**Multi-year evolutionary dynamics of West Nile virus in suburban Chicago, USA, 2005-2007**

(Article begins on next page)

22 September 2024

# UNIVERSITÀ DEGLI STUDI DI TORINO

**Multi-year evolutionary dynamics of West Nile virus in suburban Chicago, USA, 2005-2007**

Giusi Amore[1], Luigi Bertolotti[2], Gabriel L. Hamer[3,4], Uriel D. Kitron[5], Edward D. Walker[6], Marilyn O. Ruiz[7], Jeffrey D. Brawn[8], Tony L. Goldberg[3,9*]

[1] European Food Safety Authority, 43100 Parma, Italy

[2] Molecular Biotechnology Center, University of Turin, 10126 Torino, Italy

[3] Department of Pathobiological Sciences, University of Wisconsin, Madison, WI  53706

[4] Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824

[5] Department of Environmental Studies, Emory University, Atlanta, GA 30322

[6] Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824

[7] Department of Pathobiology, University of Illinois, Urbana, IL 61802

[8] Department of Natural Resources and Environmental Sciences, University of Illinois, Urbana, IL 61801

[9] Nelson Institute for Environmental Studies, University of Wisconsin, Madison, WI  53706

* Author for correspondence (tgoldberg@vetmed.wisc.edu)

**ABSTRACT**

West Nile virus has evolved in concert with its expansion across North America, but little is known about the evolutionary dynamics of the virus on local scales. We analyzed viral nucleotide sequences from mosquitoes collected in 2005, 2006, and 2007 from a known transmission "hot spot" in suburban Chicago, USA. Within this small area, the viral envelope gene has increased approximately 0.1% per year in nucleotide-level genetic diversity. In each year, viral diversity was higher in "residential" sites characterized by dense housing than in "urban green space" sites such as cemeteries and parks. Phylodynamic analyses show an increase in viral effective population size around 2005, consistent with a higher-than-average peak in mosquito and human infection rates that year. Analyses of times to most recent common ancestor suggest that WNV in 2005 and 2006 may have arisen predominantly from viruses present during 2004 and 2005, respectively, but that WNV in 2007 had an older common ancestor, perhaps indicating a predominantly mixed or exogenous origin. These results show that the population of WNV in suburban Chicago is an admixture of viruses that are both locally derived and introduced from elsewhere, containing evolutionary information aggregated across a breadth of spatial and temporal scales.

# 1. INTRODUCTION

West Nile virus (WNV) is an emerging mosquito-borne virus (family *Flaviviridae*, genus *Flavivirus*) maintained through an enzootic cycle of transmission involving birds and mosquitoes of the genus *Culex* (Weaver & Reisen 2009). The virus can cause febrile illness and neurologic disease in a range of avian and mammalian hosts, including humans. Since its introduction to North America in 1999 in the New York City area (Lanciotti *et al.* 1999), WNV has spread rapidly throughout the United States, Canada (Gancz *et al.* 2004), Mexico (Deardorff *et al.* 2006) and Latin America and the Caribbean as far south as Argentina (Petersen & Hayes 2008). Now endemic in the Americas, WNV is transmitted in different geographic regions by local species of *Culex*. In the Northeastern and Midwestern USA, where transmission is primarily by *Culex pipiens*, WNV undergoes predictable seasonal "amplification," involving accelerated bird-mosquito transmission in late summer and "spillover" to "dead-end" hosts such as humans and horses (Theophilides *et al.* 2006).

WNV has evolved in concert with its geographic spread. Notably, a fitter "WN02" variant replaced the original "NY99" variant around 2002 (Snapinn *et al.* 2007), likely due to the WN02 variant's increased efficiency of replication in *Culex* vectors (Moudy *et al.* 2007), especially at elevated temperatures (Kilpatrick *et al.* 2008). Phylodynamic analyses of viral nucleotide sequences point to a logistic dynamic of epidemic growth for WNV across North America, reflecting the WN02 variant's initially rapid spread followed by a subsequent epidemic slowdown (Snapinn *et al.* 2007). WNV has as yet shown no evidence of phylogeographic subdivisioning within its North American range but rather appears to circulate panmictically, likely because of continuous long-distance transport by migrating birds and/or mosquitoes (Bertolotti *et al.* 2007; Venkatesan & Rasgon 2010).

Although the evolution and population dynamics of WNV are well understood on continent-wide scales, spatiotemporally focused studies of WNV evolution have been rare. For example, how is viral diversity is distributed across fine spatial scales, or how do local

dynamics of seasonal WNV transmission characteristic of northern latitudes affect viral evolution? Previously, we analyzed genetic diversity and phylogenetics of WNV within in a "hotspot" of transmission in suburban Chicago, USA (Bertolotti *et al.* 2008). In this previous study, we demonstrated that viral genetic diversity was higher in vectors than in avian hosts, that viral diversity was higher in "residential" areas than in nearby urban "green spaces," and that short-term viral movement was apparently limited to relatively short distances of approximately 4 km. However, our previous study was cross-sectional, in that all viral sequences were from 2005, precluding examination of longitudinal trends in viral diversity and evolution.

Here, we analyze new viral nucleotide sequence data from the same suburban Chicago transmission "hot spot" from two subsequent years, 2006 and 2007, yielding three years of continuous data on viral genetic diversity and evolution within this important urban focus of transmission. By comparing our results to those obtained across coarser geographic scales, we provide a new perspective on the fine-scale molecular variation of WNV, including the extent to which WNV evolution in suburban Chicago reflects localized versus broader ecological processes.

## 2. METHODS

*Study area and sample collection*

The study was carried out in southwest suburban Chicago, IL (87° 44' W, 41° 42' N), a mosaic environment of residential and commercial/industrial areas interspersed with urban "green spaces" such as park and cemeteries. This area is a known "hot spot" of arboviral transmission, as evidenced by a statistically significant cluster of human WNV cases in 2002, when the virus first reached the region, and an overlapping cluster in 1975 of human cases of St. Louis encephalitis, caused by a related *Flavivirus* (Ruiz *et al.* 2004) and annually intensive episodes of transmission in the summer (Hamer *et al.* 2008). Mosquitoes were

sampled from 11 "residential" sites and four "natural" sites (urban green spaces), chosen to represent a range of land cover types within our study area, as described previously (Bertolotti *et al.* 2008; Hamer *et al.* 2008).

Mosquitoes were collected from each of the 15 study sites in 2005, 2006 and in 2007 once every two weeks from mid-May trough mid-October. We used $CO_2$-baited CDC miniature light traps, CDC gravid traps baited with rabbit pellet infusion (Lampman & Novak 1996), and battery-powered backpack aspirators (Meyer *et al.* 1983). Mosquitoes were identified to species (Andreadis *et al.* 2005) and pooled into groups of 25 or fewer, grouped by species, collection site, and date. We processed mosquitoes in the laboratory and detected WNV RNA using real-time RT-PCR as previously described (Hamer *et al.* 2008).

*Molecular analyses*

We generated full-length sequences of the WNV envelope (ENV) gene, the most variable in the WNV genome and the most commonly sequenced for phylogenetic studies. We conducted reverse transcription PCR on total RNA extracted from positive mosquito pools, and we sequenced amplicons directly following previously published protocols (Bertolotti *et al.* 2007). For mosquito pools with multiple positives, which were identified by multiple peaks in chromatograms, we cloned amplicons into plasmid vectors for sequencing and resolution of individual haplotypes. Specifically, we amplified a long fragment of 1795bp using forward primer WNV 871 (5'-CTGGTGGCAGCCGTCATTGGTTGG-3') and reverse primer WNV 2666c (5'-AAATGTGGGAAGCAGTGAAGGACG-3') (primer set and PCR conditions described in Bertolotti *et al.*, 2007). The cycling profile consisted of a denaturation step at 94 °C, followed by 35 cycles of PCR (94 °C for 30 seconds, annealing temperature of 55 °C for 30 seconds, extension temperature of 72 °C for 1.30 minutes), a terminal extension at 72 °C for 7 min, and an indefinite soak at 4 °C. We then cloned amplicons using the EPICENTRE - CopyControl[TM] cDNA, Gene & PCR Cloning Kit with

Chemically Competent TransforMax™ EPI300™ E. coli (EPICENTRE® biotechnologies, Madison, WI), following the manufacturer's protocol. Clones were sequenced following previously published protocols (Bertolotti *et al.* 2007). Multiple haplotypes and all ambiguous bases were resolved by repeat sequencing, and all sequences were hand-edited and aligned with reference to published sequences prior to analysis.

*Evolutionary inferences*

We quantified viral genetic diversity at the nucleotide level as Nei's nucleotide diversity (Nei 1987) using the computer programs MEGA4 (Tamura *et al.* 2007) and PAUP* 4.0b10 (Swofford 2003). We tested the statistical significance of differences in nucleotide diversity for different viral populations using 10,000 bootstrap resamplings of genetic distance matrices. We explored temporal trends in nucleotide diversity over the three-year sampling period using linear regression in the software R (R Development Core Team 2009). We tested the statistical significance of temporal trends using 1,000 bootstrap re-samplings of the data.

To describe evolutionary relationships among WNV sequences from our study area, we constructed phylogenetic trees, selecting models of molecular evolution using a hierarchical likelihood ratio test approach and the Akaike information criterion (Akaike 1973) implemented with the computer program Modeltest ver. 3.7 (Posada & Crandall 1998), combined with Bayesian methods in the computer program MrBayes ver. 3.1.2 (Ronquist & Huelsenbeck 2003). We calculated tree statistics using PAUP* 4.0b10 (Swofford 2003).

To describe the demographic history of WNV in suburban Chicago, we applied Bayesian coalescent methods separately for each year (2005, 2006, 2007) and for all the three years together (2005-2007) using the computer programs BEAST ver. 1.4.5 (Drummond & Rambaut 2007) and Tracer ver. 1.4.1 (Rambaut & Drummond 2007). Specifically, we tested four parametric demographic models (constant population size, exponential growth, logistic

growth and expansion population growth) and one non-parametric demographic model (Bayesian Skyline Plot, BSP) under both strict and relaxed (uncorrelated exponential and lognormal) clock models, all conducted under the HKY85 molecular substitution model (Hasegawa *et al.* 1985).  We compared models by calculating the Bayes Factor (BF), which is the ratio of the marginal likelihoods (marginal with respect to the prior) of the two models under consideration (Kass & Raftery 1995).  Marginal likelihoods were estimated using the Newton & Raftery method (Newton & Raftery 1994) with the modifications proposed by (Suchard *et al.* 2001) and were compared as twice the natural logarithm of the Bayes factor, 2ln(BF) (Kass & Raftery 1995).  Evidence against the null model (*i.e.* that the model under consideration had a lower marginal likelihood) was assessed as follows: 2ln(BF) = 2 to 6 indicates evidence against the null model; 2ln(BF) = 6 to 10 indicates strong evidence against the null model; 2ln(BF) > 10 indicates very strong evidence against the null model.

To evaluate the best-fit model describing WNV phylodynamics over the three year period, we first carried out preliminary Bayesian coalescent analyses on sixty randomly selected WNV sequences from mosquitoes collected in our study area during the 2005, 2006 and 2007 transmission seasons separately (twenty sequences per year).  The best-fit model derived from these analyses was then applied to the full set of WNV sequences from the entire three-year period.   The same best-fit models were also used to estimate viral evolutionary parameters (for each year separately and all three years together) and to evaluate WNV population dynamics.  Evolutionary parameters were estimated using the Bayesian Markov chain Monte Carlo (MCMC) method implemented in the BEAST software package.  Bayesian calculations consisted of several independent 50,000,000-generation MCMC runs for the single-year datasets separately, and four independent 100,000,000-generation MCMC runs for the multi-year dataset.  For all MCMC runs, sampling occurred every 1,000[th] generation.  The independent runs were then combined with LogCombiner version 1.4.7 (Drummond & Rambaut 2007).  Convergence of the MCMC was assessed by

calculating the Effective Sample Size (ESS) of the runs. All parameter estimates showed significant ESS (>200). Uncertainty in parameter estimates was represented as 95% highest-posterior-density (HPD) values. Results were visualized using the Tracer program.

To evaluate the strength and direction of selection in different viral populations, we calculated ratios ($\omega$) of non-synonymous ($K_A$) to synonymous ($K_S$) substitutions per site for each year of sampling using the computer program DNAsp (Rozas *et al.* 2003).

## 3. RESULTS

We analyzed 254 WNV ENV sequences of 1575 nucleotides each from *Culex* mosquito pools, including 128 WNV sequences from 2005, 80 sequences from 2006, and 46 sequences from 2007. These samples came from all 15 study sites and contained 167 unique WNV ENV sequences (68 unique sequences out of 128 in 2005, 70 unique sequences out of 80 in 2006, 29 unique sequences out of 46 in 2007). Phylogenetic analysis yielded a tree similar to those described by (Davis *et al.* 2005) and (Bertolotti *et al.* 2008), in which all newly generated samples clustered within the "WN02" clade (Figure 1). All newly generated sequences have been deposited into GenBank (accession numbers GU386758-GU386883).

Mean nucleotide diversity for WNV sequences, summarized by year and by site, are shown in Table 1. In all years, viral genetic diversity was lower among viruses from mosquitoes captured in natural sites than from mosquitoes captured in residential sites, a trend that we previously reported for 2005 (Bertolotti *et al.* 2008); however, this difference was statistically significant only for WNV sequences generated from mosquito pools collected in 2005 and 2006 but not in 2007, perhaps because of low sample sizes in the third study year (Table 1). Analyses of nucleotide diversity by year showed a significant trend of increasing genetic diversity over time in our study area ($R^2$=0.32, t=36.8, p<0.01; Figure 2). The slope of this relationship, $0.94 \times 10^{-3}$, indicates that the WNV ENV gene in our study area has diversified at a rate of 0.094% per year between 2005 and 2007.

To infer the population dynamics of WNV, we analyzed viral sequences using Bayesian coalescent methods. Because we knew exact dates of collection for all viral samples from the study area, we were able to estimate evolutionary parameters of the WNV ENV gene specifying days as units of time. In our study, the Bayes Factor comparison among parametric and non-parametric demographic models selected the Bayesian Skyline Plot uncorrelated exponential relaxed clock as the best-fit model for each year separately and for all the three years together. Therefore, WNV evolutionary parameters and population dynamics were estimated under the Bayesian Skyline Plot uncorrelated exponential relaxed clock model.

Mean rates of evolution estimated under the best-fit model varied among years, with evolution being most rapid in 2006, albeit with overlapping highest-posterior-density estimates (Table 2). The rate of evolution across all years combined was $4.17 \times 10^{-6}$ nucleotide substitutions per site per day, or $1.52 \times 10^{-3}$ nucleotide substitutions per site per year (95% HPD = $1.22 \times 10^{-3}$ – $1.84 \times 10^{-3}$ substitutions per site per year). The mean ages of the most recent common ancestors for WNV from 2005 and 2006 were 1.5 and 1.1 years, respectively, whereas the mean age of the most recent common ancestor for WNV from 2007 was 5.6 years, again albeit with overlapping highest-posterior-density estimates (Table 2).

Figure 2 shows the non-parametric trajectory of WNV effective population size over time in suburban Chicago estimated using the Bayesian Skyline Plot model. The shape of the curve shows some evidence of increasing effective population size prior to 2005 and continuing until 2006, followed by a decrease in effective population size in 2007. WNV sequences from the Chicago study area collected in the different years (2005, 2006, 2007) showed ratios of non-synonymous substitutions per site that were comparable and well below 1 ($\omega = 0.023 \pm 0.003$ in 2005, $0.032 \pm 0.005$ in 2006, $0.030 \pm 0.007$ in 2007), indicating a pattern of strong stabilizing selection, as has previously been reported for WNV (Bertolotti *et al.* 2007; Bertolotti *et al.* 2008; Jerzak *et al.* 2008).

**4. DISCUSSION**

In this study, we describe patterns of genetic diversity and evolution of West Nile virus circulating in *Culex* mosquitoes in a small geographic area in suburban Chicago (USA) over a three-year period (2005-2007). Our phylogenetic analyses indicate that WNV in this area has remained relatively stable, consisting of viral lineages within the WN02 clade that has spread across North America. The topology of our tree (Fig. 1) shows some resolution of sub-clades within the larger WN02 clade, but no evidence of phylogenetic sorting by year or sampling location. WNV in suburban Chicago therefore appears to be a relatively panmictic viral population that is roughly representative of WNV circulating in North America.

Our analyses of viral genetic diversity by year demonstrate a significant trend of increasing WNV genetic diversity in *Culex* mosquitoes from 2005 to 2007. Previously, we demonstrated a similar trend across the Midwestern USA of approximately 0.07% diversification per year from 2002 to 2005, on a similar scale of overall diversity (Bertolotti *et al.* 2008). Documenting a continuation of this trend on a much finer spatial scale than has previously been reported indicates that WNV in our suburban Chicago "hot spot" is displaying similar population genetic dynamics as WNV across the Midwest. The increasing viral diversity we have documented has probably arisen by neutral evolution, as evidenced by our analyses of non-synonymous to synonymous substitution rates that show evidence of strong stabilizing selection, and by the fact that WNV evolutionary rates in our study area were similar across years.

Our analyses confirm our previous observation from 2005 data that WNV sequences from mosquitoes captured in residential sites are more genetically diverse than are those from mosquitoes captured in natural sites (Bertolotti *et al.* 2008). The direction of this trend was consistent across all years, but did not reach statistical significance in 2007, probably due to a small sample size (2007 was characterized by relatively low transmission, and consequently only eight sequences from natural sites were available from that year). This trend may reflect

altered transmission dynamics in the two types of habitats, perhaps due to fine-scale variation in the distribution of avian hosts or microclimate. Although the ecological reasons for this trend remain obscure, it nevertheless confirms that fine-scale features of the urban landscape can influence the genetic diversity and evolution of arboviral pathogens such as WNV.

Previously, Snapinn *et al.* (2007) performed phylodynamic analyses of WNV sequences from across North America, arguing that WNV on a continent-wide scale displayed a logistic pattern of epidemic growth between 2001 and 2005, interpreted as reflecting an initial rapid expansion of the WN02 genotype followed by saturation of available hosts. In our study, we applied similar methods at a finer spatial scale (the suburban Chicago "hot spot") and temporal scale (days, versus years). As shown in Figure 3, the effective number of WNV infections increased prior to 2005 and until approximately 2006, after which it remained fairly constant before declining moderately in 2007. Importantly, we found no evidence for continued overall population expansion, consistent with the Snapinn *et al.* (2007) results, and no evidence of regular, annual population expansions and contractions, as might have been expected based on the well-known seasonal "amplification" of WNV in our study area and other northern latitude sites (Theophilides *et al.* 2006). WNV dynamics in our study area are therefore apparently "decoupled" from the local epidemic pattern of highly seasonal transmission and amplification. We note, however, that the Chicago region registered a very hot and dry summer in 2005, with temperatures that were the 3rd highest since 1958, and that the resulting combination of high temperatures and low rainfall provided favorable conditions for the reproduction of *Culex* mosquitoes, leading to a higher-than-average peak in mosquito infection rate that year (Hamer *et al.* 2008) and a correspondingly high incidence of human cases (Yoder *et al.* 2006), whereas WNV transmission in 2007 was atypically low. The expansion of WNV effective population size through 2005 and a subsequenct decline in 2007 as shown in Figure 3 is consistent with these regional epidemiological observations.

Using the same Bayesian coalescent approach, we were able to estimate the mean age of WNV (most recent common ancestor) in our study area for each year (Table 3). Our best estimates for the ages of the most recent common ancestors of WNV sampled during the 2005 and 2006 transmission seasons (1.4 and 1.1 years, respectively) suggest that WNV lineages in these years derived predominantly from viruses present the previous year, which implies a dominant evolutionary signal of overwintering of WNV in our study area in these years. However, the age of the most recent common ancestors for WNV sampled in 2007 was 5.6 years, suggesting that our 2007 sample of WNV contained lineages that had diverged much earlier than the previous year. This observation implies a stronger signal of re-introduction of exogenous and divergent WNV into our study area in 2007 than in the two previous years. In this light, we note that the 2007 winter was unusually cold in Chicago, which might have reduced the rate of overwintering and consequently increased the proportion of exogenous viruses sampled during the subsequent 2007 transmission season. Our analyses therefore suggest that WNV in our study area may derive from both local and distant sources and that the relevant proportions of viral sequences originating from these two sources may vary among years in response to winter weather conditions. We note, however, that our high posterior density estimates are wide for all years, such that we cannot at this point definitively differentiate a predominant patter of overwintering from one of reintroduction for any given year. More intensive sampling and more extensive sequencing should help clarify the relative importance of these phenomena.

We also caution that, despite our systematic and prospective sampling effort across three consecutive years, our data nevertheless suffer some limitations due to lack of phylogenetic resolution. Although we were able to resolve some sub-clades within the larger WN02 clade shown in Figure 1, relationships among many of our sequences remained unresolved. Greater phylogenetic resolution might have allowed us to focus our phylodynamic analyses on sub-clades representing clusters of locally transmitted viruses.

Because we could not resolve such sub-clades, however, it is likely that our phylodynamic analyses offer a spatially aggregated picture of WNV population dynamics that includes both locally-evolved WNV lineages and lineages introduced from elsewhere. Indeed, our analyses of times to most recent common ancestor showing TMRCAs of approximately one year for 2005 and 2006 but greater than one year for 2007 tentatively support this notion.

Overall, our results suggest that the population of WNV in suburban Chicago is broadly representative of that on coarser spatial scales, but also that it contains population genetic and phylogenetic signatures of local ecological and epidemiological processes, such as urban landscape type and climate. Moreover, WNV in suburban Chicago appears to show substantial temporal variability, both in its propensity to increase in diversity across years and in its variable evolutionary dynamics from year to year. WNV in our small "hot spot" thus appears to represent an admixture of viruses that are both locally derived and introduced from elsewhere and that reflect transmission dynamics aggregated across a breadth of spatial and temporal scales. We suggest that further studies of WNV on fine spatial and temporal scales in other regions might help clarify the importance of micro-scale processes to the transmission and evolution of this and other emerging arboviral pathogens.

**Acknowledgments**

**REFERENCES**

Akaike, H. 1973 Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrica **60**, 255-265.

Andreadis, T. G., Thomas, M. C. & Shepard, J. J. Identification Guide to the Mosquitoes of Connecticut. New Haven, CT: The Connecticut Agricultural Experiment Station; 2005.

Bertolotti, L., Kitron, U. & Goldberg, T. L. 2007 Diversity and evolution of West Nile virus in Illinois and The United States, 2002-2005. Virology **360**, 143-149.

Bertolotti, L., Kitron, U. D., Walker, E. D., Ruiz, M. O., Brawn, J. D., Loss, S. R., Hamer, G. L. & Goldberg, T. L. 2008 Fine-scale genetic variation and evolution of West Nile Virus in a transmission "hot spot" in suburban Chicago, USA. Virology **374**, 381-389.

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

Deardorff, E., Estrada-Franco, J., Brault, A. C., Navarro-Lopez, R., Campomanes-Cortes, A., Paz-Ramirez, P., Solis-Hernandez, M., Ramey, W. N., Davis, C. T., Beasley, D. W., Tesh, R. B., Barrett, A. D. & Weaver, S. C. 2006 Introductions of West Nile virus strains to Mexico. Emerg Infect Dis **12**, 314-318.

Drummond, A. J. & Rambaut, A. 2007 BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol **7**, 214.

Gancz, A. Y., Barker, I. K., Lindsay, R., Dibernardo, A., McKeever, K. & Hunter, B. 2004 West Nile virus outbreak in North American owls, Ontario, 2002. Emerg Infect Dis **10**, 2135-2142.

Hamer, G. L., Walker, E. D., Brawn, J. D., Loss, S. R., Ruiz, M. O., Goldberg, T. L., Schotthoefer, A. M., Brown, W. M., Wheeler, E. & Kitron, U. D. 2008 Rapid amplification of West Nile virus: the role of hatch-year birds. Vector Borne Zoonotic Dis **8**, 57-67.

Hasegawa, M., Kishino, H. & Yano, T. 1985 Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22, 160-174.

Jerzak, G., Brown, I., Shi, P.-Y., Kramer, L. D. & Ebel, G. D. 2008 Genetic diversity and purifying selection in West Nile virus populations are maintained during host switching. Virology **374**, 256-260,

Kass, R. E. & Raftery, A. E. 1995 Bayes factors. Journal of the American Statistical Association **90**, 773-795.

Kilpatrick, A. M., Meola, M. A., Moudy, R. M. & Kramer, L. D. 2008 Temperature, viral genetics, and the transmission of West Nile virus by *Culex pipiens* mosquitoes. PLoS Pathog **4**, e1000092.

Lampman, R. L. & Novak, R. J. 1996 Oviposition preferences of *Culex pipiens* and *Culex restuans* for infusion-baited traps. J. Am. Mosq. Control Assoc. **12**, 23-32.

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

Meyer, R. P., Reisen, W. K., Hill, B. R. & Martinez, V. M. 1983 The AFS sweeper, a battery-powered backpack mechanical aspirator for collecting adult mosquitoes. Mosquito News **43**, 346-350.

Moudy, R. M., Meola, M. A., Morin, L. L., Ebel, G. D. & Kramer, L. D. 2007 A newly emergent genotype of West Nile virus is transmitted earlier and more efficiently by *Culex* mosquitoes. Am J Trop Med Hyg **77**, 365-370.

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

Newton, M. A. & Raftery, A. E. 1994 Approximate Bayesian inference with the weighted likelihood bootstrap. Journal of the Royal Statistical Society, Series B **56**, 3-48.

Petersen, L. R. & Hayes, E. B. 2008 West Nile virus in the Americas. Med Clin North Am **92**, 1307-1322.

Posada, D. & Crandall, K. A. 1998 MODELTEST: testing the model of DNA substitution. Bioinformatics **14**, 817-818.

R Development Core Team 2009 *R: A language and environment for statistical computing* Vienna, Austria: R Foundation for Statistical Computing.

Rambaut, A. & Drummond, A. J. 2007 Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer,

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

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

Ruiz, M. O., Tedesco, C., McTighe, T. J., Austin, C. & Kitron, U. 2004 Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002. Int J Health Geogr **3**, 8.

Snapinn, K. W., Holmes, E. C., Young, D. S., Bernard, K. A., Kramer, L. D. & Ebel, G. D. 2007 Declining growth rate of West Nile virus in North America. J Virol **81**, 2531-2534.

Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. 2001 Bayesian selection of continuous-time Markov chain evolutionary models Mol Biol Evol **18**, 1001-1013.

Swofford, D. L. 2003 *PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4.* Sunderland, Massachusetts: Sinauer Associates.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. 2007 MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24**, 1596-1599.

Theophilides, C. N., Ahearn, S. C., Binkowski, E. S., Paul, W. S. & Gibbs, K. 2006 First evidence of West Nile virus amplification and relationship to human infections. Int. J. Geogr. Inf. Sci. **20**, 103-115.

Venkatesan, M. & Rasgon, J. L. 2010 Population genetic data suggest a role for mosquito-mediated dispersal of West Nile virus across the western United States. Mol Ecol in press,

Weaver, S. C. & Reisen, W. K. 2009 Present and future arboviral threats. Antiviral Res in press.

Yoder, J., Haramis, L. & Dworkin, M. 2006 West Nile surveillance in Illinois, 2005. Illinois Infectious Disease Report **3**, 1-3.

1  **Table 1.** Nucleotide diversity of West Nile virus in *Culex* mosquitoes collected during the 2005, 2006 and 2007 transmission seasons from the

2  suburban Chicago study area.

| Year | Location | N | Nucleotide Diversity (%) | Standard Error (%) | Difference [a] |
|---|---|---|---|---|---|
| 2005 | Residential sites | 76 | 0.37 | 0.022 | |
| | Natural sites | 52 | 0.26 | 0.050 | p < 0.001 |
| | All sites | 128 | 0.33 | 0.016 | |
| 2006 | Residential sites | 59 | 0.45 | 0.023 | |
| | Natural sites | 21 | 0.42 | 0.037 | p = 0.003 |
| | All sites | 80 | 0.44 | 0.019 | |
| 2007 | Residential sites | 38 | 0.50 | 0.039 | |
| | Natural sites | 8 | 0.46 | 0.063 | p = 0.107 |
| | All sites | 46 | 0.50 | 0.035 | |

3  [a] Statistical difference between nucleotide diversity for viral sequences collected in residential versus natural sites within the same year, estimated

4  from 10,000 bootstrap resamplings of the genetic distance matrices.

5    **Table 2.** Bayesian estimates of evolutionary parameters for WNV in suburban Chicago over time[a].

| Year | Number of samples | Range of dates of sample collection | Substitution rate[b] (95% HPD) | MRCA[c] (95% HPD) |
|---|---|---|---|---|
| 2005 | 128 | 21 July – 4 Oct | $6.87 \times 10^{-6}$ ($1.11 \times 10^{-6}$ – $1.50 \times 10^{-5}$) | 532 (143-1243) |
| 2006 | 80 | 13 June – 7 Sept | $13.11 \times 10^{-6}$ ($4.82 \times 10^{-7}$ – $2.74 \times 10^{-5}$) | 408 (92-1253) |
| 2007 | 46 | 5 July – 9 Sept | $4.48 \times 10^{-6}$ ($6.73 \times 10^{-8}$ – $1.38 \times 10^{-5}$) | 2057 (122-7584) |
| All years | 254 | 21 July 2005 – 9 Sept 2007 | $4.17 \times 10^{-6}$ ($3.34 \times 10^{-6}$ – $5.04 \times 10^{-6}$) | 1206 (984-1531) |

6    [a] All evolutionary parameters were estimated under the Bayesian skyline plot uncorrelated exponential relaxed clock model.

7    [b] Mean number of nucleotide substitutions per site per day. HPD: high posterior density.

8    [c] Mean age of most recent common ancestor (MRCA), in days.

9   **FIGURE LEGENDS**

10

11   **Figure 1.** Phylogenetic tree constructed using Bayesian analysis of 254 WNV envelope gene

12   sequences (length 1,575 nucleotides) from suburban Chicago, plus eight out-group

13   sequences. Taxon names generated from the Chicago study area indicate collection site

14   (described in Bertolotti et al. 2008), identification number (only for WNV sequences from

15   mosquitoes collected in 2005), dates of collection (for sequences from mosquitoes collected

16   in 2006 and 2007), and GenBank accession numbers (in parentheses). Posterior probabilities

17   of clades are indicated above branches.

18

19   **Figure 2.** WNV genetic diversity in *Culex* mosquitoes from suburban Chicago, 2005 - 2007.

20   Genetic diversity is expressed as nucleotide diversity (Nei, 1987), or mean proportion of

21   nucleotide differences among sequences within each year, corrected for multiple

22   substitutions using a model of molecular evolution derived from a hierarchical likelihood

23   ratio test approach implemented using the computer program Modeltest, version 3.7 (Posada

24   and Crandall, 2001). Lines and small error bars indicate means and standard errors,

25   respectively. Boxes indicate third quartiles, and large error bars indicate ranges.

26   The dashed line is a least squares regression line indicating the rate of increase of viral

27   genetic diversity over time ($0.94 \times 10^{-3}$ per year, $R^2=0.32$, $t=36.8$, $p<0.01$). Grey lines are

28   least-squares regression lines derived from 1000 bootstrap replicates of the data; solid lines

29   indicate 95% confidence intervals around the nucleotide diversity estimation. Sample sizes

30   for each year are indicated in parentheses.

31

32   **Figure 3.** Bayesian skyline plot of WNV in suburban Chicago, USA. The black bold dashed

33   line represents the median estimate of effective number of infections through time in

34   suburban Chicago, derived from ENV gene sequences using a Bayesian coalescent approach.

35 Genetic distances were transformed into a timescale of days taking into account the different

36 sampling times of WNV in our study area and enforcing a relaxed molecular clock (see

37 Methods). 95% HPD values shaded. Vertical dashed gray lines indicate the sampling

38 periods (July $21^{st}$ – Oct $4^{th}$ 2005, June $13^{th}$ – Sept $7^{th}$ 2006, July $5^{th}$ – Sept $9^{th}$ 2007) for the

39 WNV sequences analyzed in this study.

40

41    **Figure 1**



CpBi05 (EF632033)
CpEc01 (EF632044)
60
84    IL060717CoEN (GU386770)
IL060724CoHA (GU386780)
73  IL060727CoAL (GU386787)
IL060718CoIHb (GU386772)
IL060722CoON (GU386778)
IL060801 (GU386798)
IL070809CoALG2 (GU386864)
93  CpHs09a (EF631071)
CpHs09b (EF631072)
CpAl01 (EF631024)
CpEn04 (EF631053)
CpHs07 (EF631069)
CpOc11 (EF631093)
CpOn05 (EF631102)
CpWw13 (EF631141)
CpWw22 (EF631150)
95  IL070724CoIH (GU386845)
IL070719CoWWG1 (GU386844)
89  CpAl03 (EF631026)
84  CpEw04 (EF631063)
IL060731CoRN (GU386796)
IL060825CoPS (GU386829)
78  IL060728CoCM (GU386793)
IL070807CoMI (GU386859)
59  CpHs17 (EF631080)
IL060710CoIH (GU386763)
IL060901CoHA (GU386831)
CpOc01 (EF631083)
98  CpOc06 (EF631088)
CpOc13 (EF361095)
CpPn03 (EF631105)
96  CpOc14 (EF631096)
56
IL060823 (GU386826)
52  IL060718CoWS (GU386774)
96  IL070725CoALG2 (GU386846)
IL060822CoPFa (GU386823)
CpAl04 (EF631027)
60  CpAl05 (EF631028)
CpBi04 (EF631032)
100  CpCm03 (EF631043)
CpEn09 (EF631059)
67  CpCa01 (EF631039)
93  CpPn04 (EF631106)
CpPs11 (EF631123)
IL060718CoRN (GU386773)
IL070807CoEWG1 (GU386857)
99  IL070815CoHA (GU386869)
IL070809CoRNb (GU386866)
100  CpBi03 (EF631031)
IL060704CoON (GU386762)
86    100  CpBu04 (EF631038)
CpWw21 (EF631149)
67  CpEn08 (EF631058)
IL060825CoPN (GU386828)
52  CpHs04 (EF631067)
IL070808CoHI (GU386860)
84  CpOn02a (EF631098)
CpOn02b (EF631099)
95  IL060823CoBI (GU386825)
IL070731CoCA (GU386848)
81  IL060803CoHA (GU386801)
IL0708CoALG2 (GU386854)
69  IL060815CoHA (GU386812)
IL060817CoWW (GU386828)
IL060802CoEWa (GU386799)
IL060802CoEWb (GU386800)
85  IL060727CoSCa (GU386789)
IL060727CoSCd (GU386792)
63  IL060815CoPNa (GU386813)
IL060815CoPNb (GU386814)
IL060801CoWS (GU386797)
IL070807CoIH (GU386858)

34 sequences from 2005
42 sequences from 2006
17 sequences from 2007

100

97

66

New York 1999 (AF196835)
Israel 1998 (AF481864)
Tunisia 1997 (AY268132)
Volga 1999 (AY277252)
100  Romania 1996 (AF260969)
Russia 1999 (AY278441)
50
100  Kenya 1998 (AY262283)
Italy 1998 (AF404757)

42

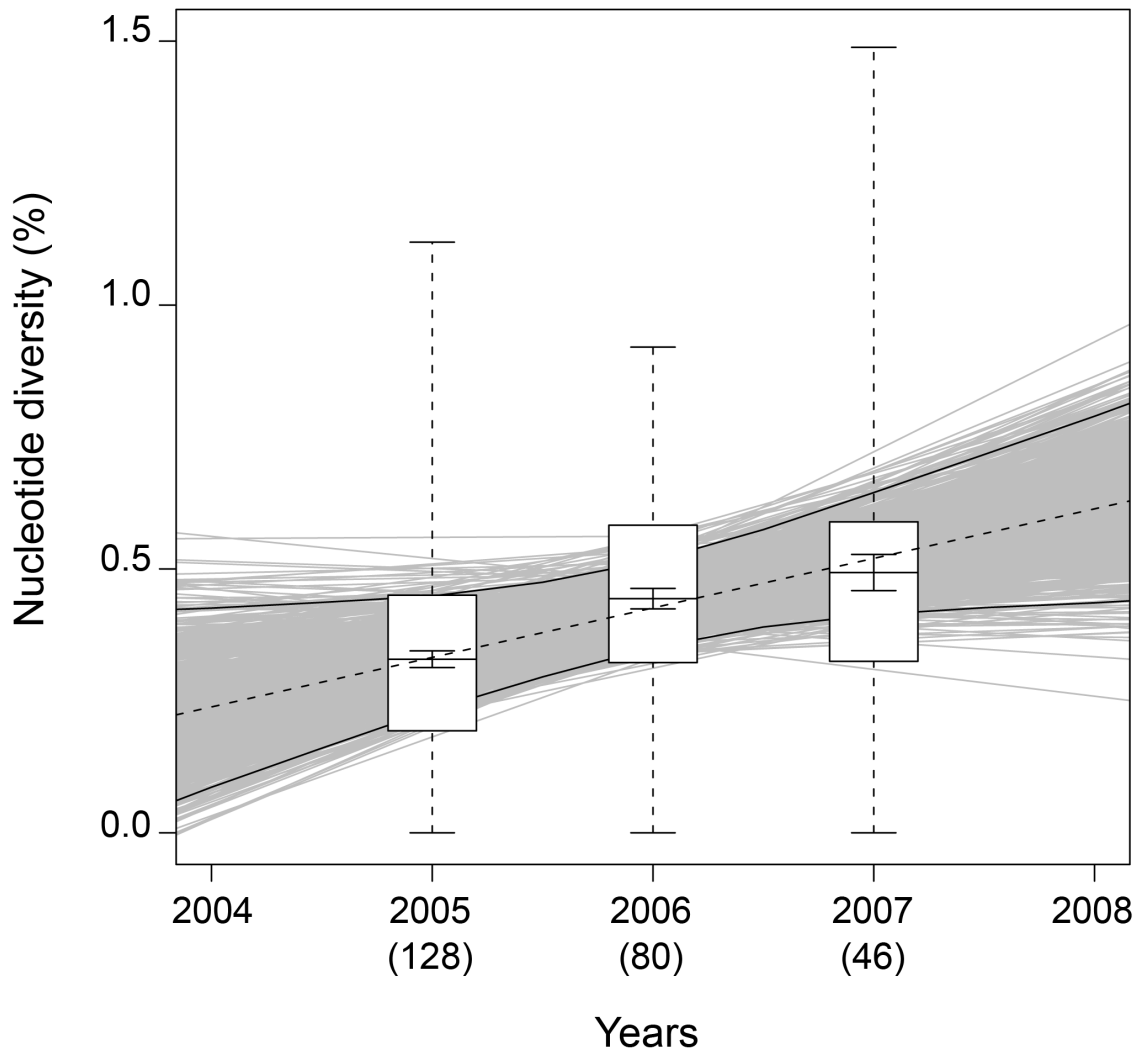0.03 substitutions per site

43

44   **Figure 2.**

47      **Figure 3.** Bayesian skyline plot of WNV in suburban Chicago, USA.



48