

Epidemic Dynamics Revealed in Dengue Evolution

S.N. Bennett,^{*1} A.J. Drummond,² D.D. Kapan,³ M.A. Suchard,⁴ J.L. Muñoz-Jordán,⁵ O.G. Pybus,⁶ E.C. Holmes,⁷ and D.J. Gubler^{1,8}

¹Department of Tropical Medicine, Medical Microbiology and Pharmacology, Asia-Pacific Institute of Tropical Medicine and Infectious Diseases, J.A. Burns School of Medicine, University of Hawaii at Manoa

²Department of Computer Science, University of Auckland, Private Bag 92019, Auckland, New Zealand

³Center for Conservation and Research Training, Gilmore 408, PBRC, University of Hawaii at Manoa

⁴Departments of Biomathematics, Biostatistics, and Human Genetics, David Geffen School of Medicine at UCLA

⁵Dengue Branch, Division of Vector-Borne Infectious Disease, Centers for Disease Control and Prevention, San Juan, Puerto Rico

⁶Department of Zoology, University of Oxford, Oxford, United Kingdom

⁷Department of Biology, Center for Infectious Disease Dynamics, The Pennsylvania State University

⁸Duke-NUS Graduate Medical School, Singapore

***Corresponding author:** E-mail: sbennett@hawaii.edu.

Associate editor: Rasmus Nielsen

Abstract

Dengue is an emerging tropical disease infecting tens of millions of people annually. A febrile illness with potentially severe hemorrhagic manifestations, dengue is caused by mosquito-borne viruses (DENV-1 to -4) that are maintained in endemic transmission in large urban centers of the tropics with periodic epidemic cycles at 3- to 5-year intervals. Puerto Rico (PR), a major population center in the Caribbean, has experienced increasingly severe epidemics since multiple dengue serotypes were introduced beginning in the late 1970s. We document the phylodynamics of DENV-4 between 1981 and 1998, a period of dramatic ecological expansion during which evolutionary change also occurs. The timescale of viral evolution is sufficiently short that viral transmission dynamics can be elucidated from genetic diversity data. Specifically, by combining virus sequence data with confirmed case counts in PR over these two decades, we show that the pattern of cyclic epidemics is strongly correlated with coalescent estimates of effective population size that have been estimated from sampled virus sequences using Bayesian Markov Chain Monte Carlo methods. Thus, we show that the observed epidemiologic dynamics are correlated with similar fluctuations in diversity, including severe interepidemic reductions in genetic diversity compatible with population bottlenecks that may greatly impact DENV evolutionary dynamics. Mean effective population sizes based on genetic data appear to increase prior to isolation counts, suggesting a potential bias in the latter and justifying more active surveillance of DENV activity. Our analysis explicitly integrates epidemiologic and sequence data in a joint model that could be used to further explore transmission models of infectious disease.

Key words: adaptive evolution, phylodynamics, dengue, virus, epidemics, population bottlenecks.

Introduction

Dengue is the most widespread vector-borne viral disease of humans. An average of almost 970,000 cases per year were reported from 63 countries between 2000 and 2007 (<http://www.who.int/csr/disease/dengue/impact/en/index.html>), but the World Health Organization (WHO) estimates the number of infections to be far higher, from 50 to 100 million per year (Gubler 1998, 2006). The large discrepancy between dengue cases reported and actual infections is due in part to the absence of a specific and clinically useful diagnostic test to distinguish dengue from a range of other nondescript febrile illnesses (Kroeger and Nathan 2006), as well as unpredictable asymptomatic infection rates (e.g., 50–87% as in Burke et al. 1988).

What was once a disease with low attack rates, moderate pathogenicity, and infrequent epidemics has become, in the last two decades, the leading arboviral cause of illness and death in humans (Gubler 2006). Unplanned urban expansion, leading to high densities of humans exposed to

anthropophilic mosquito vectors, as well as globalization and travel, has contributed to hyperendemic transmission of all four dengue serotypes throughout the tropics (Gubler 1998; Kroeger and Nathan 2006). The result has been a significant change in the modality of dengue virus transmission across its range, today characterized by periodic epidemics every few years marked by mounting numbers of severe cases, in which the predominant victims are children. The reported number of dengue cases has increased 2-fold since the last decade (Kroeger and Nathan 2006).

Dengue is endemic in all WHO regions except the European Region. Dengue virus has been hyperendemic (multiple cocirculating serotypes) in Southeast Asia since World War II and in many other tropical regions since the 1970s (Gubler 2006). DENV transmission exhibits a strong seasonal component (e.g., Morrison et al. 1998; Nisalak et al. 2003; Johansson et al. 2009) driven to some extent by short- and long-term rainfall, relative humidity, and temperature effects on mosquito vectors and vector–virus

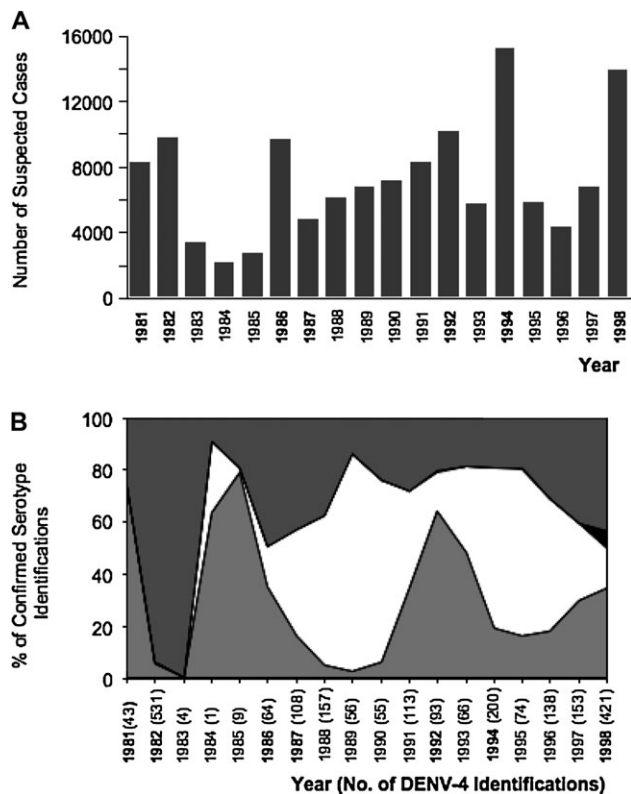


FIG. 1. Endemic and epidemic periods of dengue virus in PR. (A) Number of suspected cases of DF/DHF by year reported to the CDC-Dengue Branch between 1981 and 1998. (B) Percentage of identifications of each serotype relative to the total of positive serotype identifications by tissue culture isolation per year. Numbers in parentheses indicate the numbers of DENV-4 isolations made at the CDC San Juan, Dengue Branch each year. DENV-1 (light gray, bottom), DENV-2 (white), DENV-3 (black), and DENV-4 (dark gray, top). Isolates were made on a subset of reported cases, and the two estimates of dengue population size (suspected and confirmed cases) are strongly correlated. Hyperendemic transmission is demonstrated by the occurrence of multiple cocirculating serotypes at a given time. All four serotypes of dengue exhibit seasonal peaks in transmission as well as similar interannual but alternating periodicity in transmission; with the appearance of DENV-3 in 1998, all other serotypes underwent a substantial reduction in subsequent years (data not shown).

interactions (e.g., Li et al. 1985; Johansson et al. 2009). In addition, DENV transmission exhibits an interannual periodicity consisting of outbreaks every 3–5 years, often of alternating serotypes (Gubler 1998), a periodicity also reflected in the number of dengue hemorrhagic fever (DHF) cases (Cummings et al. 2004). The magnitude of these interannual differences in incidence can reach tenfold (Cummings et al. 2004). Hypotheses proposed and/or supported for this periodicity include 1) immune interactions amongst serotypes (Cummings et al. 2004; Adams et al. 2006; Ooi et al. 2006; Recker et al. 2009), 2) climate variability, such as warmer years exhibiting higher transmission (Johansson et al. 2009), and 3) vector variability (e.g., Strickman and Kittayapong 2002).

The Americas have seen the most dramatic epidemiologic changes, largely due to the reinvasion of the once-

eradicated vector *Aedes aegypti* from most of tropical America (Gubler 1998). When the eradication program ceased in the 1970s, only DENV-2 and -3 were present and only in a few countries (Gubler 1987). Ten years later, the region began experiencing major epidemics, soon leading to hyperendemicity and the emergence of severe forms of the disease such as DHF, 25 years behind Southeast Asia (Gubler 1987, 1998; reviewed in Gubler 2006). Puerto Rico (PR) exemplifies these dynamics—a relatively large, densely populated island in the Caribbean that has had continual hyperendemic dengue transmission since the early 1980s (Dietz et al. 1996; Gubler 1998). Outbreaks of increasing severity over the last three decades began in 1977 when DENV-3 reappeared, followed sequentially by the introduction of DENV-1 (1978), DENV-4 (1981), DENV-2 (1984), and DENV-3 (1997), thus establishing a continued pattern of epidemics in PR involving multiple serotypes, with 2–3 year cycles in the prevailing serotype (Gubler 1987; Dietz et al. 1996; Rigau-Perez et al. 2002, fig. 1). Throughout, DENV-4 has played a dominant role in PR epidemiologic dynamics (Rigau-Perez et al. 2001, 2002). Following the introduction of DENV-4 in 1981, an islandwide epidemic in 1982 was the first DENV-4 dominated outbreak and was followed by epidemics in 1986 and 1998 that were marked by the emergence of epidemic DHF in 1986 and record-high incidences of DHF/dengue shock syndrom (DSS) in subsequent years (Dietz et al. 1996; Bennett et al. 2003). Even during the DENV-2 outbreak of 1994, DENV-4 rivaled DENV-2 in numbers of infections isolated in the early stages of the epidemic (Rigau-Perez et al. 2001).

The epidemiologic dynamics of dengue virus transmission in PR reflects dramatic changes in viral population sizes, an important driver of evolutionary change (Rambaut et al. 2008; Pybus and Rambaut 2009), which partly underlies its emergence in PR (Bennett et al. 2003; Carrington et al. 2005). Indeed, the rapid evolution of dengue, and RNA viruses in general, occurs on a similar timescale as the ecological dynamics affecting its transmission and spread. Genetic diversity in viruses accumulates in real time in direct response to changes in population size or structure, etc., providing a useful tool to infer virus transmission dynamics. Thus, our phylodynamic study of DENV combines evolutionary and ecological dynamics in a common and natural timescale of months and years, to reconstruct epidemiologic events that can be directly compared with our surveillance data. Tracking changing viral population size through time using genetic diversity to complement surveillance data provides additional insight into epidemiologic and evolutionary dynamics of the pathogen (Pybus and Rambaut 2009). Carrington et al. (2005) demonstrated a dramatic expansion in DENV-4 population sizes upon introduction into PR based on genetic diversity estimates, 16-fold greater than for DENV-2 over the same period and faster than any virus previously described. Dengue epidemiologic dynamics in particular stand to gain clarification from a phylodynamic approach: passive surveillance overlooks asymptomatic infections, which can be high in dengue, and between outbreaks (or even between seasonal

high) a given DENV serotype often falls below detectable levels. The virus does not go locally extinct, because phylogenetic data indicate that strains usually persist and/or evolve locally in PR rather than by reintroduction from other islands (e.g., Bennett et al. 2003; Carrington et al. 2005). Here, we track DENV-4 population dynamics over two decades of turbulent epidemiologic history throughout its emergence into the hyperendemic arena of PR using genetic diversity of virus sequences as independent estimators of population size. These estimates of effective population size are generated over months and years on a timescale directly comparable with census population size estimates from surveillance data based on confirmed cases of DENV-4. Thus, we successfully capture phases of epidemic onset interspersed with dramatic population bottlenecks and revealing a within-serotype periodicity of 3–5 years.

Materials and Methods

A passive dengue surveillance system has been established in PR since 1969, following an explosive DENV-2 epidemic and was enhanced with active virologic surveillance beginning in 1981 (Gubler 1987). The system involves standardized case reporting of suspect dengue cases and blood specimen submission by hospitals, health centers, clinics, and private physicians throughout PR, to the San Juan Laboratories (SJL), Dengue Branch, US Centers for Disease Control and Prevention (CDC). The counts of dengue virus isolations by serotype by month between 1981 and 2001 are reported in this study.

Selected DENV-4 isolates from PR since the serotype was introduced in 1981 (Gubler 1987; Dietz et al. 1996), archived at SJL, were randomly selected and sequenced for the years 1982 ($n = 14$), 1986/1987 ($n = 19$), 1992 ($n = 15$), 1994 ($n = 14$), and 1998 ($n = 13$), as previously described (Bennett et al. 2003). To enhance this temporal coverage, we also sequenced several additional isolates from PR from 1985 ($n = 8$) and 1990 ($n = 5$). All samples were labeled by month and year of collection. All isolates sequenced had low passage histories (≤ 2). After RNA extraction using QIAamp Viral RNA Mini Kits (Qiagen GmbH), isolates were amplified by reverse-transcriptase polymerase chain reaction (RT-PCR), purified using Qiagen PCR purification kits (Qiagen GmbH), and sequenced along both strands for $2\times$ coverage in standard dye-labeling reactions for 40% of the viral genome (4,016 bp of an 11-kbp genome) spanning multiple genes. Concatenated regions included all the structural genes (capsid—C, membrane—M, and envelope—E), a subset of nonstructural genes (*NS1*, *NS2A*, and *NS4B*), and the noncoding 3' untranslated region (NTR). Sequences were assembled in Sequencher 3.1.1 (Gene Codes) and aligned using Clustal W (Megalalign v 3.1.7, Lasergene). The alignment was verified for the absence of recombination using single breakpoint analysis in GARD, HYPHY 1.0020080508b (Pond et al. 2005).

The evolutionary history of DENV-4 isolates in PR was inferred using maximum likelihood (ML) phylogenetic methods implemented in RAxML Black Box webserver

(Stamatakis et al. 2008) under the general time reversible (GTR) + I + Γ_4 model of evolution as selected by jModeltest 0.1.1 (Posada 2008) and 100 ML bootstrap replicates. Trees were rooted with the 1981 isolate from Dominica, the strain introduced into PR in 1981.

To estimate virus effective population sizes (N_e) through time, we used a coalescent-based approach that can independently estimate N_e and the rate of viral sequence evolution, given a set of sequences isolated at different points in time (Drummond et al. 2002). This approach is implemented in a Bayesian Markov Chain Monte Carlo (MCMC) inference framework in the program Bayesian evolutionary analysis sampling trees (BEAST; Drummond and Rambaut 2007), which incorporates uncertainty in the phylogeny by integrating across tree topologies to estimate relative genetic diversity ($N_e t$, where t is the generation time), an indicator of effective population size under a neutral evolutionary process. Generation time t for DENV was set at 2 weeks.

Effective population size can be defined as the census population size divided by the variance in the number of offspring among individuals. In the case of a virus, the census population size is estimated from isolate counts and case reports, whereas the effective population size is actually an estimate of the effective number of infections, that is, those that go on to produce subsequent infections. Effective population size trends were estimated in BEAST using the BEAGLE library to enable computation in greatly improved time (Suchard and Rambaut 2009) and the Bayesian Skyride model (Minin et al. 2008). In the BEAST analysis, we employed a strict molecular clock model and a codon model of substitution (Goldman and Yang 1994). Chain length for MCMC sampling was 25 million generations, sampling every 5,000 generations. We tested for an association between N_e (estimated from sequence data) and isolate counts, simultaneously estimating a scaling parameter for the amplitude and time lag between N_e (estimated from sequence data) and isolate counts, with statistical uncertainty reflected in values of the 95% highest probability density (HPD). Recall that sequence data are accompanied by date of collection, and isolate counts are recorded monthly—thus, we compare at fine temporal resolution the estimated effective population size derived from sequence diversity with census or surveillance data in terms of monthly isolations. The fit of DENV-4 isolation counts by month represented in [figure 1B](#) to the estimates of N_e from BEAST was made using least squares.

The fit of the DENV-4 isolation counts to the estimates of N_e from BEAST was made using Poisson regression, a standard approach for modeling count data. This generalized linear model (GLM) fits the logarithm of the expected number of cases in an observation window to a linear function of $\log N_e$ with parameters for any shift in time by an estimable amount and an autoregressive component. This latter component relates current counts to immediately previous counts, enabling distinction of occasional epidemic from seasonal fluctuations. The autoregressive component is standard in epidemiological modeling

(Held et al. 2006) and provides for significantly better model fit. If we let t index discrete times, C_t count the observed cases in an observation window around t , and λ_t measure the expected number of cases, then

$$C_t \sim \text{Poisson}(\lambda_t) \text{ and } \log \lambda_t = \beta_0 + \beta_1 \log N_e(t - s) + \beta_2 C_{t-1}, \quad (1)$$

where β_0 , β_1 , and β_2 are unknown regression coefficients and s is the unknown shift in time. This likelihood-based GLM is a good approximation to a biologically realistic model that assumes the isolation counts are realizations from an inhomogeneous Poisson process where the event intensity is a hypothesized function of genetic diversity. One can formally test the association and estimate the lag simultaneously. However, we note that the regression depends only on the mean N_e function and thus ignores the uncertainty in estimating N_e from the sequences. Consequently, we choose a stringent P value critical value ($< 10^{-3}$) when testing for an association between N_e and isolation counts.

Results

Two decades of dengue virus isolate counts from PR indicate interannual variation in transmission and within-year seasonal peaks in transmission from July to January across all serotypes. Isolates are routinely tested to confirm a subset of reported cases and to identify viral serotype. Isolate counts were strongly correlated with cases reported per annum ($P \ll 0.01$, $r = 0.90$, Pearson's correlation). Interannual periodicity consisted of transmission peaks separated by lower seasonal highs in adjacent years: DENV outbreaks occurred only in certain years (fig. 1A) interspersed by periods of relatively low transmission even during the peak season and were sometimes dominated by alternate serotypes: A 1981 epidemic was dominated by DENV-1, 1982 by DENV-4, 1986 by DENV-4, 1994 by DENV-2, and 1998 by DENV-4 (fig. 1B). The first DENV-4 outbreak in 1982 was followed by 2 years of low transmission, after which epidemic DENV-4 reappeared in 1986. Several years of intermediate levels of DENV-4 transmission followed until 1998 when DENV-4 dominated a severe epidemic characterized by record levels of DHF/DSS cases (Bennett et al. 2003). In 1998, DENV-3 also first reappeared in PR, followed by a rapid decline of DENV-4 and the other serotypes in the subsequent 2 years (data not shown).

The evolutionary history of DENV-4 in PR, inferred from $> 4,000$ nt sequenced for each of 89 isolates spanning almost two decades, shows considerable (up to 2%, for a single population) but well-ordered diversification since the serotype's emergence in 1982 (fig. 2). Isolates formed clear clusters according to date of collection, including distinct groupings of isolates from the initial DENV-4 outbreak (1982), the second major dengue epidemic on the island (1986–1987), and the most recent DENV-4 epidemic (1998) (fig. 2). Hence, viruses isolated from each epidemic form phylogenetically distinct clusters. The phylogenetic analysis also suggests that the 1986–1987 lineage was present as early as 1985. Similarly, the 1998 lineage was present but rare as early as 1994.

Another distinct lineage was formed by most viruses collected from 1990 to 1994 when DENV-4 prevalence was relatively low (fig. 2). Overall, the degree of genetic divergence from the oldest 1981 sequence increases proportionally with date of isolation. Nonetheless, of these temporally structured lineages, the 1998 lineage is the most genetically diverse group, reflective of the large population size of DENV-4 at the time: The 1998 epidemic was one of the most severe the island had ever seen, with 396,000–792,000 infections estimated and a record 59 DHF cases reported (Bennett et al. 2003).

The pattern of fluctuating mean relative genetic diversity ($N_e t$) of DENV-4 through time was translated into effective viral population sizes (N_e), assuming a viral generation time (t) of 2 weeks, and captured a dynamic viral demographic history over the sampling period of 17 years, from 1981 to 1998 (fig. 3). The values of N_e for DENV-4 underwent dramatic growth starting in 1981, increasing by almost 4-fold, reflecting the epidemiologic pattern of an explosive epidemic following the introduction of DENV-4 into the region in 1981. This population expansion was followed by a gradual 10-fold drop in the viral N_e that did not reach zero, despite isolate count data that shows only a single DENV-4 was isolated during the 2 years from February 1983 to September 1985. The viral N_e increased approximately 10-fold in July/August 1986, followed by a sustained low in N_e from September 1987 to November 1990. In contrast, isolate counts show a secondary peak late in 1988. Although not reflected in isolate counts, we noted a peak transmission event beginning in November 1991 to August 1992. We also observe an increase in N_e around February 1994: Rigau-Perez et al. (2001) note that DENV-4 rivaled DENV-2 in the 1994 epidemic, although isolate counts suggest the latter was dominant. Finally, virus N_e around the severe 1998 outbreak of DENV-4 suggests that virus populations began peaking earlier than reflected in isolate counts and case reports, in April 1997, 20 months prior to our last sample date of November 1998. These inferences are made on mean estimates of N_e , certainty around which is reflected by the HPDs (fig. 3). It should be noted that the “troughs” in N_e are less pronounced than troughs in isolate counts, similarly observed in simulations reported in Rambaut et al. (2008). This suggests a limitation in the inference approach of N_e to resolve the degree (amplitude) of cyclical dynamics. Nonetheless, with the exception of the 1988 increase in confirmed cases unaccompanied by an increase in N_e and the 1991/1992 increase in viral N_e unobserved in confirmed cases, the mean estimates of N_e according to genetic diversity and isolation counts reflecting census population size corresponded. The regression parameter relating $\log N_e$ and \log expected isolation counts is highly significantly nonzero (0.36, 95% confidence interval = 0.31–0.40, $P < 10^{-16}$). Poisson regression confirms the general trend of increasing viral population sizes (N_e) preceding outbreaks (isolate counts) (fig. 3A). The ML estimate of the lag between N_e and isolation

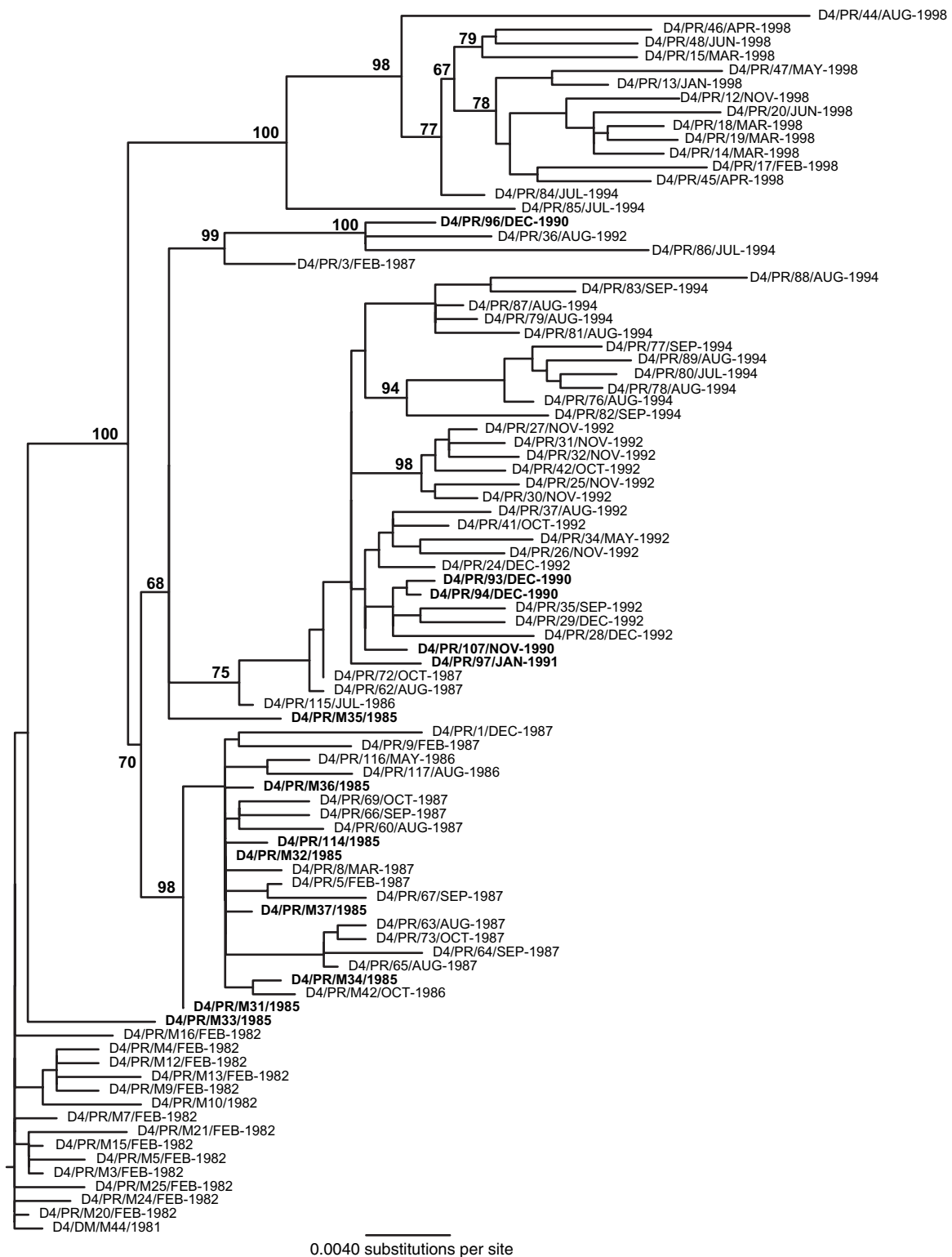


Fig. 2. ML phylogeny of DENV-4 in PR from 1981 to 1998, throughout a turbulent transmission history. ML phylogeny based on concatenated structural (Capsid, Membrane, and Envelope) and nonstructural genes (NS1 partial, NS2A, and NS4B), and the 3' NTR ($n = 4,016$), using the GTR + I + Γ model of evolution. Support given at nodes based on 100 ML replicates implemented in RAXML Black Box webserver (Stamatakis et al. 2008). Accession numbers of NCBI sequences used in this study: AY152036–AY152363 (Bennett et al. 2003) and GU318306–GU318318 (this study, shown in bold).

counts is 7 months (5–8 months, 95% confidence interval). **Figure 3B** replots lag-adjusted N_e against isolation counts. The periodicity in DENV-4 N_e was modeled with

a BEAST posterior simulation using a parametric form, a “sine” wave whose ML periodicity was estimated as 3–5 years.

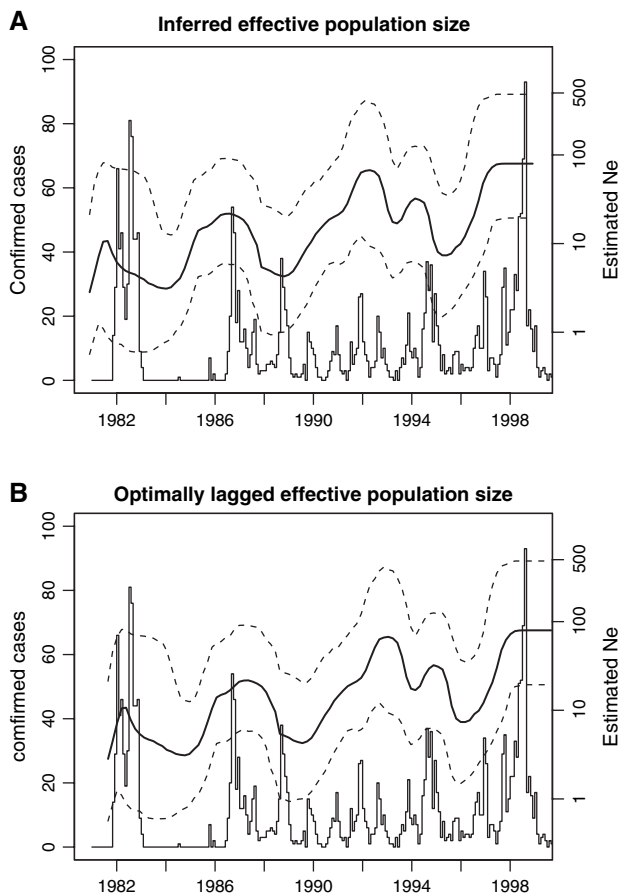


FIG. 3. Effective population size estimates in terms of effective numbers of infections per month based on viral genetic diversity and coalescent patterns, superimposed on the number of DENV-4 isolates by month (A) directly and (B) with a 7-month upward adjustment of N_e (e.g., ahead in time).

Discussion

Dengue fever is a reemerging disease worldwide in the tropics and currently poses the most significant arboviral threat to humans. Throughout this rising trend in dengue incidence and geographic range, dengue virus transmission has exhibited a striking periodicity in severe outbreaks in almost all endemic and hyperendemic populations. A dissection of these dynamics as estimated from sampled viral genes sequences, decoupled from disease syndrome-based surveillance but on a similar timescale for direct comparison, provides a powerful tool for understanding the underlying evolutionary, ecological and epidemiologic mechanisms of emergence.

Phylogenetic analysis of DENV-4 over two decades following its introduction into the region in 1981 recovered an initial dramatic 4-fold increase in mean effective population size estimate—(N_e) based sequence diversity. The growth rate of this first expansion has been regionally estimated as doubling every 2 weeks, faster than any viral population measured by coalescent methods as of 2005 (Carrington et al. 2005). Surveillance data similarly reported a rapid increase in the number of DENV-4 isolated from cases, from 43 in 1981 to 531 in 1982, over a 12-fold in-

crease. Although mean N_e and isolate counts (reflecting census population size) were highly statistically correlated, there was also a significant lag between them. That changes in genetic diversity in the population should precede changes in isolate counts is somewhat surprising and may have several alternative explanations. Increased diversity could bring on epidemics by providing the variation from which more fit, for example, epidemic-causing strains are more likely to evolve. Subsequently, during epidemics, diversity might be lost as selection for “epidemic” strains takes effect (selection was observed in some DENV-4 epidemic strains as reported in Bennett et al. 2003). Here, we observe a decrease in estimated genetic diversity (i.e., an increase in coalescent rate) during three major outbreaks, 1982, 1986–1987, and 1994 (fig. 3A, excluding the last for which we have no subsequent data). Finally, the discrepancy may be due to nonrandom and biased sampling in case counts and isolations (e.g., epidemiologic surveillance). Isolations rarely surpassed 100 per month even in epidemic years, reflecting a possible limitation in laboratory capacity. In addition, success of virus isolation rate depends on both timing in the disease course and viremia, the latter of which will presumably increase proportionally with virus population sizes. Uncertainties in passive surveillance data could be addressed with a sampling design that includes active random surveillance of the population irrespective of clinical status.

Estimates of effective viral population size in terms of effective number of infections (itself a reflection in underlying changes in genetic diversity) indicate a pattern of peak transmission events separated by lows or troughs that can be best described as population bottlenecks. These are truly bottlenecks because they represent a loss of estimated sequence diversity, which will in turn mean that genetic drift will play a more important role in the process of allele fixation, in some cases allowing low fitness mutations to reach high frequency. The repercussions on a larger evolutionary scale suggest that dengue virus emergence is not only a product of adaptive evolution (e.g., Twiddy et al. 2002; Bennett et al. 2003) but also subject to important stochastic processes.

Estimates of both the effective number of infections and epidemiologic incidence corroborate the same dynamic pattern of virus transmission. The exceptions were an increase in isolations late in 1988 not reflected in the sequence-based estimates of N_e and an increase in N_e in 1991/1992 not reflected by an increase in isolates. Incidence measures, such as isolate counts and cases reported, are rough estimates of virus census population size whose error rates are subject to accurate diagnostic rates by medical personnel involved, reporting rates, which can fluctuate in response to media coverage, for example, and the rate at which suspect cases are submitted for isolation. Estimates of the effective population size presented here, which in turn are based on sequence diversity and coalescent rates under neutral evolution, provide additional and distinct insight into the study of relative changes in population demographics that can complement passive surveillance data. Where case identification of dengue is

dependent on many factors including host response and immunologic conditioning, the sequences derived from the resultant isolates are products of the transmission arena and reflective of the transmitting population as they move from person to person via mosquitoes. Furthermore, both sequence relationships and substitution rates can be estimated precisely using the methods employed by BEAST. Concerns that oversampling during epidemic years may generate the cyclical pattern in population estimates have been addressed by simulations that show that clumping of sampling dates does not generate artefacts in the Bayesian skyline plot (Rambaut et al. 2008); gaps in sampling dates may miss peaks and troughs but will not generate peaks and troughs that do not exist. We attempted to reduce gaps in sampling dates and thus the chance of missing peaks or troughs in effective population size N_e , by sequencing additional isolates from interepidemic periods.

Sources of error in estimates of N_e stem from assumptions of the coalescent that sequences are derived from a single panmictic population not undergoing natural selection, all of which can affect genetic diversity along with population size (discussed in Carrington et al. 2005). The relationships between viruses from PR and elsewhere were analyzed phylogenetically to determine the extent of virus importation and population subdivision, both of which were minimal relative to in situ evolution (Bennett et al. 2003; Carrington et al. 2005). Imported strains will appear to maintain diversity across population level drops and thus will reduce the signal for bottlenecks. Similarly, although adaptive evolution has been identified in DENV-4 from PR, it was only confirmed in the 1998 lineage and its effect would be to reduce the diversity and therefore effective population size estimates at this time point.

We estimated the periodicity in DENV-4 transmission in PR at 3–5 years based on effective population size estimates from sequences. Gubler (1998) notes a 3- to 5-year period in peak transmission regardless of serotype in many populations. A spatiotemporal analysis of DHF cases only (without serotype breakdown) in Thailand reported a 3-year harmonic (Cummings et al. 2004). Another study noted a within serotype periodicity in transmission of 7–9 years in Thailand (Nisalak et al. 2003), whereas a Singapore-based study reported a 5- to 6-year periodicity in cases, regardless of serotype (Ooi et al. 2006). The difficulty in comparing these studies directly lies in differences in the variable measured (e.g., DHF cases, dengue cases, and serotype isolates). We argue that sequence data provide an additional and powerful tool to acquire precise estimates of relative population size change in terms of effective numbers of infections and that these estimates in the BEAST analytical framework could potentially provide a common parameter space for the testing of standard epidemiologic models based on the basic reproductive number (R_0).

Several studies have successfully applied the Bayesian coalescent methodology used here to elucidate the demographics of emerging pathogens such as influenza (Rambaut et al. 2008) and dengue (Carrington et al. 2005; reviewed in Pybus and Rambaut 2009). Here, we build

on earlier studies and include monthly resolution isolate sequences to provide a phylodynamic analysis on a common evolutionary and ecological timescale of virus population dynamics during the emergence of an important pathogen, confirming a pattern of peak transmission events capitulated in the epidemiologic data, separated by population bottlenecks that could have great evolutionary significance.

Acknowledgments

Thanks goes to Vance Vorndam, Manuela Beltran, and Maritza Chirivella for their involvement since the beginning with the dengue virus type 4 isolations and sequencing from the CDC collection. Research was supported by NIH-RR018727, NIH-AI065359, NIH-RR003061, DOD-06187000, and NSF-OIA0554657.

References

- Adams B, Holmes E, Zhang C, Mammen M, Nimmannitya S, Kalayanarooj S, Boots M. 2006. Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok. *Proc Natl Acad Sci USA*. 103(38):14234–14239.
- Bennett S, Holmes E, Chirivella M, Rodriguez D, Beltran M, Vorndam V, Gubler D, McMillan W. 2003. Selection-driven evolution of emergent dengue virus. *Mol Biol Evol*. 20(10):1650–1658.
- Burke D, Nisalak A, Johnson D, Scott R. 1988. A prospective study of dengue infections in Bangkok. *Am J Trop Med Hyg*. 38(1):172–180.
- Carrington CVF, Foster J, Pybus OG, Bennett SN, Holmes EC. 2005. Invasion and maintenance of dengue virus type 2 and type 4 in the Americas. *J Virol*. 79(23):14680–14687.
- Cummings D, Irizarry R, Huang N, Endy T, Nisalak A, Ungchusak K, Burke D. 2004. Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. *Nature* 427(6972):344–347.
- Dietz V, Gubler D, Ortiz S, Kuno G, Casta-Velez A, Sather G, Gomez I, Vergne E. 1996. The 1986 dengue and dengue hemorrhagic fever epidemic in Puerto Rico: epidemiologic and clinical observations. *Puerto Rico Health Sci J*. 15(3):201–210.
- Drummond A, Nicholls G, Rodrigo A, Solomon W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161(3):1307–1320.
- Drummond A, Rambaut A. 2007. BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 7:214.
- Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. 11(5):725–736.
- Gubler D. 1987. Dengue and dengue hemorrhagic fever in the Americas. *Puerto Rico Health Sci J*. 6(2):107–111.
- Gubler D. 1998. Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev*. 11(3):480–496.
- Gubler D. 2006. Dengue/dengue haemorrhagic fever: history and current status. *Novartis Found Symp*. 277:3–16.
- Held L, Hofmann M, Hohle M, Schmid V. 2006. A two-component model for counts of infectious diseases. *Biostatistics* 7(3):422–437.
- Johansson M, Dominici F, Glass G. 2009. Local and global effects of climate on dengue transmission in Puerto Rico. *PLoS Negl Trop Dis*. 3(2):e382.
- Kroeger A, Nathan M. 2006. Dengue: setting the global research agenda. *Lancet*. 368(9554):2193–2195.

- Li C, Lim T, Han L, Fang R. 1985. Rainfall, abundance of *Aedes aegypti* and dengue infection in Selangor, Malaysia. *Southeast Asian J Trop Med Pub Health*. 16(4):560–568.
- Minin V, Bloomquist E, Suchard M. 2008. Smooth skyline through a rough skyline: bayesian coalescent-based inference of population dynamics. *Mol Biol Evol*. 25(7):1459–1471.
- Morrison A, Getis A, Santiago M, Rigau-Perez J, Reiter P. 1998. Exploratory space-time analysis of reported dengue cases during an outbreak in Florida, Puerto Rico, 1991–1992. *Am J Trop Med Hyg*. 58(3):287–298.
- Nisalak A, Endy T, Nimmannitya S, Kalayanarooj S, Thisayakorn U, Scott R, Burke D, Hoke C, Innis B, Vaughn D. 2003. Serotype-specific dengue virus circulation and dengue disease in Bangkok, Thailand from 1973 to 1999. *Am J Trop Med Hyg*. 68(2):191–202.
- Ooi E, Goh K, Gubler D. 2006. Dengue prevention and 35 years of vector control in Singapore. *Emerg Infect Dis*. 12(6):887–893.
- Pond S, Frost S, Muse S. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21(5):676–679.
- Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol*. 25(7):1253–1256.
- Pybus OG, Rambaut A. 2009. Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet*. 10:2–13.
- Rambaut A, Pybus OG, Nelson M, Viboud C, Taubenberger J, Holmes E. 2008. The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453(7195):615–619.
- Recker M, Blyuss K, Simmons C, Hien T, Wills B, Farrar J, Gupta S. 2009. Immunological serotype interactions and their effect on the epidemiological pattern of dengue. *Proc Biol Sci*. 276(1667):2541–2548.
- Rigau-Perez J, Ayala-Lopez A, Garcia-Rivera E, Hudson S, Vorndam V, Reiter P, Cano M, Clark G. 2002. The reappearance of dengue-3 and a subsequent dengue-4 and dengue-1 epidemic in Puerto Rico in 1998. *Am J Trop Med Hyg*. 67(4):355–362.
- Rigau-Perez J, Vorndam A, Clark G. 2001. The dengue and dengue hemorrhagic fever epidemic in Puerto Rico, 1994–1995. *Am J Trop Med Hyg*. 64(1–2):67–74.
- Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol*. 57(5):758–771.
- Strickman D, Kittayapong P. 2002. Dengue and its vectors in Thailand: introduction to the study and seasonal distribution of *Aedes* larvae. *Am J Trop Med Hyg*. 67(3):247–259.
- Suchard M, Rambaut A. 2009. Many-core algorithms for statistical phylogenetics. *Bioinformatics* 25(11):1370–1376.
- Twiddy S, Woelk C, Holmes E. 2002. Phylogenetic evidence for adaptive evolution of dengue viruses in nature. *J Gen Virol*. 83(Pt 7):1679–1689.