Unifying the spatial epidemiology and molecular evolution of emerging epidemics

Unifying the spatial epidemiology and molecular evolution of emerging epidemics

Oliver G. Pybus[a,1,2], Marc A. Suchard[b,c,d,1], Philippe Lemey[e,1], Flavien J. Bernardin[f,g], Andrew Rambaut[h,i], Forrest W. Crawford[b], Rebecca R. Gray[a], Nimalan Arinaminpathy[j], Susan L. Stramer[k], Michael P. Busch[f,g], and Eric L. Delwart[f,g]

[a]Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; Departments of [b]Biomathematics, [c]Biostatistics, and [d]Human Genetics, University of California, Los Angeles, CA 90095; [e]Department of Microbiology and Immunology, Rega Institute, KU Leuven, 3000 Leuven, Belgium; [f]Blood Systems Research Institute, San Francisco, CA 94118; [g]Department of Laboratory Medicine, University of California, San Francisco, CA 94143; [h]Institute for Evolutionary Biology, Edinburgh University, Edinburgh EH9 3JT, United Kingdom; [i]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892-2220; [j]Department of Ecology and Evolution, Princeton University, Princeton, NJ 08544-2016; and [k]Scientific Support Office, American Red Cross, Gaithersburg, MD 20877

Edited by David M. Hillis, University of Texas at Austin, Austin, TX, and approved July 27, 2012 (received for review April 19, 2012)

We introduce a conceptual bridge between the previously unlinked fields of phylogenetics and mathematical spatial ecology, which enables the spatial parameters of an emerging epidemic to be directly estimated from sampled pathogen genome sequences. By using phylogenetic history to correct for spatial autocorrelation, we illustrate how a fundamental spatial variable, the diffusion coefficient, can be estimated using robust nonparametric statistics, and how heterogeneity in dispersal can be readily quantified. We apply this framework to the spread of the West Nile virus across North America, an important recent instance of spatial invasion by an emerging infectious disease. We demonstrate that the dispersal of West Nile virus is greater and far more variable than previously measured, such that its dissemination was critically determined by rare, long-range movements that are unlikely to be discerned during field observations. Our results indicate that, by ignoring this heterogeneity, previous models of the epidemic have substantially overestimated its basic reproductive number. More generally, our approach demonstrates that easily obtainable genetic data can be used to measure the spatial dynamics of natural populations that are otherwise difficult or costly to quantify.

phylogeny | phylogeography | transmission

The explanation of spatial patterns of infectious disease, particularly those of emerging pathogens, has remained a central problem of epidemiology since its inception (1). The existence and nature of traveling waves of infection were first explained in theoretical models (2, 3) and later quantified in empirical studies of rabies and the Black Death (4, 5). These and other studies highlighted the fundamental problem of spatial autocorrelation: observations of infection are statistically dependent due to transmission among proximate individuals, greatly complicating the analysis of spatiotemporal incidence. Consequently, many recent analyses of spatial epidemic behavior use detailed mathematical models of spatial structure to account for autocorrelation (6). Entirely independently, in the field of evolutionary biology there has developed a separate body of work, now termed phylogeography, which focuses on reconstructing past movement events from the genome sequences of sampled organisms (7–10). However, these evolutionary tools typically generate descriptive results that, though informative, remain divorced from epidemiological theory. Crucially neither approach can be considered complete when applied to rapidly evolving viruses, whose spatial, epidemic, and evolutionary dynamics occur on the same timescale (11), necessitating the development of methods that consider all these processes together.

Here we introduce a unique approach that integrates the disciplines of spatial epidemiology and phylogenetics. To illustrate the utility of this approach, we show how, from pathogen genomes alone, it can estimate the diffusion coefficient (D) of an epidemic as well as variation in the process of spatial spread. D is a fundamental ecological measure of the intrinsic diffusivity of infected individuals, reflecting the area that an infected host will explore per unit time (not to be confused with the area covered by the whole epidemic). It is derived from simple reaction–diffusion models of spatial spread and, together with R_0, determines the wavefront velocity of an epidemic invasion (4, 5). Despite its theoretical importance, D is exceptionally difficult to estimate in nature and rarely reported; its estimation usually requires tracking the movements of a large number of infected hosts by mark/recapture or telemetry (5, 12). As well as being time-consuming, this approach will fail to adequately capture spatial dynamics when dispersal behavior is highly variable among individuals. Alternatively, D can be inferred indirectly via its theoretical relationship to an epidemic's observed wavefront velocity (4, 13, 14); however, this requires R_0 and other transmission parameters to be known without error.

We apply our approach to the invasion of North America by the West Nile virus (WNV), an important recent example of viral spatial emergence. WNV is a mosquito-borne RNA virus whose primary host is birds, and was first detected in the United States in New York City in August 1999. The American epidemic resulted from the introduction of a single highly pathogenic lineage (15) and subsequently contributed to the decline of several North American bird species (16). Transmission from mosquitoes to humans has caused >1,200 deaths in the United States (17), although human cases are not thought to contribute to onward infection. Comprehensive records of WNV incidence in the United States demonstrate an apparent westward wave of infection that reached the country's west coast by 2004 (17), representing a mean epidemic wavefront velocity of ∼1,000 km/y during invasion. However, incidence data alone cannot determine whether the invasion resulted primarily from local, short movements of hosts and vectors, or whether east/west spread was interrupted by long-distance bird migration movements to poorly sampled tropical locations (18, 19). Despite a plethora of mathematical models, many of which consider the transmission mechanisms of WNV in great detail (13, 14, 20, 21), models of

Author contributions: O.G.P. designed research; O.G.P., M.A.S., P.L., F.J.B., A.R., F.W.C., R.R.G., N.A., S.L.S., M.P.B., and E.L.D. performed research; M.A.S., P.L., S.L.S., M.P.B., and E.L.D. contributed new reagents/analytic tools; O.G.P., M.A.S., P.L., F.J.B., A.R., F.W.C., R.R.G., and N.A. analyzed data; and O.G.P., M.A.S., and P.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database, www.ncbi.nlm.nih.gov (accession nos. GQ507468–GQ507484).

[1]O.G.P., M.A.S., and P.L. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: oliver.pybus@zoo.ox.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1206598109/-/DCSupplemental.

the epidemic's spatial dynamics have been explored only theoretically (22) or at very local scales (23, 24), and values reported for the basic reproductive number, R_0, of the epidemic vary widely (14, 21, 25). Most phylogenetic studies have revealed little about the epidemic's spatial structure due to the limited diversity of the subgenomic sequences typically used (26).

Linking Phylogeography and Spatial Ecology

This section explains how evolutionary analyses of viral spread can be formally linked with spatial ecology, enabling the estimation of spatial epidemiological variables from genomic data. The approach is based on the application of a simple yet powerful idea: phylogenies reconstructed from spatial epidemics are branching structures that record the correlated histories of transmission among sampled infections (Fig. 1 *A* and *B*), hence the phylogeny of an epidemic can be used to correct for spatial autocorrelation. More specifically, if the dates and locations of all phylogenetic nodes are known or posited, then each phylogeny branch represents a conditionally independent trajectory of viral movement, defined by a start location, end location, and duration (27) (Fig. 1 *A* and *B*). Independence is conditional on the date and location values proposed for each node; any estimation or measurement uncertainty in these can be readily incorporated by marginalization. Consequently, the spatial dynamics of an epidemic can be quantified using simple, nonparametric statistics of these displacements. This approach is analogous to that used by phylogenetic comparative methods, which convert correlated species trait values into independent observations amenable to statistical tests (28).

Although many statistics of spatial dynamics could be calculated using this framework, we introduce the approach by estimating the diffusion coefficient, D, without an explicit model of spatial autocorrelation. Given a set of n movement observations (phylogeny branches) whose durations and start and end locations are specified, D can be estimated using

$$D \approx \frac{1}{n}\sum_{i=1}^{n}\frac{d_i^2}{4t_i}, \qquad [1]$$

where t_i denotes the duration in years of branch i, during which the lineage has moved d_i km away from its start position in two dimensions (5, 12) (Fig. 1 *A* and *B*). This estimator follows the classical relationship between D and mean square displacement (29) and has been previously used to estimate the diffusivity of intentionally released rabid foxes that were subsequently tracked via telemetry (5).

Estimates of the dates and locations of internal phylogenetic nodes (ancestral infections; Fig. 1) can be readily obtained using current phylogeographic and molecular clock techniques (10). In our WNV analysis we infer the longitude and latitude of internal nodes using a 2D anisotropic random walk (*Materials and Methods*). The marginal posterior probability densities of these locations (and of D) can be estimated using standard Bayesian Markov chain Monte Carlo (MCMC) techniques; hence our procedure fully incorporates statistical uncertainty (10). Sequences sampled from the epidemic are assumed to have a single common ancestor (no recombination or introgression). Although there must be sufficient temporal information to reliably estimate the timescale of the phylogeny, the approach does not necessitate the assumption of neutral sequence evolution.

We note two key benefits of this approach: first, it will be applicable to a broad range of situations because the inference of ancestral locations is separated from the estimation of D (or other spatial variable); for each application, the most statistically appropriate model for inferring the former can be chosen. Second, the approach extends readily to more realistic, heterogeneous dispersal processes. Specifically, in this study, we use a flexible relaxed random walk that allows the rate of dispersal to vary among phylogeny branches according to some probability distribution, while constraining it to be constant along each branch (*Materials and Methods*). As a result, we can directly measure heterogeneity in

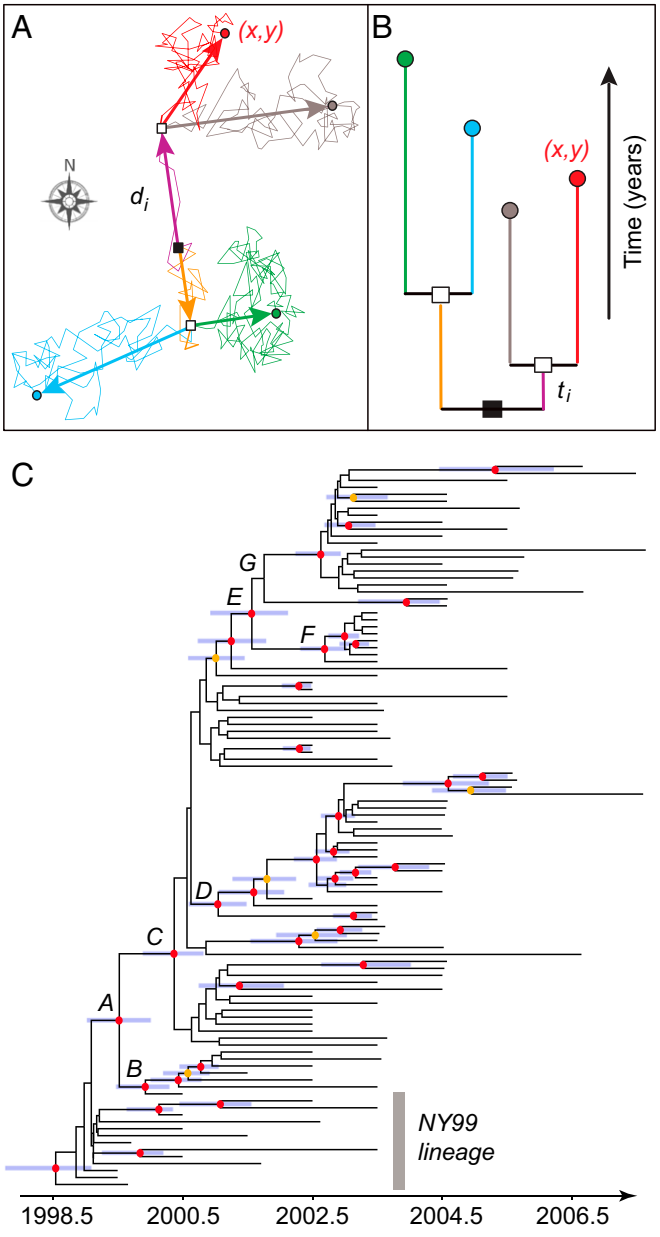

Fig. 1. (*A* and *B*) The link between spatial ecology and phylogenetics. Filled circles represent viral sequences whose locations and dates of sampling are known. Squares represent unsampled ancestral infections whose locations and dates are estimated. The black squares in *A* and *B* denote the epidemic's origin in space and time, respectively. (*A*) Colored arrows indicate the direction and distance d_i of the movement trajectory defined by each lineage. Thin colored lines show the random walk undertaken by each lineage. (*B*) The phylogeny resulting from the spatial infection process in *A*. Colored lines in *B* show the duration t_i of each lineage. Diffusivity can be inferred by combining the information in *A* and *B*. Diffusivity is low for lineages with long and winding paths that do not lead far (e.g., green), and is high for lineages that quickly move large distances (e.g., purple). (*C*) Maximum clade credibility phylogeny of the North American WNV epidemic, estimated from whole genomes under the best-fitting dispersal model (Table 1). Posterior probabilities of branching events are indicated by red ($P > 0.95$) and yellow ($P > 0.85$) circles. Blue bars show the 95% HPD credible intervals of the estimated dates of well-supported nodes. See Fig. S1 for full annotation.

epidemic spread (and in D) by evaluating the variability of dispersal paths among phylogeny branches.

Results

The commonly sequenced WNV E gene contains insufficient genetic variation to resolve the phylogeography of the North American epidemic in detail (26); therefore, we chose to analyze only whole viral genomes. However, almost all genomes available at the time of study were sampled before 2005. We therefore extended the range of sampling by fully sequencing 17 previously unreported WNV isolates sampled between 2004 and 2008 (*Materials and Methods*), thereby obtaining enough divergence to estimate a reliable WNV molecular clock. The resulting final alignment, comprising 104 genomes with defined sampling dates and locations and isolated from a variety of host and vector species (Table S1), was analyzed using the framework introduced above.

To infer the locations of ancestral infections, we used a variety of random walk models, all of which accurately recovered the epidemic's temporal and geographic origin (Table 1). However, the homogeneous model (no dispersal rate variation) was very strongly rejected in favor of heterogeneous models that permitted significant variability among lineages (Table 1) and provided more precise estimates of spatial parameters. The phylogeographic structure of WNV we obtain (Fig. 1*C* and Fig. S1) is congruent with that obtained previously using subgenomic sequences (26, 30) while providing additional resolution and dates of lineage movement. In addition to discriminating the previously defined NY99 and WN02 lineages (30), our data reveal structure in sequences sampled from western areas: the majority of Californian sequences cluster together with basal lineages from Texas [defined as clade "D" in Gray et al. (26)]. All Mexican sequences cluster together ("F") as do some sequences from the southwest ("G").

When projected through space and time (Fig. 2 and Movie S1), this phylogeny shows a westward dissemination of WNV lineages that matches the observed spatiotemporal incidence of WNV (17). Of particular note are a handful of viral lineages that exhibit atypically rapid and long-distance travel. Lineages that move north to south along the Atlantic coast (reaching Florida by 2000) and along the Rocky Mountains are consistent with bird migration corridors bounded by geographic barriers (18). Interestingly, once WNV lineages reach the eastern boundary of the Rocky Mountains, in 2001, further westward movement appears to stall (Movie S1), possibly reflecting the impediment to migration imposed by high elevations.

A key parameter of any spatial epidemic is its wavefront velocity. If we assume no variation in dispersal rates, then, as theory predicts (12), our genetic analysis reconstructs a constant invasion velocity of ~1, 000 km/y (before the western seaboard is reached; Fig. 3*A*). However, under our best-fitting heterogeneous model (Table 1), we observe an accelerating invasion: from 1999 to 2003 the origin-to-wavefront distance doubled every 0.8 y on average (Fig. 3*B*). This acceleration rate, estimated solely from viral genomic data, is almost identical to that independently estimated from large-scale patterns of spatiotemporal WNV incidence (31). Such acceleration is theoretically predicted to occur when there is high variance in dispersal among infected hosts—specifically, when the dispersal kernel is positively skewed and "fat-tailed" (32). This result implies a WNV wavefront with a long leading edge, explaining the discontinuous spread of infection into new areas.

We report empirical estimates of the diffusion coefficient, D, of the WNV epidemic, and we further quantify variability in its spatial spread (Fig. 3 *C* and *D*). Mean D under homogenous diffusion is estimated to be ~200 km^2/d. However, the best-fitting heterogeneous model indicates that WNV's spatial spread is both extraordinarily variable (coefficient of variation of D among branches ~4–8) and, on average, highly diffusive (mean D ~1,000 km^2/d; Fig. 3*D*). This exceptional mean diffusivity exceeds that estimated for the historical spread of Black Death throughout Europe (4) (~70 km^2/d) and can only be explained if some phylogeny branches represent long-distance colinear displacements (e.g., a branch representing 1,000 km unidirectional travel over 25 d would correspond to D = 10,000 km^2/d). The existence of a few rapid, long-range movements also explains the strong correlation between the mean and variation of D among branches (Fig. 3*D*). The remaining less-diffusive lineages likely represent local transmission among hosts and vectors as they move within their typical home ranges.

Discussion

We introduce a conceptual link between phylogeny and spatial ecology and demonstrate that the large-scale dynamics of biological invasions can be quantified from easily sampled and increasingly inexpensive sets of genetic data. Our framework provides a practical method for estimating the diffusion coefficient of a spatial outbreak and for measuring the variability among hosts in spatial spread. Despite being rarely reported, diffusion coefficients are practically and theoretically valuable because they quantify the intrinsic diffusivities of epidemics, analogous to the manner in which R_0 summarizes intrinsic transmission potential. Our approach will be most applicable to vector-borne viruses and to viral epizootics and epiphytotics, and is also suitable for newly emergent pathogens. Once a new pathogen has been identified, retrospective screening of available archived sera could generate a set of pathogen genomes, from

Table 1. Estimates of genetic and spatial parameters under different spatial models

	Spatial model			
		Heterogeneous dispersal*		
	Homogeneous dispersal†	Cauchy	Gamma	Lognormal
ln marginal likelihood	−643.45	−427.24	−399.43	−424.69
ln Bayes factor	244.02	27.81	Best-fitting model	25.26
Date of epidemic origin	1998.6 (1997.9–1999.3)	1998.5 (1997.7–1999.2)	1998.5 (1997.8–1999.1)	1998.6 (1997.9–1999.1)
Mean genome evolution rate (substitutions per site per year)	0.00058 (0.00049–0.00066)	0.00058 (0.00051–0.00064)	0.00057 (0.00051–0.00064)	0.00058 (0.00051–0.00064)
Variability of evolution rate among branches (SD)	0.38 (0.23–0.53)	0.33 (0.21–0.45)	0.33 (0.21–0.45)	0.33 (0.20–0.44)
Latitude of epidemic origin	40.3 (37.1, 43.7)	41.3 (40.4, 43.2)	41.1 (40.4, 43.2)	41.1 (40.3, 43.2)
Longitude of epidemic origin	−76.5 (−82.9, −70.5)	−74.4 (−76.2, −73.2)	−74.6 (−76.1, −73.3)	−74.2 (−76.1, −72.9)

*Dispersal rate varies among branches; rates for each are independently drawn from the corresponding distribution.
†Dispersal rate is equal for all branches.

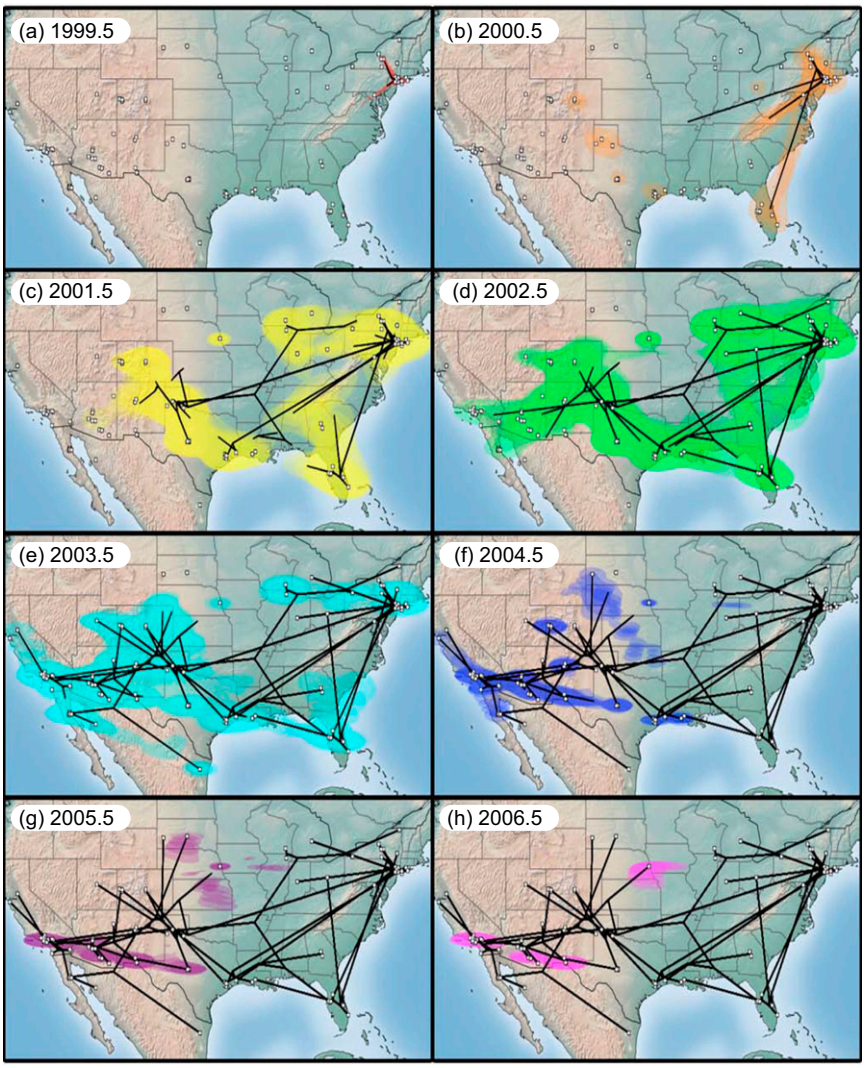

Fig. 2. The reconstructed spatiotemporal diffusion of WNV in North America, shown at annual intervals from mid-1999 onwards (*A*–*H*). White circles indicate isolate sampling locations. Black lines show a spatial projection of a representative phylogeny, with each node being mapped to its known (external node) or estimated (internal node) location. In each panel colored clouds represent statistical uncertainty in the estimated locations of WNV lineages (95% HPD regions) (42).

which the spatial dynamics of the outbreak before its date of discovery can be inferred.

Our WNV analysis shows that the epidemic cannot be adequately described by homogeneous dispersal, and instead was critically shaped by high variation in dissemination of infected hosts. The importance of such heterogeneity in determining the dynamics of spatial invasions is increasingly recognized (24, 33). Bird migrations are the most likely source of rapid, long-distance WNV movements, yet their role in the spread of WNV has been questioned (19), and our current data cannot exclude the possibility of anthropogenic transport of infected hosts or vectors. However, a key benefit of our framework is that long-range viral movements (by whatever mechanism) will leave a detectable phylogenetic footprint even when such events are too rare to be feasibly detected by direct observation. Our results demonstrate that many current mathematical models of North American WNV (13, 14) that have assumed homogenous diffusion are unrealistic, despite their use of complex transmission structures. Such studies have typically modeled host dispersal using data on the short-term home-range movements of birds, which exhibit low mean diffusion coefficients of $D < 14$ km^2/d. By ignoring the substantial variability in WNV dispersal we have uncovered, these models significantly overestimate the R_0 of the epidemic (e.g., $R_0 > 25$) (14, 21). We do not need to assume an exceptionally transmissible pathogen in a weakly diffusive host to explain the observed wavefront velocity of ~1,000 km/y. Instead, the invasion behavior of WNV is best explained by a pathogen with a lower mean R_0 that transmits among hosts whose dispersal is very variable.

Despite capturing the broad-scale spatial dynamics of the WNV invasion of North America, our spatial sampling is not comprehensive and precludes more detailed inferences—for example, whether elliptical migration and central American/Caribbean bird populations were important to WNV dissemination (18, 19). However, our main conclusions are robust to the absence of data from the tropics, because if such movements were common, then estimates of D and its variability would be even greater than those presented here. Migratory movements might explain viral reintroduction into previously colonized locations, e.g., lineages moving northeastward in 2002. More specific hypotheses could be addressed within our framework as further data (including genomes from the tropics) become available. Higher-resolution

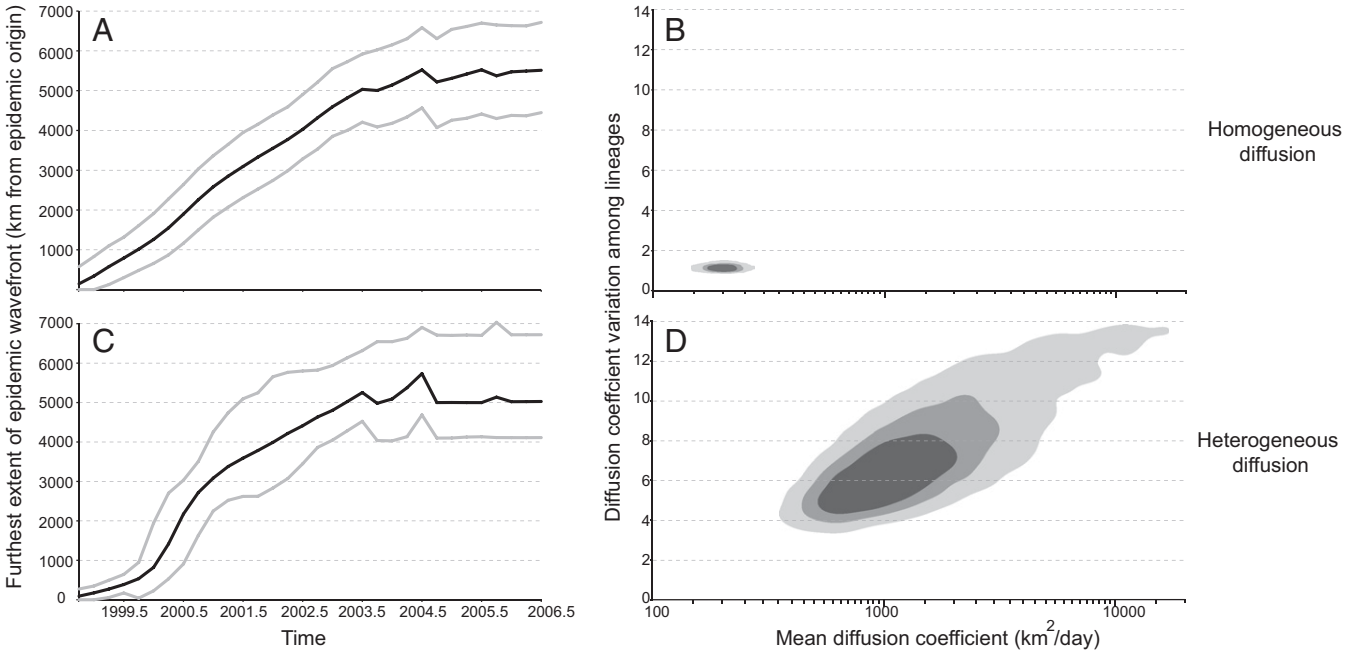

Fig. 3. Characteristics of the North American WNV invasion estimated from viral genomes. Plots *A* and *C* were estimated under a homogenous dispersal model; plots *B* and *D* under the best-fitting heterogeneous model (Table 1). Plots *A* and *B* show the reconstructed epidemic wavefront. For each point in time, the black line is the estimated distance from the epidemic wavefront to its estimated origin: the gradient of this line is thus the invasion velocity. Gray lines indicate the 95% credible regions of the estimated wavefront position. Plots *C* and *D* show kernel density estimates of the diffusion coefficient (*D*) parameters. The horizontal axis shows the estimated mean *D* among lineages; the vertical axis shows the coefficient of variation of *D* among lineages. The three contours show, in shades of decreasing darkness, the 50%, 75%, and 95% HPD regions via kernel density estimation.

sampling would also allow the application of more complex spatial processes (e.g., Lévy flights or advection-diffusion models).

The genomes of rapidly evolving pathogens are already used to estimate the date of origin and R_0 of emerging epidemics (34), most recently for pandemic H1N1/09 influenza (35). The methods introduced here could similarly enable the rate, direction, and mode of spatial spread of future emergent viruses to be inferred from genetic data. Such methods also open the door for the development of future approaches that could potentially jointly estimate R_0 and D from sampled pathogen genomes (9); however, any such approach will require a much better understanding of the effects on lineage coalescence of nonequilibrium spatial dynamics. Further, the connection between phylogeny and spatial autocorrelation exploited here could be applied to other problems in spatial ecology, such as the control of invasive species, provided that suitably diverse genetic markers for the species in question are available.

Materials and Methods

Human Samples. Only four WNV complete genomes available at the time of study were sampled after 2004. To characterize more recent isolates (and thus estimate a reliable molecular clock) we obtained 17 infected human plasma samples detected during blood donor screening at blood centers across the United States (36). The isolates reported here were sampled during 2003–2007 (Table S1). This study was approved by the University of California San Francisco Committee on Human Research and informed consent was obtained.

RT-PCR and Genome Sequencing. WNV genomes were amplified and sequenced in four fragments. Briefly, total RNA was extracted from plasma using the QIAamp Viral RNA Mini Kit (Qiagen) and eluted in 50 mL of elution buffer in the presence of 40 U Protector RNase inhibitor (Roche). First-strand cDNA synthesis was initiated using 12.5 mL of RNA and 0.5 mg of primer R1, R2, R3, or R4a (37) and 400 U of murine leukemia virus reverse transcriptase (Promega). For amplification of each portion of the genome, a nested PCR was performed using 5 mL each of cDNA and TaKaRa Ex Taq DNA polymerase (TaKaRa Bio). Primer sequences and PCR cycling conditions were identical to those in Herring et al. (37). A single 2.8- to 3.2-kb band was detected on 0.8% agarose gel. PCR products were purified with QIAquick PCR (Qiagen) and sequenced using previously reported primers (37) and the BigDye Kit on an ABI3700 capillary sequencer. After manual editing, sequences were assembled using SeqMan (GenBank accession nos. GQ507468–GQ507484).

Sequence collation and annotation. All available North American WNV near-complete genome sequences were obtained from GenBank, one of which (DQ211652) was a duplicate of AF202541 and removed; these were added to our genomes, resulting in a final data set comprising 104 genomes, 11,029 nt long. Sequences were codon aligned by hand. Host species, sampling date, and location of each sequence were obtained from the literature or provided by previous authors. ZIP code locations were converted into latitude and longitude coordinates using ZIPList5. For 27 sequences, only the US or Mexican state was known; the latitude and longitude of these was defined as the geographic centroid of the state. If only the year of sampling was known, then the sampling date was defined as the midpoint of the year (Table S1).

Model selection analyses. Model selection analyses were first undertaken to select a statistically appropriate evolutionary model (Table S2). Eight model combinations were explored, representing all permutations of (*i*) the Hasegawa-Kishino-Yano (HKY) vs. general time-reversible (GTR) substitution model, (*ii*) incorporation vs. omission of a Γ distribution of among-site rate heterogeneity, and (*iii*) strict molecular clock vs. an uncorrelated lognormal relaxed molecular clock (38). For each model, parameters were estimated using the Bayesian MCMC approach implemented in BEAST alongside a Bayesian skyline coalescent model (39). Other coalescent models were investigated but performed poorly. MCMC chains were run for 50 million states, sampled every 5,000 states. MCMC convergence was evaluated using Tracer 1.5 (http://beast.bio.ed.ac.uk). The performance of each combination was compared using Bayes factors (40). Estimated evolutionary rates and divergence times were almost identical among models. The best-fitting model was GTR + Γ with a lognormal relaxed molecular clock (Table S2), and was thus used in subsequent analyses.

Relaxed random-walk models. We extended the phylogeographic approach in BEAST 1.7 (10) and used the BEAGLE library to accelerate computation (41). Movement in two dimensions was modeled as a scaled-mixture generalization of a Brownian motion process (*SI Text*). This model is motivated by formal Lévy flight models while not strictly enforcing dispersal kernels with power-law tails. Realized dispersal path lengths were corrected for the Earth's curvature using great circle distances. As in Lemey et al. (10), diffusion rate variation was implemented by rescaling the diffusion process along each phylogeny branch, with the scalars for each being drawn from

a specified distribution: these scaled mixtures generate a wide range of relaxed random walks. We evaluated different probability distributions (Cauchy, gamma, lognormal) to accommodate among-branch diffusion rate variation and compared their fit to a homogeneous process. To aid computation, we developed unique analytical solutions to the marginalization of unobserved multivariate traits at internal nodes under relaxed random-walk models (*SI Text*). Methods are implemented in BEAST 1.7 (source code available from http://beast-mcmc.googlecode.com).

Postprocessing and visualization. MCMC chains were run for 250 million states, sampled every 50,000 states. The posterior distribution of phylogenies was summarized using maximum clade credibility (MCC) trees in TreeAnnotator. MCC trees and 95% highest posterior density (HPD) contours were visualized using SPREAD (42). Various statistics (e.g., the wavefront velocity) were extracted from the posterior distribution by sampling each rooted phylogeny at multiple time points and summarizing the resulting distributions.

ACKNOWLEDGMENTS. We thank Eddie Holmes, Mike Bonsall, Sunetra Gupta, John Drake, Robert May, and Paul Harvey for discussion. Support for this work was provided by the Royal Society (O.G.P. and A.R.); National Institutes of Health Grant R01 GM086887 (to M.A.S. and F.W. C.); Centers for Disease Control/National Center for Infectious Diseases Grant R01-CI-000214 (to F.J.B, M.P.B., and E.L.D.); UK Medical Research Council (R.R.G.); European Research Council Seventh Framework Programme Grant 260864 (to P.L.); and the Institute for Mathematical Sciences, National University of Singapore (M.A.S. and P.L.).

1. Snow J (1854) The cholera near Golden Square and at Deptford. *Med Times Gazette* 9: 321–322.
2. Skellam JG (1951) Random dispersal in theoretical populations. *Biometrika* 38: 196–218.
3. Mollison D (1977) Spatial contact models for ecological and epidemic spread. *J Roy Stat Soc B* 39:283–326.
4. Noble JV (1974) Geographic and temporal development of plagues. *Nature* 250: 726–729.
5. Murray JD, Stanley EA, Brown DL (1986) On the spatial spread of rabies among foxes. *Proc R Soc Lond B Biol Sci* 229:111–150.
6. Grenfell BT, Bjørnstad ON, Kappey J (2001) Travelling waves and spatial hierarchies in measles epidemics. *Nature* 414:716–723.
7. Fitch WM (1996) The variety of human virus evolution. *Mol Phylogenet Evol* 5:247–258.
8. Bourhy H, et al. (1999) Ecology and evolution of rabies virus in Europe. *J Gen Virol* 80: 2545–2557.
9. Biek R, Henderson JC, Waller LA, Rupprecht CE, Real LA (2007) A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proc Natl Acad Sci USA* 104:7993–7998.
10. Lemey P, Rambaut A, Welch JJ, Suchard MA (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol* 27:1877–1885.
11. Grenfell BT, et al. (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303:327–332.
12. Shigesada N, Kawasaki K (1997) *Biological Invasions: Theory and Practice* (Oxford Univ Press, London).
13. Lewis M, Rencławowicz J, van den Driessche P (2006) Traveling waves and spread rates for a West Nile virus model. *Bull Math Biol* 68:3–23.
14. Maidana NA, Yang HM (2009) Spatial spreading of West Nile Virus described by traveling waves. *J Theor Biol* 258:403–417.
15. Lanciotti RS, et al. (1999) Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science* 286:2333–2337.
16. LaDeau SL, Kilpatrick AM, Marra PP (2007) West Nile virus emergence and large-scale declines of North American bird populations. *Nature* 447:710–713.
17. Centers for Disease Control (2011) Statistics, Surveillance and Control Archive. Available at http://www.cdc.gov/ncidod/dvbid/westnile.
18. Reed KD, Meece JK, Henkel JS, Shukla SK (2003) Birds, migration and emerging zoonoses: West Nile Virus, Lyme disease, influenza A and enteropathogens. *Clin Med Res* 1:5–12.
19. Rappole JH, et al. (2006) Modeling movement of West Nile virus in the Western hemisphere. *Vector Borne Zoonotic Dis* 6:128–139.
20. Bowman C, Gumel AB, van den Driessche P, Wu J, Zhu H (2005) A mathematical model for assessing control strategies against West Nile virus. *Bull Math Biol* 67:1107–1133.
21. Wonham MJ, Lewis MA, Rencławowicz J, van den Driessche P (2006) Transmission assumptions generate conflicting predictions in host-vector disease models: A case study in West Nile virus. *Ecol Lett* 9:706–725.
22. Liu R, Shuai J, Wu J, Zhu H (2006) Modeling spatial spread of West Nile virus and impact of directional dispersal of birds. *Math Biosci Eng* 3:145–160.
23. Yiannakoulias NW, Schopflocher DP, Svenson LW (2006) Modelling geographic variations in West Nile virus. *Can J Public Health* 97:374–378.
24. Magori K, Bajwa WI, Bowden S, Drake JM (2011) Decelerating spread of West Nile virus by percolation in a heterogeneous urban landscape. *PLOS Comput Biol* 7: e1002104.
25. Cruz-Pacheco G, Esteva L, Montaño-Hirose JA, Vargas C (2005) Modelling the dynamics of West Nile virus. *Bull Math Biol* 67:1157–1172.
26. Gray RR, Veras NM, Santos LA, Salemi M (2010) Evolutionary characterization of the West Nile Virus complete genome. *Mol Phylogenet Evol* 56:195–200.
27. Felsenstein J (1985) Phylogenies and the comparative method. *Am Nat* 125:1–15.
28. Harvey PH, Pagel MD (1991) *The Comparative Method in Evolutionary Biology* (Oxford Univ Press, London).
29. Einstein A (2003) *Investigations on the Theory of the Brownian Movement*, ed Furth R (Dover, New York).
30. Davis CT, et al. (2005) Phylogenetic analysis of North American West Nile virus isolates, 2001–2004: Evidence for the emergence of a dominant genotype. *Virology* 342: 252–265.
31. Mundt CC, Sackett KE, Wallace LD, Cowger C, Dudley JP (2009) Long-distance dispersal and accelerating waves of disease: Empirical relationships. *Am Nat* 173: 456–466.
32. Kot M, Lewis MA, Van den Driessche P (1996) Dispersal data and the spread of invading organisms. *Ecology* 77:2027–2042.
33. Melbourne BA, Hastings A (2009) Highly variable spread rates in replicated biological invasions: Fundamental limits to predictability. *Science* 325:1536–1539.
34. Pybus OG, et al. (2001) The epidemic behavior of the hepatitis C virus. *Science* 292: 2323–2325.
35. Fraser C, et al.; WHO Rapid Pandemic Assessment Collaboration (2009) Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science* 324:1557–1561.
36. Busch MP, et al. (2005) Screening the blood supply for West Nile virus RNA by nucleic acid amplification testing. *N Engl J Med* 353:460–467.
37. Herring BL, et al. (2007) Phylogenetic analysis of WNV in North American blood donors during the 2003–2004 epidemic seasons. *Virology* 363:220–228.
38. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4:e88.
39. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29:1969–1973.
40. Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol* 18:1001–1013.
41. Suchard MA, Rambaut A (2009) Many-core algorithms for statistical phylogenetics. *Bioinformatics* 25:1370–1376.
42. Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: Spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* 27:2910–2912.