

Inferring the Rate and Time-Scale of Dengue Virus Evolution

S. Susanna Twiddy, Edward C. Holmes, and Andrew Rambaut

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

Dengue is often referred to as an emerging disease because of the rapid increases in incidence and prevalence that have been observed in recent decades. To understand the rate at which genetic diversification occurs in dengue virus and to infer the time-scale of its evolution, we employed a maximum likelihood method that uses information about times of virus sampling to estimate the rate of molecular evolution in a large number of viral envelope (E) gene sequences and to place bounds around the dates of appearance of all serotypes and specific genotypes. Our analysis reveals that dengue virus generally evolves according to a molecular clock, although some serotype-specific and genotype-specific rate differences were observed, and that its origin is more recent than previously suggested, with the virus appearing approximately 1000 years ago. Furthermore, we estimate that the zoonotic transfer of dengue from sylvatic (monkey) to sustained human transmission occurred between 125 and 320 years ago, that the current global genetic diversity in the four serotypes of dengue virus only appeared during the past century, and that the recent rise in genetic diversity can be loosely correlated both to human activities such as population growth, urbanization, and mass transport and to the emergence of dengue hemorrhagic fever as a major disease problem.

Introduction

Dengue is the most common vector-borne viral disease of humans, infecting in excess of 50 million people in tropical and subtropical regions each year (World Health Organization, 1999). Where infection results in overt disease, the most common form is an acute febrile disease similar to influenza (classical dengue fever [DF]). However, in a minority of cases, this disease progresses to spontaneous hemorrhaging (dengue hemorrhagic fever [DHF]) and, most seriously, to dengue shock syndrome (DSS), characterized by a lack of detectable blood pressure and/or pulse. Case-fatality rates for the latter syndromes can be as high as 5% (World Health Organization, 2001).

The agent of these diseases, dengue virus, is an RNA virus with a positive-sense genome of approximately 11 kb, belonging to the genus *Flavivirus* and existing as four genetically distinct serotypes (DEN-1 to DEN-4). There are two distinct transmission cycles for dengue. The first generally involves transmission to human hosts in urban areas by the “domesticated” mosquito, *Aedes aegypti*, while the second, the sylvatic, or jungle, transmission cycle involves nonhuman primates as the major vertebrate hosts (although humans living or working in the forest or its fringes are occasionally infected), with the main vectors being canopy-dwelling *Aedes* mosquitoes (Rudnick 1986).

The prevalence of all four dengue serotypes has risen dramatically in recent years, accompanied by an increase in genetic diversity within each serotype. This diversity was previously demonstrated in an analysis of the branching structure of dengue phylogenies, which revealed a simultaneous increase in the number of viral lineages in all serotypes over the past 200 years (Zanotto et al. 1996). Clearly, such a rise in genetic diversity may also have important phenotypic implications, such as the emergence of viruses with altered antigenicity, virulence, or tissue tropism. Therefore, information about the rates of

nucleotide substitution in dengue virus not only provides information about its epidemiological history but also is crucial to our understanding of the processes controlling viral evolution and, consequently, to predicting responses to drug treatments or vaccination programs (Holmes and Burch 2000).

Previous studies of the rate and time frame of dengue evolution have employed several different methods. Zanotto et al. (1996) compared phylogenetically independent pairs of dated sequences with an outgroup sequence, determining the rate for each pair of sequences by dividing the divergence that has occurred between their sampling by the difference in isolation times. In this manner, a rate estimate for nonsynonymous substitutions in the dengue envelope (E) gene of mosquito-borne flaviviruses (most of which were dengue) of 7.5×10^{-5} subs/site/year was obtained. With this rate in hand, the origin of the four serotypes was placed at around 1500 to 2000 years ago, with the rapid increase in genetic diversity taking place within the past 200 years, coincident with the increased size and mobility of the human host population over this period. Lanciotti, Gubler, and Trent (1997) employed linear regression to a scatter plot of genetic distance from tree root to tip against date of isolation to estimate an overall substitution rate (i.e., combining both synonymous and nonsynonymous rates) in the E gene of DEN-4 of 8.3×10^{-4} subs/site/year. More recently, Wang et al. (2000) used both these methods to estimate substitution rates in DEN-2, producing an overall rate estimate of approximately 6×10^{-4} subs/site/year. This rate was extrapolated to serotypes 1 and 4, placing the divergence between sylvatic and human strains of DEN-1 at 200 ± 100 years before the present, with those of DEN-2 and DEN-4 at 1000 ± 500 years ago and 600 ± 300 years ago, respectively.

However, none of these methods can be considered statistically rigorous. For pairwise comparison methods, the number of suitable pairs of sequences in any data set is usually limited. Furthermore, due to the stochastic nature of the substitution process, sequences sampled earlier may exhibit more divergence from the outgroup than those sampled later, resulting in a negative rate estimate (Suzuki et al. 1999). Such cases cannot be easily included in the

Key words: dengue virus, substitution rate, divergence time, maximum likelihood, E gene.

E-mail: susanna.twiddy@zoo.ox.ac.uk.

Mol. Biol. Evol. 20(1):122–129. 2003

DOI: 10.1093/molbev/msg010

© 2003 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

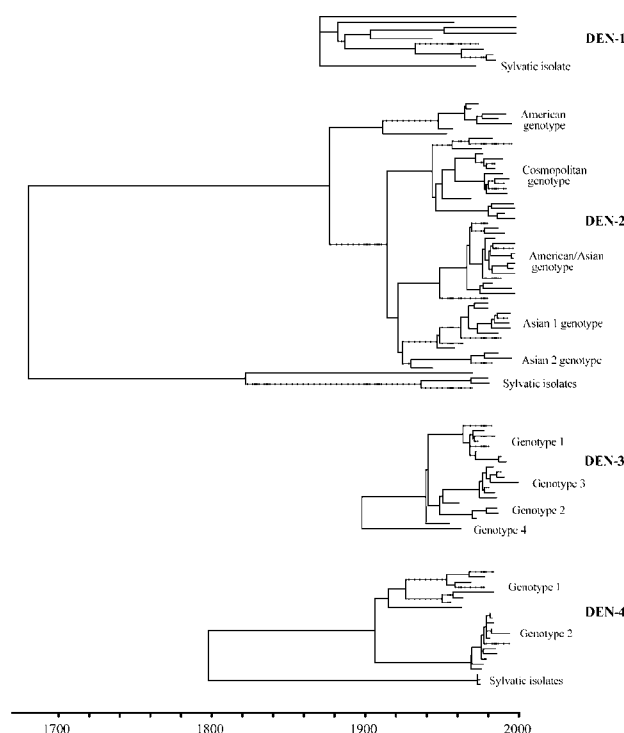


FIG. 1.—Maximum likelihood (ML) phylogenies for the dengue virus data sets used in this study, aligned according to age of their MRCA. Horizontal branch lengths are proportional to time.

analysis, but excluding them will bias the rate estimate upwards. In the case of fitting a regression line to a scatter plot, the data points cannot be considered to be independent, as any phylogenetic structure present in the data is ignored. Finally, both methods assume a constant rate of evolution across the entire data set but are unable to test this assumption.

The aim of our study was to use the dengue virus sequence data available on GenBank to estimate substitution rates and divergence dates, with appropriate confidence intervals, for all four serotypes of dengue and for the genotypes within each serotype where they were readily identifiable. To achieve this, we employed a maximum likelihood (ML) method (Rambaut 2000), which incorporates phylogenetic relationships and information about sampling times among sequences isolated over a range of up to 60 years, thereby allowing calibration of the long-term evolutionary history of dengue virus. The advantages of this ML approach are that it (1) allows assumptions of the model (in this case, that of a single substitution rate) to be tested directly using the likelihood ratio test (LRT), (2) uses all the sequences within the data set and takes into account phylogenetic relationships among them to produce appropriate confidence intervals around estimates of substitution rate and divergence time (more specifically, the age of the most recent common ancestor [MRCA]), and (3) allows the use of complex models of nucleotide substitution, including those incorporating rate heterogeneity among sites, which can bias the estimate of divergence times if ignored.

Materials and Methods

A primary data set was compiled consisting of an alignment of all available E gene sequences on GenBank. This was further divided into four subsets comprising the serotypes. Where sylvatic (monkey or jungle mosquito) strains were available (DEN-1, DEN-2, and DEN-4 only), the data set for each serotype was analysed with and without the sylvatic strains. This enabled us to obtain estimates for the age of the MRCAs of the human strains of dengue, as well as the divergence time of the human and sylvatic strains, thereby giving an upper bound on the date of emergence of DEN as a human disease. In addition, where sufficient dated sequences were available, the distinct genotypes previously identified in DEN-2, DEN-3, and DEN-4 were analyzed separately.

The following criteria were used when compiling the data sets: (1) all sequences had a known date of isolation; (2) only a single sequence was taken from any one patient; (3) all known recombinant sequences (identified in Holmes, Worobey, and Rambaut 1999; Worobey, Rambaut, and Holmes 1999; Toulou et al. 2001) were excluded; (4) where data sets were very large (>50 sequences), all strains that were $>99\%$ similar to any other strain in the data set were removed from the analysis. Full details of the sequences used in the analysis are available as supplementary information at <http://www.molbiolvol.org>.

The PAUP* program (Swofford 2001) was used to infer ML trees for each data set. The model of nucleotide substitution used was the general time-reversible (GTR) model with a different substitution rate for each codon position. This model generally provided a better fit to the data than those using the gamma distribution to model among-site rate heterogeneity. The relative rates for the codon positions and the GTR substitution matrix were coestimated with the ML tree, while the frequencies of the four nucleotides were estimated empirically from the data. In all subsequent analyses, the same model of substitution was used with the relevant parameters reestimated on each occasion. All parameter values are available from the authors on request.

ML trees were rooted using well-supported outgroup sequences. Where sylvatic sequences were available, these were used to root the ML trees of individual dengue serotypes, since a previous phylogenetic study revealed that sylvatic sequences were basal within their respective serotypes (Wang et al. 2000). For the human-only data sets, outgroups consisting of two human strains of each of the serotypes not under analysis were utilized, and for the genotypes of DEN-2, which have previously been demonstrated to form well-supported monophyletic groups (Twiddy et al. 2002), outgroups comprised strains from the genotypes not under analysis. Finally, for the phylogeny of all four serotypes, the ML root was estimated because there are no close outgroup sequences from other flaviviruses with which the ML tree can be reliably rooted. The ML trees obtained for each data set are shown in figures 1 and 2.

Three phylogenetic models of sequence evolution were used in this study. The first is that described by Felsenstein (1981) and labeled the “different rate” (DR)

model by Goldman (1993), as each branch has its own rate. The second assumes that all branches have the same rate of evolution (i.e., a molecular clock) and that all sequences are contemporaneous. This is referred to as the “single rate” (SR) model. The third model is an extension of the SR model that relaxes the assumption of contemporaneous tips. This model, designated the “single rate dated tip” (SRDT) model (Rambaut 2000), uses information about the date of isolation of each sequence to estimate the substitution rate for each data set and is at the heart of our analysis.

An LRT was used to examine the fit of each model to the data. The DR model is the most general, allowing a different substitution rate for each branch of the unrooted ML tree, so that for a tree of n taxa, this model has $2n-3$ free parameters. The SR model is the most specific, with a single rate for all the branches, giving $n-1$ free parameters (one for each internal node in the rooted tree). The SRDT model adds an additional free parameter to the SR model (the substitution rate). A particular model can be rejected in favor of a more general one if twice the difference in the log likelihood between the models is greater than the χ^2 critical value, with df equal to the difference in free parameters between the models. This is the test of the molecular clock suggested by Felsenstein (1981), based on a standard statistical result (Wilks 1938). The χ^2 critical value for the LRTs was corrected for multiple tests using the Dunn-Sidak method (Ury 1976). This correction was applied for sets of independent tests such as the four serotypes or the genotypes within each serotype.

The rate of nucleotide substitution was estimated for each data set using the SRDT model. Confidence intervals (CIs) were estimated by finding the hypothesized substitution rates above and below the ML rate that had log likelihoods that were 1.92 less than that of the ML rate. The value of 1.92 is the χ^2 critical value ($P = 0.05$) with 1 df . All these analyses were undertaken using the program TipDate, version 1.1 (Rambaut 2000).

The use of these models involves an assumption about the topology of the ML tree for each data set: When analyzing the data under the SR and SRDT models, we are using the tree shape (branching order) that was estimated using the DR model. This can be justified for the following reasons: (1) the estimates of substitution rate under the SRDT model have been shown in simulations to be robust to errors in tree reconstruction (Rambaut 2000), (2) the tree shape was estimated using a model (DR) that makes fewer assumptions about the rate of substitution than either the SR or the SRDT model, and (3) if there exists, for the SR or SRDT model, a tree shape which has a higher likelihood than that estimated under the DR model, then the LRT will be liberal, being more likely to reject the molecular clock.

Results

Tests of Fit for Models of Sequence Evolution

The likelihood ratio tests of fit for the SR and SRDT models (compared with the DR model) in each data set are given in table 1. In all cases, the SR model is strongly

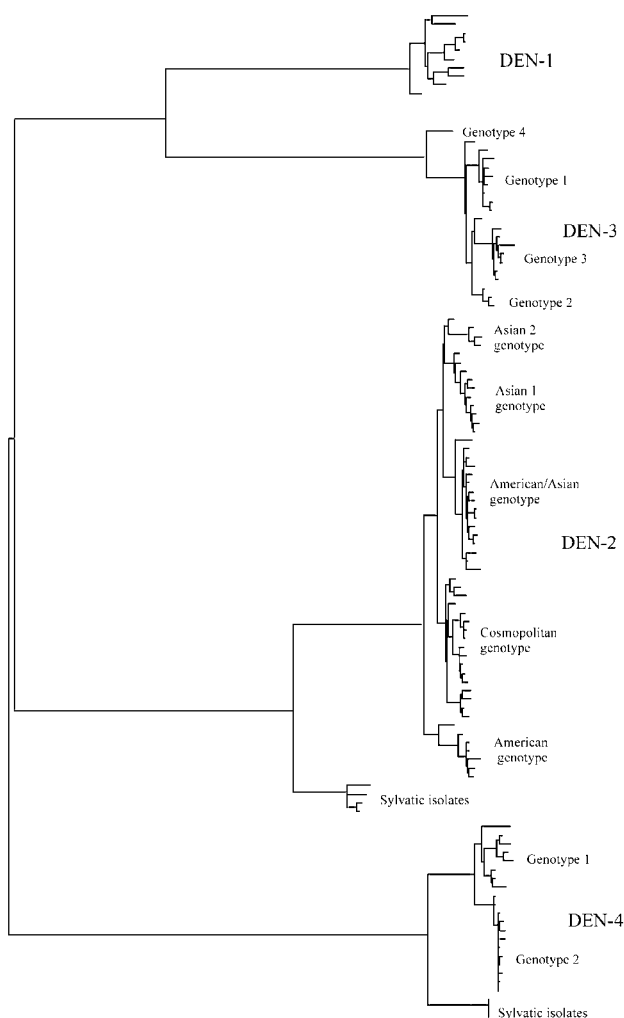


FIG. 2.—ML phylogeny of all four dengue serotypes estimated using representative taxa.

rejected as an adequate description of the evolution of dengue virus, presumably because this model does not accommodate for the different dates of sampling of the isolates.

For both DEN-1 data sets, the SRDT model is not significantly worse than the DR model, indicating that this model adequately describes the substitution process. In contrast, in both DEN-2 data sets, the DR model is significantly favored over the SRDT model ($P = 0.03$ in both cases). However, when the Dunn-Sidak correction for multiple tests is applied, the clock cannot be rejected for either DEN-2 data set. Removal of the sylvatic strains in the analysis of the human-only DEN-2 data set makes very little difference to either the fit of the model or the estimated substitution rate (table 2).

Of the five genotypes of DEN-2, only four are suitable for our rate analysis, having sufficient numbers of taxa isolated over a range of dates: American, Asian 1, American/Asian, and Cosmopolitan. Among these, in the American, American/Asian, and Cosmopolitan genotypes, the SRDT model is not significantly worse than the DR

Table 1
Likelihood Ratio Tests for Different Models of Sequence Evolution in Dengue Virus

Serotype	Genotype	Number of Sequences	LRT (SR vs. DR)	df	$P(\chi^2)$	LRT (SRDT vs. DR)	df	$P(\chi^2)$	Significant (Clock Rejected)?	
DEN-1	...	All	10	28.29	8	4.23×10^{-4}	9.28	7	9.28	No
DEN-2	...	All	58	222.16	56	1.19×10^{-21}	77.30	55	0.03	Yes ^a
DEN-4	...	All	23	93.47	21	4.04×10^{-11}	36.76	20	0.01	Yes
DEN-1	...	All human	9	25.85	7	5.36×10^{-4}	3.10	6	0.80	No
DEN-2	...	All human	54	218.51	52	2.71×10^{-22}	72.29	51	0.03	Yes ^a
DEN-3	...	All human	21	134.06	19	2.48×10^{-19}	22.23	18	0.22	No
DEN-4	...	All human	20	103.40	18	5.26×10^{-14}	33.83	17	0.01	Yes
DEN-2	...	Asian 1	31	82.43	29	5.07×10^{-7}	45.17	28	0.02	Yes ^a
		American/Asian	29	153.97	27	9.72×10^{-20}	24.10	26	0.57	No
		Cosmopolitan	25	57.45	23	8.85×10^{-5}	26.53	22	0.23	No
		American	10	66.36	8	2.60×10^{-11}	7.76	7	0.35	No
DEN-3	...	1	8	28.97	6	6.16×10^{-5}	6.92	5	0.23	No
		3	8	47.77	6	1.32×10^{-8}	6.00	5	0.31	No
DEN-4	...	1	12	44.90	10	2.27×10^{-6}	17.69	9	0.04	Yes ^a
		2	8	61.12	6	2.66×10^{-11}	9.50	5	0.09	No
All	112	5094.71	110	0.00	166.91	110	3.03×10^{-4}	Yes	

^a Not significant (clock cannot be rejected) when Dunn-Sidak correction for multiple tests is applied to the χ^2 critical value.

model ($P = 0.35, 0.57$, and 0.28 , respectively). However, the LRT for the Asian 1 data set indicates that the substitution rate within this genotype is probably not constant, as the SRDT model is rejected in favor of the DR model, although once again when the Dunn-Sidak correction is applied, the likelihood difference between the two models is no longer significant.

DEN-3 is the only serotype for which no sylvatic strains are currently available, although DEN-3 antibodies have been discovered in nonhuman primates in Malaysia, suggesting that a sylvatic DEN-3 cycle also exists (Rudnick 1986). Consequently, the single data set for DEN-3 comprises strains isolated only from humans. In these data, the SRDT model is not significantly worse than the DR model, revealing an approximately constant rate of substitution in this serotype. Four genotypes have been

identified in an E gene phylogeny of DEN-3 (Lanciotti et al. 1994). In our analysis, there are insufficient sequences from genotype 2 for analysis, and of the three sequences making up genotype 4, two (Puerto Rico 1977 and Tahiti 1965) were excluded from the analysis as suspected recombinants (Worobey, Rambaut, and Holmes 1999). However, analysis of genotypes 1 and 3 reveals clocklike evolution in both ($P = 0.23$ for genotype 1 and 0.31 for genotype 3), with similar rates of substitution in each genotype (table 2).

For both the human-only and sylvatic DEN-4 data sets, the SRDT model is rejected in favor of the DR model, with $P = 0.01$ in each case. Two genotypes of DEN-4 were identified in a previous study of DEN-4 (Lanciotti, Gubler, and Trent 1997) and are also discernible in the ML tree for this serotype. Separate analysis of these reveals

Table 2
Maximum Likelihood Estimates of Substitution Rates and Divergence Times in Dengue Virus

Serotype	Genotype	ML Rate ($\times 10^{-4}$ subs/site/year)	CI ($\times 10^{-4}$ subs/site/year)	MRCA (years)	CI (MRCA-years)
DEN-1 All	4.55	2.54–6.54	126.21	96.37–208.58
DEN-2 All	6.07	5.08–7.09	320.90	280.15–378.00
DEN-4 All	6.02	4.29–8.07	195.32	152.25–264.80
DEN-1 All human	5.11	3.08–7.16	111.81	89.17–170.28
DEN-1 (Zanotto et al.) ^a All human			170	140–200
DEN-2 All human	6.16	5.15–7.19	121.78	110.10–138.75
DEN-2 (Zanotto et al.) ^a All human			215	181–249
DEN-3 All human	9.01	7.27–10.84	100.89	90.80–116.22
DEN-3 (Zanotto et al.) ^a All human			178	147–209
DEN-4 All human	6.89	5.09–9.00	76.32	64.26–95.56
DEN-4 (Zanotto et al.) ^a All human			119	94–134
DEN-2 Asian 1	5.42	3.43–7.47	51.76	46.31–65.33
 American/Asian	10.05	7.82–12.59	35.12	31.79–40.35
 Cosmopolitan	6.20	4.00–8.45	51.00	42.44–71.33
 American	6.54	4.61–8.92	82.63	71.10–101.79
DEN-3 1	7.48	4.47–10.72	29.00	24.08–41.24
 3	11.58	7.76–15.86	46.67	41.47–57.28
DEN-4 1	5.42	3.03–8.21	26.03	22.47–37.11
 2	8.91	6.32–11.79	54.91	47.01–69.74
All	6.48	n/a ^b	1114.80	n/a ^b

^a Estimates provided by Zanotto et al. (1996).

^b Not available.

that while for genotype 1 the SRDT model is rejected ($P = 0.04$; however this rejection of the clock is lost on application of the Dunn-Sidak correction), genotype 2 displays more clocklike behavior, and the SRDT model is not significantly worse than the DR model ($P = 0.09$). This genotype-specific difference may explain the failure of the clock model in the data sets containing all strains, although the substitution rates in the two genotypes are not significantly different.

Our final data set consisted of strains from all four viral serotypes. Here the SRDT model is strongly rejected in favor of the DR model, indicating that we cannot assume that the substitution rate is the same in the four dengue serotypes.

Rates of Nucleotide Substitution in Dengue Viruses

Table 2 shows the estimated rate of nucleotide substitution in each of the dengue data sets analyzed. These rates range from 4.55×10^{-4} for the DEN-1 data set including the sylvatic strains, to 11.58×10^{-4} for genotype 3 of DEN-3. The CIs for the DEN-1, DEN-2, and DEN-4 data sets overlap considerably, indicating that rates of substitution do not differ significantly among these serotypes. However, while the CIs for DEN-3 overlap with those for the DEN-4 data set, the lower CI for DEN-3 (7.27×10^{-4}) is higher than the upper CI for both the DEN-1 (7.09×10^{-4}) and DEN-2 (7.19×10^{-4}) data sets. This suggests that DEN-3 is evolving significantly faster than both DEN-1 and DEN-2, but not DEN-4. Similarly, the substitution rate in the American/Asian genotype is significantly higher than that of the Asian 1 genotype, but not the American or Cosmopolitan genotypes. A graphical representation of these rate differences is shown in figure 3.

The Time-Scale of Dengue Virus Evolution

The estimated age of the MRCA of each dengue data set, with appropriate confidence intervals, is shown in table 2. The dates for the data sets that combine human and sylvatic strains are approximately 320 years ago for DEN-2, 200 years ago for DEN-4, and 125 years ago for DEN-1. These dates constitute the earliest estimate for the date of the zoonosis in each serotype (that is, the transfer of virus from monkeys to humans) and predate those of human-only data sets in all cases by approximately 15 to 200 years. The estimated ages of the human-only strains can be taken as the latest dates for the zoonotic transmission of the dengue serotypes to humans, and hence the earliest dates for the beginning of the epidemic spread of the viruses. For all four serotypes, these estimates place the beginning of epidemic transmission of dengue in humans near to the end of the nineteenth and the beginning of the twentieth century (fig. 1). Finally, our estimate for the MRCA of all four dengue serotypes is 1,115 years ago, although it is not possible to estimate CIs in this case.

Discussion

Our analysis of rates of nucleotide substitution in dengue virus reveals that the four serotypes, considered separately, exhibit clocklike or near-clocklike behavior.

Although caution must clearly be exercised when inferring divergence times from rate estimates where the DR model significantly outperforms the SRDT model, a previous simulation study has shown that even these nonclock rates should be reliable indicators of average substitution rate providing that data sets are large, sequence length is >100 bp, and the inclusion of isolation dates into a single-rate model improves its likelihood (Jenkins et al. 2002). This is the case for all our data sets; even though six of 12 data sets rejected the molecular clock, in all cases the likelihood for the SRDT model was higher than that of the SR model. Furthermore, in no case was the SRDT model rejected for all comparisons within a serotype, and when the Dunn-Sidak correction for multiple tests was applied, in six of seven data sets in which the clock had initially been rejected, the likelihood differences between DR and SRDT models were no longer significant.

There are a number of reasons why some dengue data sets might reject the molecular clock. Obviously, the rate variation observed among lineages could be real, as appears to be the case in DEN-4, indicating that its mechanistic basis warrants further investigation. However, it is also possible that the observed rate variation is artificial. One possible cause is recombination. Although all known recombinants were removed from the analysis, the possibility remains that some sequences have a recombinant history that is too complex to be easily detected by a simple phylogenetic survey (Worobey, Rambaut, and Holmes 1999). Another possible cause is artificial selection; to provide as wide a range of isolation dates as possible, sequences isolated as long ago as 1944 were included in the analysis. In some cases, these sequences have been repeatedly passaged through several cell types, which may have artificially accelerated the substitution rate through *in vitro* adaptation (Wain-Hobson 1992; Woelk et al. 2001). In this context, it is notable that in a reanalysis of the DEN-2 Asian 1 genotype but excluding two early and extensively passaged isolates (strains 16681 and Th-36), the DR model could no longer reject the SRDT model ($P = 0.24$). However, the removal of these sequences resulted in only a small difference in the estimated substitution rates.

In general, our rate estimates are similar to those previously reported for dengue virus (Lanciotti, Gubler, and Trent 1997; Wang et al. 2000; Jenkins et al. 2002), although the latter two studies used less statistically rigorous methods. Moreover, the mean substitution rates fall within the range seen in other RNA viruses (Jenkins et al. 2002), indicating that rates of mutation and replication must be similar among a diverse array of viruses that replicate using RNA polymerase. Despite this pattern of overall similarity, there was some localized rate variation, with DEN-3 and the DEN-2 American/Asian genotype both having significantly higher substitution rates than some other dengue strains. Why this should be so is unclear, particularly in the case of DEN-3, about which relatively little is known. For the American/Asian genotype of DEN-2, a possible explanation lies in this genotype's recent introduction into the Americas, where there were large numbers of susceptible humans, providing the necessary conditions for explosive viral transmission

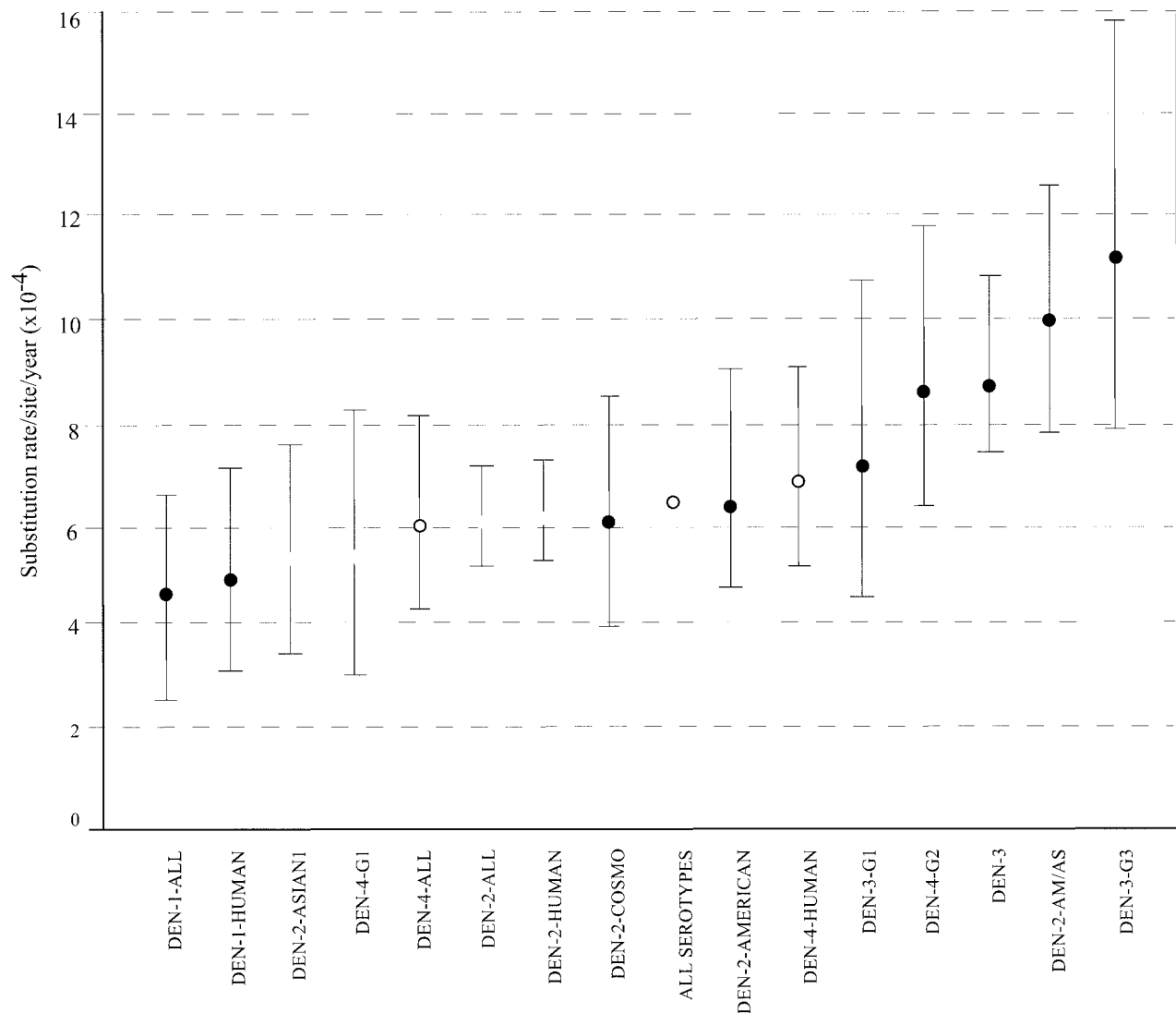


FIG. 3.—Nucleotide substitution rates per site per year for the dengue virus data sets ranked in ascending order. Filled circles indicate clock accepted; empty circles indicate clock rejected; grey circles indicate clock accepted when Dunn-Sidak correction is applied. DEN-1-ALL indicates DEN-1 all isolates; DEN-1-HUMAN indicates DEN-1 human isolates only; DEN-2-ASIAN1 indicates DEN-2 Asian 1 genotype; DEN-4-G1 indicates DEN-4 genotype 1; DEN-4-ALL indicates DEN-4 all isolates; DEN-2-ALL indicates DEN-2 all isolates; DEN-2-HUMAN indicates DEN-2 human isolates only; DEN-2-COSMO indicates DEN-2 Cosmopolitan genotype; DEN-2-AMERICAN indicates DEN-2 American genotype; DEN-4-HUMAN indicates DEN-4 human isolates only; DEN-3-G1 indicates DEN-3 genotype 1; DEN-4-G2 indicates DEN-4 genotype 2; DEN-3 indicates DEN-3 all isolates; DEN-2-AM/AS indicates DEN-2 American/Asian genotype; DEN-3-G3 indicates DEN-3 genotype 3.

(Guzman et al. 1984a, 1984b). Under epidemic conditions, each infected individual is likely to transmit the virus to multiple contacts, thus increasing the overall level of diversity in the population. After a period of sustained epidemic transmission, randomly sampled sequences are therefore expected to show more diversity than would be the case under endemic conditions, where transmission rates are lower.

Our analysis of divergence times has also provided a more recent time-scale of dengue virus evolution. Using a far smaller sample of sequences, Zanotto et al. (1996) estimated the date for the beginning of rapid cladogenesis in the four serotypes to be approximately 200 years ago, which they suggested coincided with the beginning of sustained human transmission. For comparative purposes,

the estimates of Zanotto et al. (1996) are also shown in table 2. In each case, our estimated MRCAs for the human-only data sets are more recent than those provided by Zanotto et al. and significantly so in the case of DEN-2 and DEN-3 (i.e., with no overlap of CIs). In the case of DEN-1 and DEN-4, there is a slight overlap in the confidence limits, although our estimates still suggest a substantially later MRCA for these serotypes. Our dates further suggest that the explosion of dengue activity in humans occurred within a short period of time in all four serotypes. The estimated ages of the MRCAs of the human strains from the four serotypes are within 50 years of each other, whereas in the Zanotto et al. study, the range of dates is almost twice as great. In this context, it is important to note that the earliest reports of significant

numbers of DHF cases are at the end of the nineteenth and beginning of the twentieth centuries (Gubler 1997), coinciding with our estimates for the start of widespread human transmission. At this time, human population growth rates were accelerating worldwide, world trade was increasing rapidly, and industrialization and urbanization were taking place on a global scale, creating large populations of susceptible hosts and fertile habitats for anthropophilic mosquito vectors. At the same time, the transport revolution began the process of global dispersal of both virus and vector (O'Brien 1999). Similarly, our estimates of the age of the MRCA of the four genotypes of DEN-2 analyzed ranged from 35 to 83 years ago, averaging 55 years ago. These dates coincide with World War II and its aftermath, in which mass population movement, ecological destruction, and the urbanization of Southeast Asia provided dengue virus with new opportunities for infection, sustained transmission, and worldwide dissemination. Indeed, the mass movement of people during and immediately after World War II has been cited as a reason for the large-scale emergence of DHF/DSS, the risk of which is thought to increase if sequential infections with heterogeneous serotypes occur (Halstead 1988).

Less attention has been directed toward determining the time of the most recent zoonotic transfers of dengue from monkeys to humans (i.e., the transfers giving rise to the current diversity of dengue strains in humans and monkeys). Wang et al. (2000) estimate that these events occurred at approximately 200 ± 100 , 1000 ± 500 , and 600 ± 300 years ago for DEN-1, DEN-2, and DEN-4, respectively. These dates are roughly three times greater than our estimates for DEN-2 and DEN-4 (320.90 and 195.32 years ago), although there is again some overlap in the case of DEN-1 (126 years ago, CI = 96.37–208.58). Our estimates for the age of the MRCAs of the sylvatic and human strains also suggest that DEN-2 was the first serotype to develop sustained transmission cycle in humans, significantly earlier than either DEN-1 or DEN-4, although this inference may be affected by small sample size.

Finally, we also provide a new estimate for the age of MRCA of all four serotypes at 1115 years ago. Again, there is some tentative epidemiological evidence to support this time-scale, with the first reported outbreak of possible denguelike illness occurring in China in the year 992 (Gubler 1997). Although this is substantially earlier than our estimates for the earliest human transmission of dengue, such sporadic outbreaks could occur after the introduction of a sylvatic strain of dengue into populations sufficiently large to transiently sustain viral transmission. Such "spillover" strains would eventually go extinct when the supply of susceptible hosts was exhausted and hence would not be represented in a sample of more recent dengue isolates. This is likely to have been the case for most outbreaks of dengue before urbanization and world travel became common, although a greater number of samples, especially from sylvatic strains, is again needed to confirm this hypothesis.

Overall, our study shows that the evolutionary history of dengue, in particular as a pathogen with sustained transmission in human populations, is more recent than was previously supposed, with the majority of the genetic

diversity we see today having appeared only during the past century, contemporaneous with its emergence as a global disease problem. Although little is known about the evolution of virulence in dengue virus, it is clear that an understanding of the rate at which genetic diversity is increasing must be beneficial in predicting possible changes in virulence or cell tropism, as well as responses to future vaccination programs. Additionally, by placing the evolutionary history of dengue within an appropriate time frame, we can better appreciate the factors contributing to its emergence, thereby providing a firm foundation for the design of effective control programs.

Acknowledgments

This work was supported by The Wellcome Trust and the Royal Society.

Literature Cited

- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**:368–376.
- Goldman, N. 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* **36**:182–198.
- Gubler, D. J. 1997. Dengue and dengue hemorrhagic fever: its history and resurgence as a global public health problem. Pp. 1–22 in D. J. Gubler and G. Kuno, eds. *Dengue and dengue hemorrhagic fever*. CAB International, New York.
- Guzman, M. G., G. P. Kouri, J. Bravo, M. Calunga, M. Soler, S. Vazquez, and C. Venereo. 1984a. DHF in Cuba. I. Serological confirmation of clinical diagnosis. *Trans. Roy. Soc. Trop. Med. Hyg.* **78**:235–238.
- Guzman, M. G., G. P. Kouri, J. Bravo, M. Soler, S. Vazquez, M. Santos, R. Villaescusa, P. Basanta, G. Indan, and J. M. Ballester. 1984b. DHF in Cuba. II. Clinical investigations. *Trans. Roy. Soc. Trop. Med. Hyg.* **78**:239–241.
- Halstead, S. B. 1988. Pathogenesis of dengue: challenges to molecular biology. *Science* **239**:476–80.
- Holmes, E. C., and S. S. Burch. 2000. The causes and consequences of genetic variation in dengue virus. *Trends Microbiol.* **8**:74–77.
- Holmes, E. C., M. Worobey, and A. Rambaut. 1999. Phylogenetic evidence for recombination in dengue virus. *Mol. Biol. Evol.* **16**:405–409.
- Jenkins, G. M., A. Rambaut, O. G. Pybus, and E. C. Holmes. 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J. Mol. Evol.* **54**:152–161.
- Lanciotti, R. S., D. J. Gubler, and D. W. Trent. 1997. Molecular evolution and phylogeny of dengue-4 viruses. *J. Gen. Virol.* **78**:2279–2286.
- Lanciotti, R. S., J. G. Lewis, D. J. Gubler, and D. W. Trent. 1994. Molecular evolution and epidemiology of dengue-3 viruses. *J. Gen. Virol.* **75**:65–75.
- O'Brien, P. K. 1999. *Philip's atlas of world history*. George Philip Limited, London.
- Rambaut, A. 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16**:395–399.
- Rudnick, A. 1986. Dengue virus ecology in Malaysia. *Bull. Inst. Med. Res. Malaysia* **23**:51–153.
- Suzuki, Y., K. Katayama, S. Fukushi, T. Kageyama, A. Oya, H. Okamura, Y. Tanaka, M. Mizokami, and T. Gojobori. 1999. Slow evolutionary rate of GB virus C/hepatitis G virus. *J. Mol. Evol.* **48**:383–389.

- Swofford, D. L. 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Mass.
- Toulou, H., P. Coussinier-Paris, J. P. Durand, V. Mercier, J. J. de Pina, P. de Micco, F. Billoir, R. N. Charrel, and X. de Lamballerie. 2001. Evidence for recombination in natural populations of dengue virus type 1 based on the analysis of complete genome sequences. *J. Gen. Virol.* **82**:1283–1290.
- Twiddy, S. S., J. J. Farrar, N. Vinh Chau, B. Wills, E. A. Gould, T. Gritsun, G. Lloyd, and E. C. Holmes. 2002. Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. *Virology* **298**:63–72.
- Ury, H. K. 1976. A comparison of four procedures for multiple comparisons among means (pairwise) for arbitrary sample sizes. *Technometrics* **18**:89–97.
- Wain-Hobson, S. 1992. Human immunodeficiency virus type 1 quaspecies in vivo and ex vivo. *Curr. Top. Microbiol. Immunol.* **176**:181–193.
- Wang, E., H. Ni, R. Xu, A. D. T. Barrett, S. J. Watowich, D. J. Gubler, and S. C. Weaver. 2000. Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. *J. Virol.* **74**:3227–3234.
- Wilks, S. S. 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* **9**:60–62.
- Woelk, C. H., L. Jin, E. C. Holmes, and D. W. G. Brown. 2001. Immune and artificial selection in the hemagglutinin (H) glycoprotein of measles virus. *J. Gen. Virol.* **82**:2463–2474.
- World Health Organization. 1999. Strengthening implementation of the global strategy for DF/DHF prevention and control. <http://www.who.int/vaccines-diseases/research/virus2.shtml>.
- . 2001. Vaccines, immunization, and biologicals: dengue and Japanese encephalitis vaccines. <http://www.who.int/vaccines-diseases/research/virus2.htm>.
- Worobey, M., A. Rambaut, and E. C. Holmes. 1999. Widespread intra-serotype recombination in natural populations of dengue virus. *Proc. Natl. Acad. Sci. USA.* **96**:7352–7357.
- Zanotto, P. M. de A., E. A. Gould, G. F. Gao, P. H. Harvey, and E. C. Holmes. 1996. Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc. Natl. Acad. Sci. USA.* **93**:548–553.

Keith Crandall, Associate Editor

Accepted September 17, 2002