Available online at www.sciencedirect.com
SCIENCE DIRECT®
Infection, Genetics and Evolution 3 (2003) 87–95
Infection, Genetics and Evolution
www.elsevier.com/locate/meegid

Comparative population dynamics of mosquito-borne flaviviruses

S.Susanna Twiddy*, Oliver G. Pybus, Edward C. Holmes

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Received 14 July 2002; received in revised form 27 November 2002; accepted 27 November 2002

Abstract

Among the members of the genus *Flavivirus* are several important human pathogens, including the dengue (DEN) and Japanese encephalitis (JE) viruses. From the analysis of gene sequence data of samples of these virus populations it is possible to infer phylogenetic relationships, which in turn can yield important epidemiological information, including their demographic history in humans. In this study, we use a recently developed method, based on coalescent theory, to infer the population dynamics of a variety of mosquito-borne flaviviruses. Our study involves the testing of alternative hypotheses, the estimation of confidence intervals around demographic model parameter values, and the placing of the maximum likelihood (ML) demographic model into a "real time" epidemiological history. We reveal that all the *Flavivirus* populations studied are growing at an exponential rate, with the rates of population growth of dengue virus serotypes 2 and 3 increasing rapidly in the recent past, and that of Japanese encephalitis virus changing from constant population size to exponential growth within the last century. We therefore demonstrate that the use of these coalescent methods may be extremely valuable in monitoring responses to interventions such as vaccination or vector control.

Keywords: *Flavivirus*; Population dynamics; Coalescent theory; E gene; Phylogeny

1. Introduction

The genus *Flavivirus* comprises a group of positive-sense, single-stranded RNA viruses, most of which are either mosquito- or tick-borne, and several of which are important human pathogens, notably yellow fever (YF; the prototype virus of the genus), dengue (DEN) and Japanese encephalitis (JE). Many of these are also considered 'emerging viruses', classified as such due to recent increases in prevalence and/or geographic distribution, with the likelihood of continued increases in incidence in the future (Morse, 1994). This was most recently documented in West Nile (WN) virus, which was the agent of a highly publicised outbreak of human encephalitis in the United States in 1999 (Petersen and Roehrig, 2001). Likewise, the global morbidity and mortality due to dengue virus has grown dramatically in recent decades (WHO, 2001), and the geographical distribution of JE, which has an annual incidence of up to 50,000 and a case fatality rate of 33%, (Solomon et al., 2000) is continuing to advance westwards from Southeast Asia.

Phylogenetic relationships within *Flavivirus* populations can be inferred from viral gene sequences and yield important evolutionary information, such as the geographical origin of epidemic strains of the virus (Rico-Hesse, 1990; Rico-Hesse et al., 1997), the rate and variability of nucleotide substitution among strains (Rambaut, 2000), the time-frame of the origin and evolution of the virus (Zanotto et al., 1996; Wang et al., 2000; Twiddy et al., in press), whether recombination between virus strains has occurred in the past (Holmes et al., 1999; Worobey et al., 1999; Toulou et al., 2001; Twiddy and Holmes, in press), and, most pertinent to this study, the demographic history of populations, that is, whether the viral population has remained at a constant size or has grown over time, and if the latter, what the pattern of that growth has been.

An earlier method of inferring population dynamics from molecular phylogenies was developed by Nee et al. (1995) and applied to virus populations by Holmes et al. (1995). The rationale behind this method was that the branching pattern of a phylogeny provides information about the frequency of transmission events over time. This branching pattern can be graphically represented using a 'lineages-through-time plot', in which the logarithm of the number of lineages in the phylogeny is plotted against the time at which they appear. Such plots differ for populations with a history of constant population size and those with a history of exponential growth. Using this technique, Zanotto et al. (1996) showed that the epidemic history of the four serotypes of

* Corresponding author. Tel.: +44-1865-281987; fax: +44-1865-310447.
E-mail address: susanna.twiddy@zoo.ox.ac.uk (S.S. Twiddy).

1567-1348/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
doi:10.1016/S1567-1348(02)00153-3

dengue virus consisted of two stages of exponential growth, with a low rate of growth changing at a particular point in time to one of rapid population expansion. Zanotto et al. (1996) dated this change to approximately 200 years before the present, and speculated that it could be correlated with the concurrent rapid growth of human population size. A similar result was obtained for JE in the same study.

In the current study, we use a new method for inferring demographic history from sequence data (Pybus et al., 2000), which, like the above method, is based on coalescent theory, in this case the variable population size model (Griffiths and Tavare, 1994). However, unlike previous approaches, it allows explicit testing of alternative demographic hypotheses within a maximum likelihood (ML) framework, as well as measures of statistical confidence around parameter estimates. This method has recently been applied to hepatitis C virus (HCV), revealing in this case different patterns of demographic history in different subtypes of HCV that could be correlated with differences in modes of transmission (Pybus et al., 2001). In addition, the method has recently been extended to incorporate the time at which viral isolates were sampled, and as a result is able to place results of analysis into a natural time-scale of years.

Here, for the first time, we use this method to make inferences about the demographic history of vector-borne viruses. Unlike previous attempts to infer the demographic history of dengue viruses (Zanotto et al., 1996), we consider each of the four dengue serotypes separately in our analysis. This is a more appropriate approach, since each serotype is the result of a separate zoonotic transfer from non-human primates to humans ('sylvatic' isolates group at the root of each serotype; Wang et al., 2000) and, therefore represents a single, distinct, epidemic. We demonstrate that populations of at least three serotypes of dengue, the most important of the *Flaviviridae* in terms of human morbidity and mortality, continue to grow exponentially, with a change to a higher rate of population growth in the recent past (approximately 30–35 years ago). We also reveal that a population of dengue virus serotype 2 (DEN-2) recently introduced into Venezuela has been exponentially growing at a faster rate than any virus population yet analysed, and therefore that this strain should be considered to have very high epidemic potential given suitable environmental conditions. Finally, our analysis reveals that, despite the presence of an efficient vaccine, JE populations are also experiencing exponential growth.

2. Materials and methods

2.1. Datasets

Gene sequences of members of the genus *Flavivirus* were collected from GenBank. Where possible, the entire envelope (E) gene (or a larger genome fragment containing the whole E gene) was used in the analysis. However, in some cases sufficient data for analysis (20 or more taxa) was not available if the complete E gene was used. Where this was the case, the fragment of the E gene sequence that maximised the number of sequences available was used (such as a fragment of the E/non-structural protein 1 (E/NS1) sequence used in several phylogenetic studies of dengue (Rico-Hesse, 1990; Rico-Hesse et al., 1997)). Since data availability was further restricted by the need for sequences from dated isolates, the length of sequence in some datasets was as short as 240 nucleotides (for example, dengue serotype 1 (DEN-1)). Known recombinant sequences (identified from Holmes et al., 1999; Worobey et al., 1999; Toulou et al., 2001; Twiddy and Holmes, in press) were excluded from all datasets, since a key assumption of our coalescent model is that no recombination occurs in the population studied. The effects of violation of this assumption are hard to predict; some branching events may be 'pushed' toward the present, whilst others may be 'pulled' backwards in time. Such effects will clearly bias estimates of demographic history parameters. Where there were many available sequences in a dataset, such as for the DEN-2 E gene, all sequences that bore a greater than 99% similarity to any other sequence were excluded, since their strong similarity to other sequences in the same dataset is likely to be indicative of close epidemiological linkage. Inclusion of such linked isolates would violate another important assumption of the coalescent model, namely that of random sampling. The effect of such a violation is that taxa are more closely related to one another than would be expected by chance, resulting in an increase of apparently recent coalescent events.

The resulting datasets were as follows (parentheses show abbreviation of virus name; gene(s) analysed; length of alignment; number of taxa for each dataset): dengue serotype 1 (DEN-1; E/NS1; 240 nucleotides; 43 taxa), dengue serotype 2 (DEN-2; E; 1485 nucleotides; 53 taxa), DEN-2 (E; samples isolated in Venezuela during 1997–2000; 1485 nucleotides; 23 taxa (Uzcategui et al., 2001), dengue serotype 3 (DEN-3; E gene fragment; 385 nucleotides; 34 taxa), JE (E; 1500 nucleotides; 35 taxa), St. Louis encephalitis virus (SLE; M (membrane protein)/E/NS1; 1617 nucleotides; 53 taxa), and WN (E gene fragment; 1278 nucleotides; 42 taxa). Insufficient numbers of dengue serotype 4 (DEN-4) and YF viruses were available for analysis. The tick-borne encephalitis (TBE) group was not included in the analysis for two reasons. First, the transmission cycle of these viruses is fundamentally different to that of the mosquito-borne flaviviruses, consisting of tick-to-tick transmission via co-feeding on minimally infected vertebrate hosts (Randolph et al., 1996; Labuda et al., 1993), a transmission route that has been shown to result in population dynamics not comparable to that of the mosquito-borne flaviviruses (Zanotto et al., 1996). Second, previous studies of the TBE group have shown that populations of this virus are strongly geographically sub-divided, so that Far Eastern TBE, Western TBE, and louping ill are phylogenetically distinct (Zanotto et al., 1995; Kuno

et al., 1998), probably due to the relative immobility of both host and vector species. This property violates one of the assumptions of the coalescent model used in this study, namely that populations are approximately panmictic. The effect of population sub-division cannot at present be predicted in analyses such as ours, since the GENIE program (see below) does not incorporate any models of population structure, although this is a future goal.

2.2. Phylogenetic tree estimation

Maximum likelihood trees for each dataset were estimated using PAUP* (Swofford, 2001). The model of nucleotide substitution used was the general time-reversible (GTR) model, with a different substitution rate for each codon position. The relative rates for the codon positions and the substitution matrix were co-estimated with the tree, while the frequencies of the four nucleotides were estimated empirically from the data. Trees were rooted using known outgroups. All parameter values are available from the authors on request. After inspection of the resulting phylogenies, the West Nile dataset was excluded from further analysis, since it was clear that it violated the assumption of random sampling implicit in the coalescent model. Specifically, out of 42 sequences, 32 fall into a single clade of very little genetic diversity consisting entirely of strains from the Israel–New York outbreaks of 1997–2000 (tree not shown, available from the authors on request). Thus, the shape of this tree reflects the strongly biased sampling of WN strains, rather than the true population dynamics of the virus. The coalescent methods used require that the strains are at least approximately randomly sampled with respect to phylogeny.

2.3. Testing of the molecular clock and substitution rate estimation

Our analytical method requires a genealogy estimated under the assumption of a constant substitution rate across all branches, that is a ‘molecular clock’, and transformation of parameter estimates for models of demographic history into a natural time-scale (i.e. years before present) requires knowledge of this substitution rate. Therefore, we used the program TipDate (Rambaut, 2000) to estimate the branch lengths of a phylogeny for each of our datasets using the single rate dated tip (SRDT) model. This model assumes that all branches have the same substitution rate, but does not require that all sequences are contemporaneous. To test the molecular clock hypothesis in each dataset, the likelihood of the phylogeny obtained under the SRDT model was compared with that of a ML tree obtained using PAUP (which represents a model where each branch is allowed a different substitution rate) using a likelihood ratio test. The rate of nucleotide substitution was also estimated for each of our datasets under the SRDT model.

2.4. Population dynamics

For this analysis, we used a statistical framework based on coalescent theory (Pybus et al., 2000). The coalescent model used incorporates changes in population size, but has various assumptions that must be taken into consideration for the viruses analysed here: (i) random sampling from a large population; (ii) no recombination, selection or population sub-division; and (iii) constant substitution rate in all branches of the genealogy.

The statistical framework referred to above implements three methods: first, a non-parametric estimate of demographic history called the skyline plot is used to give a graphical representation of estimated effective population size against time. In this context, effective population size is a function of the number of infected individuals and is directly proportional to the census population size and to the generation time (from one human infection to the next; in years if the tree length is measured in years), and inversely proportional to the variance in reproductive success (where the reproductive success of an infection is measured by the number of human infections arising from it). Since this variance is likely to be high in a virus with a vector stage such as dengue, and since the generation time is likely to be substantially less than 1 year, the effective population size may be very much smaller than the census population size. This ‘classic’ skyline plot (Pybus et al., 2000) is extended by allowing multiple coalescent events to be grouped together, giving a ‘generalised’ skyline plot (Strimmer and Pybus, 2001), in which much of the stochastic ‘noise’ present in the data is smoothed (Fig. 1). Second, we obtained the maximum likelihood estimates of various demographic parameters and their confidence limits for a number of different models of demographic history, each of which describes effective population size through time and which may describe the epidemiological history of viral populations. These models include constant population size, exponential growth, expansion growth, piecewise expansion (constant population size changing to exponential growth at time t (measured in years from the present and estimated from the data)), and two-phase exponential growth, in which an initial exponential growth rate changes at time t to a second exponential growth rate. Finally, demographic models for each dataset were tested using likelihood ratio tests; a demographic model A can be rejected in favour of alternative model B, provided that A is a special case of B. For the models in this study, the constant size model is a special case of all the other models, while the exponential model is a special case of the expansion, piecewise expansion, and two-stage expansion models. In addition, the piecewise expansion model is a special case of the two-stage exponential model, in which the exponential growth rate for the first stage is constrained to zero. All these methods were implemented using the program GENIE (Pybus and Rambaut, 2002) and all parameter values are available from the authors on request.

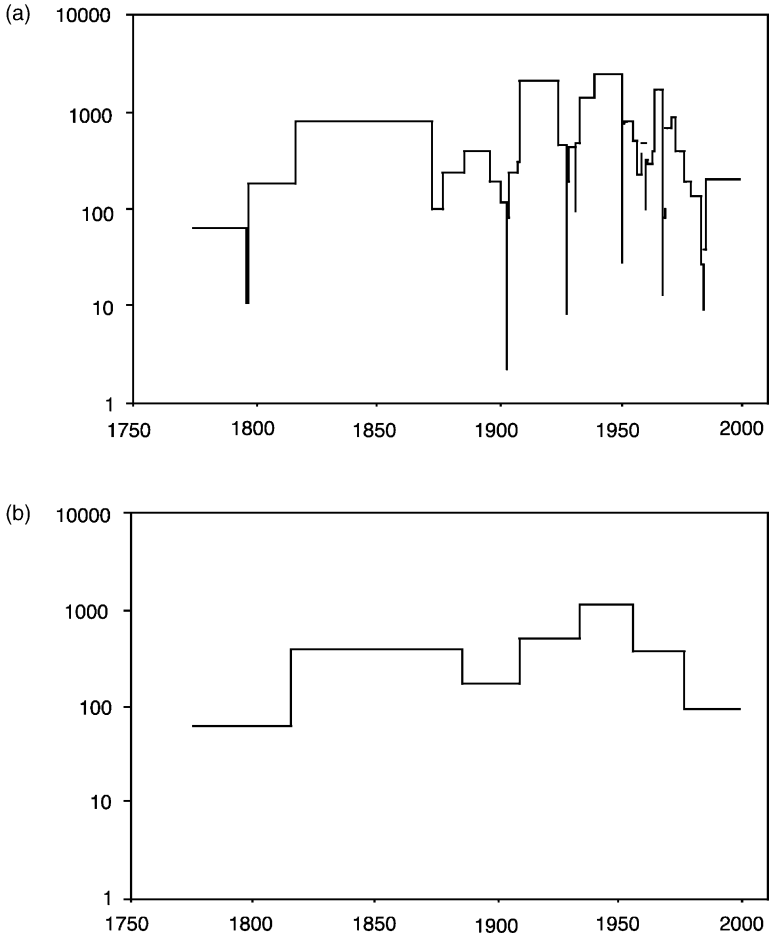

Fig. 1. (a) Classic and (b) generalised skyline plots for the St. Louis encephalitis (SLE) dataset, showing (i) the lack of an obvious demographic trend in this dataset, and (ii) the 'smoothing' effect of the generalised skyline plot. Horizontal axes represent date, vertical axes give effective population size (logarithmic scale).

3. Results

3.1. Clock tests

The assumption of a constant rate of evolution or 'molecular clock' was rejected in all of the datasets analysed, with the exception of the DEN-2 Venezuela 1997–2000 isolates. Although this is an important assumption of our models, a previous study has shown that substitution rates estimated from large datasets where the sequence length is >100 nucleotides should be reliable indicators of average rates of evolution so long as the inclusion of isolation dates into the SRDT model improves the likelihood of the model (Jenkins et al., 2002). This is the case in all datasets analysed in this study, indicating that it is possible to infer population dynamics from these data.

3.2. Skyline plots, demographic hypothesis testing, and ML parameter estimation

In the SLE dataset, no clear demographic trend could be discerned in either the classic or the generalised skyline plot, and no model was significantly better than the constant population size model (Fig. 1). However, in the remaining five datasets, one or more of the models implemented in GENIE were able to account for the observed data. This is depicted in Fig. 2, where the ML estimates of effective population size over time are represented graphically in comparison with the non-parametric estimates (i.e. generalised skyline plots). In each case, the skyline plots show increases in population size, and the constant size demographic model is strongly rejected in all except the DEN-3 dataset. The ML models of demographic history and their parameter value estimates for these five datasets are shown in Table 1.

The demographic models with the highest likelihoods varied between datasets. For the DEN-1 E/NS1 genealogy, the ML model was that of exponential growth, with a growth rate of 0.047, similar to the estimates for DEN-3 and JE, but lower than those for the global DEN-2 population and significantly lower than the estimated growth rate in the DEN-2 Venezuela population (see below). The ML model for the dataset of complete DEN-2 E genes was the two-phase expansion model, which suggested a very low rate of population growth in the past, with a change to a

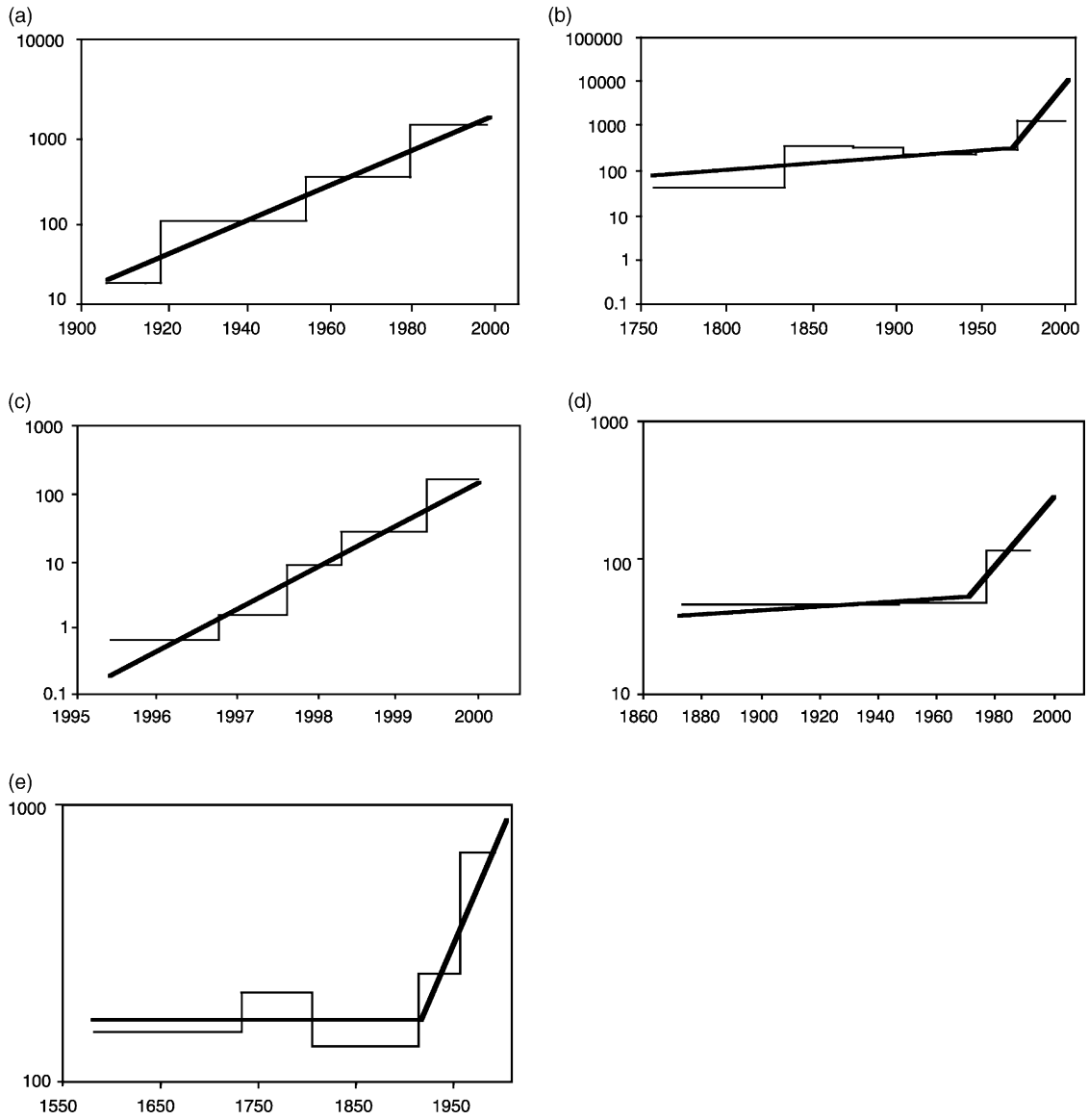

Fig. 2. Generalised skyline plots for each of the *Flavivirus* datasets for which one or more demographic hypotheses fitted the data with ML demographic models superimposed (thick lines): (a) DEN-1 E/NS1; (b) DEN-2 E gene global dataset; (c) DEN-2 E gene Venezuela 1997-2000 isolates; (d) DEN-3 E gene fragment; (e) JE E gene. Horizontal axes represent date, vertical axes give effective population size (logarithmic scale).

higher rate of growth approximately 30–35 years before the present. However, although this model was significantly better than the single-phase exponential growth model, it could not reject the piecewise expansion model, that is, we cannot reject the hypothesis that the early growth rate was zero. The parameter estimates for the two-phase exponential and piecewise expansion models were very similar, with identical estimates for the exponential growth rate at present, rate-change time estimates within 3 years of one another, and estimated current effective population sizes similar, and substantially larger, than those for any other dataset.

The DEN-2 Venezuelan dataset represents a sample of the virus population that was circulating in Venezuela during the years 1997–2000 (Uzcategui et al., 2001). The skyline plot shows rapid growth, with the population size increasing by three orders of magnitude over 5 years, this is confirmed by the estimate of the exponential growth rate (1.424 per year), which is more than 10 times that of any other dataset. No alternative model performed better than the exponential model for this dataset, which forms a clade within the American/Asian genotype (Twiddy et al., 2002), and is thought to have been introduced into the Americas from Southeast Asia in the early 1980s (Rico-Hesse, 1990; Rico-Hesse et al., 1997; Leitmeyer et al., 1999; Uzcategui et al., 2001).

As with the DEN-2 complete E gene dataset, the ML model for the DEN-3 dataset was the two-phase exponential model. This time, however, no model was able to unambiguously reject the constant population size model. The

Table 1
Parameter values for the highest likelihood models in each *Flavivirus* dataset

Dataset	Gene			ML demographic model	Exponential model rejected?	ML estimates of[a]			
	Tips	Length	Clock?			N_e	$r/r1$[b]	Change time t	$r2$
DEN-1 (E/NS1)	43	240	No	Exponential	N/A	1217.920 (678.87, 2280.75)	0.047 (0.032, 0.060)	N/A	N/A
DEN-2 (E)	53	1485	No	Two-phase exponential	Yes	9265.860 (1795.4, 89,000)	0.107 (0.045, 0.196)	31.810(25.075, 42.208)	0.007 (0, 0.001)
DEN-2 (E, Venezuela)	23	1485	Yes	Exponential	N/A	134.207 (57.629, 345.952)	1.424 (1.021, 1.796)	N/A	N/A
DEN-3 (E, 385 bp)	35	385	No	Two-phase exponential	No[c]	277.148	0.059	29.075	0.003
JE (E)	35	1500	No	Piecewise expansion	No[d]	860.455 (343.1, 12,813)	0.020 (0.002, 0.925)	83.805 (38.189, ∞)	N/A

N_e: estimated current effective population size; $r/r1$: rate of exponential growth at the present; $r2$: past exponential growth rate for two-phase exponential growth model.
[a] Confidence intervals are in parentheses; no confidence intervals are given for the DEN-3 dataset for the reasons given in the text.
[b] $r1$ For the two-phase expansion model is the rate of exponential growth at the present (i.e. the second of the two phases).
[c] The likelihood for this model was not significantly greater than either the exponential growth or the constant population size models.
[d] This model was not able to reject the exponential growth model, however, it was significantly better than the constant population size model.

two-phase model and the piecewise expansion model both suggested a change (to exponential growth from slow or zero growth) approximately 30 years ago, similar to the estimates of change time for DEN-2. The estimated effective number of infections at present was somewhat lower than might be expected for a virus that is known to have a global distribution and may mean that none of the models implemented in this study is a good fit to the data in this case, or that there is insufficient demographic signal in the data to distinguish between the models, although the skyline plot appeared to support the piecewise expansion/two-phase exponential growth hypotheses (Fig. 2d). For this combination of model and data, the likelihood surface was too complex for the numerical algorithms implemented in GENIE to find the confidence intervals.

Finally, the ML model for the JE dataset was also that of piecewise expansion, although in this case the constant size model was strongly rejected, the exponential growth model could not be rejected and the very large confidence intervals surrounding each of the estimates for this model reflect this uncertainty. The parameter estimates suggest a constant effective population size in the past, experiencing a change to relatively slow exponential growth ($r = 0.02$), approximately 85 years ago. The estimate for the current effective number of infections is less than those for DEN-1 and DEN-2, as might be expected for a virus that does not have the global distribution of the four dengue virus serotypes.

4. Discussion

In this study, we make use of a novel analytical technique based on coalescent theory to make inferences and test hypotheses about the demographic history of a group of viruses that represent important human pathogens (Pybus et al., 2000; Strimmer and Pybus, 2001). For all the virus populations in which the analysis was successful, the demographic history appeared to be one of exponential increase in population growth towards the present, either single-phase or two-phase. In the case of JE, there was no growth in the first phase.

Our analysis suggests that both DEN-2 and DEN-3 have a demographic history of two-phase exponential growth. According to this model, for the majority of their history these virus populations have been experiencing a low rate of exponential growth, however, approximately 30 years before the present, this rate of growth suddenly increased by a factor of between 15 and 20. Significantly, our estimate of the time-frame of the acceleration of dengue population growth suggests that in DEN-2 and DEN-3 populations this occurred in the order of 150 years after the increase in human population growth postulated by Zanotto et al. (1996) to be the cause of the increase in dengue incidence (as reflected in the increased rate of cladogenesis in all four serotypes). As discussed earlier, Zanotto et al.'s analysis of all dengue serotypes together is not valid with the coalescent models

employed here. However, the correlation between the expansion of human populations and the origin of each serotype (that is, the point of the zoonotic transfer from non-human primates to humans) that was suggested in the earlier study could be more than coincidental, since with increasing human population size encroachment on pre-existing sylvatic cycles of dengue would be more likely, increasing the opportunity for zoonotic transfer.

The recent expansion of dengue population size indicated by our analysis is also supported by available evidence. Although there is little information about the incidence of dengue in the first half of the 20th century, it is generally agreed that there has been a dramatic rise in the numbers of dengue cases in the last half century (Monath, 1994; WHO, 2001; Gubler, 2002). The change from a low to a much more rapid rate of population growth in dengue could therefore have been stimulated by human activities in the last half century, such as increases in global travel and mass urbanisation (encouraging geographical spread of virus and vector and combining large host populations with ideal breeding conditions for an urban mosquito such as *Aedes aegypti*). The second-phase rate of exponential growth for DEN-2 (0.11 per year) is of note in that it is similar to that estimated for hepatitis C virus subtype 1a, a virus noted for its "recent and swift" global spread through efficient transmission networks, namely injecting drug users and infected blood products, and which appears to have an epidemic history of exponential growth (Pybus et al., 2001).

Unlike the other two dengue serotypes analysed, the epidemic history of DEN-1 appears to consist of a single stage of exponential growth. The reason for the difference between these dengue serotypes is unclear—there may be a genuine difference in demographic history between the serotypes, or the difference may be an artefact of the sampling process. Similarly, the demographic history of the population of DEN-2 viruses isolated in Venezuela between 1997 and 2000 also appears to consist of a single stage of exponential growth. However, this dataset represents an interesting case, all of these isolates are descended from one or a few Asian strains of DEN-2, which were introduced into the Americas in the late 1970s or early 1980s and may have been responsible for the first major outbreak of dengue haemorrhagic fever in the Americans (Rico-Hesse, 1990; Rico-Hesse et al., 1997; Uzcategui et al., 2001). The demographic history of this dataset, therefore, reflects the growth of a single population under epidemic conditions, so that rapid exponential growth would be expected. However, the rate of growth in this population is exceptionally high (more than an order of magnitude greater than that of the world-wide population of DEN-2). It has been suggested that strains from the American/Asian genotype (in which the sequences in this dataset form a clade) have greater potential to cause dengue haemorrhagic fever/dengue shock syndrome than strains from the American genotype (Leitmeyer et al., 1999; Watts et al., 1999); if this is the case, the very high population growth rate estimated here is alarming. However, it is also possible that this high growth rate is due to particularly good conditions for transmission in Venezuela during these years. As yet, few estimates for other viruses are available for comparison, although a recent study of HIV-1 group M in the Democratic Republic of the Congo revealed a demographic history of expansion with the exponential growth rate at the present of 0.168 per year, significantly lower than our rate for the DEN-2 Venezuela population (Yusim et al., 2001).

As mentioned above, the epidemic history of JE also appears to have been a two-phase process, however, in this case the first stage was one of constant population size, with a change to a relatively low rate of exponential growth approximately 85 years ago. This hypothesis is plausible, given the well-documented westward spread of JE (Solomon et al., 2000; Leake, 1990). Although an effective vaccine is available and has reduced the incidence of JE in developed countries such as Japan, Taiwan, and South Korea (Peiris et al., 1992; Monath and Heinz, 1996), JE transmission has increased in rural and under-developed areas. This is for the most part due to human activities such as deforestation, irrigation schemes for rice cultivation (providing ideal breeding conditions for *Culex* mosquitoes, the principal vector of this virus), and agricultural development including integrated swine-waterfowl farming, which brings natural and amplifying hosts, as well as vectors and humans, into close contact, thereby providing perfect conditions for human epidemics. The low rate of growth is possibly due to the 'damping' effect of vaccination on the population growth of the virus. Whether mass vaccination campaigns or changes to agricultural practices can slow or halt future expansion remains to be seen.

Finally, we attempted to identify a model among those implemented in GENIE that would fit the demographic signal in our SLE sequence data. However, none of the models tested were significantly better than the constant size model, which itself did not appear to be a good fit to the data as displayed in the skyline plot (Fig. 1). There are two major reasons why our analysis may have failed in this case. First, the data may break one or more of the assumptions of the variable population size coalescent model. In particular, the SLE phylogeny shows a clear geographical distinction between strains isolated in United States and those originating in Central or South American countries (Kramer and Chandler, 2001), so that the viruses do not constitute a single panmictic population. It may be that the transmission dynamics in these two regions are different, indeed, it is likely that transmission characteristics will differ between temperate regions of the United States and subtropical and tropical Central and South America. It is also thought that there are as many as four distinct transmission cycles within the United States itself, each with a different mosquito vector and possibly involving viruses with different virulence characteristics (Monath and Tsai, 1987; Day, 2001)—so that combining sequence data from both regions will give a misleading result. However, the JE phylogeny also shows evidence of geographical sub-division, with viruses from the

southern regions of the virus' range phylogenetically distinct from those in the northern range; again, it has been proposed that different transmission cycles occur in these regions, with large seasonal epidemics among humans in the northern region, reflecting seasonal transmission in the natural hosts, whilst in the south transmission tends to be endemic, with sporadic cases in humans (Vaughn and Hoke, 1992; Solomon et al., 2000).

A second possible explanation for the lack of a good epidemic model for SLE is that the history of this virus is such that none of the models implemented in GENIE at present can adequately describe it. For example, the population may have undergone repeated bottlenecks, resulting in the loss of genetic information, or the demographic history may consist of a combination of the GENIE models (for example, an exponential growth phase followed by a logistic growth phase). Alternatively, it may be that the available data is not sufficient (either in volume or quality, there is clearly a noticeable bias towards isolates from the United States at present) to reveal what may be a complex demographic history. Hence, in order to fully understand the demographic history of SLE it is likely that more complicated models of population dynamics are required, together with a larger dataset including more samples from Central and South American transmission cycles.

In summary, our study demonstrates that the statistical approach developed by Pybus et al. (2000) is a useful tool for inferring the demographic history of virus populations, providing that certain key assumptions are upheld. Our results demonstrate that populations of important human pathogens such as dengue and JE continue to grow exponentially, in the case of JE despite widespread use of an effective vaccine. The growth rate of the American/Asian DEN-2 population in Venezuela is the highest yet estimated for any virus population and clearly demonstrates the epidemic potential of dengue in fully or partially susceptible populations. Further research in this area should be valuable not only in predicting future trends in the growth rates of virus populations, but also in the long-term monitoring responses to interventions such as vaccination campaigns or vector control programmes.

Acknowledgements

This work was supported by The Wellcome Trust (grants 50275 and 54231) and the Royal Society.

References

Day, J.F., 2001. Predicting St. Louis encephalitis virus epidemics: lessons from recent, and not so recent outbreaks. Annu. Rev. Entomol. 46, 111–138.

Griffiths, R.C., Tavare, S., 1994. Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. London, Ser. B 344, 403–410.

Gubler, D.J., 2002. Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. Trends Microbiol. 10, 100–103.

Holmes, E.C., Nee, S., Rambaut, A., Garnett, G.P., Harvey, P.H., 1995. Revealing the history of infectious disease epidemics through phylogenetic trees. Philos. Trans. R. Soc. London, Ser. B 349, 33–40.

Holmes, E.C., Worobey, M., Rambaut, A., 1999. Phylogenetic evidence for recombination in dengue virus. Mol. Biol. Evol. 16, 405–409.

Jenkins, G.M., Rambaut, A., Pybus, O.G., Holmes, E.C., 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. 54, 152–161.

Kramer, L.D., Chandler, L.J., 2001. Phylogenetic analysis of the envelope gene of St. Louis encephalitis virus. Arch. Virol. 146, 2341–2355.

Kuno, G., Chang, G.-J.J., Tsuchiya, K.R., Karabatsos, N., Cropp, C.B., 1998. Phylogeny of the genus *Flavivirus*. J. Virol. 72, 73–83.

Labuda, M., Nuttall, P.A., Kozuch, O., Eleckova, E., Williams, T., Zuffova, E., Sabo, A., 1993. Non-viraemic transmission of tick-borne encephalitis virus: a mechanism for arbovirus survival in nature. Experientia 49, 802–805.

Leake, C.J., 1990. Japanese encephalitis. Parasitol. Today 6, 38.

Leitmeyer, K.C., Vaughn, D.W., Watts, D.M., Salas, R., Villalobos de Chacon, I., Ramos, C., Rico-Hesse, R., 1999. Dengue virus structural differences that correlate with pathogenesis. J. Virol. 73, 4738–4747.

Monath, T.P., 1994. Dengue: the risk to developing and developed countries. Proc. Natl. Acad. Sci. U.S.A. 91, 2395–2400.

Monath, T.P., Heinz, F.X., 1996. Flaviviruses. In: Fields, B.N., Knipe, D.M., Howley, M. (Eds.), Fields Virology. Lippincott-Raven, Philadelphia, pp. 961–1034.

Monath, T.P., Tsai, T.F., 1987. St. Louis encephalitis: lessons from the last decade. Am. J. Trop. Med. Hyg. 37, 40S–59S.

Morse, S.S., 1994. The viruses of the future? Emerging viruses and evolution. In: Morse, S.S. (Ed.), The Evolutionary Biology of Viruses. Raven Press, New York, pp. 325–335.

Nee, S., Holmes, E.C., Rambaut, A., Harvey, P.H., 1995. Inferring population history from molecular phylogenies. Philos. Trans. R. Soc. London, Ser. B 349, 25–31.

Peiris, J.S.M., Amerasinghe, F.P., Amersinghe, P.H., Ratnayake, C.B., Karunaratne, S.H.P.P., Tsai, T.F., 1992. Japanese encephalitis in Sri Lanka—the study of an epidemic: vector incrimination, porcine infection and human disease. Trans. R. Soc. Trop. Med. Hyg. 86, 307–313.

Petersen, L.R., Roehrig, J.T., 2001. West Nile virus: a reemerging global pathogen. Emerg. Infect. Dis. 7, 611–614.

Pybus, O.G., Rambaut, A., 2002. GENIE: estimating demographic history from molecular phylogenies. Bioinformatics 18, 1404–1405.

Pybus, O.G., Rambaut, A., Harvey, P.H., 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. Genetics 155, 1429–1437.

Pybus, O.G., Charleston, M.A., Gupta, S., Rambaut, A., Holmes, E.C., Harvey, P.H., 2001. The epidemic behavior of the hepatitis C virus. Science 292, 2323–2325.

Rambaut, A., 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. Bioinformatics 16, 395–399.

Randolph, S.E., Gern, L., Nuttall, P.A., 1996. Co-feeding ticks: epidemiological significance for tick-borne pathogen transmission. Parasitol. Today 12, 472–479.

Rico-Hesse, R., 1990. Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. Virology 174, 479–493.

Rico-Hesse, R., Harrison, L.M., Salas, R.A., Tovar, D., Nisalak, A., Ramos, C., Boshell, J., de Mesa, M.T.R., Nogueira, R.M.R., da Rosa, A.T., 1997. Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. Virology 230, 244–251.

Solomon, T., Dung, N.M., Kneen, R., Gainsboroug, M., Vaughn, D.W., Khanh, V.T., 2000. Japanese encephalitis. J. Neurol. Neurosurg. Psychiatry 68, 405–415.

Strimmer, K., Pybus, O.G., 2001. Exploring the demographic history of DNA sequences using the generalized skyline plot. Mol. Biol. Evol. 18, 2298–2305.

Swofford, D.L., 2001. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Sinauer Associates, Sunderland, MA.

Toulou, H.J.G., Coussinier-Paris, P., Durand, J.-P., Mercier, V., de Pina, J.-J., de Micco, P., Billoir, F., Charrel, R.N., de Lamballerie, X., 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., Holmes, E.C, in press. The extent of homologous recombination in the genus *Flavivirus*, J. Gen. Virol, in press.

Twiddy, S.S., Holmes, E.C. in press, Rambaut, A. Inferring the rate and time-scale of dengue virus evolution. Mol. Biol. Evol., in press.

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

Uzcategui, N.Y., Camacho, D., Comach, G., Cuello de Uzcategui, R., Holmes, E.C., Gould, E.A., 2001. Molecular epidemiology of dengue type 2 virus in Venezuela: evidence for in situ virus evolution and recombination. J. Gen. Virol. 82, 2945–2953.

Vaughn, D.W., Hoke, C.H., 1992. The epidemiology of Japanese encephalitis: prospects for prevention. Epidemiol. Rev. 14, 197–221.

Wang, E., Ni, H., Xu, R., Barrett, A.D.T., Watowich, S.J., Gubler, D.J., Weaver, S.C., 2000. Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. J. Virol. 74, 3227–3234.

Watts, D.M., Porter, K.R., Putvatan, P., Vasquez, B., Calampa, C., Hayes, C.G., Halstead, S.B., 1999. Failure of secondary infection with American genotype dengue 2 to cause dengue haemorrhagic fever. Lancet 354, 1431–1434.

WHO, 2001. WHO Report on Global Surveillance of Epidemic-prone Infectious Diseases, Chapter 6: Dengue and dengue haemorrhagic fever. WHO, Geneva.

Worobey, M., Rambaut, A., Holmes, E.C., 1999. Widespread intra-serotype recombination in natural populations of dengue virus. Proc. Natl. Acad. Sci. U.S.A 96, 7352–7357.

Yusim, K., Peeters, M., Pybus, O.G., Bhattacharya, T., Delaporte, E., Mulanga, C., Muldoon, M., Theiler, J., Korber, B., 2001. Using human immunodeficiency virus type 1 sequences to infer historical features of the acquired immune deficiency syndrome epidemic and human immunodeficiency virus evolution. Philos. Trans. R. Soc. London, Ser. B 356, 855–866.

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

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