Journal of General Virology (2001), **82**, 1867–1876. Printed in Great Britain

Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography

Michael W. Gaunt,[1,5] Amadou A. Sall,[2,5] Xavier de Lamballerie,[3,5] Andrew K. I. Falconar,[4] Tatyana I. Dzhivanian[6] and Ernest A. Gould[5]

[1] Pathogen Molecular Biology and Biochemistry Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

[2] Institut Pasteur de Dakar, Dakar, Senegal

[3] Unité de Virus Emergents, Faculté de Médecine, Boulevard Jean Moulin, 13005 Marseille, France

[4] Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3FY, UK

[5] Centre for Ecology and Hydrology (formerly Institute of Virology and Environmental Microbiology), Mansfield Road, Oxford OX1 3SR, UK

[6] Institute of Poliomyelitis and Viral Encephalitis, Moscow, Russia

Phylogenetic analysis of the *Flavivirus* genus, using either partial sequences of the non-structural 5 gene or the structural envelope gene, revealed an extensive series of clades defined by their epidemiology and disease associations. These phylogenies identified mosquito-borne, tick-borne and no-known-vector (NKV) virus clades, which could be further subdivided into clades defined by their principal vertebrate host. The mosquito-borne flaviviruses revealed two distinct epidemiological groups: (i) the neurotropic viruses, often associated with encephalitic disease in humans or livestock, correlated with the *Culex* species vector and bird reservoirs and (ii) the non-neurotropic viruses, associated with haemorrhagic disease in humans, correlated with the *Aedes* species vector and primate hosts. Thus, the tree topology describing the virus–host association may reflect differences in the feeding behaviour between *Aedes* and *Culex* mosquitoes. The tick-borne viruses also formed two distinct groups: one group associated with seabirds and the other, the tick-borne encephalitis complex viruses, associated primarily with rodents. The NKV flaviviruses formed three distinct groups: one group, which was closely related to the mosquito-borne viruses, associated with bats; a second group, which was more genetically distant, also associated with bats; and a third group associated with rodents. Each epidemiological group within the phylogenies revealed distinct geographical clusters in either the Old World or the New World, which for mosquito-borne viruses may reflect an Old World origin. The correlation between epidemiology, disease correlation and biogeography begins to define the complex evolutionary relationships between the virus, vector, vertebrate host and ecological niche.

Introduction

The *Flavivirus* genus contains many viruses associated with emerging and re-emerging human diseases, including dengue haemorrhagic fever, Kyasanur Forest haemorrhagic disease, Japanese encephalitic disease, Rocio virus encephalitis (Monath & Heinz, 1990) and West Nile fever (Lanciotti *et al.*, 1999). Elucidating the evolution of viruses is particularly valuable for understanding the origin and spread of emerging and re-emerging diseases (Holmes, 1998).

Flaviviruses are a useful model for studying the evolution of vector-borne virus diseases, since they comprise mosquito-borne, tick-borne and no-known-vector (NKV) viruses (Porterfield, 1980). The genus contains about 70 recognized flaviviruses that are antigenically related and have a widespread geographical dispersion. They are positive-stranded RNA viruses with a genome of approximately 10·5 kb. Virions contain three structural proteins, capsid (C), membrane (M) and

Author for correspondence: Michael Gaunt (at the London School of Hygiene and Tropical Medicine).
Fax +44 20 7636 5739. e-mail michael.gaunt@lshtm.ac.uk

Table 1. Flaviviruses analysed in this study

All of the flaviviruses that were analysed in this study are classified by their virus group (Heinz *et al.*, 2001).

Virus group	Flavivirus analysed, excluding flavivirus E gene sequences (presented separately)	Abbreviation	Flavivirus partial E genes sequenced in this study	Abbreviation	Accession no.
Louping ill	Louping ill	LI			
	Spanish sheep encephalitis	SSE			
	Turkish sheep encephalitis	TSE			
	Greek goat encephalitis	GGE			
Tick-borne encephalitis	Far-east Asian subtype	FETBE			
	West European subtype	WTBE			
Mammalian tick-borne	Kyasanur Forest disease	KFD	Gadgets Gulley	GGY	AF372408
	Langat	LGT	Kadam	KAD	AF372420
	Omsk haemorrhagic fever	OHF			
	Powassan	POW			
	Royal Farm	RF			
	Karshi	KSI			
Seabird tick-borne	Tyuleniy	TYU	Meaban	MEA	AF372423
	Saumarez Reef	SRE			
Aroa	Iguape	IGU	Aroa	AROA	AF372413
			Bussuquara	BSQ	AF372410
			Naranjal	NJL	AF372411
Dengue	Dengue types 1–3, 4	DEN1–3, 4			
	Kedougou	KED			
Japanese encephalitis	Japanese encephalitis	JE	Alfuy	ALF	AF372406
	Murray Valley encephalitis	MVE	Cacipacore	CPC	AF372417
	Saint Louis encephalitis	SLE	West Nile (tick)	WN (tick)	AF372405
	Usutu	USU			
	Kunjin	KUN			
	Yaounde	YAO			
Ntaya	Tembusu	TMU	Bagaza	BAG	AF372407
	THCr*	THCr*	Ilheus	ILH	AF372414
			Rocio	ROC	AF372409
			Israel turkey meningoencephalo-myelitis	IT	AF372415
			Ntaya	NTA	AF372416
Spondweni			Spondweni	SPO	AF372412
			Zika	ZIK	AF372422
Yellow Fever	Banzi	BAN	Edge Hill	EH	AF372419
	Bouboui	BOU	Jugra	JUG	AF372418
	Potiskum	POT	Saboya	SAB	AF372421
	Sepik	SEP			
	Uganda S	UGS			
	Yellow Fever	YF			
Kokobera	Kokobera	KOK			
	Stratford	STR			
Entebbe bat	Entebbe bat	ENT			
	Sokoluk	SOK			
	Yokose	YOK			
Modoc	Apoi	APOI			
	Cowbone Ridge	CR			
	Jutiapa	JUT			
	Modoc	MOD			
	Sal Vieja	SV			
	San Perlita	SP			

Table 1 (*cont.*)

Virus group	Flavivirus analysed, excluding flavivirus E gene sequences (presented separately)	Abbreviation	Flavivirus partial E genes sequenced in this study	Abbreviation	Accession no.
Rio Bravo	Rio Bravo	RB			
	Bukalasa	BUK			
	Carey Island	CI			
	Dakar Bat	DK			
	Montana myotis leukoencephalitis	MML			
	Phnom Penh	PPB			
	Batu Cave	BC			

* Additional flavivirus analysed by Kuno *et al.* (1998).

envelope (E), and infected cells have been shown to contain seven non-structural (NS) proteins, NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 (Rice *et al.*, 1985; Rice, 1996).

The evolution, dispersal patterns and epidemiological characteristics of flaviviruses are believed to have been determined through a combination of constraints imposed by the arthropod vector, the vertebrate hosts, the associated ecology and the influence of human commercial activity. For example, the clinal evolution of the tick-borne encephalitis (TBE) complex viruses across the Euro-Asian land mass reflects the life-cycle and feeding habits of the ixodid tick (Zanotto *et al.*, 1995) combined with the appropriate rodent host species and climatic conditions (Randolph *et al.*, 2000). Similarly, the introduction of goats and sheep onto the hillsides of Turkey, Greece, Spain, Ireland, Norway and the British Isles was followed by the appearance of louping ill (LI) virus (Reid, 1984; Gao *et al.*, 1993; Gaunt *et al.*, 1997; McGuire *et al.*, 1998). The emergence and expansion of dengue haemorrhagic fever in the tropics has followed an increase in human and mosquito population densities brought about by urbanization and industrialization (Zanotto *et al.*, 1996). Finally, the trans-Atlantic dispersal of yellow fever (YF) virus, and possibly many other flaviviruses, was thought to have coincided with the transportation of people and mosquitoes from Africa to the Americas on slave ships during the past few centuries (Strode, 1951; Gould *et al.*, 1997).

Using detailed molecular phylogenetic analyses, we have attempted to bring all these factors together in order to understand the nature of flavivirus evolution, epidemiology and dispersal. In this study, maximum likelihood (ML) phylogenetic analyses of virtually all of the recognized flaviviruses were performed using partial NS5 gene sequences and the tree was compared with one based on the E gene sequences. The epidemiological and aetiological characteristics of each flavivirus have been mapped onto the phylogeny to reveal a striking pattern of coincidence between the topological arrangement of the viruses and their associated epidemiological characteristics.

Methods

The flavivirus cDNA sequences determined in this study and the flavivirus abbreviations used hereafter are described in Table 1. Viral RNA extraction and RT–PCR procedures have been described previously (Gritsun & Gould, 1995; Gaunt *et al.*, 1997). A nested set of primers encompassing the species-specific amino acid motifs defined previously (Marin *et al.*, 1995) was designed to amplify a partial E gene locus for the genus *Flavivirus*. The PCR primers amplified 85% of the flaviviruses described by Calisher *et al.* (1989) and will be described elsewhere. Amplified cDNA was cloned into pGEM T vectors (Promega) and sequenced using the Thermo-Sequenase Cycle Sequencing kit (Amersham), according to the manufacturer's instructions. cDNA sequences were determined from two recombinant plasmids prepared from each virus and any differences were resolved using a third clone.

Phylogenetic analyses were performed using the partial NS5 sequences described previously by Kuno *et al.* (1998), with the addition of LI virus, strain 369 (Y07863). The flavivirus E gene sequences were obtained in this study and also included E gene sequences from previous studies (Marin *et al.*, 1995; Billoir *et al.*, 2000). Partial NS5 gene nucleotide sequences were aligned using CLUSTAL X and edited manually from their amino acid alignment (Thompson *et al.*, 1994). The partial E gene sequences were exclusively aligned using their amino acid sequences and the final alignment comprised 308 amino acids (including gaps).

The robustness of the data sets was examined for deviations in nucleotide or amino acid base composition between taxa, as nucleotide or amino acid base homogeneity is a prerequisite for the mutation models used in this study. The NS5 nucleotide alignment required the removal of the third codon position to obtain nucleotide base homogeneity (χ^2-test, $P < 0{\cdot}05$ for all three codon positions and $P > 0{\cdot}99$ for codon positions one and two only) (PAUP*, version 4.0; Swofford, 1999). The E gene amino acid alignment required the removal of cell fusing agent virus, which had sometimes been used in previous analyses, to obtain amino acid homogeneity under the same test ($P > 0{\cdot}05$ for all other species) (Puzzle; Strimmer & von Haeseler, 1996). Nucleotide variation was examined using a sliding window analysis (SWAN; Proutski & Holmes, 1998). The NS5 gene was further investigated for possible saturation

using a 10 bp sliding window analysis, which estimated as an entropy function of the nucleotide variation (*Var*). The sliding window analysis of the mosquito-borne flaviviruses, and each mosquito-borne flavivirus clade described later, identifies a region between nucleotides 342 and 392 of the 693 bp NS5 gene alignment showing the highest level of variation (*Var* = 0·72–1·21). This region coincided a large alignment gap and proved difficult to align using amino acids; therefore, these nucleotides were subsequently removed. The ML model for NS5 gene nucleotide substitution was determined by testing 40 models of nucleotide substitutions (MODELTEST; Possado & Krandall, 1998), which described an eight parameter model consisting of the general time reversible (GTR) model of nucleotide substitution (six parameters), an invariant rate parameter (PINVAR) and the alpha parameter of a four category discrete gamma distribution (Γ; one parameter). The ML model parameters were estimated by an automated reiterative ML heuristic search (PAUP*) and from Jukes–Cantor distances (MODELTEST). Parameter estimates were incorporated into a full heuristic ML search for ten replications. The level of phylogenetic support was determined by bootstrap re-sampling using ML distances incorporating each parameter estimate and a full heuristic search for 1000 replications (PAUP*). In addition, 100 Monte Carlo DNA sequence simulations were performed from the final NS5 gene tree topology and reconstructed using the ML parameter estimates, as described previously (Seq-Gen; Rambaut & Grassly, 1997).

Evolutionary reconstructions of partial E gene amino acid alignments were performed by ML quartet puzzling using the JTT substitution matrix (Jones *et al.*, 1992) for 10000 quartet puzzling steps (Puzzle; Strimmer & von Haeseler, 1996). Likelihoods were analysed for 0, 4, 8 and 16 discrete Γ categories and the presence or absence of an PINVAR was assessed using a likelihood ratio test.

The phylogenies obtained from the partial NS5 and E gene sequences of each flavivirus were tested for their association with particular arthropod vectors, disease associations (haemorrhagic and encephalitic) and their geographical distribution using the International Catalogue of Arboviruses (Karabatsos, 1995).

Results

Phylogenetic reconstruction

All NS5 gene phylogenies produced a single likelihood regardless of the taxa order in each full heuristic search performed ($n \geqslant 10$ in all cases; Table 2). Specific details for building the ML NS5 gene phylogeny and the model parameters used are described in Table 2.

The robustness (convergence) of the E gene ML phylogeny was observed after 10000 quartet puzzling steps from two key indicators. Firstly, only 3% of the quartets for all of these searches was unresolved. Secondly, although six models of rate substitution were tested (see Methods), the quartet puzzling support differed by a maximum of 1%. Thus, no difference in branching order between the E gene models was observed, regardless of the combination of rate parameters used. The phylogeny for the significantly highest likelihood, as determined by likelihood ratio tests, was an eight category Γ distribution with PINVAR.

Figs 1 and 2 present the phylogenies for the NS5 gene and the E gene, respectively. The topologies showed congruence at all levels where bootstrap or quartet puzzling support was

Table 2. ML tree building method for NS5 amino acid-based nucleotide alignment

A six parameter GTR (a–f), PINVAR and a four category discrete Γ distribution were used. J–C denotes the Jukes–Cantor distance method, ML denotes an automated reiterative maximum likelihood parameter estimation. Re-estimation of the model of nucleotide substitution for the truncated alignment was identical to the original and the resulting parameter estimates were assessed for convergence at two decimal places (d.p.). The parameters of the GTR model showed no convergence between phylogenies at 2 d.p. and a second ML automated reiterative search was conducted to re-estimate the six parameters, incorporating the values obtained for Γ and PINVAR.

Calculation	Nucleotide region removed	GTR						Γ, α	PINVAR	Replications per estimate	No. full searches including parameters*	Likelihood (-ln)
		a	b	c	d	e	f					
1. J–C	None	2·2536	1·8041	1·0767	1·3668	4·3338	1	0·8346	0·2625	–	30	19085·5
2. ML	None	4·3028	2·3345	1·2051	2·0317	5·7642	1	0·8401	0·2489	310	10	19040·2
3. J–C	Hyper-variable	2·4503	1·7691	1·0382	1·4249	4·5630	1	0·835	0·274	–	–	–
Convergence at 2 d.p.	–	–	–	–	–	–	–	0·84	0.25			
4. ML	Hyper-variable	4·1869	2·1711	1·1514	1·9341	5·8155	1	0·84	0·25	1620	10	16577·8

* A full heuristic ML search incorporating the parameter estimates.

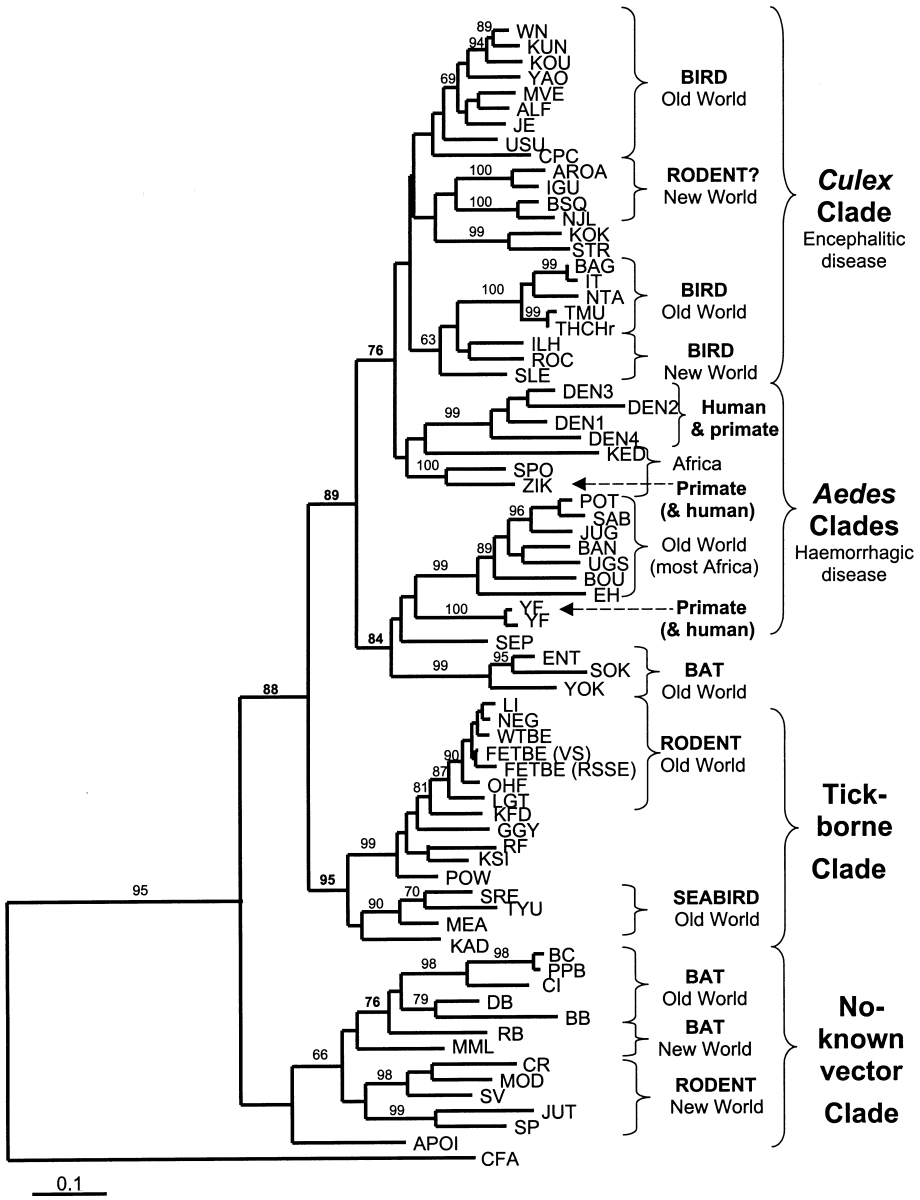

Fig. 1. Phylogeny tree of the genus *Flavivirus* based on the NS5 gene sequence constructed using an eight parameter model (GTR, Γ and PINVAR). Brackets denote the principle vector clade or the principle vertebrate host clade for the viruses therein. The numbers above each lineage show the percentage bootstrap support for that branch; numbers below 60 are not shown. Unequivocal vertebrate host clades are designated for monophyletic groups containing two or more virus species.

greater than 60%. Equivocal incongruence was observed for SLE, AROA, BSQ and NJL virus monophyly and the DEN viruses between the two phylogenies presented in Figs 1 and 2.

As demonstrated previously (Kuno *et al.*, 1998), the NS5 gene phylogeny defined three major groups comprising the mosquito-borne, tick-borne and NKV flaviviruses. However, both analyses showed that the three NKV bat-associated viruses, ENT, SOK and YOK viruses, were grouped within the mosquito-borne virus clades forming a basal lineage with YF and SEP viruses.

Aedes clades and the *Culex* monophyly

Mapping epidemiological and disease characteristics of the individual mosquito-borne viruses onto the phylogenetic trees revealed a correlation between the principal vector genera (*Culex* and *Aedes* species), the principal vertebrate hosts (birds and/or mammals) and the virus tropisms in humans and livestock (neurotropic versus non-neurotropic) (Figs 1 and 2). The mosquito-borne viruses could be divided into two epidemiologically distinct vector groups, those that were primarily isolated from *Aedes* species and those that were

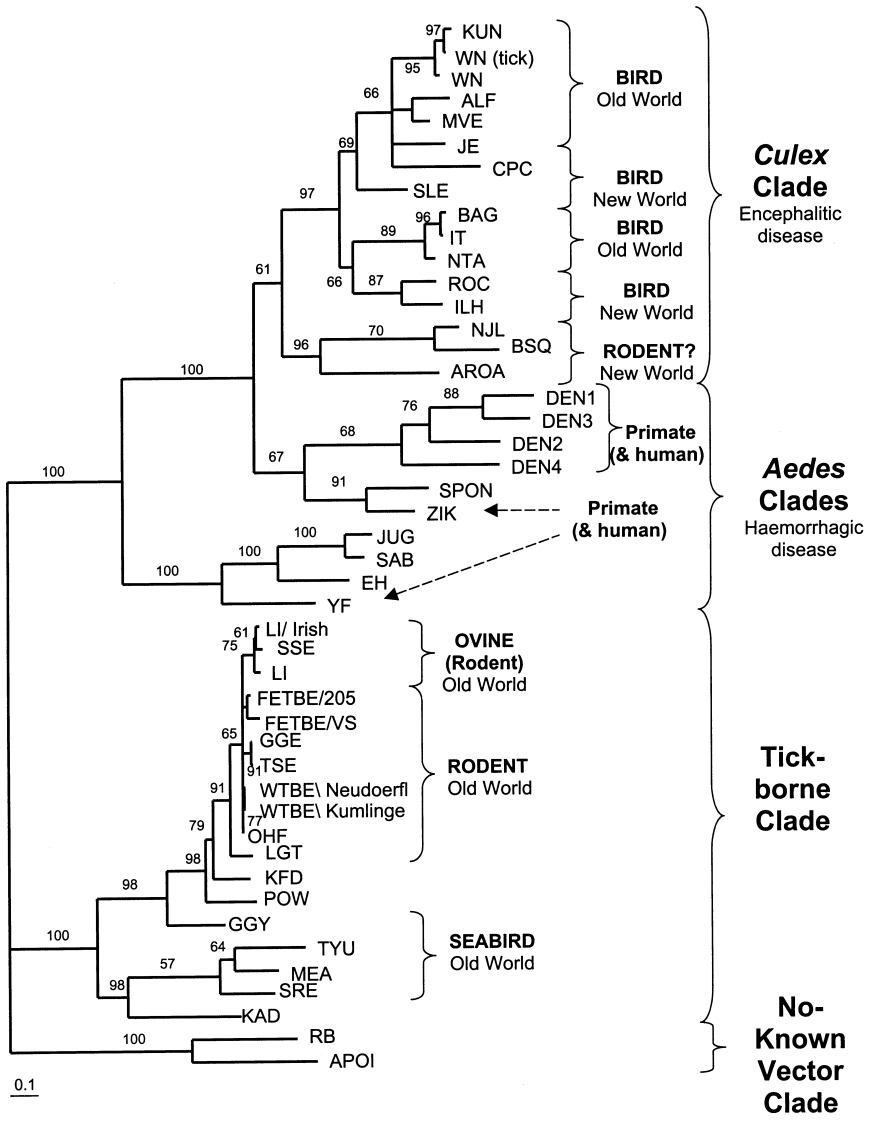

Fig. 2. Phylogeny tree of the genus *Flavivirus* based on the E gene amino acid sequence constructed using the JTT substitution model and incorporating an additional two parameters (Γ and PINVAR). Nomenclature is identical to that in Fig. 1.

primarily isolated from *Culex* species. The 17 flaviviruses that were primarily isolated from *Aedes* species formed two paraphyletic groups, one containing YF virus and the other containing the DEN viruses. In these two paraphyletic groups, 82% of mosquito-borne flaviviruses are known to be associated with *Aedes* species (14/17), hereafter denoted as the *Aedes* clades. The other viruses in these clades, i.e. SEP, POT, BAN and SAB, have been primarily associated with *Mansonia* species, rodents, *Culex* species and sandflies, respectively. The major exception is the mosquito species *Haemagogus*, the sylvatic vector of YF virus in South America, although this does not apply to the urban vector. The flaviviruses primarily isolated from *Culex* species formed a single clade of 23 viruses, which included JE, SLE and WN viruses as notable examples. The mosquito vector was identified for at least 16 of these viruses (Karabatsos, 1995). Of these 16, 89% of mosquito-borne flaviviruses in the *Culex* species monophyly are known to be isolated from *Culex* species (14/16), hereafter denoted as the *Culex* clade.

Figs 1 and 2 show that there is a clear correlation between the virus, mosquito vector species and associated host. Both *Aedes* clades contained viruses that were maintained in sylvatic primate cycles, namely YF, DEN and ZIK viruses, or other mammals, while birds were not strongly associated with any of the viruses in the *Aedes* clades. Although *Aedes* clade viruses may infect birds, they are believed to be 'dead-end hosts'. In contrast, none of the *Culex* clade viruses are maintained in primate cycles. Moreover, a high proportion of flaviviruses in the *Culex* clade are associated with mosquito–bird cycles (at least 12/23 viruses) (Figs 1 and 2). However, mammals are additionally involved in the persistence of *Culex* clade viruses, although many are considered dead-end hosts. For example,

pigs play a role in the maintenance of JE viruses, while bats could be involved in WN virus persistence (Monath & Heinz, 1990). Furthermore, AROA, IGU, BSQ and NJL viruses, which form a single clade, could be maintained by rodents.

Mosquito-borne viruses normally associated with neurological disease in humans or livestock, leading to encephalitis in severe cases, were found in the *Culex* clade and were generally associated with viruses that cycled between mosquitoes and birds. DEN virus from the *Aedes* clades was the exception, since there are rare cases of DEN encephalitis (Lum *et al*., 1996; Hommel *et al*., 1998; Solomon *et al*., 2000). In contrast, the mosquito-borne flaviviruses that are normally associated with haemorrhagic disease were exclusive to the *Aedes* clades and were associated with viruses that cycle between mosquitoes and primate hosts.

Clade robustness

The *Culex* and *Aedes* clades showed robust quartet support in the partial E gene phylogenetic tree and robust bootstrap support using the partial NS5 gene sequence for a single *Aedes* clade containing YF virus. The robustness of the distinction between the *Aedes* clade that contained the DEN viruses and the *Culex* clade was separately assessed using a Monte Carlo simulation. The observed phylogeny was subject to 100 Monte Carlo simulations, reconstructed using ML (reconnection limit = 1) and manually assessed for alternative topologies. Distinct *Culex* and *Aedes* clades were observed in 98% of all simulations, while the SPO and ZIK virus sister group formed a trifurcation with the *Aedes* clade containing DEN virus and the *Culex* clade in 9% of these simulations.

There are also significant differences between the relative positions of some flaviviruses, as presented in the NS5-derived tree shown in Fig. 1 and those presented by Kuno *et al*. (1998). For example, ZIK and SPO viruses are positioned together with other *Aedes*-associated viruses (Fig. 1); they had previously been placed in different positions among the mosquito-borne viruses by Kuno *et al*. (1998). Secondly, and in contrast with the previous analysis (Kuno *et al*., 1998), KED virus showed close phylogenetic relationships with SPO and ZIK viruses (Fig. 1), confirming the published serological data (Karabatsos, 1995).

Tick-borne and NKV flaviviruses

Vertebrate host clades were also observed in the tick-borne and NKV flaviviruses for both the NS5 and E gene trees. In the NS5 gene phylogeny, NKV viruses, for which APOI virus was the basal lineage, were subdivided into rodent and bat clades. The rodent clade showing robust bootstrap support contained CR, JUT, MOD, SV and SP viruses, whereas the bat clade showing robust bootstrap support contained BUK, CI, DK, PPB and RB viruses (Figs 1 and 2). APOI virus is also associated with rodents and, therefore, could be included in the rodent clade. The bat MML virus was also included in the bat clade, despite low bootstrap support. The remaining three NKV (bat) viruses, ENT, SOK and YOK, in the NS5 gene phylogeny form a sister group with the *Aedes* clade containing YF virus and are maintained by robust bootstrap support. From the point of view of their evolutionary origins, it is important to note that the rodent clade NKV viruses, with the exception of APOI virus, have only been isolated in the New World, whereas bat-associated viruses have been isolated from the Old and the New World, although none has been isolated in both regions of the world.

The TBE complex viruses were primarily associated with ixodid (hard) ticks, mainly *Ixodes* species and rodent hosts. The second group, consisting of tick-borne seabird-associated viruses (MEA, TYU and SRE), were most frequently isolated from *Ornithodorus* species or *Ixodes uriae*. KAD virus, which is associated with *Rhipicephalus appendiculatus* in Africa and *Hyalomma pravus* in Saudi Arabia, and GGY virus, which is associated with seabirds and *Ixodes uriae*, represent early lineages in the two tick-borne clades and possibly indicate a genetic link between the two tick-borne flavivirus groups.

Geographical distribution

The geographical distribution of the mosquito-borne flaviviruses was also examined to see whether or not virus dispersal correlated with either the *Aedes* clades or the *Culex* clade. With the exception of YF virus and DEN virus, which are believed to have originated in the Old World but can now also be found in the New World, all other viruses in the *Aedes* clades are only found in the Old World. On the other hand, the viruses in the *Culex* clade show geographical clustering, but genetically closely related viruses in the *Culex* clade have been widely dispersed to the Americas, Africa, Asia and Australasia, i.e. the Old and the New World.

Discussion

Early attempts to define flavivirus interrelationships and their evolutionary characteristics were based on antigenic cross reactivity in neutralization, complement fixation and haemagglutination inhibition tests (de Madrid & Porterfield, 1974; Porterfield, 1980; Calisher *et al*., 1989). Classification schemes based on these criteria have proved helpful in understanding the flaviviruses, but many of the viruses have subsequently been shown to be incorrectly assigned within the schemes. Molecular sequencing and phylogenetic reconstructions have largely overcome these problems and have provided important insights into the taxonomy (Heinz *et al*., 2000) and dispersal of flaviviruses (Gould *et al*., 1997). The association of specific flaviviruses with particular arthropod vectors and vertebrate hosts has been defined precisely and a list of these characteristics for each virus is available in the International Catalogue of Arboviruses (Karabatsos, 1995). Despite these extensive data, there have been few previous attempts to correlate molecular evolution with epidemiological

and ecological features of the flaviviruses. The phylogenetic trees presented here have extended previous analyses of the flavivirus NS5 (Kuno *et al*., 1998; Billoir *et al*., 2000) and E gene phylogenetic trees (Marin *et al*., 1995; Zanotto *et al*., 1995). By mapping these biological characteristics onto the trees, the phylogenetic analyses presented in this paper demonstrate a striking series of correlations between molecular phylogenetic and ecological/epidemiological characteristics.

It was demonstrated previously (Marin *et al*., 1995; Kuno *et al*., 1998) that the *Flavivirus* genus was monophyletic and three distinct groups of viruses, namely tick-borne, mosquito-borne and NKV viruses, diverge at the deepest nodes. We have now demonstrated that the mosquito-borne viruses are subdivided into the *Culex* clade and *Aedes* clades. Moreover, the evolution of the *Culex* clade appears to have occurred after the separation of the mosquito-borne viruses from the tick-borne and NKV viruses. These observations were supported by the congruence between the NS5 and E gene phylogenies as well as by Monte Carlo simulation and quartet puzzling support.

The dominance of *Aedes* and *Culex* species (subfamily Culicinae) in flavivirus transmission is explained by the species prevalence of each of the genera, which contain 975 and 769 species, respectively, and comprise more species than all other mosquito genera combined (1522 species). *Aedes* and *Culex* mosquitoes are also among the small number of genera that are globally dispersed. Blood-meal data obtained for *Aedes* species suggest that mammals are the primary hosts of most species, which could explain the *Aedes* clades–primate/mammal association (Mitchell, 1988; Christensen *et al*., 1996; Clements, 1999). The feeding patterns of only relatively few species of *Culex* mosquito are known, although a small number of bird- or mammal-specific species have been identified. Many *Culex* species feed indiscriminately on both mammals and birds and they include the principal vectors for several flaviviruses in the *Culex* clade, such as *C. annulirostris* (MVE virus), *C. tritaeniorhynchus* (JE virus), *C. tarsalis* (SLE virus) and *C. univittatus* (WN virus) (Robertson *et al*., 1993; Christensen *et al*., 1996; Clements, 1999). The difference in feeding behaviour between *Aedes* and *Culex* mosquitoes provides a clear explanation for the associations between *Aedes*-borne flaviviruses and mammals or between *Culex*-borne flaviviruses and birds. Moreover, it explains why the association between the *Aedes* clades and mammals appears to be unequivocal, while the association between the *Culex* clade and birds contains a number of notable exceptions.

The second major correlation was between the type of disease produced and the mosquito clade in which each virus appeared. In general, severe infections caused by some *Aedes* species viruses result in haemorrhagic disease, whereas many *Culex* species viruses cause encephalitic disease; however, there have been reported cases of DEN (*Aedes* species-associated) encephalitis, but these seem to be very rare (Lum *et al*., 1996; Hommel *et al*., 1998; Solomon *et al*., 2000). Until the precise basis of flavivirus pathogenicity has been defined at the molecular level, it is not possible to understand why these different disease associations can be seen in the phylogenetic tree. In contrast with the mosquito-borne flaviviruses, different viruses in the tick-borne virus groups produce encephalitic disease, but OHF and KFD viruses may also produce haemorrhagic disease in humans and this does not appear to correlate with either their phylogenetic or their geographical characteristics.

Phylogenetic divisions between Old and New World flaviviruses were seen throughout the NS5 and E gene phylogenies. In some instances, dispersal of flaviviruses could be readily linked with the vertebrate host, providing evidence of the importance of the host in flavivirus evolution. In the case of viruses that established infections in bats, it is easy to imagine dispersal to remote sites, as the Old World bats from which flaviviruses have been isolated are known to migrate hundreds of kilometres (Shilton *et al*., 1999). On the other hand, individual rodent-associated NKV viruses might be expected to show a more restricted distribution and this is demonstrated by their detection almost exclusively in the New World and by their localized or niche-like distribution.

Virtually all of the tick-borne flaviviruses are exclusively Old World, with the exception of POW virus. The seabird-associated tick-borne viruses were dispersed to geographical areas where they established niches in seabird colonies in both the Northern and the Southern hemispheres (TYU, SRE and MEA viruses) (Chastel *et al*., 1985). At the early period of their evolution, the TBE complex viruses appear to have been dispersed either by seabirds or by rodents and their associated ticks. As they reached the forests of Asia, they became established predominantly in *Ixodes* species, where they continued their clinal evolution into Europe (Gao *et al*., 1993; Zanotto *et al*., 1995; Gould *et al*., 1997).

The earliest evolutionary lineages in the mosquito-borne virus clades appear to have radiated to geographically distant parts of the Old World and to a wide variety of species, i.e. bats, *Aedes* species, sandflies and large animals, including simians and humans. Only YF virus and the four DEN virus serotypes, which cause human epidemics, are found in the New World. There is strong evidence to support the notion that YF virus was introduced to the Americas from the Old World during the past few centuries when slaves were transported across the Atlantic Ocean (Strode, 1951; Monath & Heinz, 1990; Gould *et al*., 1997).

There are also reasons to believe that DEN viruses have an African ancestry. The other members of the *Aedes* clade containing the DEN, ZIK, SPO and KED viruses were all isolated from Africa and formed two paraphyletic lineages to the DEN viruses. In addition, the E gene phylogenies of endemic/epidemic and sylvatic DEN viruses show a basal position for Old World sylvatic lineages of DEN1, DEN2 (Africa and Malaysia) and DEN4 (Wang *et al*., 2000). The vector of DEN virus, *Aedes aegyti*, is also believed to have originated in Africa (Tabachnick, 1991). There is no reason to

believe that DEN virus could not have been shipped to the Americas from the Old World in the same way as YF virus. Therefore, as most of the other *Aedes* species-associated viruses are found solely in Africa and since the *Culex* species-associated viruses appear to be descendants of the *Aedes* species-associated viruses, the mosquito-borne flaviviruses appear to have evolved out of Africa.

In conclusion, the flaviviruses that are recognized today represent a diverse group of viruses that could have emerged and dispersed during the past 10000 years, i.e. since the most recent ice age (Zanotto *et al.*, 1996). The characteristic epidemiological groupings of the viruses that are apparent in the phylogenetic trees illustrate the significant influence of the invertebrate vectors, the vertebrate hosts and the particular ecological niches into which these species have evolved.

We gratefully acknowledge the financial support from the NERC and the Wellcome Trust. We thank E. C. Holmes (University of Oxford) for help and advice as well as A. Lilley (CEH), G. Clarke and M. A. Miles (LSHTM). We thank Dr R. Shope (University of Texas) and Dr J. S. Porterfield (previously University of Oxford) for supplying most of the virus stocks used in this study.

References

Billoir, F., de Chesse, R., Tolou, H., de Micco, P., Gould, E. A. & de Lamballerie, X. (2000). Phylogeny of the genus *Flavivirus* using complete coding sequences of arthropod-borne viruses and viruses with no known vector. *Journal of General Virology* **81**, 781–790.

Calisher, C. H., Karabatsos, N., Dalrymple, J. M., Shope, R. E., Porterfield, J. S., Westerway, E. G. & Brandt, W. E. (1989). Antigenic relationships between flaviviruses as determined by cross-neutralisation tests with polyclonal antisera. *Journal of General Virology* **70**, 37–43.

Chastel, C., Main, A. J., Guiguen, C., le Lay, G., Quillien, M. C., Mannat, J. Y. & Beaucournu, J. C. (1985). The isolation of Meaban virus, a new flavivirus from the seabird tick *Ornithodoros* (Alecorobius) *maritimus* in France. *Archives of Virology* **83**, 129–140.

Christensen, H. A., de Vasquez, A. M. & Boreham, M. M. (1996). Host-feeding patterns of mosquitoes (Diptera: *Culicidae*) from central Panama. *American Journal of Tropical Medicine and Hygiene* **55**, 202–208.

Clements, A. N. (1999). *The Biology of Mosquitoes: Sensory, Reception, and Behaviour*, 2nd edn. Wallingford, Oxford: CABI Publishing.

de Madrid, A. T. & Porterfield, J. S. (1974). The flaviviruses (group B arboviruses): a cross-neutralisation study. *Journal of General Virology* **23**, 91–96.

Gao, G. F., Hussain, M. H., Reid, H. W. & Gould, E. A. (1993). Classification of a new member of the TBE flavivirus subgroup by its immunological, pathogenetic and molecular characteristics: identification of subgroup-specific pentapeptides. *Virus Research* **30**, 129–144.

Gaunt, M. W., Jones, L. D., Laurenson, K., Hudson, P. J., Reid, H. W. & Gould, E. A. (1997). Definitive identification of louping ill virus by RT–PCR and sequencing in field populations of *Ixodes ricinus* on the Lochindorb estate. *Archives of Virology* **142**, 1181–1191.

Gould, E. A., Zanotto, P. M. A. & Holmes, E. C. (1997). The genetic evolution of the flaviviruses. In *Factors in the Emergence of Arboviruses Diseases*, pp. 51–63. Edited by J.-F. Saluzzo & B. Dodet. Paris: Elsevier.

Gritsun, T. S. & Gould, E. A. (1995). Infectious transcripts of tick-borne encephalitis virus, generated in days by RT–PCR. *Virology* **214**, 611–618.

Heinz, F. X., Collett, M. S., Purcell, R. H., Gould, E. A., Howard, C. R., Houghton, M., Moormann, R. J. M., Rice, C. M. & Thiel, H. J. (2000). Family *Flaviviridae*. In *Virus Taxonomy. Seventh International Committee for the Taxonomy of Viruses*, pp. 859–878. Edited by M. H. V. van Regenmortel, C. M. Fauquet & D. H. L. Bishop. San Diego: Academic Press.

Holmes, E. C. (1998). Molecular epidemiology and evolution of emerging infectious diseases. *British Medical Bulletin* **54**, 533–543.

Hommel, D., Talarmin, A., Deubel, V., Reynes, J. M., Drouet, M. T., Sarthou, J. L. & Hulin, A. (1998). Dengue encephalitis in French Guiana. *Research in Virology* **149**, 235–238.

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* **8**, 275–282.

Karabatsos, N. (1995). *International Catalogue of Arboviruses*. San Antonio: American Society of Tropical Medicine and Hygiene.

Kuno, G., Chang, G.-J. J., Tsuchiya, K. R., Karabatsos, N. & Cropp, S. B. (1998). Phylogeny of genus *Flavivirus*. *Journal of Virology* **72**, 73–83.

Lanciotti, R. S., Roehrig, J. T., Deubel, V., Smith, J., Parker, M., Steele, K., Crise, B., Volpe, K. E., Crabtree, M. B., Scherret, J. H., Hall, R. A., MacKenzie, J. S., Cropp, C. B., Panigrahy, B., Ostlund, E., Schmitt, B., Malkinson, M., Banet, C., Weissman, J., Komar, N., Savage, H. M., Stone, W., McNamara, T. & Gubler, D. J. (1999). Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science* **286**, 2333–2337.

Lum, L. C. S., Lam, S. K., Choy, Y. S., George, R. & Harun, F. (1996). Dengue encephalitis: a true entity? *American Journal of Tropical Medicine and Hygiene* **54**, 256–259.

McGuire, K., Holmes, E. C., Gao, G. F., Reid, H. W. & Gould, E. A. (1998). Tracing the origins of louping ill virus by molecular phylogenetic analysis. *Journal of General Virology* **79**, 981–988.

Marin, M. S., Zanotto, P. M., Gritsun, T. & Gould, E. A. (1995). Phylogeny of TYU, SRE, and CFA virus: different evolutionary rates in the genus *Flavivirus*. *Virology* **206**, 1133–1139.

Mitchell, C. J. (1988). Occurrence, biology and physiology of diapause in overwintering mosquitoes. In *The Arboviruses: Epidemiology and Ecology*, vol. 1. Edited by T. P. Monath. Boca Raton: CRC Press.

Monath, T. P. & Heinz, F. X. (1990). Flaviviruses. In *Fields Virology*, 2nd edn, pp. 763–814. Edited by B. N. Fields & D. M. Knipe. New York: Raven Press.

Porterfield, J. S. (1980). Antigenic characteristics and classification of *Togaviridae*. In *The Togaviruses*, pp. 13–46. Edited by R. W. Schlesinger. New York: Academic Press.

Possado, D. & Krandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Computer Applications in the Biosciences* **14**, 817–818.

Proutski, V. & Holmes, E. C. (1998). SWAN: a new Macintosh application for the sliding window analysis of nucleotide sequence variability. *Computer Applications in the Biosciences* **14**, 467–468.

Rambaut, A. & Grassly, N. C. (1997). Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer Applications in the Biosciences* **13**, 235–238.

Randolph, S. E., Green, R. M., Peacey, M. F. & Rogers, D. J. (2000). Seasonal synchrony: the key to tick-borne encephalitis foci identified by satellite data. *Parasitology* **121**, 15–23.

Reid, H. W. (1984). Epidemiology of louping ill. In *Vectors in Biology*, pp. 161–178. London: Academic Press.

Rice, C. M. (1996). *Flaviviridae*: the viruses and their replication. In *Fields Virology*, 3rd edn, pp. 931–959. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott–Raven.

Rice, C. M., Lenches, E. M., Eddy, S. R., Shin, S. J., Sheets, R. L. & Strauss, J. H. (1985). Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. *Science* **229**, 726–733.

Robertson, L. C., Prior, S., Apperson, C. S. & Irby, W. S. (1993). Bionomics of *Anopheles quadrimaculatus* and *Culex erraticus* (Diptera: *Culicidae*) in the Falls Lake basin, North Carolina: seasonal changes in abundance and gonotropic status, and host-feeding patterns. *Journal of Medical Entomology* **30**, 689–698.

Shilton, L. A., Altringham, J. D., Compton, S. G. & Whittaker, R. J. (1999). Old World fruit bats can be long-distance seed dispersers through extended retention of viable seeds in the gut. *Proceedings of the Royal Society of London Series B Biological Scienes* **266**, 219–223.

Solomon, T., Dung, N. M., Vaughn, D. W., Kneen, R., Thao, L. T. T., Raengsakulrach, B., Loan, H. T., Day, N. P. J., Farrar, J., Myint, K. S. A., Warrell, M. J., James, W. S., Nisalak, A. & White, N. J. (2000). Neurological manifestations of dengue infection. *Lancet* **335**, 1053–1059.

Strimmer, K. & von Haeseler, A. (1996). Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution* **13**, 964–969.

Strode, G. K. (1951). *Yellow Fever*. New York: McGraw–Hill.

Tabachnick, W. J. (1991). Evolutionary genetics and arthropod-borne disease: the yellow fever mosquito. *American Entomologist* **37**, 14–24.

Swofford, D. L. (1999). PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods), version 4.0. Sinauer Associates, Sunderland, MA, USA.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

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

Zanotto, P. M. 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. A., Gould, E. A., Gao, G. F., Harvey, P. H. & Holmes, E. C. (1996). Population dynamics of flaviviruses revealed by molecular phylogenies. *Proceedings of the National Academy of Sciences, USA* **93**, 548–553.

Received 28 February 2001; Accepted 25 April 2001