Journal of General Virology (2005), **86**, 2175–2183 DOI 10.1099/vir.0.81015-0

Genetic variation in *West Nile virus* from naturally infected mosquitoes and birds suggests quasispecies structure and strong purifying selection

Greta Jerzak,[1] Kristen A. Bernard,[1,2] Laura D. Kramer[1,2] and Gregory D. Ebel[1,2]

Correspondence
Gregory D. Ebel
ebel@wadsworth.org

[1]The Arbovirus Laboratories, Wadsworth Center, New York State Department of Health, 5668 State Farm Road, Slingerlands, NY 12159, USA

[2]Department of Biomedical Sciences, School of Public Health, The University at Albany, State University of New York, Albany, NY 12144-3456, USA

Intrahost genetic diversity was analysed in naturally infected mosquitoes and birds to determine whether *West Nile virus* (WNV) exists in nature as a quasispecies and to quantify selective pressures within and between hosts. WNV was sampled from ten infected birds and ten infected mosquito pools collected on Long Island, NY, USA, during the peak of the 2003 WNV transmission season. A 1938 nt fragment comprising the 3′ 1159 nt of the WNV envelope (E) coding region and the 5′ 779 nt of the non-structural protein 1 (NS1) coding region was amplified and cloned and 20 clones per specimen were sequenced. Results from this analysis demonstrate that WNV infections are derived from a genetically diverse population of genomes in nature. The mean nucleotide diversity was 0·016 % within individual specimens and the mean percentage of clones that differed from the consensus sequence was 19·5 %. WNV sequences in mosquitoes were significantly more genetically diverse than WNV in birds. No host-dependent bias for particular types of mutations was observed and estimates of genetic diversity did not differ significantly between E and NS1 coding sequences. Non-consensus clones obtained from two avian specimens had highly similar genetic signatures, providing preliminary evidence that WNV genetic diversity may be maintained throughout the enzootic transmission cycle, rather than arising independently during each infection. Evidence of purifying selection was obtained from both intra- and interhost WNV populations. Combined, these data support the observation that WNV populations may be structured as a quasispecies and document strong purifying natural selection in WNV populations.

Received 7 March 2005
Accepted 24 April 2005

INTRODUCTION

West Nile virus (WNV) (family *Flaviviridae*, genus *Flavivirus*) was first detected in North America in 1999 in the New York City area, where it caused an avian and equine epizootic and 68 cases of human disease (CDC, 1999; Lanciotti *et al.*, 1999). It subsequently spread throughout the United States and into Canada, Mexico and the Caribbean (Blitvich *et al.*, 2003; Dupuis *et al.*, 2003; Austin *et al.*, 2004). Recently, several studies have examined the course of WNV evolution since its introduction and have concluded that WNV remains a relatively homogeneous virus population, with the most divergent strains containing only a few nucleotide and/or amino acid substitutions (Anderson *et al.*, 2001; Ebel *et al.*, 2001a, 2004; Lanciotti *et al.*, 2002; Beasley *et al.*, 2003; Davis *et al.*, 2003). However, a single WNV genotype that differs from the introduced strain has arisen since 1999 and has become dominant, largely displacing previously circulating strains throughout North America (Davis *et al.*, 2003; Ebel *et al.*, 2004). Therefore, although WNV remains relatively genetically homogeneous, it appears to be undergoing a process of adaptation to local transmission cycles. To date, no published studies have quantitatively examined the role of positive and/or purifying selection in WNV since its introduction into North America.

The mechanisms that lead to population-level genetic change in WNV and other arthropod-borne viruses (arboviruses) are poorly understood. RNA viruses such as WNV are thought to exist within a host as a genetically heterogeneous mixture of variants that differ from a

The GenBank/EMBL/DDBJ accession numbers for consensus sequences reported in this paper are DQ010338–DQ010357.

Supplementary material is available in JGV Online.

consensus nucleotide sequence: the term 'quasispecies' has come to refer to the complex mutant spectrum that surrounds the master viral nucleotide sequence (Eigen & Biebricher, 1988; Eigen, 1993; Domingo, 1998). Quasispecies result from the high error rates of most RNA virus-encoded RNA-dependent RNA polymerases (RDRP) (Holland *et al.*, 1982), as well as their short generation times and large population sizes. A genetically diverse virus population would seem to have adaptive advantages due to the pre-existence of variants within the mutant spectrum that may be more fit in a novel and/or changing environment. This genetic diversity may have important implications for virus populations and hosts. For example, quasispecies structure has been shown to be critical in the failure of hepatitis C treatment (Farci *et al.*, 2000, 2002), AIDS disease progression (Essajee *et al.*, 2000) and the persistence of virus infections at the cellular, organism and population levels (Domingo *et al.*, 1998). The diversity of the viral mutant spectrum has been shown to be both host- and virus-dependent (Schneider & Roossinck, 2000, 2001) and a critical determinant of virus fitness (Martínez *et al.*, 1991; Ruiz-Jarabo *et al.*, 2002). Further, quasispecies structure provides a virus population with a molecular memory that exists as minority genotypes within the quasispecies mutant distribution (Ruiz-Jarabo *et al.*, 2000, 2002; Domingo *et al.*, 2002). Quasispecies population structure may therefore be critical to RNA virus perpetuation in nature.

It is unclear, however, whether mosquito-transmitted RNA viruses, which evolve more slowly than other RNA viruses and are constrained by replication in taxonomically diverse hosts (Weaver *et al.*, 1992; Zanotto *et al.*, 1995), are best described and understood as quasispecies (Smith *et al.*, 1997; Holmes & Moya, 2002). In particular, quasispecies theory states that natural selection acts on the whole of the mutant spectrum rather than on the individual genomes composing it (Eigen & Biebricher, 1988). Intrahost genetic variation is thus necessary for a quasispecies population structure, but not sufficient. If intrahost genetic diversity arises anew following infection by a clonal or highly bottlenecked virus population, the possibility for natural selection to act on the entire mutant spectrum at critical points in its life cycle (and for the transmission of the proposed molecular memory) is negated. This problem is compounded in arboviruses such as WNV that are maintained in nature in an enzootic cycle between arthropods and vertebrates. In complex transmission cycles, the genetic bottlenecks that may accompany transmission between host types may be frequent and the duration of an infection in any given host may be comparatively short. Few studies have examined the relevance of quasispecies to naturally occurring arbovirus populations. The hypothesis that WNV is maintained in nature as a quasispecies was therefore evaluated in this study. In particular, WNV was sampled from naturally infected mosquitoes and birds and intrahost genetic diversity was quantified within each host type. By using these data, we determined whether patterns of intrahost genetic diversity are host-dependent and whether minority (i.e. non-consensus) WNV haplotypes may be preserved throughout the transmission cycle, and also assessed the role of natural selection in shaping intrahost and interhost genetic diversity.

METHODS

Specimen selection and collection. WNV-infected specimens obtained from the New York State Department of Health's arbovirus surveillance programme were used to analyse intrahost genetic diversity. A series of infected bird kidneys and mosquito pools were selected from samples submitted from Suffolk County, NY, USA, during the transmission season in 2003. Specimens were selected to represent species likely to be important to WNV perpetuation in nature. Sampling focused on Brookhaven Township during late summer 2003 to minimize interhost variation in virus strain, mosquito and avian genetics, and environmental factors that may influence virus populations indirectly through impacts on mosquito and avian health. Kidney was selected from WNV-infected birds due to the availability and reliably high virus titres of this tissue. Mosquitoes had been collected in CDC miniature light traps, identified to species, pooled in 2 ml safe-lock microfuge tubes (Eppendorf) containing a 4·5 mm diameter zinc-plated steel ball bearing (Daisy Brand) and shipped to the Wadsworth Center Arbovirus Laboratories on dry ice. To assess the likelihood that mosquito pools may be infected with more than one infected mosquito, minimum infection rates (MIR) were calculated for each collection site during the month of collection by using Wadsworth Center surveillance data. The number of WNV-positive pools of *Culex* mosquitoes collected in each township in Suffolk County was divided by the total number of individual mosquitoes tested and multiplied by 1000. The MIR can thus be interpreted as the minimum number of infected mosquitoes per 1000 in a given township during the period of collection. Dead birds were submitted to the arbovirus-surveillance programme by Suffolk County health officials and necropsies were performed at the Wildlife Pathology Unit of the New York State Department of Environmental Conservation; tissues were sent to the Arbovirus Laboratories in individual jars on dry ice.

WNV detection and quantification. The infection status of each specimen was determined by quantitative, real-time (TaqMan) RT-PCR as described previously (Shi *et al.*, 2001; Kauffman *et al.*, 2003). To determine whether low virus titres may bias estimates of genetic diversity via repeated sampling of DNA amplified from a single RNA molecule, WNV RNA and infectious virus were quantified in each specimen by real-time RT-PCR as described previously (Shi *et al.*, 2001; Kauffman *et al.*, 2003) and by plaque assay on African Green Monkey kidney (Vero) cells according to standard protocols. Briefly, confluent cell monolayers in six-well culture plates were inoculated with 0·1 ml serial tenfold dilutions of WNV-containing material diluted in BA-1 diluent [M-199H, 1 % bovine serum albumin, 0·05 M Tris (pH 7·6), 0·35 g sodium bicarbonate l^{-1}, 100 μg penicillin ml^{-1}, 100 μg streptomycin sulfate ml^{-1}, 1 μg fungizone ml^{-1}]. Plates were incubated for 1 h at 37 °C with 5 % CO_2. A primary overlay containing 0·6 % agar in Eagle's minimal essential medium with 10 % fetal bovine serum (FBS) was applied and plates were incubated as described above. After 48 h incubation, a second overlay, as described above except that it contained 2 % FBS and neutral red, was applied to each well. Plates were returned to the incubator and plaques were counted at 72 h post-infection.

***Culex* species identification.** Females of *Culex pipiens* and *Culex restuans* are difficult to differentiate on the basis of morphological characters. To determine the species composition of the WNV-infected mosquito pools that had been identified in the field as either *C. pipiens* or *C. restuans*, species present in the pools were

identified by using PCR. Genomic DNA was extracted from mosquito-pool homogenate by using a DNeasy Tissue kit (Qiagen) according to the manufacturer's protocol. PCRs containing primers targeting ITS1 and ITS2 sequences specific for *C. pipens*, *C. restuans* and *Culex salinarius* were performed as described by Crabtree *et al.* (1995). Reactions were amplified in an MJ Research PTC 2000 thermocycler programmed for one cycle at 95 °C for 5 min, 35 cycles at 95 °C for 15 s, 55 °C for 30 s and 72 °C for 1 min, and one cycle at 72 °C for 5 min. Products were visualized on a 1·5 % agarose gel and inspected for band sizes of 698, 506 and 175 bp, representing *C. pipiens*, *C. restuans* and *C. salinarius*, respectively.

High-fidelity RT-PCR, cloning and sequencing. RNA was extracted from infected specimens by using RNeasy spin columns (Qiagen) and RT-PCR was conducted by using primers designed to amplify the 3′ 1159 nt of the WNV envelope (E) coding region and the 5′ 779 nt of the WNV non-structural protein 1 (NS1) coding region [forward primer WNV1311 (5′-ATGCGCCAAATTTGCC-TGCTCTAC-3′); reverse primer WNV3248 (5′-ATGGGCCCTGG-TTTTGTGTCTTGT-3′)]. RT of 5 µl RNA was performed with M-MLV-RT (Ambion) and Sensiscript RT (Qiagen) at 45 °C for 40 min. RT reactions were followed by heat inactivation at 95 °C for 5 min. The resulting cDNA was used as a template for PCR amplification. To minimize misincorporations introduced during RT-PCR by *Taq* polymerase, the resulting WNV cDNA was then amplified with a 'high-fidelity' protocol using *Pfu*Ultra (Stratagene), according to the manufacturer's specifications. Amplification was carried out for 40 cycles at 94 °C for 30 s, 50 °C for 30 s and 72 °C for 4 min, and one cycle at 72 °C for 10 min. PCR products were visualized on a 1·5 % agarose gel, amplicons of the appropriate size were excised and DNA was recovered by using a MinElute Gel Extraction kit (Qiagen) as specified by the manufacturer. The recovered DNA was ligated into the cloning vector pCR-Script Amp SK(+) and transformed into XL10-Gold Ultracompetent cells (Stratagene) according to the manufacturer's protocol. The blue–white colour-screening method was used to select transformed colonies. White colonies were screened by direct PCR using primers specific for the insert of interest. Plasmid DNA was purified by using either a Wizard Plus Miniprep kit (Promega) or a QIAprep Spin Miniprep kit (Qiagen) as specified by the manufacturers. Sequencing was carried out by using five pairs of overlapping primers (sequences available from the authors upon request) and the T3 reverse primer. Sequencing was performed at the Wadsworth Center Molecular Genetics Core (WCMGC) using ABI 3700 and 3100 automated sequencers (Applied Biosystems). Twenty clones per positive bird and per mosquito pool were sequenced. Cloning and plasmid DNA extraction were performed on separate days for each specimen to reduce the likelihood of between-specimen contamination. In addition, each set of clones was sent to the WCMGC separately so that no two sets of clones would be included on the same sequencing run.

Sequence analysis. Sequences were compiled and edited by using the SeqMan module of the DNAStar software package and a minimum of twofold redundancy throughout each clone was required for sequence data to be considered complete. Twenty clones from each individual bird or mosquito pool were aligned by using MegAlign within DNAStar. The consensus sequence for each sample was determined and the sequence of each clone was compared to the consensus. The percentage of nucleotide mutations (total number of mutations divided by total number of bases sequenced) and the percentage of mutant clones were used as indicators of genetic diversity. Statistics were performed by using the STATA software package and GraphPad Prism version 4.00.

Analysis of divergence and selection. Analysis of intrahost WNV populations was conducted by using alignments of the nucleotide sequences of 20 clones for each of the 20 WNV specimens described above. Interhost measures of divergence and selection were obtained from an alignment of the 20 consensus sequences obtained from these intrahost populations (designated NY-Suffolk-03 in Table 6) and from an alignment of 67 WNV E coding region sequences obtained during previous studies of WNV in New York described elsewhere (designated NY-99-03 in Table 6) (Ebel *et al.*, 2001a, 2004). To determine the nucleotide divergence present in each WNV population (both intra- and interhost alignments), the mean pairwise nucleotide distance between sequences (π) was computed by using DnaSP (www.ub.es/dnasp) (Rozas & Rozas, 1999). In addition, the proportion of mutations in each alignment that were non-synonymous [designated *pN* by Holmes (2003)] was computed. Finally, the number of non-synonymous (d_N) to synonymous (d_S) substitutions per site (d_N/d_S) was compared. The mean number of synonymous and non-synonymous sites per sequence in each alignment was calculated by using the Nei–Gojobori method implemented in DnaSP and d_N/d_S ratios were calculated by using Microsoft Excel. d_N/d_S ratios were set as 'undefined' in alignments without any mutations and as 1·000 in alignments with no synonymous mutations.

RESULTS

WNV-infected samples were taken from bird and mosquito species that are important in the enzootic transmission cycle of WNV. Birds included seven American crows, two blue jays and one fish crow (Table 1). Mosquito pools that were identified morphologically as *C. pipiens* and/or *C. restuans* consisted of *C. pipiens*, *C. restuans* and *C. salinarius*, all important WNV vectors in the north-eastern United States (Bernard *et al.*, 2001) (Table 2). PCR amplification for mosquito-species identification failed on a single pool that had been identified morphologically as *Culex* species and the identity of the individuals that composed that pool could not be determined precisely. All infected specimens were collected from a focus of WNV transmission that was centred on Brookhaven Township during late summer 2003 (Tables 1 and 2). Mean MIR was 2·3 (range, 0·5–5·7) and the mean number of individuals in mosquito pools was 22 (range, 10–50). Examination of infectious WNV titres and WNV RNA copy number demonstrated the presence of infectious virus in most (16 of 20) specimens and high RNA copy numbers in all specimens analysed (Tables 1 and 2).

Analysis of intrahost genetic diversity within individual specimens revealed that WNV is composed of a genetically heterogeneous mixture of variants, or a quasispecies, in most specimens. The number of unique variants present in the quasispecies of each sample was one to six in birds and one to ten in mosquitoes. The percentage of clones and the percentage of bases sequenced within each specimen that differed from consensus were similarly variable (Tables 1 and 2). In two infections, one from a mosquito pool and one from an American crow, no evidence of a quasispecies distribution was obtained in the sample of clones analysed. Overall, 121 of 771 600 nt sequenced (0·016 %) and 78 of 400 (19·5 %) genomes sampled differed from consensus and 31 % of nucleotide substitutions resulted in amino acid changes (data not shown). Control observations using a WNV population derived from an infectious cDNA clone

Table 1. Avian specimens included in this study: source, infection status and nucleotide diversity

Specimen no.	Species*	Collection		Titre†	RNA copies‡	Percentage with mutation§	
		Township	Date			Nucleotides	Clones
03001492	American crow	Brookhaven	24 July 2003	<1·0	4·9	0·025	25
03001502	Blue jay	Brookhaven	25 July 2003	3·7	5·9	0·013	25
03001707	Blue jay	Brookhaven	31 July 2003	<1·0	5·1	0·013	25
03001722	American crow	Brookhaven	1 August 2003	4·7	5·0	0·002	5
03001723	Fish crow	Brookhaven	30 July 2003	6·2	6·4	0·005	10
03001798	American crow	Brookhaven	4 August 2003	1·8	5·4	0·013	10
03001802	American crow	Brookhaven	5 August 2003	3·2	5·3	0·010	20
03001807	American crow	Brookhaven	6 August 2003	4·6	5·4	0·000	0
03001810	American crow	Brookhaven	5 August 2003	4·3	6·3	0·021	20
03001855	American crow	Brookhaven	6 August 2003	<1·0	5·7	0·007	15

*American crow, *Corvus brachyrhynchos*; blue jay, *Cyanocitta cristata*; fish crow, *Corvus ossifragus*.
†Log_{10} p.f.u. per 0·1 ml tissue suspension used for RNA isolation and copy-number determination.
‡Log_{10} RNA copy number per 0·1 ml tissue suspension used for RNA isolation and virus titration.
§From each infection, 38 620 nt from 20 clones (1931 nt per clone) was sequenced.

of the strain introduced into New York in 1999 (Shi *et al.*, 2002) estimated the rate of misincorporations introduced by our experimental method (RT-PCR, cloning and sequencing). Three of 75 309 bases sequenced (0·004 %) and three of 39 clones (7·7 %) differed from the consensus. All three mutations detected in the control WNV population were C to A transversions and all resulted in changes in the predicted amino acid sequence.

To evaluate the possibility that estimates of intrahost genetic diversity may differ when structural and non-structural coding regions were examined, these estimates were compared independently for the portions of the E and NS1 coding regions that were co-amplified during RT-PCR (data not shown). Estimates of nucleotide diversity within the NS1 coding region were slightly higher than those from the E coding region and estimates of predicted amino acid diversity were slightly higher within the predicted E protein sequence. However, these estimates did not differ significantly when considered either independently for mosquitoes and birds or for the dataset as a whole.

Examination of estimates of intrahost genetic diversity derived from mosquito- and bird-derived populations of WNV showed that the percentage of substituted bases and mutant clones was significantly greater in WNV sequences

Table 2. Mosquito specimens included in this study: source, infection status and nucleotide diversity

Specimen no.	Species*	Collection		Pool				Percentage with mutation§	
		Township	Date	MIR	Size	Titre†	RNA copies‡	Nucleotides	Clones
32030356	NA	Islip	6 August 2003	1·1	21	3·2	4·9	0·000	0
34030291	P	Babylon	20 August 2003	5·7	25	1·3	4·8	0·002	5
35030311	P, R, S	Huntington	27 August 2003	3·3	19	2·6	4·6	0·005	10
35030318	R, S	Smithtown	27 August 2003	3·0	31	2·3	4·8	0·065	45
35030321	P	Babylon	27 August 2003	5·7	10	3·1	4·7	0·021	30
36030191	P, S	Brookhaven	3 September 2003	0·5	19	2·2	5·0	0·036	45
36030227	P	Riverhead	3 September 2003	1·1	14	<0·7	4·1	0·013	20
37030094	S	Brookhaven	10 September 2003	0·5	50	2·3	4·7	0·019	35
39030148	P, S	Riverhead	24 September 2003	1·1	17	3·0	5·0	0·016	20
34030297	P, R	Riverhead	20 August 2003	1·1	13	2·1	4·0	0·028	25

*Species were identified by examining adult morphology and by PCR. All mosquitoes were identified as *Culex pipiens/restuans* by examining adult female morphology. Identification using PCR is shown: P, *Culex pipiens*; R, *Culex restuans*; S, *Culex salinarius*; NA, no amplification.
†Log_{10} p.f.u. per 0·1 ml clarified mosquito homogenate used for RNA isolation.
‡Log_{10} RNA copies per 0·1 ml clarified mosquito homogenate used for RNA isolation and virus titration.
§From each infection, 38 620 nt from 20 clones (1931 nt per clone) was sequenced.

Table 3. Mean nucleotide and amino acid diversity in WNV populations in birds and mosquitoes

WNV source	Nucleotide substitution (%)		Amino acid substitution (%)	
	Bases (*n*)	Clones (*n*)	Amino acids (*n*)	Clones (*n*)
Avian kidney	0·011 (386 200)	15·5 (200)	0·012 (128 800)	6·5 (200)
Mosquito pool	0·020 (386 200)	23·5 (200)	0·017 (128 800)	11 (200)
Infectious clone	0·004 (75 309)	7·7 (39)	0·012 (25 116)	7·7 (200)
P value*	0·001	0·043	0·330	0·111

*χ^2 test for difference between avian and mosquito values.

Table 4. Types of mutation observed in WNV populations in birds and mosquitoes

WNV source	Substitutions		Insertions	Deletions	NTR*	Total mutations
	Transitions	Transversions				
Avian	35	5	1	2	0	43
Mosquito	60	9	1	6	2	78
Total†	95	14	2	8	2	121

*NTR, Non-translated mutations.
†Proportion of particular mutations was not dependent on WNV source (χ^2 test, $P=0·783$).

from mosquito pools than from birds, with mosquito populations of WNV approximately twofold more genetically diverse than WNV from birds (Table 3). Although predicted amino acid diversity was also greater in mosquitoes than in birds, the difference was not statistically significant. Comparison of the mutation types observed in mosquitoes and birds failed to identify a host-dependent bias for a particular class of mutation (Table 4). Transitions and transversions were the most abundant mutations observed. However, insertions, deletions and non-translated mutations were also present in a small proportion of clones.

Comparison of clones from all 20 specimens included in this study identified two avian specimens, 03001798 and 03001810, that shared non-consensus components of their mutant spectra (i.e. minority haplotypes) (Table 5). Three clones shared five nucleotides that differed from the consensus sequence obtained from either or both specimens. Three shared mutations were detected in the E coding region and two were in the NS1 coding region. All five mutations were synonymous.

Patterns of genetic divergence and non-synonymous variation were examined in intra- and interhost WNV populations (Table 6). The mean genetic distance (π) in all of the alignments was characteristically low, with the highest value of π occurring in interhost WNV populations. The mean value of π in intrahost WNV populations was approximately tenfold lower than the genetic distance in interhost sequences ($P<0·0001$, *t*-test for unpaired samples, equal variances). The π value was slightly higher in mosquito- than in avian-derived WNV ($P>0·05$). The proportion of mutations that resulted in amino acid substitutions, *pN*, tended to be lower than the ~70 % that would be expected if mutations occurred at random (Holmes, 2003) and was similar ($P=0·4261$, *t*-test for unpaired samples, equal variances) in intra- and interhost populations. *pN* was slightly, but not significantly statistically, higher ($P>0·05$) in mosquito-derived WNV populations than in WNV from birds. d_N/d_S ratios were uniformly $\leqslant 1·000$ and were lower in interhost WNV populations than in intrahost populations ($P=0·0177$, *t*-test for unpaired samples, unequal variances). d_N/d_S was higher in WNV from mosquitoes, but the difference was not statistically significant.

Table 5. Mutations shared by non-consensus clones in multiple specimens

Sequence	Nucleotide position				
	1599	1728	2280	2538	3138
03001810 consensus	C	A	U	U	C
03001810 clone 24	C	A	A	C	U
03001810 clone 25	C	A	A	C	U
03001798 consensus	U	G	U	U	U
03001798 clone 23	C	A	A	C	U

Table 6. Inter- and intrahost variation in WNV

Alignments without mutations were set as undefined (UND).

Host	WNV population	π*	pN†	d_N/d_S‡
Intrahost				
Avian	03001492	0·00041	0·125	0·045
	03001502	0·00026	0·600	0·471
	03001707	0·00026	0·400	0·209
	03001722	0·00005	0·000	0·000
	03001723	0·00010	0·000	0·000
	03001798	0·00026	0·200	0·078
	03001802	0·00021	0·500	0·314
	03001807	0·00000	UND	UND
	03001810	0·00035	0·250	0·000
	03001855	0·00015	0·500	0·314
Avian mean		0·00021	0·286	0·159
Mosquito	32030356	0·00000	UND	UND
	34030291	0·00005	0·000	0·000
	35030311	0·00010	1·000	1·000
	35030318	0·00118	0·438	0·242
	35030321	0·00033	0·000	0·000
	36030191	0·00070	0·500	0·313
	36030227	0·00010	1·000	1·000
	37030094	0·00025	0·000	0·000
	39030148	0·00031	0·333	0·157
	34030297	0·00041	0·500	0·296
Mosquito mean		0·00034	0·419	0·334
Intrahost mean		0·00027	0·353	0·247
Interhost				
	NY-Suffolk-03	0·00232	0·111	0·038
	NY-99-03	0·00241	0·220	0·064
Interhost mean		0·00237	0·166	0·051

*Mean pairwise distance among the sequences from each specimen.
†Proportion of mutations in each alignment that were non-synonymous.
‡The number of synonymous and non-synonymous mutations divided by the number of synonymous and non-synonymous sites per specimen (d_N/d_S).

DISCUSSION

Intrahost genetic diversity has been amply demonstrated in infections with several RNA viruses (Domingo *et al.*, 1992; Plyusnin *et al.*, 1996; Bonneau *et al.*, 2001; Alves *et al.*, 2002; Farci *et al.*, 2002; Jones *et al.*, 2002). It is not clear, however, whether RNA virus populations in nature are true quasispecies (Jenkins *et al.*, 2001; Holmes & Moya, 2002). An observational approach was used to evaluate the hypothesis that WNV exists in its enzootic cycle as a quasispecies by determining (a) whether individual WNV infections are genetically heterogeneous and (b) whether this genetic heterogeneity may be continuous within the virus population as a whole, in addition to existing within individual infections.

Experimental error was monitored in this study by using a control WNV population that is expected to be highly genetically homogeneous. Experimental errors were introduced at a low but measurable rate that is consistent with the error rate for *Pfu*Ultra provided by the manufacturer, $4{\cdot}3 \times 10^{-7}$ misincorporations per base copied. All three mutations detected in the control studies were C to A transversions, which were detected in our study samples only infrequently (data not shown) and may have been introduced into control sequences during the *in vitro* transcription process required for the generation of infectious RNA transcripts from the WNV cDNA clone. Estimates for nucleotide diversity in the control WNV population were approximately fourfold lower than in mosquitoes and birds, suggesting that the observed quasispecies diversity was not an experimental artefact. Analysis of predicted amino acids, however, revealed a more significant impact; all of the nucleotide substitutions in the control WNV resulted in amino acid substitutions, leading to artificially high estimates of protein diversity. Monitoring of experimental error through the use of a control WNV population allowed us to conclude that primary nucleotide sequence data are reliable, but failed to confirm the reliability of data on predicted amino acids.

Approximately 0·016 % of bases sequenced, and 19·5 % of clones, differed from the consensus. These estimates of nucleotide diversity are approximately tenfold lower than previous reports of intrahost genetic diversity for *Dengue virus* (DENV) (Wang *et al.*, 2002; Lin *et al.*, 2004) and WNV (Beasley *et al.*, 2003; Davis *et al.*, 2003). Several factors may explain this lack of concordance. *Taq* polymerase, which introduces errors at a higher rate than proofreading polymerases such as *Pfu* (Malet *et al.*, 2003), was used in the PCR step of several published studies, potentially inflating estimates of nucleotide and amino acid diversity. Importantly, our estimates of intrahost genetic diversity are consistent with studies of RNA viruses conducted by using a similar methodology (Schneider & Roossinck, 2000, 2001; Bonneau *et al.*, 2001). Virological and ecological differences between DENV and WNV may also partially explain the differences between the findings reported here and those for DENV. Although DENV and WNV are both flaviviruses, they belong to different antigenic groups and may have diverged sufficiently to differ in the basic fidelity of their respective RDRPs. Furthermore, DENV and WNV perpetuate in divergent natural transmission cycles: DENV perpetuates between *Aedes* species mosquitoes and human hosts, whereas WNV perpetuates in nature between birds and *Culex* species mosquitoes. The implications of divergent transmission cycles on intrahost genetic diversity are not clear at present. However, their impact on the evolution of arbovirus consensus sequences is well described (Weaver *et al.*, 1994; Zanotto *et al.*, 1996; Ebel *et al.*, 2001b).

WNV was more diverse in mosquitoes than in birds. A greater measure of genetic diversity in mosquito compared with bird specimens could be the result of multiple infected

mosquitoes within a single pool. To evaluate this possibility, the MIR was determined for *C. pipiens/restuans* at each collection site. The MIRs reported, and the number of individual mosquitoes in each infected pool, are generally low. Combined, these site- and time-specific MIRs and small pools support the assumption that only one WNV-infected mosquito was analysed in each pool. Greater genetic diversity in mosquito specimens than in bird specimens could also be explained by our use of a single avian tissue as a source of WNV from birds and whole-body homogenates as a source of WNV from mosquitoes. The hypothetical impact of tissue-specific selection and compartmentalization of particular variants within the quasispecies distribution was minimized by limiting our analysis to corvids, which are susceptible to particularly high viraemias (Komar *et al.*, 2003). At the time of death, experimentally inoculated American crows have viraemias up to 10^7 p.f.u. ml^{-1} (K. A. Bernard, unpublished data). A large proportion of virus detected in the kidney is thus attributable to viraemia and is reflective of circulating virus produced in tissues throughout the bird's body. As WNV infections in mosquitoes appear to be chronic, whereas infections in birds tend to be acute and either (i) resolve upon the initiation of an immune response or (ii) terminate fatally, the increase in genetic diversity in mosquitoes could also be explained by the longer duration of WNV infection in these hosts, even though the underlying mutation rate may be approximately constant. WNV infections in birds may produce extremely high titres compared to WNV infections in mosquitoes. Therefore, although WNV is proportionally more diverse in mosquitoes, a similarly large pool of variants may be present in birds, particularly corvids. The finding of higher quasispecies diversity in mosquitoes than in birds contrasts with a recent study of DENV in mosquitoes and patients by Lin *et al.* (2004), who found that DENV sequences in human sera were more diverse than those obtained from a mosquito pool. As discussed above, the lack of concordance in results may be related to methodological, virological and ecological factors that require further study. These data show, however, that WNV usually exists as a genetically diverse population within hosts and that it is more diverse within mosquitoes than birds. Mosquitoes may therefore provide a source for WNV genetic variation in nature.

In addition to nucleotide substitutions, insertions, deletions and non-translated mutations (substitutions that occurred 3′ of an insertion or deletion) were observed in 2·25 % (nine of 400) of the genomes sampled. Insertions and deletions invariably disrupted open reading frames and introduced premature stop codons, resulting in non-viable genomes that are highly likely to be removed rapidly by natural selection and contribute little to the WNV mutant spectrum.

Evidence that WNV may be transmitted between hosts as a genetically diverse population was sought by sampling specimens for analysis from a single transmission focus on Long Island, NY, USA, during late summer 2003. Three clones in two avian samples shared non-consensus mutations. Convergent evolution in the same host type (American crows) might have produced the observed pattern of genetic similarity. The lack of amino acid changes associated with the mutations, however, renders this scenario unlikely. Rather, the several mutations that genetically link the three clones suggest strongly that they are identical by descent and not the product of rapid convergent evolution. The sampling methods used in this study allow for only the most common viral variants within a single infection to be sampled reliably – at most, approximately 0·2 % (20 of at least 10 000) of genomes present in each infection was sampled. It is therefore not surprising that a relatively small proportion of the clones in this study had genetic signatures suggesting that minority components of the WNV mutant spectrum might occur in multiple specimens. Conversely, it is surprising that these minority genomic variants were detected at all. Experiments with *Bluetongue virus* have documented transmission of minor variants between vertebrates and *Culicoides sonorensis* (Bonneau *et al.*, 2001). To our knowledge, however, this is the first published report providing evidence that an arbovirus may be transmitted as a genetically diverse population in nature.

Data on intrahost genetic diversity in the 20 specimens analysed in this study permitted examination of the selective pressures acting on WNV within and between hosts. The mean genetic distance in each of the intrahost alignments was quite low and was approximately tenfold lower than the mean genetic distance of each interhost WNV population considered. WNV sequences are therefore more diverse between hosts than within an individual host. The proportion of non-synonymous mutations (*pN*) and d_N/d_S ratios tended to be low in both intra- and interhost populations, suggesting the action of purifying selection in both groups of sequences. Combined, the low values of π and the relative rarity of non-synonymous variation in all WNV alignments examined support the observation of tight constraints on arbovirus sequence variation (Weaver *et al.*, 1992; Zanotto *et al.*, 1995; Twiddy *et al.*, 2002; Holmes, 2003). The slight elevation in values of π, *pN* and d_N/d_S obtained from mosquito infections suggests that constraints on WNV sequence variation may be looser in mosquitoes than in birds. *In vivo* experimental studies currently in progress will address this issue directly.

ACKNOWLEDGEMENTS

This work was supported by NIH grant AI055609. The authors gratefully acknowledge the Wadsworth Center Molecular Genetics Core Facility for sequencing, the Wadsworth Center Tissue Culture Core for preparation and maintenance of Vero cell cultures, the New York State Division of Epidemiology for coordinating collection of the specimens included in this study, Ward Stone and the New York State Wildlife Pathology Unit for performing avian necropsies, the Suffolk County Department of Health for collecting mosquito and bird specimens, the New York State Department of Health for establishing and funding arbovirus surveillance, Pei-Yong Shi for

donating clone-derived WNV as method controls, David Young, Susan Jones and Mary Franke for providing technical assistance, Linda Styer for providing thoughtful comments on the manuscript and Sean Philpott for helpful discussions.

REFERENCES

Alves, K., Canzian, M. & Delwart, E. L. (2002). HIV type 1 envelope quasispecies in the thymus and lymph nodes of AIDS patients. *AIDS Res Hum Retroviruses* **18**, 161–165.

Anderson, J. F., Vossbrinck, C. R., Andreadis, T. G., Iton, A., Beckwith, W. H., III & Mayo, D. R. (2001). A phylogenetic approach to following West Nile virus in Connecticut. *Proc Natl Acad Sci U S A* **98**, 12885–12889.

Austin, R. J., Whiting, T. L., Anderson, R. A. & Drebot, M. A. (2004). An outbreak of West Nile virus-associated disease in domestic geese (*Anser anser domesticus*) upon initial introduction to a geographic region, with evidence of bird to bird transmission. *Can Vet J* **45**, 117–123.

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

Bernard, K. A., Maffei, J. G., Jones, S. A. & 13 other authors (2001). West Nile virus infection in birds and mosquitoes, New York State, 2000. *Emerg Infect Dis* **7**, 679–685.

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. *Emerg Infect Dis* **9**, 853–856.

Bonneau, K. R., Mullens, B. A. & MacLachlan, N. J. (2001). Occurrence of genetic drift and founder effect during quasispecies evolution of the VP2 and NS3/NS3A genes of bluetongue virus upon passage between sheep, cattle, and *Culicoides sonorensis*. *J Virol* **75**, 8298–8305.

CDC (1999). Update: West Nile virus encephalitis – New York, 1999. *Morb Mortal Wkly Rep* **48**, 944–955.

Crabtree, M. B., Savage, H. M. & Miller, B. R. (1995). Development of a species-diagnostic polymerase chain reaction assay for the identification of *Culex* vectors of St. Louis encephalitis virus based on interspecies sequence variation in ribosomal DNA spacers. *Am J Trop Med Hyg* **53**, 105–109.

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

Domingo, E. (1998). Quasispecies and the implications for virus persistence and escape. *Clin Diagn Virol* **10**, 97–101.

Domingo, E., Escarmis, C., Martinez, M. A., Martinez-Salas, E. & Mateu, M. G. (1992). Foot-and-mouth disease virus populations are quasispecies. *Curr Top Microbiol Immunol* **176**, 33–47.

Domingo, E., Baranowski, E., Ruiz-Jarabo, C. M., Martín-Hernández, A. M., Sáiz, J. C. & Escarmís, C. (1998). Quasispecies structure and persistence of RNA viruses. *Emerg Infect Dis* **4**, 521–527.

Domingo, E., Ruiz-Jarabo, C. M., Sierra, S., Arias, A., Pariente, N., Baranowski, E. & Escarmís, C. (2002). Emergence and selection of RNA virus variants: memory and extinction. *Virus Res* **82**, 39–44.

Dupuis, A. P., II, Marra, P. P. & Kramer, L. D. (2003). Serologic evidence of West Nile virus transmission, Jamaica, West Indies. *Emerg Infect Dis* **9**, 860–863.

Ebel, G. D., Dupuis, A. P., II, Ngo, K. & 8 other authors (2001a). Partial genetic characterization of West Nile virus strains, New York State, 2000. *Emerg Infect Dis* **7**, 650–653.

Ebel, G. D., Spielman, A. & Telford, S. R., III (2001b). Phylogeny of North American Powassan virus. *J Gen Virol* **82**, 1657–1665.

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**, 493–500.

Eigen, M. (1993). Viral quasispecies. *Sci Am* **269**, 42–49.

Eigen, M. & Biebricher, D. K. (1988). Sequence space and quasispecies distribution. In *RNA Genetics*, vol. 3, pp. 211–245. Edited by E. Domingo, J. J. Holland & P. Ahlquist. Boca Raton, FL: CRC Press.

Essajee, S. M., Pollack, H., Rochford, G., Oransky, I., Krasinski, K. & Borkowsky, W. (2000). Early changes in quasispecies repertoire in HIV-infected infants: correlation with disease progression. *AIDS Res Hum Retroviruses* **16**, 1949–1957.

Farci, P., Shimoda, A., Coiana, A. & 9 other authors (2000). The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288**, 339–344.

Farci, P., Strazzera, R., Alter, H. J. & 10 other authors (2002). Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome. *Proc Natl Acad Sci U S A* **99**, 3081–3086.

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

Holmes, E. C. (2003). Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. *J Virol* **77**, 11296–11298.

Holmes, E. C. & Moya, A. (2002). Is the quasispecies concept relevant to RNA viruses? *J Virol* **76**, 460–462.

Jenkins, G. M., Worobey, M., Woelk, C. H. & Holmes, E. C. (2001). Evidence for the non-quasispecies evolution of RNA viruses. *Mol Biol Evol* **18**, 987–994.

Jones, L. R., Zandomeni, R. & Weber, E. L. (2002). Quasispecies in the 5′ untranslated genomic region of bovine viral diarrhoea virus from a single individual. *J Gen Virol* **83**, 2161–2168.

Kauffman, E. B., Jones, S. A., Dupuis, A. P., II, Ngo, K. A., Bernard, K. A. & Kramer, L. D. (2003). Virus detection protocols for West Nile virus in vertebrate and mosquito specimens. *J Clin Microbiol* **41**, 3661–3667.

Komar, N., Langevin, S., Hinten, S., Nemeth, N., Edwards, E., Hettler, D., Davis, B., Bowen, R. & Bunning, M. (2003). Experimental infection of North American birds with the New York 1999 strain of West Nile virus. *Emerg Infect Dis* **9**, 311–322.

Lanciotti, R. S., Roehrig, J. T., Deubel, V. & 21 other authors (1999). Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science* **286**, 2333–2337.

Lanciotti, R. S., Ebel, G. D., Deubel, V. & 9 other authors (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.

Lin, S.-R., Hsieh, S.-C., Yueh, Y.-Y., Lin, T.-H., Chao, D.-Y., Chen, W.-J., King, C.-C. & Wang, W.-K. (2004). Study of sequence variation of dengue type 3 virus in naturally infected mosquitoes and human hosts: implications for transmission and evolution. *J Virol* **78**, 12717–12721.

Malet, I., Belnard, M., Agut, H. & Cahour, A. (2003). From RNA to quasispecies: a DNA polymerase with proofreading activity is highly recommended for accurate assessment of viral diversity. *J Virol Methods* **109**, 161–170.

Martínez, M. A., Carrillo, C., González-Candelas, F., Moya, A., Domingo, E. & Sobrino, F. (1991). Fitness alteration of foot-and-mouth disease virus mutants: measurement of adaptability of viral quasispecies. *J Virol* **65**, 3954–3957.

Plyusnin, A., Cheng, Y., Lehväslaiho, H. & Vaheri, A. (1996). Quasispecies in wild-type tula hantavirus populations. *J Virol* **70**, 9060–9063.

Rozas, J. & Rozas, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175.

Ruiz-Jarabo, C. M., Arias, A., Baranowski, E., Escarmís, C. & Domingo, E. (2000). Memory in viral quasispecies. *J Virol* **74**, 3543–3547.

Ruiz-Jarabo, C. M., Arias, A., Molina-París, C., Briones, C., Baranowski, E., Escarmis, C. & Domingo, E. (2002). Duration and fitness dependence of quasispecies memory. *J Mol Biol* **315**, 285–296.

Schneider, W. L. & Roossinck, M. J. (2000). Evolutionarily related Sindbis-like plant viruses maintain different levels of population diversity in a common host. *J Virol* **74**, 3130–3134.

Schneider, W. L. & Roossinck, M. J. (2001). Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. *J Virol* **75**, 6566–6571.

Shi, P.-Y., Kauffman, E. B., Ren, P. & 10 other authors (2001). High-throughput detection of West Nile virus RNA. *J Clin Microbiol* **39**, 1264–1271.

Shi, P.-Y., Tilgner, M., Lo, M. K., Kent, K. A. & Bernard, K. A. (2002). Infectious cDNA clone of the epidemic West Nile virus from New York City. *J Virol* **76**, 5847–5856.

Smith, D. B., McAllister, J., Casino, C. & Simmonds, P. (1997). Virus 'quasispecies': making a mountain out of a molehill? *J Gen Virol* **78**, 1511–1519.

Twiddy, S. S., Farrar, J. J., Vinh Chau, N., Wills, B., Gould, E. A., Gritsun, T., Lloyd, G. & Holmes, E. C. (2002). Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. *Virology* **298**, 63–72.

Wang, W.-K., Lin, S.-R., Lee, C.-M., King, C.-C. & Chang, S.-C. (2002). Dengue type 3 virus in plasma is a population of closely related genomes: quasispecies. *J Virol* **76**, 4662–4665.

Weaver, S. C., Rico-Hesse, R. & Scott, T. W. (1992). Genetic diversity and slow rates of evolution in New World alphaviruses. *Curr Top Microbiol Immunol* **176**, 99–117.

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

Zanotto, P. M. de A., Gao, G. F., Gritsun, T., Marin, M. S., Jiang, W. R., Venugopal, K., Reid, H. W. & Gould, E. A. (1995). An arbovirus cline across the northern hemisphere. *Virology* **210**, 152–159.

Zanotto, P. M. de A., Gould, E. A., Gao, G. F., Harvey, P. H. & Holmes, E. C. (1996). Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc Natl Acad Sci U S A* **93**, 548–553.