B are rich in transmembrane domains and lack any significant functional motifs as judged confidently by these protein portals. Interestingly, NS2A harbored six transmembrane spanning regions, located between residues 37 and 59, 74 and 96, 103 and 125, 140 and 162, 169 and 187, and 202 and 224. NS2B encompassed three transmembrane domains (residues 7 to 24, 28 to 47, and 103 to 125). WNV NS3 is known to harbor multiple functions such as NTPase, helicase and protease activities. The presence of DEXDc and HELICc domains on NS3 support these functions. DEXDc domain spans the region of residues 167 to 355 in NS3

and this domain is characteristic of a superfamily of proteins of protein containing DEAD/H-Box motifs. HELICc domain refers to the helicase activity of NS3 and is located between residues 370 and 466. NS4A protein has three transmembrane domains, spanning the positions at 50 to 72, 77 to 99, and 104 to 121. In NS4B, other than a low complexity region between residues 106 and 118, no other significant functional motifs were noted. In NS5, an eIF1a (eukaryotic translation initiation factor 1A) domain, spanning the positions 625 to 706, was observed. The gene product of NS5 mediates RNA-dependent RNA polymerase that is important during WNV replication.

Expression of Merion strain genes in vitro

The integrity of WNV Merion strain expression plasmids was tested by *in vitro* translation using ^{35}S-labeled methionine. The radio-labeled protein products were subsequently immunoprecipitated using anti-His antibody that target a polyhistidine epitope fused to the carboxyl-terminal end of WNV proteins. Autoradiographic analysis revealed that the molecular mass of these gene products corresponded to the length of their open reading frames (Figure 2A). The molecular weight of the protein C and preM were judged to be 14 kDa and 22.0 kDa, respectively. E protein is the largest among the structural proteins, as revealed as 55 kDa in mass. Of the nonstructural genes, NS2B appeared to encode for the smallest protein of 15.0 kDa in mass, as evidenced by its mobility (Figure 3). The *NS5* gene being the longest open reading frame among WNV genes, encoded for a protein of about 115.0 kDa in mass.

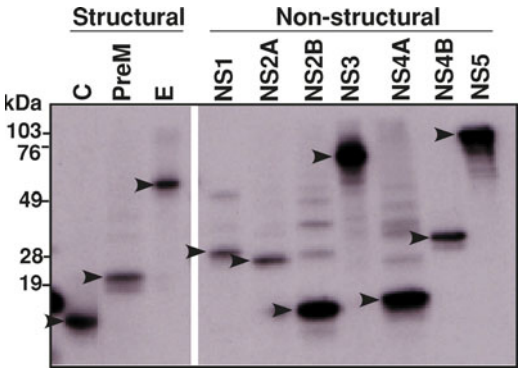

Figure 3 Expression of Merion strain WNV genes *in vitro*. *In vitro* translated radiolabeled WNV gene products were immunoprecipitated using anti-His antibody that target polyhistidine epitope fused to the C-terminal end of the WNV genes, and resolved in a 15% gel. The dried gel was analyzed by autoradiography. The capsid, preM, and E gene products were visualized as proteins of 14 kDa, 20 kDa, and 55 kDa, in mass respectively, as shown in the left panel. The right side panel reveals the mobility of the nonstructural gene products. NS5 is the largest protein with a mass of 115 kDa among WNV gene products. The mobility of individual gene products corresponds to the length of their open reading frames. The WNV gene products with the expected mass are arrow marked in order to distinguish them from nonspecific signals.

Localization of structural and nonstructural proteins of WNV Merion strain in human neuroblastoma cells

The interaction of a virus with a cellular receptor initiates a chain of dynamic events that enable entry of the virus into the cell. Upon viral entry, capsid proteins are exposed to the cellular environment and mediate transport of the viral RNA through the cytoplasmic environment (Brinton, 2002). During WNV replication cycle, virus assembly and budding occur mostly in the cytoplasm of the infected cells. We sought to see whether all of the viral antigens are expressed in the cytoplasm of the infected cells. To address this question, SY5Y cells were transfected with WNV expression constructs encoding individual open reading frames. The protein C exhibited a solid nuclear straining (Figure 4B); in addition to the

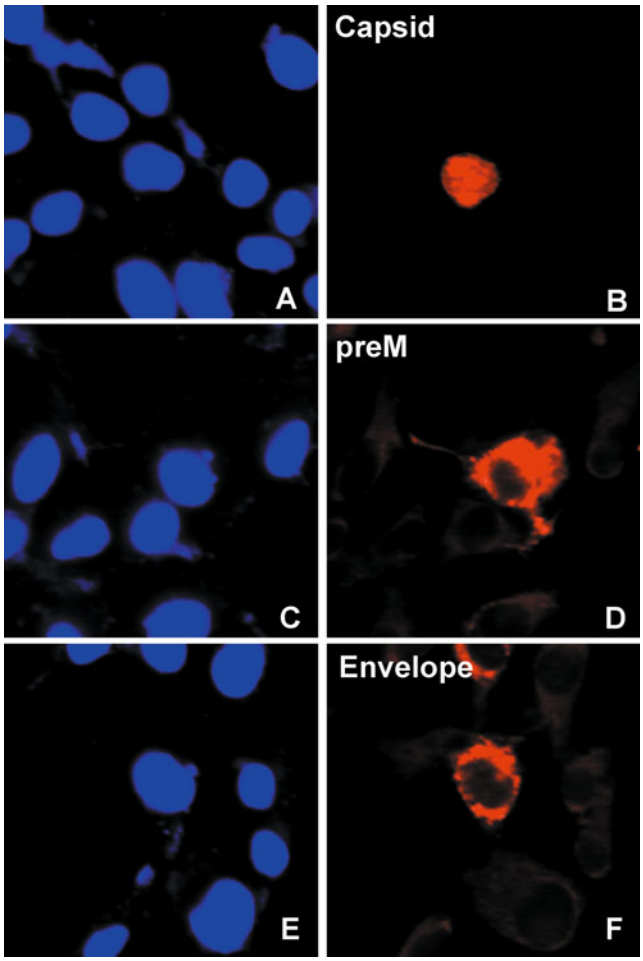

Figure 4 Localization of structural genes in neuronal cells. The human neuroblastoma cells (SY5Y) transfected with individual WNV gene products were analyzed at 36 h post transfection by immunofluorescent assay. The cells were stained with anti-His– and Alexa-546-conjugated anti-mouse secondary antibodies. The nuclear content of the samples were counter stained with DAPI to indicate the available number of cells in the selected field. The full-length capsid protein (**A** to **B**) localized in the nucleus. Both preM (**C**, **D**) and E proteins (**E**, **F**) are distributed in the cytoplasm of the transfected cells.

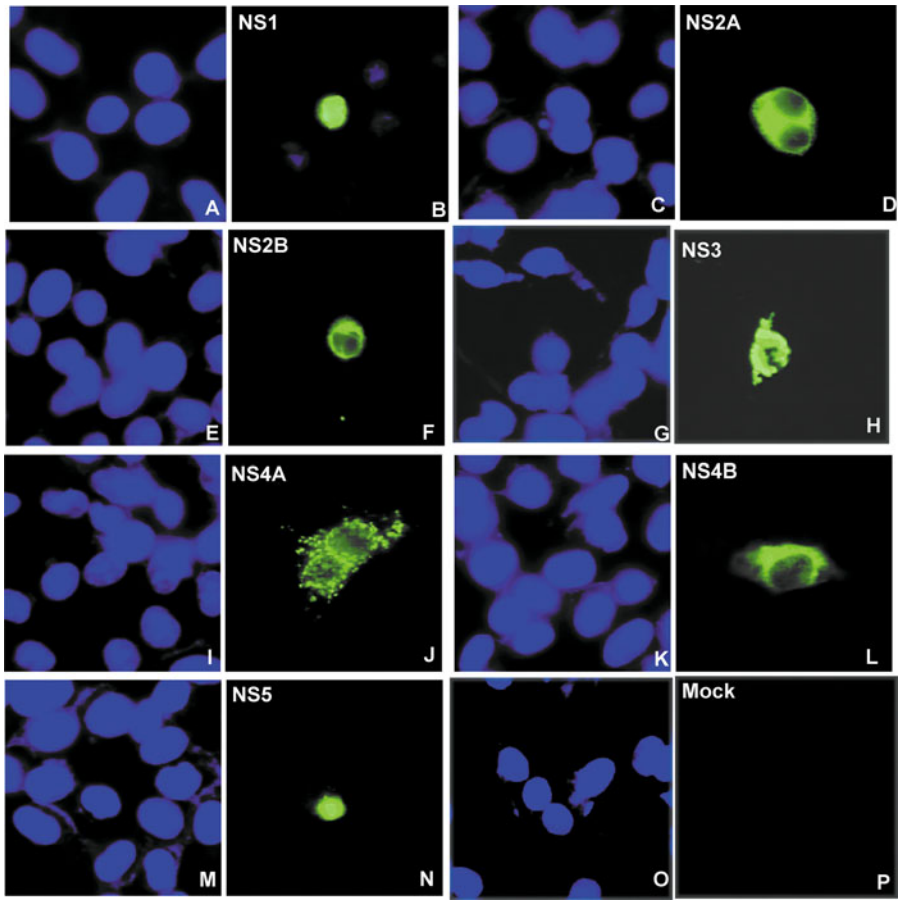

Figure 5 Localization of nonstructural genes in neuronal cells. SY5Y cells were transfected with WNV nonstructural genes from NS1 through NS5. At 36 h post transfection, the cells were fixed and incubated with anti-His- and subsequently with anti-mouse Alexa-488-conjugated secondary antibodies. The staining with DAPI indicates the total nuclear content of the cells in the field. NS1 (**A**, **B**) and NS5 (**M**, **N**) exhibited a strong karyophilicity; whereas NS2A (**C**, **D**), NS2B (**E**, **F**), NS3 (**G**, **H**) NS4A (**I**, **J**), and NS4B (**K**, **L**) localized in the cytoplasm of the neuronal cells. Mock-transfected controls (**O**, **P**) did not yield any specific staining pattern, suggesting the specificity of this assay.

dense nuclear-straining, the expression of the capsid protein resulted in conspicuous punctate staining patterns inside the nucleus in majority of the cells. preM (Figure 4D) and E (Figure 4F) proteins were localized to the cytoplasm of the cell. Figure 5 reveals the intracellular distribution pattern of WNV nonstructural antigens in the transfected cells. NS1 (Figure 5B) and NS5 (Figure 5N) were highly karyophilic as they were localized in the nucleus. The rest of the non-structural proteins such as NS2A (Figure 5D), NS2B (Figure 5F), NS3 (Figure 5H), NS4A (Figure 5J), and NS4B (Figure 5L) were expressed as cytoplasmic proteins. In particular, NS4A expression yielded distinct punctate structures scattered throughout the cytoplasm of the transfected cells (Figure 5J). Mock-transfected cells (Figure 5P) did not yield any specific staining pattern demonstrating the specificity of this assay.

Phylogenetic significance of the Merion strain

ModelTest selected the General Time Reversible model with gamma-distributed rates (GTR + G) as the best-fitting model for the global set and selected the Tamura-Nei model with a constant rate and a class of invariant sites (TN93 + I) as the best-fitting model for the North American set. The greater number of segregating sites in the global set probably accounts for the parametrically richer model selected for this data. For both sets, the transition rate is much greater than the transversion rate, but the HKY85 model is inadequate because the rate of C T transitions is also much greater than the rate of AG transitions (r_{CT} = 29.31 versus r_{AG} = 8.44 for the global set; r_{CT} = 24.11 versus r_{AG} = 9.19 for the North American set). The shape parameter of the gamma distribution of rates was estimated to be a = 0.222 for the global WN virus set, whereas the proportion of invariant sites in the North American WN virus set was estimated to be pI = 0.877. Both results are indicative of substantial rate of heterogeneity between sites, with most sites evolving slowly and a few evolving more rapidly. Because the North American WN virus sequences are all closely related, little time has elapsed between the root and the leaves of the tree and so the slowly

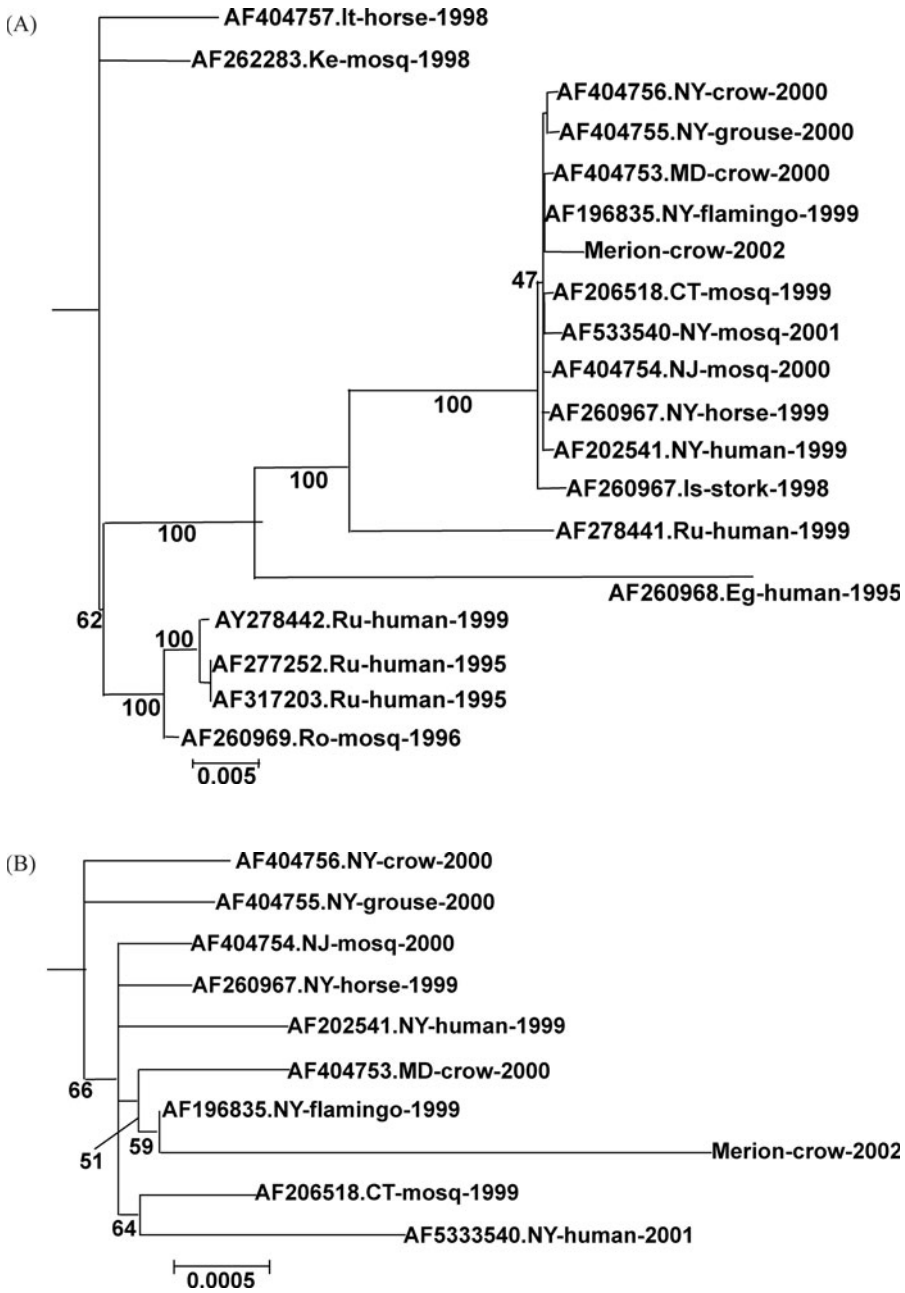

Figure 6 Evolutionary tree indicating the phylogenetic status of the Merion strain in the context of reported strains from global (**A**) as well as North American (**B**) entries.

evolving sites can be accommodated as a class of invariant sites without having to invoke a continuous rate distribution.

Maximum likelihood trees for the global and North American WN virus sets are shown in Figure 6A and B, respectively. Similar to previous studies (Lanciotti *et al*, 2002), there is strong support for a clade consisting of North American WN virus sequences, including the Merion sequence and the Israeli WN virus isolates (bootstrap value = 100%), but only weak support for monophyly of the North American sequences alone (bootstrap value = 47%). The North American sequences have undergone only limited diversification since 1999 (note the order of magnitude difference in the scale of the global and North American phylogenies) and there is little phylogenetic structure evident in this group. Bootstrap values provide weak support for two clades containing the Merion sequence, one allying the Merion sequence to the sequence isolated from a Chilean Flamingo in a New York zoo in 1999 (bootstrap value = 59%) and a second allying those two viruses to a WN isolate from an American Crow found dead in Maryland in 2000 (bootstrap value = 51%). The Merion sequence

Table 1 Mean divergence from the North American WN virus consensus by year of isolation

Year	*Number of sequences*	*Mean number of nucleotide differences*	*Mean number of amino acid differences*
1999	4	5.5	1.25
2000	4	8.75	1.25
2001	1	15	2
2002	1	31	13

is the most divergent member of the North American group. The branch connecting the Merion isolate to the rest of the North American group is the longest ($P = .002875$ substitutions/site) and this isolate also shows the greatest numbers of nucleotide and amino acid differences from the consensus sequence for the North American isolates (Table 1). Analysis of the North American WN virus sequences using LDhat provides no evidence of recombination. The Watterson estimate of Q was 33.94 for the entire coding portion of the genome (10302 sites), whereas the maximum composite-likelihood estimate of r was 0.00. Inspection of the composite likelihood surface (not shown) indicates that $r = 0.00$ is the sole local minimum (at least to a resolution of 0.01) for values ranging between 0.00 and 50.00. Permutation tests of the significance of correlations between three measures of linkage disequilibrium and physical distance between segregating sites are shown in Table 2. None of these correlations are significant. Table 3 shows the parameter estimates for the prior distributions on w, along with the mean of the prior distribution, for each of the separately analyzed regions of the WNV genome. In all regions, the prior mean of w is less than 0.10, with at least 80% of the mass assigned to the w = 0 category and not more than 0.1% of the mass assigned to the w = 1 and w = 3 categories. These results imply that the regions are all broadly subject to purifying selection and that most nonsynonymous mutations are deleterious and hence removed from the population. The posterior distributions on w at individual codons (results not shown) corroborate this inference. At every codon, both the posterior mean and mode of w are less than 0.60. Indeed, only two of the 3433 codons comprising the WN virus genome are segregating more than two distinct amino acids in the global data set: isoleucine, methionine, and valine are segregating at codon 449 (Env 158) and valine, threonine, and alanine are segregating at codon 2209 (NS4A 84).

Table 2 Correlation between linkage disequilibrium and distance

Statistic (LD)	*Value*	*P-value*
G4	41731	0.435
corr(r2 D)	0.02119	0.832
corr(D'd)	−0.00282	0.422

Table 3 Distribution of dN/dS (w) values for global WN virus set

Region	*p1*	*p2*	*p3*	*p4*	*p5*	*Mean*
w	0	1/3	2/3	1	3	
Capsid + preM	0.92908	0.04046	0.03039	0.00007	0.00000	0.0339
Envelope	0.87549	0.12442	0.00008	0.00001	0.00000	0.0415
NS1	0.89266	0.10734	0.00000	0.00000	0.00000	0.0358
NS2 A + B	0.81338	0.14724	0.03924	0.00014	0.00000	0.0752
NS3	0.95074	0.03363	0.01487	0.00074	0.00003	0.0218
NS4 A + B	0.88081	0.11919	0.00000	0.00000	0.00000	0.0397
NS5	0.92121	0.07879	0.00000	0.00000	0.00000	0.0263

Discussion

The complete sequence of a new strain of WNV, named as WNV Merion strain, together with the expression analysis of its individual open reading frames is presented in this work. Phylogenetically WNV Merion strain appears to be distinct from the previously reported WNV strains. In this study, brain tissues of an infected crow from Merion Station, a suburban region of Philadelphia, PA, served as a source for the amplification of WN viral genes through RT-PCR.

Flaviviral replication in host cell is reported to occur mostly in the cytoplasm of infected cells; however, viral assembly and budding occur in the peripheral region of the cells. Currently little is known about the localization pattern of individual WNV antigens. We observed that the full-length protein C, NS1, and NS5 antigens are localized in the nucleus though the pathological relevance of their nuclear localization is not understood. Consistently we observed that the full-length C is imported into the nucleus in variety of cell lines (data not shown). Interestingly, this protein appears to harbor frequent leucine repeats, a cryptic bipartite nuclear localization signal and a homeobox-associated leucine zipper (HALZ) domain, though the physiological significance of these domains is unclear at this stage. HALZ regions are mostly plant-specific leucine zippers that are always found associated with homeobox proteins. Mostly HALZ-containing proteins involve DNA-binding and regulatory functions. The capsid protein species isolated from virus particles were shown to be shorter than the actual capsid molecule cleaved from WNV polyprotein. Thus, during the Flaviviral life cycle, capsid protein should exist in two forms: a full-length nuclear-localized form and a truncated form that participates in viral assembly.

We observed that NS1 was also localized in the nucleus. However, the expression of NS1 has been observed on the surface of WNV-infected cells and served as a potential marker for WNV infection in a recent report (Shrestha *et al*, 2003). This suggests that other WNV antigens modulate NS1 trafficking during natural infection. NS2A and NS2B exhibited a clear cytoplasmic staining pattern. In contrast, the observed intracellular distribution of these antigens

correlates with their putative functions. NS2B has enjoyed significant attention as it acts a cofactor for NS3 to form the NS2B-NS3 proteolytic complex, which is vital in the cleavage of flaviviral polyprotein into its individual gene products (Chambers *et al*, 1993, 1995; Falgout *et al*, 1993). The expression of NS3 resulted in membrane blebbing and the appearance of clusters of vesicular structures near the perinuclear portion of the cells. Similar observations were made in Dengue and Tick-borne encephalitis viruses. The amino terminus of NS3 encodes a serine protease (Wengler, 1993) whereas the carboxyl end mediates its helicase and NTPase activities. NS3 comes under the superfamily of helicases that are characterized by the presence of DEAD/H box domains. Several human proteins also contain DEAH/H domains and they exhibit highly diversified functions. It is also to be noted that although NS3 appears to possess a strong bipartite nuclear targeting sequence (Figure 2), expression of WNV NS3 is observed only in the cytoplasmic portion (Figure 5) of the cells and none of transfected cells displayed a nuclear accumulation of NS3 in the transfected cells. The actual cause behind this disparity is yet to be analyzed.

In the transfected cells, NS5 is expressed as a nuclear-localized protein, though its amino acid sequence does not appear to encode any of the known nuclear targeting sequences. It is the largest as well as the most highly conserved protein among flaviviruses. RNA polymerase activity of this antigen has been demonstrated *in vitro* (Grun and Brinton, 1986, 1987). Sequence analysis shows that NS5 encompasses domains that are characteristic of RNA-dependent RNA polymerases (RdRps) and methyltransferase. Such methyltransferase activity likely plays key role in the methylation of the type I cap (Koonin, 1993). In the nucleus, three activities such an RNA triphosphatase, guanylyltransferase, and methyltransferase are required to synthesize the viral cap structure. But flaviviruses replicate in the cytoplasm and the synthesis of viral type I cap is essential for the production of infectious virus particles. Although WNV mediates triphosphatase and methyltransferase activities, the action of guanylyltransferase has not yet been identified as being encoded by the virus. The host cell machinery may complement the guanylyltransferase activity. Because the triphosphatase activity is provided by NS3 (Wengler, 1993), an interaction between NS3 and NS5 would facilitate the coordination of the helicase, polymerase, and capping activities. However, as presented in this study, both NS3 and NS5 clearly differ in their intracellular distribution patterns. Hence, this suggests not only a coordination between the cellular enzymes and viral enzymes, but also that the nucleocytoplasmic shuttling and reorganization of these enzymatic-viral genome complexes within the host cell are mediated in a more complex fashion during WNV infection. An elucidation of this pathway will be importance for understanding WNV pathogenesis.

The sequence of Merion strain was aligned with various WNV strains from the database collected from divergent hosts and divergent geographical distributions. From a phylogenetic standpoint, the WNV Merion strain appears to be a typical, if slightly more divergent member of the North American group of WNVs. Because the WNV Merion strain is the most recent member of this group, it is possible that its greater distance from the consensus sequence simply reflects the neutral accumulation of mutations over time. The mean numbers of nucleotide and amino acid differences from the North American consensus isolates is listed by year in Table 1. Both show the expected increase with time, although sample sizes are small, and the number of amino acid differences between the Merion sequence and the consensus is exceptionally large, even relative to the number of nucleotide differences. This extreme value could be due to either chance substitution, to a novel selective regime along the lineage leading to the Merion virus, or to sequencing error.

The limited phylogenetic structure evident in the tree constructed for the North American WNVs is consistent with neutral evolution in an expanding epidemic. Rapid population growth, such as has been witnessed in the WNV epidemic in North America, is expected to lead to star-like genealogies in which most coalescences occur near the root (Slatkin and Hudson, 1991). This is evidently the geometry of the North American WNV tree shown in Figure 4A. However, recombination can produce similar patterns (Schierup and Hein, 2000). Twiddy and Holmes (2003) examined diversity plots for a set of 37 WNV envelope sequences from both the Old and the New World and found no evidence of recombination. Visual examination of sequences for breakpoints is likely to reveal recombination only when the parental strains have diverged substantially (Twiddy and Holmes, 2003). We utilized an alternative approach, based on the ancestral recombination graph, which gives an estimate of the population recombination rate, r. The maximum composite-likelihood estimate of r for the coding portion of the North American WN virus sequences is 0.00. Because these sequences are not contemporaneous, the coalescent process implemented in Pairwise does not strictly apply to this data set and so this estimate should be treated with some caution. However, this result is consistent with the lack of significance of the linkage disequilibrium analysis (Table 2) and also corroborates the findings reported in Twiddy and Holmes (2003). Taken together, these analyses all suggest that the WN virus is evolving clonally and that the geometry of the North American WN virus tree reflects the demographics of the epidemic.

Host-parasite interactions coupled with high mutation rates for RNA viruses sometimes leads to locally enhanced substitution rates for this important group of pathogens (Conway and Polley, 2002; Frank, 2002) For example, if host immune responses target

different regions of the viral genome, then each viral lineage may represent a succession of different environments favoring different viral genotypes. Estimation of the relative rates of nonsynonymous and synonymous substitutions can be used to identify both regions and individual codons subject to diversifying selection (Goldman and Yang, 1994). Curiously, we found no evidence for diversifying selection anywhere in the coding regions of the WNV genome. Although distribution of w ratios differed slightly between the different regions, this rate ratio was estimated to be less than one in all regions and at all codons. Indeed, only two codons in the entire genome are segregating more than two amino acids in the global WNV set. These results suggest that escape mutation is playing only a limited role in the current evolution of WNV.

The ability of WNV to replicate in a wide variety of cells and a broad range of hosts as well as vectors support the vulnerability of this virus. Although there are a few reports on the use of antiviral agents to control WNV replication (Anderson and Rahal, 2002; Jordan *et al*, 2000), effective control measures against WNV specifically have yet to be developed. Understanding the mechanism of actions of WNV genes in the host cell and their roles in viral replication will offer important insight into how this virus successfully replicates in the cells of diverse hosts and vectors while evading host-defense strategies. Identification of host cellular factors used by structural as well as nonstructural proteins of WNV in cells will offer valuable tools for the development of targeted therapeutic strategies against WNV. The open reading frames of WNV Merion strain may be of use in this regard.

Materials and methods

Cells, reagents, and plasmids
Human neuroblastoma cell line (SY5Y) was obtained from ATCC (Manassas, VA). Culture media and other standard tissue culture reagents were from Life Technologies (Rockville, MD). The mammalian expression vector, pcDNA3.1/His A, Top 10F *Escherichia coli* chemically treated competent cells, and monoclonal anti-His antibody were obtained from Invitrogen (Carlsbad, CA). Alexa-488- and Alexa-546-conjugated anti-mouse secondary antibodies were purchased from Molecular Probes (Engene, OR).

Cloning and sequence analysis of WNV genes
Total RNA was isolated from brain tissue samples of a dead Crow (Rnasy Kit, Qiagen) to prepare cDNA by using MMLV reverse transcriptase (Clontech, CA). The reaction for cDNA preparation was set as follows: In a 0.5-ml RNASE-free microcentrifuge tube (Ambion), 1 μg of RNA and 1 μl of random primers were added and diluted to the volume of 12.5 μl with DEPC (diethyl pyrocarbonate)-treated water. This mix was incubated for 2 min at 70°C and quenched on ice. To this, diluted RNA-primers mix, 4.0 μl of 5 × reaction buffer, 1.0 μl of dNTP mix (10 mM each), 0.5 μl RNase inhibitor, and 1.0 μl of MMLV reverse transcriptase were added, mixed, and incubated for one hour at 34°C. The cDNA synthesis and DNase activity were terminated by boiling the reaction mix for 5 min at 94°C. Typically 1 μl of the cDNA product was used as template in the PCR reaction. Advantage Taq polymerase (Clontech, CA) with a broad range temperature regime of annealing temperatures (54°C to 59°C) (Stratagene ThermoCycler) was employed. The amplified fragments were separated in a 1.0% agarose gel and based on the expected length of PCR products, the appropriate bands were eluted from the gel. The eluted gel products were ligated into pcDNA3.1 HisA/TOPO cloning vector (Invitrogen) and the colonies were screened onto Luria broth (LB)-amphicillin agar plates. The positive clones were confirmed by automated sequencing and the genetic sequence was confirmed by two independent reactions to verify the consistency.

Protein architectural analysis
The deduced amino acid sequence derived from WNV Merion strain open reading frames were analyzed using online modular programs including SMART (Simple Modular Architecture Research Tool) program at (http://www.embl-heidelberg.de/predictprotein/predictprotein.html) EMBL site, PORT II server (http://psort.nibb.ac.jp), and Prosite-ExPaSy (http://us.expasy.org/) to identify various structural motifs and functional domains existing on WNV proteins(Copley *et al*, 1999; Letunic *et al*, 2002; Ponting *et al*, 1999; Schultz *et al*, 1998, 2000).

In vitro expression of WNV gene products
The WN viral gene expression plasmids were used to generate ^{35}S-labeled protein samples by using TNT coupled *in vitro* transcription/translation system (Promega Corporation, Madison, WI). The radioactive protein samples were immunoprecipitated by using anti-His antibody as described (Ramanathan *et al*, 2002). The immunoprecipitated samples were resolved by sodium dodecyl sulfate–Polyacrylamide gel electro phoresis (SDS-PAGE) and subsequently analyzed by autoradiography.

Indirect immunofluoresencent assay
The intracellular distribution pattern of WNV gene products in transfected cells were monitored by indirect immunofluoresecent assay (Ramanathan *et al*, 2002). Neuronal cells, grown in slide chambers (Beckton-Dickenson, Franklin Lakes, NJ), were transfected with WNV expression plasmids. Thirty-six hours post transfection, the cells were fixed using methanol and incubated with monoclonal anti-His antibody (1:200 dilution) for 90 min Subsequently, the slides were incubated with either Alexa-488–conjugated or Alexa-546-conjugated secondary antibody (Molecular Probes) for 45 min, after gentle rinsing of the samples with phosphate-buffered

saline (PBS). DAPI (4′-6′-Diamidino-2-phenylindole) (Sigma, St. Louis, MO) was added to counterstain the nuclear content of the cells. The slides were rinsed gently to remove residual DAPI and secondary antibodies, and mounted with an antifading mounting solution (Molecular Probes). The samples were analyzed under Nikon T-800 fluorescent microscope and photographs were taken with appropriate filters and with a cooled CCD camera (SensiCam High Performance). IPLab (3.6.5) software was used to acquire and analyze the images.

Phylogenetic analysis

Upon thorough verification of the genetic sequences using MacVector (Oxford Probes, UK) program, the deduced amino acid sequences were determined and aligned. The open reading frame of the Merion genome was aligned manually to a set of 18 WNV genomes sequenced by other groups and deposited in GenBank (Table 4). This set includes viral sequences previously isolated in North America, Europe, Africa, and the Middle East, from avian, mammalian, and dipteran hosts. We conducted all phylogenetic analyses using PAUP* (version 4.0b10) (Swofford, 2002) through heuristic searches (tree-bisection-reconnection) with maximum likelihood as the tree optimality criterion. The program ModelTest (version 3.6) Posada and Crandall,1998), was used initially to select an appropriate model of DNA sequence evolution for each data set; ModelTest uses PAUP* to estimate the likelihood of the sequence data under a series of 56 reversible Markov models and then identifies a best-fitting model subject to the Aikake Information Criterion.We parameterized the resulting model iteratively by first finding a tree for the sequence data under the current set of parameter values and then reoptimizing the parameters with respect to this new tree. The initial tree was inferred using the HKY85 model with kappa set equal to 2 and the iterations were halted when the parameter estimates and the tree no longer changed between iterations. Support for each clade was assessed by bootstrapping with 100 replicates.

The population mutation ($Q = 2N_em$) and recombination $r = 2N_er$) rates were estimated for the North American WNV sequences using the program Pairwise from the package Ldhat(Mc Vean *et al*, 2002). Pairwise constructs a composite-likelihood surface *r* using an importance sampler derived from the two-locus Kingman coalescent (with effective population size N_e) and allowing for recurrent mutation and recombination. Q is estimated by a finite-sites approximation of the Watterson estimate. Because this method assumes contemporaneous sampling of the terminal sequences and lack of geographical structure, we retained only the North American sequences for this analysis. We also used the LDhat program Permute to assess the significance of the correlation between linkage disequilibrium and the physical distance between segregating sites. At intermediate levels of recombination, these correlations are expected to be negative. The program Codeml (version 3.13) (Yang, 1997) was used to quantify the relative rates of nonsynonymous and synonymous substitution for the global set of the WNV sequences. The rate ratio, $w = dN/dS$, can be interpreted as a measure of the strength and kind of selection operating on a coding sequence. Values of W that are less than or greater than 1 are indicative of purifying or diversifying selection, respectively. When $w = 1$, neutral evolution can be inferred. Because even adjacent codons may be subject to different selective regimes, it is desirable to obtain estimates of w at each codon. Codeml implements an empirical Bayesian strategy to obtain such estimates. The data at all sites is first used to parameterize a prior distribution on w. Then, Bayes formula is used with that prior and with the data at individual sites to obtain the codon-specific estimates of w. For these analyses, we used the frequencies prior ($M = 4$) implemented in Codeml, which allows for five values of w, fixed at 0, 1/3, 2/3, 1, and 3, with weights to be estimated with respect to the data. We divided the WNV genome into seven regions, consisting of Capsid and preM, envelope (E), NS1, NS2A and -B, NS3, NS4A and -B, and NS5, and analyzed each region separately. The ML tree for the complete genome was used for all of these analyses, with branch lengths reoptimized by Codeml.

Table 4 West Nile virus sequences used for phylogenetic analysis

Accession no.	*Location*	*Host*	*Year*	*Ref.*
AY262283	Kenya	*Culex univittatus*	1998	1
AF404757	Italy	Horse	1998	2
AF260969	Romania	*Culex pipiens*	1996	3
AY278442	Volgograd, Russia	Human	2000	4
AY277252	Volgograd, Russia	Human	1999	5
AF317203	Volgograd, Russia	Human	1999	6
AF260968	Egypt	Human	1951	7
AY278441	Astrakhan, Russia	Human	1999	8
AF481864	Israel	White stork	1998	9
AF196835	New York	Chilean flamingo	1999	10
AF260967	New York	Horse	1999	7
AF206518	Connecticut	*Culex pipiens*	1999	11
AF202541	New York	Human	1999	12
AF404756	New York	American crow	2000	2
AF404755	New York	Ruffed grouse	2000	2
AF404754	New Jersey	*Culex pipiens*	2000	2
AF404753	Maryland	American crow	2000	2
AF533540	New York	Human	2001	13
	Merion, PA	American crow	2002	

1. AC Brault and X. de Lamballerie. 2003. Unpublished.
2. Lanciotti *et al*, 2002.
3. Savage *et al*, 1999.
4. GK Sadykova *et al.* 2003. Unpublished.
5. AG Prilipov *et al.* 2003. Unpublished.
6. AE Platonov *et al.* 2000. Unpublished.
7. M Bowen *et al.* 2000. Unpublished.
8. AG Voronina *et al.* 2003. Unpublished.
9. Malkinson *et al*, 2002.
10. Lanciotti *et al*, 1999.
11. Anderson *et al*, 1999.
12. Jia *et al*, 1999.
13. Huang *et al*, 2002.

References

Alpert SG, Fergerson J, Noel LP (2003). Intrauterine West Nile virus: ocular and systemic findings. *Am J Ophthalmol* **136:** 733–735.

Anderson JF, Andreadis TG, Vossbrinck CR, Tirrell S, Wakem EM, French RA, Garmendia AE, Van Kruiningen HJ (1999). Isolation of West Nile virus from mosquitoes, crows, and a Cooper's hawk in Connecticut. *Science* **286:** 2331–2333.

Anderson JF, Rahal JJ (2002). Efficacy of interferon alpha-2b and ribavirin against West Nile virus in vitro. *Emerg Infect Dis* **8:** 107–108.

Brinton MA (2002). The molecular biology of West Nile virus: a new invader of the western hemisphere. *Annu Rev Microbiol* **56:** 371–402.

Campbell GL, Marfin AA, Lanciotti RS, Gubler DJ (2002). West Nile virus. *Lancet Infect Dis* **2:** 519–529.

Chambers TJ, Hahn CS, Galler R, Rice CM (1990). Flavivirus genome organization, expression and replication. *Ann Rev Microbiol* **44:** 649–688.

Chambers TJ, Nestorowicz A, Amberg SM, Rice CM (1993). Mutagenesis of the yellow fever virus NS2B protein: effects on proteolytic processing, NS2B-NS3 complex formation, and viral replication. *J Virol* **67:** 6797–6807.

Chambers TJ, Nestorowicz A, Rice CM (1995). Mutagenesis of the yellow fever virus NS2B/3 cleavage site: determinants of cleavage site specificity and effects on polyprotein processing and viral replication. *J Virol* **69:** 1600–1605.

Conway DJ, Polley SD (2002). Measuring immune selection. *Parasitol Today* **S5:** S3–S16.

Copley RR, Schultz J, Ponting CP, Bork P (1999). Protein families in multicellular organisms. *Curr Opin Struct Biol* **9:** 408–415.

Couzin J, (2003). West Nile virus. Blood banks in a 'race against the mosquitoes.' *Science* **299:** 1824.

DeSalvo D, Roy-Chaudhury P, Peddi R, Merchen T, Konijetti K, Gupta M, Boardman R, Rogers C, Buell J, Hanaway M, Broderick J, Smith R, Woodle ES (2004). West Nile virus encephalitis in organ transplant recipients: another high-risk group for meningoencephalitis and death. *Transplantation* **77:** 466–469.

Falgout B, Miller RH, Lai CJ (1993). Deletion analysis of dengue virus type 4 nonstructural protein NS2B: identification of a domain required for NS2B-NS3 protease activity. *J Virol* **67:** 2034–2042.

Frank SA (2002). *Immunology of infectious diseases*. Princepton, NJ: Prinston University Press.

Goldman N, Yang Z (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* **11:** 726–736.

Grun JB, Brinton MA (1986). Characterization of West Nile virus RNA-dependent RNA polymerase and cellular terminal adenylyl and uridylyl transferases in cell-free extracts. *J Virol* **60:** 1113–1124.

Grun JB, Brinton MA (1987). Dissociation of NS5 from cell fractions containing West Nile virus-specific polymerase activity. *J Virol* **61:** 3641–3644.

Huang C, Slater B, Rudd R, Parchuri N, Hull R, Dupuis M, Hindenburg A (2002). First Isolation of West Nile virus from a patient with encephalitis in the United States. *Emer, Infect Dis* **8**(12): 1367–1371.

Iwamoto M, Jernigan DB, Guasch A, Trepka MJ, Blackmore CG, Hellinger WC, Pham SM, Zaki S, Lanciotti RS, Lance-Parker SE, DiazGranados CA, Winquist AG, Perlino CA, Wiersma S, Hillyer KL, Goodman JL, Marfin AA, Chamberland ME, Petersen LR (2003). Transmission of West Nile virus from an organ donor to four transplant recipients. *N Engl J Med* **348:** 2196–2203.

Jia XY, Briese T, Jordan I, Rambaut A, Chi HC, Mackenzie JS, Hall RA, Scherret J, Lipkin WI (1999). Genetic analysis of West Nile New York 1999 encephalitis virus. *Lancet* **354:** 1971–1972.

Jordan I, Briese T, Fischer N, Lau JY-N, Lipkin WI (2000). Ribavirin inhibits West Nile Virus replication and cytopathic effect in neural cells. *J Infect Dis* **182:** 1214–1217.

Kiberd BA, Forward K (2004). Screening for West Nile virus in organ transplantation: a medical decision analysis. *Am J Transplant* **4:** 1296–1301.

Koonin EV (1993). Computer-assisted identification of a putative methyltransferase domain in NS5 protein of flaviviruses and lambda 2 protein of reovirus. *J Gen Virol* **74(Pt 4):** 733–740.

Kusne S, Smilack J (2005). Transmission of West Nile virus by organ transplantation. *Liver Transpl* **11:** 239–241.

Lanciotti R, Ebel GD, Deubel V, Kerst AJSM, Meyer R, Bowen M, McKinney N, Morrill WE, Crabtree MB, Kramer LD, Roehrig JT (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.

Lanciotti RS, Roehrig JT, Deubel V, Smith J, Parker M, Steele K, Crise B, Volpe KE, Crabtree MB, Scherret JH, Hall RA, MacKenzie JS, Cropp CB, Panigrahy B, Ostlund E, Schmitt B, Malkinson M, Banet C, Weissman J, Komar N, Savage HM, Stone W, McNamara T, Gubler DJ. (1999). Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science* **286:** 2333–2337.

Lefrere JJ, Allain JP, Prati D, Sauleda S, Reesink H (2003). West Nile virus and blood donors. *Lancet* **361:** 2083–2084.

Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Ciccarelli F, Copley RR, Ponting CP, Bork P. (2002). Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res* **30:** 242–244.

Malkinson M, Banet C, Weisman Y, Pokamunski S, King R, Drouet MT, Deubel V. (2002). Introduction of West Nile virus in the Middle East by migrating white storks. *Emerg Infect Dis* **8:** 392–397.

Mc Vean G, Awadalla P, Fearnhead P (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160:** 1231–1241.

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:** 1807–1814.

Ponting CP, Schultz J, Milpetz F, Bork P (1999). SMART: identification and annotation of domains from signalling and extracellular protein sequences. *Nucleic Acids Res* **27:** 229–232.

Posada D, and Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

Ramanathan MP, Curley E, 3rd, Su M, Chambers JA, Weiner DB (2002). Carboxyl Terminus of hVIP/mov34 Is Critical for HIV-1-Vpr Interaction and Glucocorticoid-mediated Signaling. *J Biol Chem* **277:** 47854–47860.

Savage HM, Ceianu C, Nicolescu G, Karabatsos N, Lanciotti R, Vladimirescu A, Laiv L, Ungureanu A, Romanca C, Tsai TF. (1999). Entomologic and avian investigations of an epidemic of West Nile fever in Romania in 1996, with serologic and molecular characterization of a virus isolate from mosquitoes [published erratum appears in Am J Trop Med Hyg 2000 Jan;62(1):162]. *Am J Trop Med Hyg* **61:** 600–611.

Schierup MH, Hein J (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156:** 879–891.

Schultz J, Copley RR, Doerks T, Ponting CP, Bork P (2000). SMART: a Web-based tool for the study of genetically mobile domains. *Nucleic Acids Res* **28:** 231–234.

Schultz J, Milpetz F, Bork P, Ponting CP (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc Natl Acad Sci U S A* **95:** 5857–5864.

Shrestha B, Gottlieb D, Diamond MS (2003). Infection and injury of neurons by West Nile encephalitis virus. *J Virol* **77:** 13203–13213.

Slatkin M, Hudson RR (1991). Pairwise comparisons of mitochondrial-DNA sequences in stable and exponentially growing populations. *Genetics* **129:** 555–562.

Swofford DL (2002). *PAUP—phylogenetic analysis using parsimony*. Version 4. Sunderland, Massachusetts: Sinauer Associates.

Twiddy SS, Holmes EC (2003). The extent of homologous recombination in members of the genus Flavivirus. *J Gen Virol* **84:** 429–440.

Wengler G (1993). The NS 3 nonstructural protein of flaviviruses contains an RNA triphosphatase activity. *Virology* **197:** 265–273.

Williams K (2004). Modes of transmission for West Nile virus. *Clin Lab Sci* **17:** 56.

Yamshchikov VF, Compans RW (1994). Processing of the intracellular form of the west Nile virus capsid protein by the viral NS2B-NS3 protease: an in vitro study. *J Virol* **68:** 5765–5771.

Yang J-S, Kim J, Hwang D, Choo AY, Dang K, Maguire H, Kudchodkar S, Ramanathan MP, Weiner DB (2001). Induction of potent Th1-type immune responses from a novel DNA vaccine for West Nile Virus New York Isolate (WNV-NY1999). *J Infect Dis* **184:** 809–816.

Yang J-S, Ramanathan MP, Muthumani K, Jin SH, Yu, QC, Choo AY, Hwang DH, Lee MD, Dang K, Dang W, Kim JJ, Weiner, DB (2002). Induction of inflammation by West Nile Virus capsid through the caspase-9 apoptotic pathway. *Emerg Infect Dis* **8:** 1379–1384.

Yang Z (1997). PAML: a program package for phylogenetic analyses by maximum likelihood. *Appl Biosci* **13:** 555–556.