The extent of homologous recombination in members of the genus Flavivirus

S. Susanna Twiddy and Edward C. Holmes

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

The family Flaviviridae includes important human pathogens, such as dengue (DEN) virus, yellow fever (YF) virus and hepatitis C virus, many of which have emerged or re-emerged in recent years. Until recently, flavivirus evolution was thought to proceed in a clonal manner, with diversity generated mainly through the accumulation of mutational changes. However, this assumption has now been shown to be invalid, with homologous recombination demonstrated in all three genera of the Flaviviridae. Since recombination has important implications for the study of virus evolution, a survey of recombination in the viruses of the genus Flavivirus was carried out. Using envelope gene sequence data and a combination of graphical and phylogenetic analyses, hitherto unreported recombination in Japanese encephalitis virus and St Louis encephalitis virus was detected, as well as further recombinants in DEN virus. However, no evidence for recombination was found in West Nile or YF viruses, or in the tick-borne flavivirus group. It is proposed that the difference between the mosquito- and tick-borne viruses can be accounted for by their differing modes of transmission, whilst the variation among the mosquito-borne flaviviruses reflects both the ecology of the particular host and vector species and also bias in the sampling process.

INTRODUCTION

The family Flaviviridae comprises a group of positive-sense, single-stranded RNA viruses and consists of three genera: Flavivirus, Pestivirus and Hepacivirus. A number of these viruses are important human pathogens; the prototype virus, yellow fever (YF) virus, for example, causes an estimated 30 000 deaths annually (Vainio & Cutts, 1998), whilst 2·8 billion people live in areas at risk for Japanese encephalitis (JE) virus, which has a case fatality rate of 33 %, with 50 % of survivors suffering severe neuropsychiatric sequelae (Solomon et al., 2000). Many flaviviruses are also classified as emerging or re-emerging. As a case in point, West Nile (WN) virus has been described as undergoing ‘disturbing’ epidemiological trends since the mid-1990s, with an increase in frequency and severity of outbreaks, such as the highly publicized outbreak in New York in 1999 (Petersen & Roehrig, 2001). Similarly, JE virus activity is continuing to spread westwards, with epidemics now occurring in Pakistan and Nepal (Solomon et al., 2000). In many cases, human activity is largely responsible for the emergence of these diseases; this is particularly the case for dengue (DEN) virus, which has emerged since World War II to become globally the most important arbovirus agent of human disease. Its worldwide dispersal was aided by mass movement of troops to and from endemic regions of Southeast Asia during and after the war, whilst mass urbanization, resulting in large numbers of people living in close quarters with no running water or sanitary facilities, provided breeding grounds for the anthropophilic vector mosquito Aedes aegypti.

Although RNA recombination was shown to be possible two decades ago, until recently it had only been demonstrated in a small number of RNA viruses (King et al., 1982; Lai, 1992). Poliovirus was the first RNA virus in which homologous recombination was demonstrated, and since then this process has been detected in a variety of other RNA viruses, including Western equine encephalitis virus (Hahn et al., 1988), human immunodeficiency virus (Robertson et al., 1995) and foot-and-mouth disease virus (King et al., 1985). Until 1999, however, there was no evidence for recombination in flaviviruses, although the possibility was considered (Blok et al., 1992; Kuno et al., 1997; Monath, 1994). Accordingly, the vast majority of work on flaviviruses, including vaccine studies and phylogenetic analyses in which genotypes were identified and sometimes correlated with disease severity (Chen et al., 1990; Leitmeyer et al., 1999; Rico-Hesse, 1990), has rested on the implicit assumption that evolution in the family Flaviviridae is clonal, with diversity generated largely by the accumulation of mutational changes. Recent studies have shown this assumption to be invalid, as homologous recombination has now been demonstrated in pestiviruses (bovine viral diarrhoea virus) (Becher et al., 2001), flaviviruses (all four serotypes of DEN virus) (Holmes et al., 1999; Tolou et al., 2001; Uzcategui et al., 2001; Worobey & Holmes, 1999), hepaciviruses (GB virus C/hepatitis G virus) (Worobey & Holmes, 1999), and in even more surprising places.
Holmes, 2001) and most recently in hepatitis C virus (Kalinina et al., 2002). Given the implications of recombination for virus evolution (Worobey et al., 1999) and the development of vaccines or virus control programmes, it is clearly important to determine the extent to which recombination plays a role in flavivirus evolution.

For the purposes of this study, we carried out a survey of recombination in the viruses comprising the genus Flavivirus. We chose to use sequence data from the envelope (E) gene because it encodes the most important antigen with regards to virus biology and humoral immunity. Therefore, large-scale genetic changes in this region, as might be brought about by recombination, could have significant impact on virus phenotype. Secondly, the vast majority of published gene sequence information for flaviviruses in GenBank consists of E gene sequences. The methods used were the graphical detection of conflicting phylogenetic signals (sliding-window analysis), followed by maximum-likelihood (ML) phylogenetic analysis to define recombination breakpoints and to test their significance. In most cases, the results obtained were consistent with what is known of the biology of the different viruses within the genus Flavivirus.

**METHODS**

**Data sets.** Where possible, complete E gene sequences of all members of the genus Flavivirus for which there were sufficient data for analysis (a minimum of eight sequences) were collected from GenBank. In the case of St Louis encephalitis (SLE) virus, a large number of sequences comprising the C terminus of the M protein (39 bp), the E protein and the N-terminal 74 bp of the NS1 protein were available (total of 1617 bp) and were therefore used here. Similarly, for WN virus, a larger data set was available if only 1278 bp of the partial E gene sequence were used in the analysis. The data sets used were therefore as follows (numbers of sequences in parentheses): DEN type 1 (DEN-1) (n=19), DEN-2 (human sequences only, n=99), DEN-3 (n=32), DEN-4 (human sequences only, n=23), JE virus (n=49), SLE virus (n=61), WN virus (n=37), YF virus (n=20), tick-borne encephalitis (TBE) complex (n=40), of which 12 were looping ill virus, 14 were Western tick-borne encephalitis (WTBE) virus and 8 were Far Eastern tick-borne encephalitis (FETBE) virus.

ML trees were constructed for each data set using PAUP* (Swofford, 1998), assuming the HKY85 model of nucleotide substitution and incorporating a gamma distribution of among-site rate variation with eight rate categories (parameter values available from the authors on request). Figs 1–3 show the resultant phylogenetic trees for JE, WN and SLE viruses (as trees for the other viruses have been published recently they are not included here but are available from the authors on request).

**Exploratory tree analysis.** Initial ‘exploratory’ tree analysis was accomplished using a 300 bp window slid along the alignment in 150 bp increments, generating trees for the different regions. These trees were reconstructed using the neighbour-joining (NJ) method with distances estimated under the HKY85 model. Methods for correcting for site-specific rate heterogeneity were not employed at this stage, as they would have increased the duration of the analysis without contributing significantly to the quality of the results obtained. Isolates that clearly changed topological position over different regions of the genome were earmarked as putative recombinants. These isolates were subsequently examined using sliding-window diversity plots, in which the pairwise percentage difference between query sequences and other sequences in the alignment was determined by sliding a window of 150 bp along the alignment in 3 bp increments. The resulting diversity profiles were used to determine which of the other sequences in the data set were most closely related to the putative recombinant and could therefore be identified as the closest ‘parental’ sequences.

**Breakpoint analysis.** Once the closest parental sequences were identified, we used an ML method to estimate recombination breakpoints and assess their significance (program LARD) (Holmes et al., 2002). This method finds the optimal recombination breakpoints, given a sequence alignment of a putative recombinant and its two parents, by breaking the alignment into two regions at each possible breakpoint and reconstructing a separate ML tree for each region. The two likelihoods obtained from each region are then combined to give a ‘recombination model’ likelihood score for that particular partitioning of the alignment. The highest combined likelihood score is expected when the alignment is broken at the actual recombination position, since the two trees reconstructed either side of the breakpoint, in this case, accurately reflect the true phylogenetic history of the separate recombinant regions. Breakpoints estimated in this manner can be tested for their statistical significance by comparing the recombination model to the likelihood obtained from the unbroken alignment (i.e. the ‘no-recombination model’) using a likelihood ratio (LR) test and a Monte Carlo approach using sequences simulated without recombination under the appropriate model of nucleotide substitution.

**Phylogenetic trees and bootstrap support.** After breakpoints were identified, separate ML trees were constructed for each putative recombinant region and the phylogenetic position of the recombinant was compared between regions. Phylogenetic conflicts were assessed using the percentage of bootstrap replicates supporting the conflicting phylogenetic positions. Bootstrapping (1000 replicates) was conducted using NJ trees with distances estimated using the ML model of nucleotide substitution defined previously.

**RESULTS**

The putative recombinants uncovered in this analysis and their ‘parental’ sequences are listed in Table 1, together with the results of the ML breakpoint analysis in each case. A diversity plot is shown for one putative recombinant only (see below); the remainder are available from the authors on request. There was no evidence for recombination in WN, YF or the TBE group viruses. Similarly, exploratory tree analysis provided no evidence for recombination in DEN-1 or DEN-3, beyond those recombinants already identified by Holmes et al. (1999), Worobey et al. (1999) and Tolou et al. (2001).

However, evidence for recombination was identified in a number of other cases. First, a DEN-2 strain, hitherto unidentified as a recombinant, exhibited shifts in phylogenetic position in the sliding-window analysis. The diversity profile (Fig. 4) shows that different regions of the E gene of a DEN-2 isolate from Thailand (D80-038) appear to alternate between similarity to the Asian 1 genotype (typified by D2-Thailand 80, D80-100) and the American/Asian genotype (represented by D2-Thailand 80, D80-141) [genotypes of DEN-2 defined by Twiddy et al. (2002)]. It can be seen from the figure that there are a large number of recombination
crossover points (at least seven over the length of the gene, according to the diversity plot, giving very short sequences either side of some of the breakpoints), so it was not possible to estimate them all using LARD. Therefore, to evaluate the statistical support for the hypothesis that recombination had occurred in this strain, we used LARD to estimate the position of only the first breakpoint in the E gene. The ML estimate agreed closely with the first crossover point in the diversity plot and the LR for the putative recombinant was greater than all but 10 of the 200 LRs produced using Monte Carlo simulation, giving a P value of 0.05. This suggests that the conflicting phylogenetic relationships of DEN-2 strains with D80-038, at least in the first 500 nt of the E gene, are unlikely to be due to chance.

**Fig. 1.** ML phylogenetic tree showing the evolutionary relationships between the JE virus E gene sequences used in this study. The recombinant strains Korea 82 (K82PO1), Korea 91 (K91P55) and Thailand 82 (KPPO34-35CT) are indicated with boxes. Roman numerals signify the genotype classifications used. The tree is unrooted and horizontal branch lengths are drawn to scale.
To demonstrate further that regions of the E gene of the putative recombinant D80-038 alternate between similarity to the American/Asian genotype and the Asian 1 genotype, we constructed representative ML phylogenetic trees either side of the breakpoints to illustrate the changes in topological position in different regions of the gene. Eight trees were constructed in all but, as some of these were constructed using very short lengths of sequence (<100 bp), only the trees for the four longest sections are shown (Fig. 5a, nt 1–296, 525–735, 960–1265 and 1266–1392). These trees show clearly the conflicting phylogenetic relationships in different regions of the E gene of DEN-2 D80-038 and the strong bootstrap support for these relationships.

The second case of recombination involves a strain of DEN-4. According to the diversity profile, the majority of the E
gene sequence of DEN-4 strain D4-Philippines 56 (H241) closely resembles strain D4-Philippines 64, which falls into genotype I of the two genotypes identified previously for DEN-4 (Lanciotti et al., 1997). However, approximately 70 bp before the 3’ end of the E gene, H241 diverges from D4-Philippines 64 and becomes more similar to genotype II strains, such as DEN-4 strain D4-Tahiti 85. LARD analysis highlighted a breakpoint in a similar position to that shown in the diversity plot; the $P$ value for the LR test was <0.005, suggesting that the conflicting phylogenies are the result of recombination rather than chance. Fig. 5(b) depicts the phylogenetic relationships of H241 either side of the breakpoint and shows the high bootstrap values for the clustering of H241 with genotype I at the 5’ end of the E

Fig. 3. ML phylogenetic tree showing the evolutionary relationships between the SLE virus sequences used in this study. The recombinant strain Guatemala 69 (GMO94) is boxed. The tree is unrooted and horizontal branch lengths are drawn to scale.
Table 1. Recombinant strains, their closest ‘parental’ sequences and the results of ML breakpoint analysis

<table>
<thead>
<tr>
<th>Virus</th>
<th>Isolate</th>
<th>Accession no.</th>
<th>Region of E gene</th>
<th>‘Parents’</th>
<th>Subtype</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEN-2</td>
<td>Thailand 80 (D80-038)</td>
<td>M24448</td>
<td>1–296</td>
<td>D2-Thailand 80 (D80-100)</td>
<td>Asian 1</td>
<td>1st breakpoint: P=0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>525–735</td>
<td>D2-Thailand 80 (D80-141)</td>
<td>American/Asian</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>960–1265</td>
<td>D2-Thailand 80 (D80-100)</td>
<td>Asian 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1266–1485</td>
<td>D2-Thailand 80 (D80-141)</td>
<td>American/Asian</td>
<td></td>
</tr>
<tr>
<td>DEN-4</td>
<td>Philippines 56 (H241)</td>
<td>S66064</td>
<td>1–1412</td>
<td>D4-Philippines 64</td>
<td>I</td>
<td>P&lt;0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1413–1485</td>
<td>D4-Tahiti 85</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>JE</td>
<td>Korea 82 (K82PO1)</td>
<td>U34926</td>
<td>1–759</td>
<td>Japan 88 (JaOArK6688)</td>
<td>I</td>
<td>P&lt;0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>760–1095</td>
<td>Korea 94 (K94P05)</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1096–1500</td>
<td>Japan 88 (JaOArK6688)</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>JE</td>
<td>Korea 91 (K91P55)</td>
<td>U34928</td>
<td>1–412</td>
<td>Korea 87 (K87P39)</td>
<td>I</td>
<td>P&lt;0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>413–1500</td>
<td>Japan 98 (Ishikawa)</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>JE</td>
<td>Thailand 82 (KPPO34-35CT)</td>
<td>U03693</td>
<td>1–308</td>
<td>Thailand 93 (ThCMAr6793)</td>
<td>I</td>
<td>P&lt;0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>309–1500</td>
<td>China 54 (SA.14)</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>SLE</td>
<td>Guatemala 94 (GMO94)</td>
<td>AF205513</td>
<td>1–428</td>
<td>Argentina 66 (CorAn9124)</td>
<td>NA</td>
<td>P&lt;0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>429–1617</td>
<td>USA 74 (TNM4-711K)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NA, Not applicable.

Three putative recombinants were identified in the analysis of the JE virus E gene. There have as yet been few rigorous studies of the phylogenetic relationships among JE virus strains or attempts to classify them into ‘genotypes’ (Chen et al., 1990; Ritchie et al., 1997; Tsarev et al., 2000). In the absence of an accepted convention for classification of JE virus strains, we reconstructed an ML tree from our data set and identified four major clusters of strains, which we have denoted I–IV and which are represented in the trees of Fig. 5(c–e). Interestingly, there appears to be some correlation between the phylogenetic groupings in this tree and the regions of either epidemic or endemic activity reported, as suggested previously by Chen et al. (1990). Group I consisted mainly of strains from China, India and Taiwan, areas where JE virus transmission follows a pattern of seasonal epidemics, whilst groups II, III and IV were largely made up of strains from Thailand, Indonesia,
Malaysia and Australia, more Southern regions where JE virus transmission is endemic (Burke & Leake, 1988). All of the putative recombinant sequences described in this paper are mosaics of genotypes I and II. Breakpoint analysis for all of the putative JE virus recombinants confirmed the recombination crossover points from the diversity plots; \( P \) values were <0.005 in each case.

Two of the putative JE virus recombinants, Korea 82 (K82P01) and Korea 91 (K91P55) appear to have parental strains originating in Japan and Korea. In the case of K82P01, the first half and the final 400 bp of the E gene sequence are related to genotype I strains isolated from the 1980s in Japan and Korea, with the central section bearing greater similarity to Korean and Japanese isolates from the 1990s (genotype II). For K91P55, approximately the first 400 nt resemble the genotype I Korean and Japanese strains, with the rest of the sequence being more closely related to the 1990s genotype II isolates. The third putative JE virus recombinant was isolated from Thailand (Thailand 82,
KPP034-35CT). At the 5' end of the E gene this strain is closely related to another genotype I Thai 93 strain (ThCMAr6793); however, after approximately 300 bp it becomes more similar to a genotype II strain originating in China (SA-14). The phylogenetic relationships for the different regions of the E genes in JE virus are shown in Fig. 5(c–e) and the topological shifts for each of the recombinants have strong bootstrap support.

Finally, a putative recombinant was identified among the SLE virus strains analysed. Specifically, Guatemala 94 (GMO94), a Central American strain, is identical to an Argentinean isolate (CorAn9124) at the 5' end of the E gene; however, after approximately 350 bp it diverges from this South American strain and, after a putative recombination crossover point between nt 427 and 428, it bears a strong similarity to an isolate from the southern United
States (TNM4-711K). The hypothesis that recombination occurred in this case is strongly supported using LARD (P<0.005). Again, no generally accepted classification of subtypes of SLE virus exists (Kramer & Chandler, 2001), although it is clear that there are genetic differences between strains isolated in the United States, Central America and South America (Fig. 3). In the ML trees showing phylogenetic shifts, we attempted to include strains that represented the full diversity observed in natural isolates of SLE virus (Fig. 5f). These trees show clearly the different topological position of Guatemala 94 (GMO94) in the different regions of the gene, with the topological shift supported in each case by very high bootstrap values.

DISCUSSION

That recombination can occur in flaviviruses has now been demonstrated on a number of occasions (Becher et al., 2001; Kalinina et al., 2002; Worobey & Holmes, 2001; Worobey et al., 1999). In this study we undertook a wider survey of recombination in those flaviviruses for which sufficient E gene sequence data were available. Our results show that demonstrable recombination (that is, between sufficiently diverse sequences and of sufficient length for it to be detectable by sliding-window analysis) occurs not only in the DEN viruses but also in other members of the mosquito-borne flavivirus group.

Phylogenetic and biological consequences of recombination have been discussed in depth in previous studies (Worobey et al., 1999; Worobey & Holmes, 1999) and therefore will not be addressed here. In the current study, the most notable result was that recombination was detected in most, but not all, of the mosquito-borne flaviviruses and in none of the tick-borne flaviviruses. Although these differences could be caused by ascertainment bias, as recombination was detected most frequently in those viruses for which most data were available, it is also possible that they reflect biological and ecological factors related to virus transmission.

Conditions for recombination

For recombination to occur, several conditions must be met. Principally, either vector or host must be co-infected with greater than one strain of the virus in question. This requires either that hosts are fed upon by more than one infected vector and are productively infected by at least two strains of virus at the same time, or that vectors engage in multiple feeding on viraemic hosts, allowing the possibility of vector co-infection with greater than one strain of virus (although the nature of TBE virus transmission is such that the criteria for vector co-infection with TBE virus are somewhat different; see below). The limitations of the methods used to detect recombination impose a further set of conditions; it is extremely difficult to identify recombinant sequences in cases where the sequences analysed are very similar to one another. In practice, this means that only recombinant sequences whose parents are from different genotypes are readily identifiable. Therefore, for detectable recombination to occur, co-circulation of different genotypes must take place, usually requiring mobility of host and/or vector

---

**Fig. 5.** ML trees showing shifts in phylogenetic position with strong bootstrap support for different regions of each putative recombinant sequences: (a) D2-Thailand 80 (D80-038) (abbreviated to D038), (b) D4-Philippines 56 (H241), (c) JE-Korea 82 (K82), (d) JE-Korea 91 (K91), (e) JE-Thailand 82 (KPP) and (f) SLE-Guatemala 94 (GMO94). The putative recombinant in each case is highlighted with a box on the trees. Subtypes for each virus are indicated with shading and are labelled (where appropriate) in bold next to the subtype. All trees are unrooted and branch lengths are drawn to scale.
species. In addition, sampling effects can play a major role. In general, the longer the sequence used in the analysis and the wider (and denser) the geographical and temporal sampling, the more likely it is that a recombinant sequence will be identified.

**Mosquito-borne flaviviruses in which recombination was detected**

Of the mosquito-borne flaviviruses studied, DEN viruses have the highest frequency of recombination. As expected, they fulfil most or all of the criteria outlined above. The principal vector of DEN virus, the *A. aegypti* mosquito, has been shown to engage in multiple feeding, allowing co-infection of more than one strain in the vector (Gould et al., 1970). In addition, as in all mosquito-borne flaviviruses, virus replication takes place in both the vector and the vertebrate hosts (principally humans in the case of DEN virus) to a sufficient level to allow mosquito-human-mosquito transmission to occur. In all four serotypes, distinct ‘genotypes’ are observable. Furthermore, the contemporaneous circulation of different genotypes in a particular geographical region has been documented in DEN-2 (Twiddy et al., 2002) and must presumably have occurred in the other three serotypes, as recombination has been demonstrated in each (Holmes et al., 1999; Tolou et al., 2001; Uzcategui et al., 2001; Worobey et al., 1999). Given the global distribution of DEN virus and the exceptional mobility of the human host species, the relatively high frequency of recombination in DEN virus is not surprising.

Although JE virus is also the agent of an important human disease, its natural transmission cycle is between *Culex* mosquitoes and wild and domestic birds, with pigs as an amplifying host and humans a dead-end host (Burke & Leake, 1988). As with the *Aedes* vector of DEN virus, the genus *Culex* has been demonstrated to engage in multiple feeding (Wekesa et al., 1997), so there is a potential for co-infection of hosts and/or vectors. Furthermore, the JE virus phylogeny presented here shows co-circulation of genotypes III and IV in Southeast Asia in the early 1980s and of genotypes I and II in Korea in the late 1980s/early 1990s, the period in which two of the three JE virus recombinants identified in this study were isolated in Korea. The presence of strains isolated in Japan and Korea in genotype II also suggests that strains may have been introduced into the northern ‘epidemic’ region from the southern ‘endemic’ region, probably by birds, the likely agents of the continuing westward spread of JE virus (Solomon et al., 2000). Finally, the available sample of JE virus sequences is biased towards genotype I, which comprises sequences from the northern range of JE virus transmission and a study conducted using a more representative sample from both transmission regions might be expected to uncover more recombinant JE virus strains.

Recombination was also detected in SLE virus, which, like JE virus, is transmitted by *Culex* mosquitoes, with birds being the principal vertebrate host. Although all of the sequences are very similar (<10 % nucleotide difference between any two strains), there are discernible differences between strains circulating in the United States and those isolated from Central and South America (Fig. 3). Isolates that cluster in the United States have been recovered from Brazil and Panama, indicating that long-distance transfer of strains is possible, with migrating birds being the most likely agent of geographical dispersal (Monath & Tsai, 1987). Indeed, the recombinant virus identified here – GMO94 – does itself appear to be the product of a strain from Central America and a strain from the southern region of the United States. As yet, there are no isolates from the United States that clearly have a South American origin, lending some support to the hypothesis that maintenance of the virus in North America is predominantly local, with virus overwintering in either chronically infected birds or diapausal mosquitoes (Kramer & Chandler, 2001).

**Mosquito-borne flaviviruses in which recombination was not detected**

WN virus, another member of the JE virus complex, has a transmission cycle similar to that of both SLE and JE viruses (i.e. avian-*Culex*-avian) and has been the subject of considerable interest since the New York outbreak of 1999 (Ebel et al., 2001; Lanciotti et al., 1999). Our analysis failed to find any evidence of recombination in WN virus, although detectable recombination could, in theory, occur, as the requirements for vector co-infection are fulfilled and the close relationship of the New York 1999 strains of WN virus to those isolated in Israel in 1998 and 2000 bears witness to the ability of the virus to be transmitted over long distances (Petersen & Roehrig, 2001). In addition, distinct genotypes identifiable on the tree of the WN virus E gene (Fig. 2) appear to have been co-circulating in Israel in 2000 and in Romania in 1996. However, the data set of WN virus sequences is heavily biased towards strains from Israel and New York from 1998 to 2000, among which there is little diversity. Inclusion of sequences from more of the regions where WN virus occurs [including Kunjin virus sequences, which have been suggested to be a subtype of WN virus rather than a separate virus species (Savage et al., 1999), and for which no E gene sequence data of >300 bp is currently available] may lead to the discovery that recombination does occur in this virus.

Last among the mosquito-borne flaviviruses, we analysed the available E gene sequences from YF virus. Again, there is no reason to suppose that co-infection in host or, more likely, vector could not occur and there is sufficient genetic diversity within the YF virus population thus far sampled to identify at least three major genotypes, previously dubbed I, IIa and IIb (Chang et al., 1995). These three groups are geographically distinct, with two (I and IIa) found in sylvatic cycles in East/Central and West Africa, respectively, and the third (IIb) occurring in the forests of Central and South America, with many Caribbean islands also at risk (Vainio & Cutts, 1998). More recently, Mutebi et al. (2001) suggested that each of the African clades could be divided further, with
Tick-borne flaviviruses

The final group of flaviviruses analysed were the TBE viruses. No evidence of recombination was detected in this group. However, the biology of these viruses is such that the likelihood of recombination identifiable by the methodologies used in this survey is low. The principal mechanism of transmission for the TBE virus group is tick-to-tick transmission via co-feeding on a host with low or undetectable viraemia (Jones et al., 1997; Labuda et al., 1993; Randolph et al., 1996). Therefore, the frequency of co-infection in ticks is dependent upon the likelihood of a previously infected tick co-feeding with at least one other tick that is infected with a different TBE virus strain. As Ixodes ticks feed only infrequently (in general, once as larva, nymph and adult) (Kettle, 1995) and as TBE virus infection prevalence in I. ricinus ticks in Central Europe has been reported to be in the region of 0.2 % (Randolph et al., 1999), the probability that any tick is co-infected with more than one virus strain must be low. Similarly, co-infection in the vertebrate host would require transmission to an already viraemic host by a tick infected with a different strain. This is a relatively unlikely prospect as in those vertebrate hosts that develop a systemic viraemia, its duration is short and mortality is high (Gilbert et al., 2000; Randolph et al., 1996).

Second, although co-infection may occur, it is highly unlikely that any resulting recombinant strains would be detectable using the methods employed in this survey. Zanotto et al. (1995) observe that there is very little variation in strains of TBE virus in any single geographical area. Ticks themselves are essentially immobile in terms of geographical dispersal of virus strains and although they may be carried some distance by larger vertebrate hosts such as deer (or humans), these tend to be ‘dead-end’ hosts. The natural hosts of TBE virus, at least in Central Europe, are small rodents, with correspondingly small geographical ranges, so that co-circulation of genetically diverse strains of TBE virus is likely to be extremely rare. When these factors are considered, it is not surprising that, even with a relatively large sample of virus strains, no putative recombinants were identified.

Overall, our study demonstrates that recombination, at least in the mosquito-borne flaviviruses, is a relatively frequent occurrence. In addition, we propose that aspects of the biology of the flaviviruses, such as life-history of vector or vertebrate host, might explain the patterns of recombination frequency that we have observed. Given the drastic effect that recombination can have on phylogenetic studies, it is highly desirable that sequence analysis should be carried out on whole genomes wherever possible, so that putative recombinants anywhere in the genome can be identified and the possibility of misclassification of strains reduced. At present, there is a clear shortage of such whole genome data; this needs to be addressed in the immediate future.

ACKNOWLEDGEMENTS

This work was supported by the Wellcome Trust and The Royal Society. The authors thank Dr M. Worobey for his valuable advice.

REFERENCES


http://vir.sgmjournals.org


