Arch Virol (2006) 151: 309–325
DOI 10.1007/s00705-005-0626-6

Archives of
Virology
Printed in Austria

A multigene analysis of the phylogenetic relationships among the flaviviruses (Family: *Flaviviridae*) and the evolution of vector transmission

S. Cook[1] and **E. C. Holmes**[2]

[1]Department of Zoology, University of Oxford, Oxford, U.K.
[2]Department of Biology, The Pennsylvania State University,
University Park, PA, U.S.A.

Received May 12, 2005; accepted July 11, 2005
Published online September 20, 2005 © Springer-Verlag 2005

Summary. The genus *Flavivirus* (family *Flaviviridae*) presently comprises around 70 single-strand positive-sense RNA viruses. These replicate in a range of vertebrate and invertebrate cells and may be mosquito-borne, tick-borne or have no-known-vector. Since transmission mode correlates strongly with phylogeny, the flaviviruses constitute a valuable model for the evolution of vector-borne disease. Attempts to resolve the higher-level taxonomic relationships of the flaviviruses through molecular phylogenetics have thus far proved inconclusive because of conflicting positions for the three main transmission groups. We conducted the most comprehensive phylogenetic study to date, involving maximum likelihood analyses of the NS3 and NS5 genes and the entire genome sequences available at present. For the first time, we use and test a variety of more robust methods of sequence alignment and appropriate models of amino acid replacement to study these highly divergent sequences, and explicitly test specific hypotheses of tree topology. We show that (i) the NS5 gene contains insufficient phylogenetic signal to choose between competing topological hypotheses, (ii) the NS3 gene and whole genome data indicate that the mosquito-borne flaviviruses represent an outgroup to the remaining flaviviruses, and (iii) that tick-borne transmission is probably a derived trait within the genus.

Introduction

The genus *Flavivirus* currently consists of approximately 70 single-strand, positive-sense RNA viruses. The genus is classified within the family *Flaviviridae*, which also contains the *Pestivirus* and *Hepacivirus* genera [4, 26]. A number of the flaviviruses are associated with human disease. For example, dengue virus, present as four serotypes (DENV-1 to DENV-4), is prevalent in over 100 countries and 2.5 million people live in dengue-endemic areas [12], while yellow fever virus (YFV) affects 200,000 persons annually [23], with a case fatality rate of around 20 percent [22]. Flaviviruses infect a range of hosts and many are capable of replicating in both vertebrate and invertebrate cells. Since the genus includes viruses that are mosquito-borne, tick-borne and those with no-known-vector (NKV), the flaviviruses represent a useful model to study the evolution of vector-borne disease and of transmission modes. In addition, understanding the evolution of these viruses may provide valuable general insights into the origin and spread of emerging and re-emerging viruses [14].

Early attempts to define taxonomic relationships within the genus were based on antigenic cross-reactivity in neutralization, complement fixation and haemagglutination tests [4, 26]. More recently, explicitly phylogenetic studies have aimed to infer the evolutionary history of the flaviviruses from the comparative analysis of amino acid and nucleotide sequences. Flaviviruses have an average genome size of around 11 kb. Virions contain three structural proteins, the capsid (C), membrane (M) and envelope (E), and infected cells contain seven non-structural (NS) proteins, namely NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 [30, 31]. Early phylogenetic work used E gene, NS3, NS4, and NS5 sequences from vector-borne flaviviruses alone and suggested a major split between the tick-borne and mosquito-borne strains, with YFV then diverging from the mosquito-borne lineage and a subsequent split between DENV and the mosquito-borne viruses associated with encephalitis [2]. Inclusion of E gene and NS5 sequences from additional tick-borne viruses also supported an early split between mosquito- and tick-borne viruses [19, 20]. However, none of the NKV group of viruses were represented in these studies. Zanotto et al. [40] expanded this work using E gene sequence data from tick-borne and mosquito-borne flaviviruses and revealed important differences in the mode of evolution of the two groups of vector-borne flaviviruses. Specifically, the tick-borne viruses were characterised by a continual branching pattern that is correlated with geographical distance, indicating a clinal mode of dispersal and evolution. Transmission patterns comprise (a) traditional horizontal transmission among viremic hosts, and (b) tick-to-tick transmission via co-feeding on non-viremic hosts. In either case, this may be followed by long periods during which the ticks do not feed. Hence, viral lineages may survive for relatively long periods of time. In contrast, phylogenetic trees revealed a "discontinuous" evolutionary pattern in the mosquito-borne flaviviruses with little geographical structure and frequent lineage extinction. This may reflect the fact that vector lifespan in this case is significantly shorter, typically measured in days. Evolutionary dynamics will also be affected by other differences between these two vector groups, including the number of blood-feeds during the arthropod lifespan, the number of different hosts, the volume of blood-feeds, the mobility of the vector and the likelihood of vertical transmission [40].

Sequences from the NKV group were first included in phylogenetic studies by Kuno et al. [18] in a study which analysed virtually all flaviviruses described at that time. Using partial NS5 sequences and rooting the phylogeny on the highly divergent sequence from Cell Fusing Agent Virus (CFAV), they showed that the

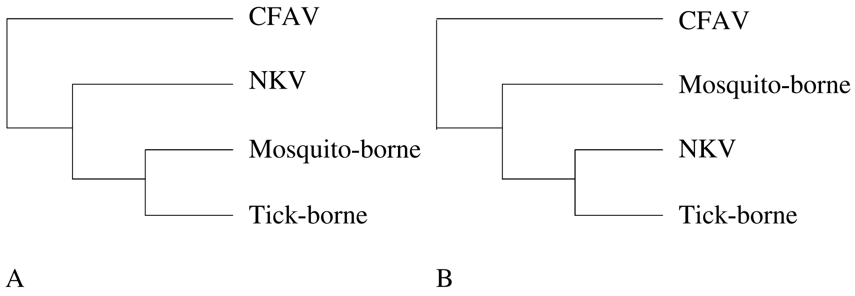

Fig. 1. Alternative phylogenetic relationships of the genus *Flavivirus* derived from past studies. (**A**) Mosquito and tick-borne viruses are sister groups – the "NS5-like" pattern of Billoir et al., 2000, (**B**) NKV (no known vector) and tick-borne viruses are sister groups – the "NS3-like" of Billoir et al., 2000

NKV viruses appeared to diverge before the vector-borne viruses. Of particular interest was the presence of some non-vector viruses within the mosquito-borne clade, indicating a secondary loss of vector-borne transmission [18]. This general topology is shown in Fig. 1A, and may be thought of as the "NS5-like" pattern. The flavivirus phylogeny of Gaunt et al. [11] agreed with this general topology using the NS5 gene, with CFAV as an outgroup. To take into account the substantial variation in base composition among the three main groups of viruses, Jenkins et al. [16] constructed a phylogeny using the first and second codon positions only with CFAV as an outgroup and also observed that the NKV flaviviruses were the most divergent group.

However, conflicting phylogenetic positions have been observed in other studies. In particular, Billoir et al. [1], determined the first two complete ORF sequences for NKV viruses (namely RBV and APOIV) and observed that the tick-borne and NKV viruses formed a sister-group to the mosquito-borne members of the genus in trees of the NS3 gene and a data set containing the entire ORF sequences of the available flaviviruses (see Fig. 1B). This "NS3-like" pattern was obtained using both amino acid and nucleotide sequences. The only exception occurred when CFAV was included in a complete ORF alignment for first and second codon positions only, which resulted in a phylogeny in which the NKV viruses diverged separately from the arthropod-borne viruses as seen in the NS5 gene trees. Hence, the respective positions of the NKV, mosquito- and tick-borne clades differ according to what gene is used and the phylogenetic relationships within the genus *Flavivirus* remain unresolved at present.

De Lamballerie et al. [7] recently determined the sequence of Tamana Bat virus (TABV), a hitherto unclassified flavivirus originally isolated in 1973 from the insectivorous bat *Peteronotus parnelli* [27]. TABV was found to share many characteristics with the flaviviruses, including similar genomic organisation, hydropathy plots, conserved polyprotein cleavage sites and enzyme domains. In addition, phylogenetic analysis of the structural genes indicated that although TABV was clearly related to the flaviviruses, it was highly genetically divergent such that little further phylogenetic resolution could be achieved, a notion supported by a lack of serological cross-reactivity [18]. Indeed, TABV also exhibited a variety of unique characteristics, including a short polyprotein and non-conserved cysteine residues in NS1.

Also of importance was the recent isolation and identification of a new flavivirus, Kamiti River virus (KRV) from *Ae. macintoshi* mosquitoes in Kenya [5, 32]. In terms of both nucleotide sequence and growth kinetics in culture, KRV was most similar to the only other known insect-only flavivirus, CFAV. Notably, whereas CFAV was isolated from insect cells in the laboratory, KRV was isolated from a wild mosquito population. In addition, it has recently been shown that sequences related to the flaviviruses persist in DNA form integrated into the genome of some *Aedes* mosquito species [6]. Specifically, an ORF of 1557 amino acids closely related to the NS1-NS4A genes of CFAV and KRV was observed in both laboratory-bred and wild *Aedes albopictus* and the cell line C6/36. Similarly, in the *Aedes aegypti* cell line A20 and laboratory-bred and wild *Aedes aegypti* samples, a 492 amino acid ORF related to the NS5 of CFAV and KRV was detected. Other flaviviral-like sequences, in which genes were truncated or contained multiple stop codons were also found. These sequences most likely resulted from two or more independent integration events, following infection of each mosquito species by a virus (or viruses) related to the CFAV group. These findings raise questions regarding the possible existence of further members of the CFAV group in the wild that are as yet unidentified.

Members of the genus *Flavivirus* are highly genetically divergent, a combination of the intrinsically high mutation rates of RNA viruses [8] coupled with an extended period of independent evolution. As a consequence, one of the main obstacles to the higher-level analysis of the flaviviruses is the accurate alignment of highly divergent amino acid sequences. To determine the evolutionary relationships among the flaviviruses with as much accuracy as possible we undertook a comprehensive phylogenetic analysis involving multiple genes (NS3, NS5 and complete genomes), a variety of new and more robust methods of amino acid sequence alignment [9, 24] and appropriate models of amino acid replacement. Moreover, using a maximum likelihood approach, we explicitly tested the competing phylogenetic hypotheses for the phylogenetic positions of the NKV, mosquito- and tick-borne groups.

Materials and methods

Taxa

Amino acid sequence data sets for the NS5 gene (73 sequences), the NS3 gene (30 sequences), and the entire genome (23 sequences from the coding region only) were compiled for all available sequences for the flaviviruses to date. These are listed in Table 1.

Data analysis

For all data sets, sequence alignments were produced using three different protocols; (i) ClustalW [13], (ii) T-Coffee [24] and (iii) MUSCLE [9]. ClustalW is the most widely-used heuristic multiple alignment method, based on a progressive-alignment strategy [10]. This

Table 1. Flaviviruses analysed in this study, classified according to virus group (Heinz et al., 2001)

		NS5	NS3	Genome	Virus group
Alfuy virus, ALFV	M	AF013360	N/A	N/A	Japanese encephalitis
Alkhurma virus, ALKV	T	NC_004355	NC_004355	NP_722551	Mammalian tick-borne
Apoi virus, APOIV	N	NC_003676	NC_003676	NP_620045	Modoc
Aroa virus, AROAV	M	AF013362	N/A	N/A	Aroa
Bagaza virus, BAGV	M	AF013363	N/A	N/A	Ntaya
Banzi virus, BANV	M	L40951	N/A	N/A	Yellow Fever
Batu Cave virus, BCV	N	AF013369	N/A	N/A	Rio Bravo
Bouboui virus, BOUV	M	AF013364	N/A	N/A	Yellow Fever
Bukalasa Bat virus, BKV	N	AF013365	N/A	N/A	Rio Bravo
Bussuquara virus, BSQV	M	AF013366	N/A	N/A	Aroa
Cacipacore virus, CPCV	M	AF013367	N/A	N/A	Japanese encephalitis
Carey Island virus, CIV	N	AF013368	N/A	N/A	Rio Bravo
Cell Fusing Agent, CFAV	N	NC_001564	NC_001564	NP_041725	Unclassified
Cowbone Ridge virus, CRV	N	AF013370	AF297461	N/A	Modoc
Dakar Bat virus, DBV	N	AF013371	AF297462	N/A	Rio Bravo
Deer Tick, DTV	T	NC_003218	NC_003218	NP_476520	Mammalian tick-borne
Dengue virus 1, DENV1	M	M87512	M87512	M87512	Dengue
Dengue virus 2, DENV2	M	M19197	M19197	NP_056776	Dengue
Edge Hill virus, EHV	M	AF013372	N/A	N/A	Yellow Fever
Entebbe bat virus, ENTV	N	AF013373	AF295069	N/A	Entebbe bat
Gadgets Gully virus, GGYV	T	AF013374	N/A	N/A	Mammalian tick-borne
Iguape virus, IGUV	M	AF013375	N/A	N/A	Aroa
Ilheus virus, ILHV	M	AF013376	N/A	N/A	Ntaya
Israel Turkey Meningoencephalitis virus, ITV	M	AF013377	N/A	N/A	Ntaya
Japanese Encephalitis virus, JEV	M	M55506	M55506	AAA81554	Japanese encephalitis
Jugra virus, JUGV	M	AF013378	N/A	N/A	Yellow Fever
Jutiapa virus, JUTV	N	AF013379	N/A	N/A	Modoc
Kadam virus, KADV	T	AF013380	N/A	N/A	Mammalian tick-borne
Kamiti River virus, KRV	M	NC_005064	NC_005064	NP_891560	Unclassified
Karshi virus, KSIV	T	AF013381	AF297463	N/A	Mammalian tick-borne
Kedougou virus, KEDV	M	AF013382	N/A	N/A	Dengue
Kokobera virus, KOKV	M	AF013383	N/A	N/A	Kokobera
Koutango virus, KOUV	M	AF013384	N/A	N/A	Japanese encephalitis
Kunjin virus, KUNV	M	D00246	D00246	BAA00176	Japanese encephalitis
Kyasanur Forest disease virus, KFDV	T	AF013385	N/A	N/A	Mammalian tick-borne
Langat virus, LGTV	T	M86650	NC_003690	NP_620108	Mammalian tick-borne
Louping ill virus, LIV	T	Y07863	Y07863	NP_044677	Louping ill
Meaban virus, MEAV	T	AF013386	N/A	N/A	Seabird tick-borne
Modoc virus, MODV	N	AF013387	NC_003635	NP_619758	Modoc
Montana Myotis Leucoencephalitis virus, MMLV	N	AF013388	NC_004119	NP_689391	Rio Bravo
Murray Valley encephalitis virus, MVEV	M	AF013389	NC_000943	NP_051124	Japanese Encephalitis
Naranjal virus, NJLV	M	AF013390	N/A	N/A	Aroa

(*continued*)

Table 1 (*continued*)

		NS5	NS3	Genome	Virus group
Negishi virus, NEGV	T	AF013391	N/A	N/A	Tick-borne encephalitis
Ntaya virus, NTAV	M	AF013392	N/A	N/A	Ntaya
Omsk Haemorrhagic Fever virus, OHFV	T	AF013393	NC_005062	NP_878909	Mammalian tick-borne
Phnom Penh Bat virus, PPBV	N	AF013394	N/A	N/A	Rio Bravo
Potiskum virus, POTV	M	AF013395	N/A	N/A	Yellow Fever
Powassan virus, POWV	T	NC_003687	NC_003687	NP_620099	Mammalian tick-borne
Rio Bravo virus, RBV	N	AF013396	NC_003675	NP_620044	Rio Bravo
Rocio virus, ROCV	M	AF013397	N/A	N/A	Ntaya
Royal Farm virus, RFV	T	AF013398	N/A	N/A	Mammalian tick-borne
Russian spring summer encephalitis, RSSEV	T	AF013399	N/A	N/A	Tick-borne encephalitis
Saboya virus, SABV	M	AF013400	AF295070	N/A	Yellow Fever
Sal Vieja virus, SVV	N	AF013401	AF297460	N/A	Modoc
San Perlita virus, SPV	N	AF013402	N/A	N/A	Modoc
Saumaraez Reef virus, SREV	T	AF013403	N/A	N/A	Seabird tick-borne
Sepik virus, SEPV	M	AF013404	N/A	N/A	Yellow Fever
Sokoluk virus, SOKV	N	AF013405	N/A	N/A	Entebbe bat
Spondweni virus, SPOV	M	AF013406	N/A	N/A	Spondweni
St Louis Encephalitis virus, SLEV	M	AF013416	N/A	N/A	Japanese encephalitis
Stratford virus, STRV	M	AF013407	N/A	N/A	Kokobera
Tamana bat virus, TABV	N	NC_003996	NC_003996	NP_658908	Unclassified
Tembusu virus, TMUV	M	AF013408	N/A	N/A	Ntaya
Tick-borne Encephalitis virus, TBEV	T	U39292	NC_001672	NP_043135	Tick-borne encephalitis
Tyuleniy virus, TYUV	T	AF013410	N/A	N/A	Seabird tick-borne
Uganda S virus, UGSV	M	AF013411	N/A	N/A	Yellow Fever
Usutu virus, USUV	M	AF013412	AY453412	AAS59401	Japanese encephalitis
Wesselsbron virus, WESSV	M	N/A	AF295072	N/A	Yellow Fever
Western Tick-borne Encephalitis virus, WTBEV	T	U27495	U27495	AAA86870	Tick-borne encephalitis
West Nile virus, WNV	M	M12294	M12294	NP_041724	Japanese encephalitis
Yaounde virus, YAOV	M	AF013413	N/A	N/A	Japanese encephalitis
Yellow Fever virus, YFV	M	X03700	X03700	NP_041726	Yellow Fever
Yokose virus, YOKV	N	AB114858	NC_005039	NP_872627	Entebbe bat
Zika virus, ZIKV	M	AF013415	N/A	N/A	Spondweni

N/A: Sequence not available, or too short for inclusion in this study
M: Mosquito-borne, T: Tick-borne, N: No-known-vector

approach involves gradually building up an alignment from an initial, approximate phylogeny, following the order of the tree. However, errors made in the early stages of alignment are not rectified and the program attempts to align sequences along their full length i.e. it is a "global" alignment method. In contrast, T-Coffee computes a primary library of both global, via ClustalW and local, via Lalign [15, 25], pairwise alignments of all input sequences, which are weighted according to consistency of sequence identity before being "stacked". This is then extended into a multiple alignment using a position-specific scoring scheme. The MUSCLE algorithm involves three stages incorporating fast distance estimation using *k*mer counting, progressive alignment using a new profile function and refinement using tree-dependent restriction partitioning [9].

Phylogenetic trees for the alignment produced under each method were estimated using the maximum likelihood (ML) method available in TREE-PUZZLE [34] with 10,000 puzzling steps. To choose the model of amino acid replacement that best fitted the empirical data, the likelihood scores of trees produced by all the six models of amino acid replacement available in TREE-PUZZLE were compared for the full genome data set, both with equal rates of substitution and with a gamma distribution of rate heterogeneity with a shape parameter (alpha, α) of 1.0. The model that produced the phylogeny with the highest likelihood score for this data set was then used for further analyses. For the alignment method that gave the tree with the highest likelihood under these conditions for each individual data set, the ML value of the α parameter was then estimated from the empirical data. Analyses for the NS5 data set were conducted using two data sets, one including TABV and one with this highly divergent sequence removed.

To test alternative phylogenetic hypotheses for the evolutionary history of the genus *Flavivirus*, specifically the relationships among the tick-borne, mosquito-borne and NKV groups, we employed the Kishino and Hasegawa (KH) test which compares, statistically, the likelihoods of competing tree topologies [17]. First, we used TreeView (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html) to modify the ML trees for each data set to generate new phylogenies with branching orders consistent with three competing hypotheses; (i) that the mosquito-borne flaviviruses are the most divergent group, (ii) that the NKV flaviviruses are the most divergent group, and (iii) that the tick-borne flaviviruses are the most divergent group. These trees were then compared using the KH test in TREE-PUZZLE. We also calculated the GC content for the NKV, mosquito- and tick-borne groups to determine whether base composition had an effect on the degree of tree congruence.

Results

Sequences

The length and number of variable sites of final data sets used for phylogenetic analyses varied according to alignment method due to the differential insertion of gaps. For all alignment methods, the final data set for the NS5 gene comprised 352 amino acid sites, with 76.4% of sites being variable. For the full genome sequences, the final data set varied between 3556 (ClustalW), 3629 (T-COFFEE) and 3686 amino acid sites (MUSCLE), with 89.1% of sites being variable in all alignments. For the NS3 gene, the final alignment comprised 614 amino acid sites with 90.6% of sites being variable using MUSCLE, 610 amino acid sites

Table 2. Summary of alignment methods and α values of among-site rate variation used to infer the phylogenetic tree with the highest likelihood for each data set

Data set	Alignment method	α	−ln L	
NS5 (incl. TABV)	MUSCLE	0.5	−16958.00	
NS5 (excl. TABV)	T-COFFEE	1.0	−16172.68	Fig. 2
NS3	MUSCLE	1.0	−15409.45	Fig. 3
Genome	MUSCLE	1.0	−93589.53	Fig. 4

ln L, log likelihood

with 91.5% of sites being variable using ClustalW, and 622 amino acid sites with 91.0% of sites being variable using the T-COFFEE alignment.

Phylogenetic analyses

Using the full genome data set, the Whelan and Goldman (WAG) model of amino acid replacement, with a gamma distribution of rate heterogeneity (with 8 rate

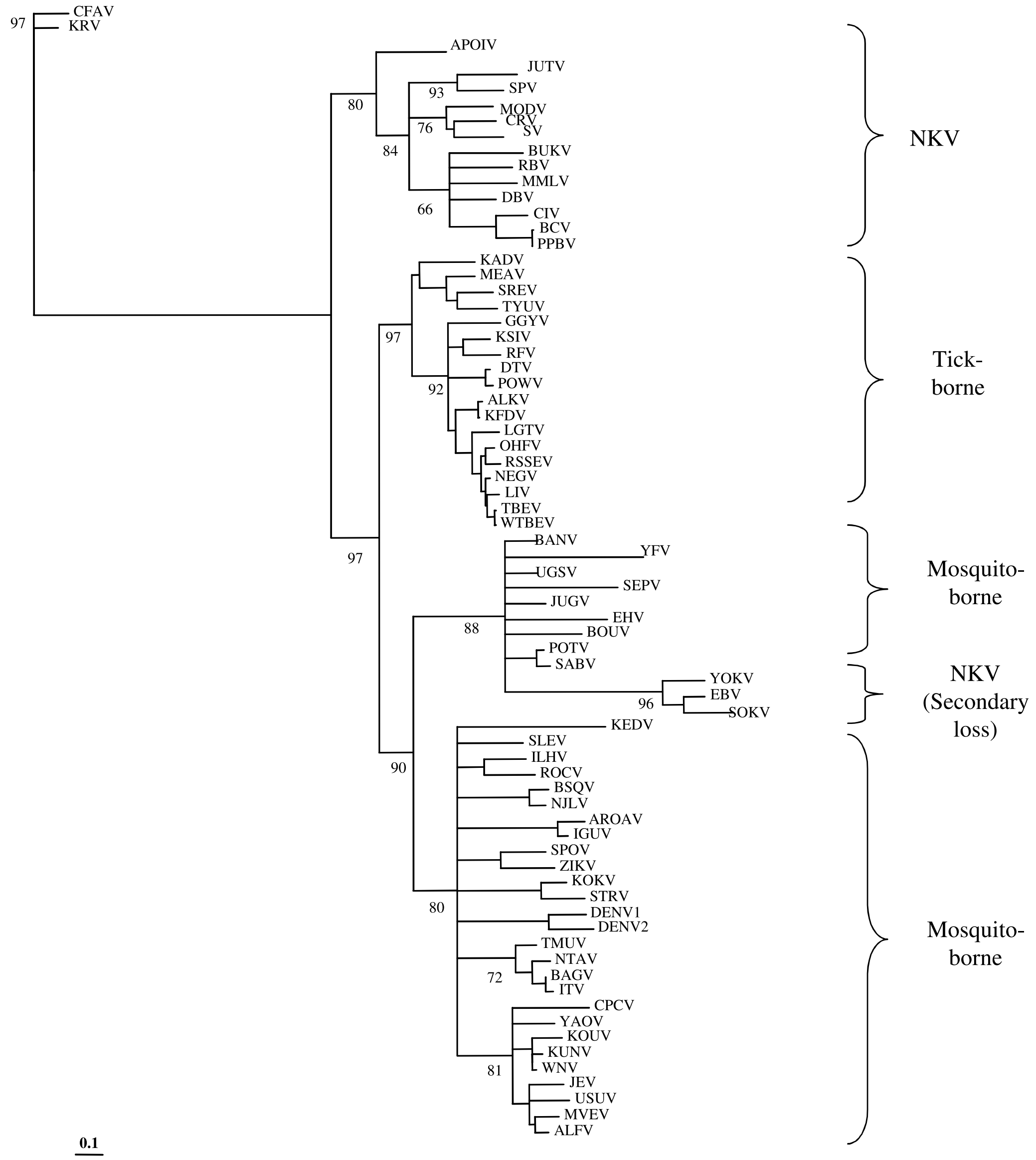

Fig. 2. Phylogenetic relationships of the genus *Flavivirus* inferred using the NS5 gene. Numbers next to branches depict quartet puzzling support values for main clades of interest, which give an indication of the robustness of each node on the current data, with 100 representing maximum support for the branch in question. Puzzling support values for all nodes were above 50

categories), gave the phylogeny with the highest likelihood. This model was then used to infer phylogenetic trees for all data sets since there was no evidence for different substitution processes between genes. Results of these analyses are summarised in Table 2. This reveals that MUSCLE was the best alignment method, in that it was associated with the highest likelihood in the resultant phylogenetic trees for all data sets except for the NS5 data set excluding TABV.

Preliminary analysis of the NS5 data set with TABV included demonstrated that this sequence represented a highly divergent outgroup even for this gene, which is the most strongly conserved among the flaviviruses. The phylogenetic position of TABV is not in question and its overly divergent nature means that it is unsuitable for use as an outgroup since we can no longer be certain of positional homology and an increase in the number of multiple substitutions may induce phylogenetic error. Therefore, all further analyses proceeded with the TABV sequence excluded. Similar reasoning precluded any divergent member of the Flaviviridae as a suitable outgroup. Hence, all trees are rooted on CFAV and KRV.

The ML tree for the NS5 gene is shown in Fig. 2. The phylogeny suggests that the NKV viruses form a monophyletic outgroup to a clade containing the tick-borne and mosquito-borne flaviviruses, the latter lineage including the monophyletic group of "secondary NKV" viruses in which vector-borne transmission has been lost. However, KH tests conducted on this gene indicate that the ML tree

Table 3. Results of the Kishino-Hasegawa (KH) test

Tree	−ln L	δ	KH test
NS5			
1. (Tick-borne, NKV) mosquito-borne	−16183.54	10.86	No significant difference
2. (Tick-borne, mosquito-borne), NKV	−16172.68	BEST	ML tree (Fig. 2)
3. (NKV, mosquito-borne), tick-borne	−16183.54	10.86	No significant difference
NS3			
1. (Tick-borne, NKV) mosquito-borne	−15409.45	BEST	ML tree (Fig. 3)
2. (Tick-borne, mosquito-borne), NKV	−15501.12	91.67	Significantly worse
3. (NKV, mosquito-borne), tick-borne	−15431.77	22.32	Significantly worse
Genome			
1. (Tick-borne, NKV) mosquito-borne	−93589.53	BEST	ML tree (Fig. 4)
2. (Tick-borne, mosquito-borne), NKV	−93608.02	18.49	No significant difference
3. (NKV, mosquito-borne), tick-borne	−93613.32	23.79	Significantly worse

For each data set, Hypothesis 1 comprises the tick-borne and NKV viruses as a sister-group to the mosquito-borne viruses ("NS3-like", Fig. 1B). Hypothesis 2 suggests that the two arthropod-borne clades are sister groups, with the NKV viruses being a divergent outgroup ("NS5-like" pattern, Fig. 1A). In Hypothesis 3, we tested the possibility that the NKV and mosquito-borne viruses were sister groups

ln L, log likelihood; δ difference in log likelihood from best tree. Tests are at the 5% level of significance is not significantly better than trees in which either of the vector-borne groups of viruses is made the most divergent clade (Table 3).

The ML phylogeny for the NS3 data set in shown in Fig. 3. In contrast to the NS5 tree, the mosquito-borne flaviviruses are now an outgroup to a clade comprising the tick-borne and NKV groups. Importantly, the KH test also significantly

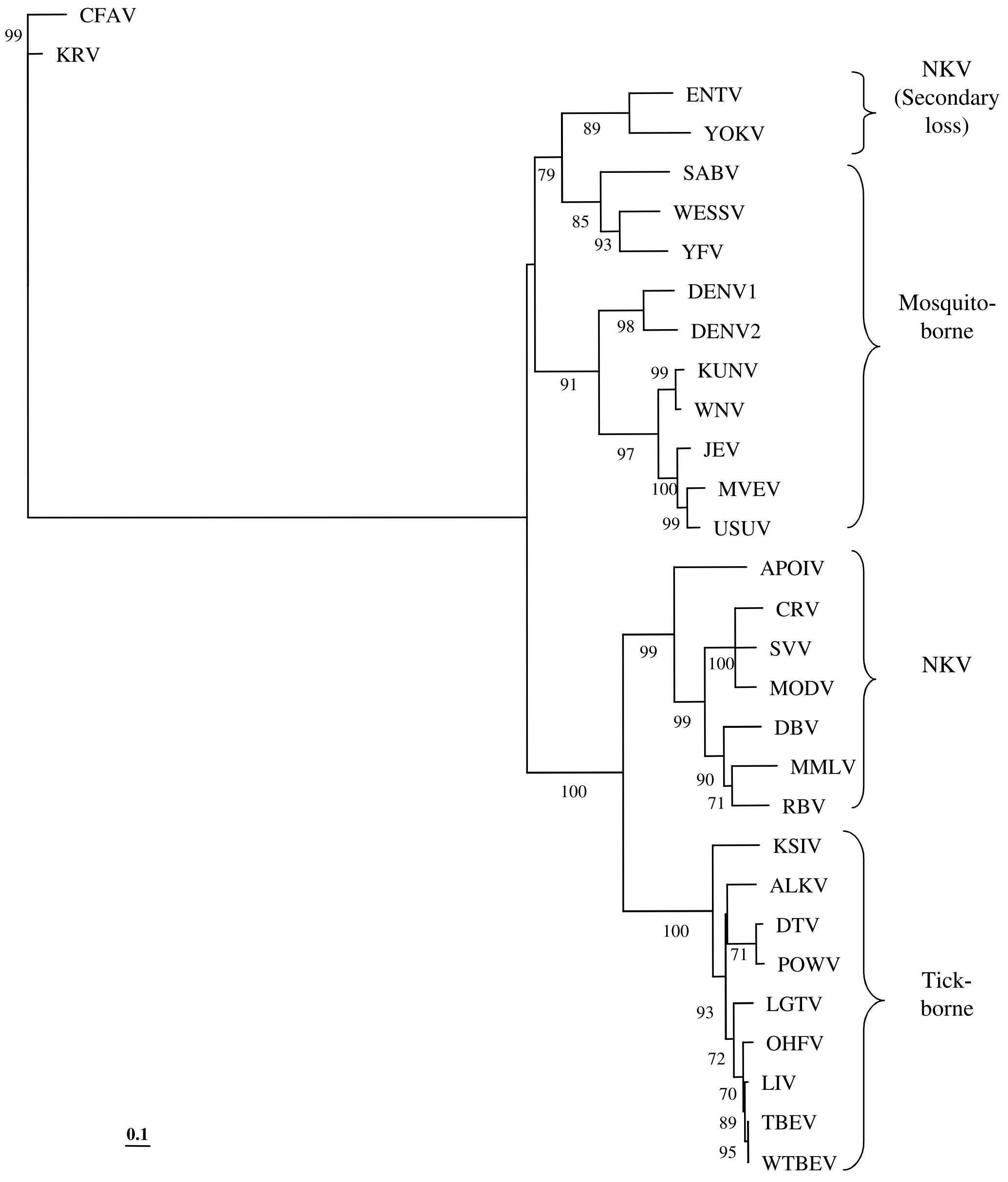

Fig. 3. Phylogenetic relationships of the genus *Flavivirus* inferred using the NS3 gene. Numbers next to branches depict quartet puzzling support values for clades, which give an indication of the robustness of each node on the current data, with 100 representing maximum support for the branch in question

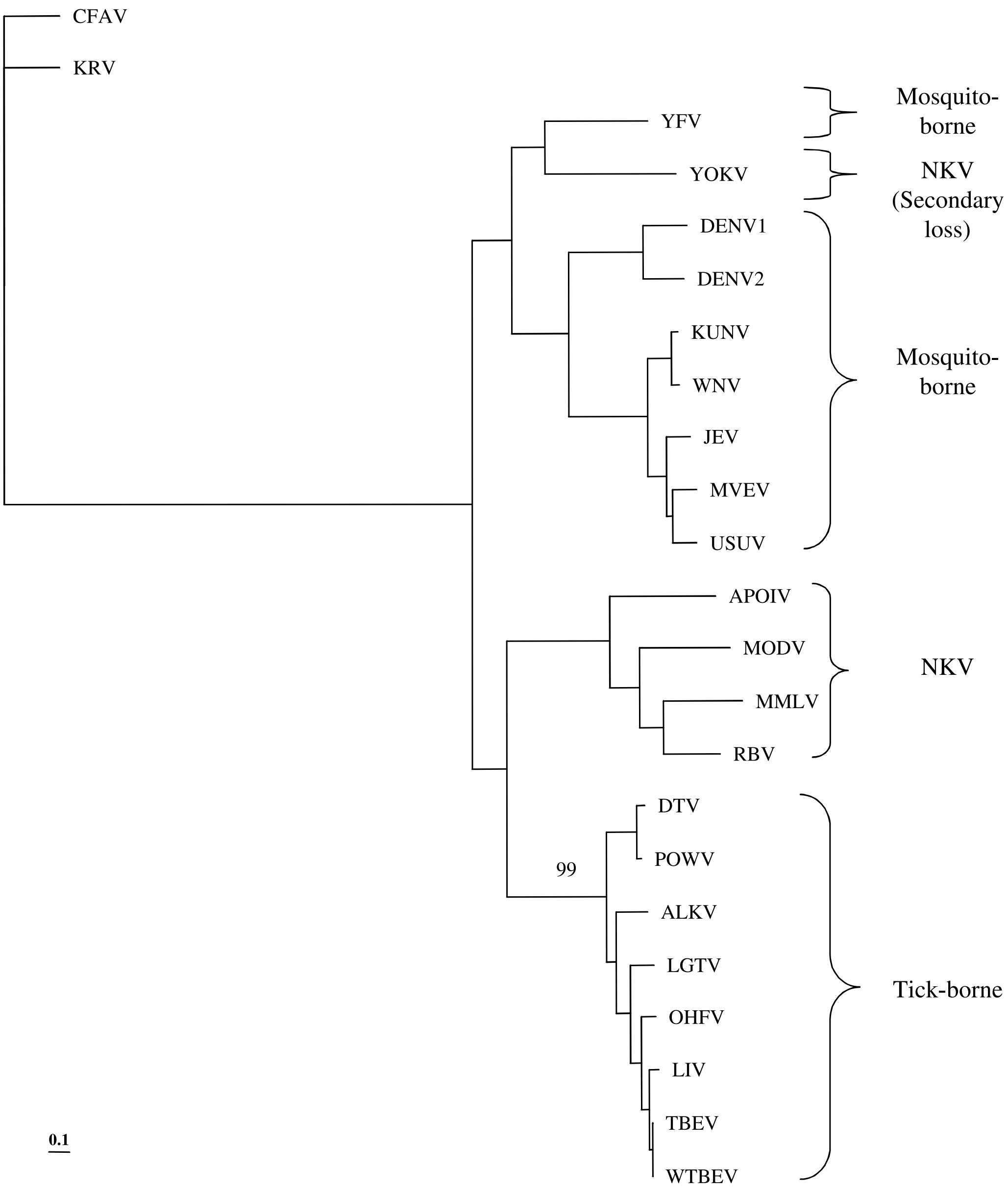

Fig. 4. Phylogenetic relationships of the genus *Flavivirus* inferred using the entire genome. Quartet puzzling support values for all clades were 100, with the exception of the clade shown, with a value of 99

supports the hypothesis that the NKV and tick-borne clades are sister-groups and that the mosquito-borne viruses represent a divergent outgroup (Table 3).

Finally, the ML phylogeny for the full genome data set is shown in Fig. 4. As in the case of the NS3 gene this trees supports an early divergence of the mosquito-borne viruses. The KH tests reveal that this topology has a significantly higher likelihood score than one comprising an early divergence of the tick-borne viruses

Table 4. GC content among different members of the genus Flavivirus

Group	No. taxa	Sequence length	%G	%C	%G + C
NS5					
Mosquito-borne	36	1038	29.5	20.7	50.2
NKV	14	1011	28.3	18.2	46.5
Tick-borne	18	1026	32.0	22.0	54.0
NS3					
Mosquito-borne	10	1803	27.6	21.9	49.5
NKV	7	1791	26.6	19.0	45.6
Tick-borne	9	1800	31.3	22.6	53.9

(Table 3). However, a phylogenetic tree with the NKV clade as an outgroup could not be rejected by the KH test on these data.

Finally, to determine whether changes in base composition have influenced our phylogenetic analysis we measured GC content among all the viruses in our data set (Table 4). The mosquito-borne group has a GC content intermediate between that of the NKV (lowest GC content) and tick-borne sequences (highest GC content). This is in agreement with Jenkins et al. [16] who determined that significant differences in GC content only existed between the NKV and tick-borne groups. However, as the NKV and tick-borne lineages group together in both the NS3 and full genome phylogenetic trees, we conclude that changes in base composition have not had a major effect on phylogenetic accuracy.

Discussion

Our analysis represents the most comprehensive phylogenetic study of the genus *Flavivirus* undertaken to date. With respect to the mosquito-borne flaviviruses, all analyses thus far, including the current study, suggest a clear division between the YFV clade including the "secondary loss" NKV flaviviruses and a sister-group containing the remaining mosquito-borne members. Billoir et al. [1] first mapped mosquito vector species onto the NS5 and NS3 phylogenies and proposed that the *Aedes*-associated viral lineages were paraphyletic whereas the *Culex*-associated clade was monophyletic, although the number of representatives from each group was low. This idea was supported by the NS5 phylogeny of Jenkins et al. [16]. Gaunt et al. [11] further suggested that the mosquito-borne flaviviruses could be split into two distinct epidemiological groups: (i) the neurotropic viruses often associated with *Culex* species and bird reservoirs, and (ii) the non-neurotropic viruses, associated with haemorrhagic disease in humans, correlated with *Aedes* mosquitoes and primate hosts. In fact, the original NS5 phylogeny of Kuno et al. [18] suggested both the *Aedes*- and *Culex*-associated flaviviral clades were paraphyletic due to the presence of SPOV and ZIKV nested within the *Culex* lineage. The analysis of NS5 undertaken here is equivocal since the DENV serotypes, KEDV, SPOV and ZIKV all appear to be more closely related to the *Culex*-associated flaviviruses than to the other *Aedes*-associated members and the *Culex* clade is not well-resolved. In contrast, the NS3 and whole genome data sets support the hypotheses of Gaunt et al. [11] more strongly, but the sample size is significantly smaller and sequences for KEDV, SPOV and ZIKV are not available for these regions. Taken together, it is evident that further studies of the vector competence, host specificity, host range and disease aetiology and pathogenesis of each mosquito-borne virus are required before the suggestions made by Gaunt et al. [11] can be fully tested. Similarly, it is essential to clarify the phylogenetic relationships of the *Aedes* and *Culex* mosquitoes. Recently, Reinert [29] used morphological characters to suggest that the genus *Aedes* as a composite genus separate from a second genus, *Ochlerotatus*. However, since Reinert's work, Savage and Strickman [33] have argued for the restoration of the traditional usage of the genus *Aedes* and subgenus *Ae.* (*Ochlerotatus*) since female adult specimens of *Ochlerotatus* and *Aedes* as defined by Reinert cannot be identified morphologically without dissection and no distinct biological, behavioural or ecological differences seem to distinguish the two groups. Hence, current research refers to *Ochlerotatus* as a subgenus of *Aedes* even though no molecular studies to date have examined the status of these taxa. Clearly, in order to determine virus host specificity, an accurate system for the delineation of *Aedes* species is first required.

For the tick-borne viruses, most work to date points to the existence of two main clades, one containing the flaviviruses infecting seabird colonies (KADV, MEAV, SREV and TYUV) and the other primarily associated with rodents (e.g. LIV). Both Gaunt et al. [11], and Jenkins et al. [16] found that POWV occupied a basal position within this second clade. However, our study is in agreement with Kuno et al. [18] whose NS5 phylogeny suggested that this virus did not represent an outgroup to the other members of the lineage. In addition to equivocal evidence regarding the ancestry of this clade, the geographic range and host range of the tick-borne viruses in general is not clear. For example, RSSEV has been confirmed in the wild outside Russia, in Japan [35]. Therefore, although the characteristics and likely mechanism of dispersal seems clear, few conclusions can be drawn about the early origin and spread of those flaviviruses associated with ticks based on the data in hand.

The majority of previous studies of flavivirus evolution have suggested that arthropod-mediated transmission is a derived trait within the genus, with the ancestral condition being non-vector transmission [2, 18]. Various observations have been cited as evidence in support of this hypothesis. First, none of the NKV viruses tested by Varelas-Wesley and Calisher [38] replicated in mosquito cell culture. In contrast, some flaviviruses from the mosquito-borne group have been isolated from ticks, such as WNV, YFV and SLEV, whereas POWV is the only tick-borne flavivirus that has been isolated in mosquitoes (however, it should be noted that isolation of a virus from a hematophagous arthropod does not automatically imply infection or replication). Further, the Tyuleniy group of tick-borne flaviviruses displays some properties typical of mosquito-borne viruses including the absence of a hexapeptide insertion, possession of a common glycosylation site in the E gene and ability to replicate in mosquito cell culture. These properties have on occasion been suggested to represent a vestigial trait found in mosquito-borne flaviviruses as a result of a past association with ticks [18]. Second, the majority of the other members of the *Flaviviridae*, namely the pestiviruses and the hepaciviruses, are not associated with vector-borne transmission, although there are some very limited examples of laboratory transmission of bovine viral diarrhea virus by bloodfeeding flies [37] and equivocal evidence for transmission of hepatitis C virus by ticks [39]. Therefore, based on current evidence, it is most parsimonious to assume that the absence of a vector is the ancestral condition for this family of viruses.

More direct evidence for the transition from non-vector to vector-borne transmission was presented by phylogenetic analyses of the NS5 gene which suggested that the NKV group diverged before the arthropod-borne flaviviruses [1, 11, 16]. However, our study shows that the NS5 data set possesses insufficient phylogenetic signal to discriminate between topological hypotheses regarding the relationships of the three main transmission groups of flaviviruses. In contrast, our NS3 analysis provides statistically significant evidence that the mosquito-borne viruses are a divergent outgroup to the NKV and tick-borne clades. The phylogeny estimated from the full genome data is also compatible with this hypothesis, although the possibility of a NKV outgroup cannot be rejected.

Taken together, two working hypotheses are consistent with the phylogenetic trees presented here: (i) that the NKV group diverged before the arthropod-borne flaviviruses, a possibility that cannot be ruled out by the full genome or NS5 data sets but that is rejected using the NS3 data alone, and (ii) that the mosquito-borne flaviviruses diverged first, as strongly suggested by the NS3 data set and compatible with the analysis of the full genome data. The latter hypothesis conflicts strongly with traditional views regarding the evolution of the genus, but is best supported by the flavivirus sequence data currently available. In either case, the acquisition of tick-borne transmission is clearly a derived trait within the flaviviruses as in every analysis the tick-borne group was rejected as the most divergent clade.

Importantly, some aspects of previous studies do support the early divergence of the vector-borne viruses, such as the "NS3-like" phylogenies determined by Billoir et al. [1] for the NS3 gene and entire ORF sequences of the flaviviruses available at that time. Indeed, although these authors did not regard their phylogenetic study as conclusive, they suggested that the NS3 region was most appropriate for determining phylogenetic relationships within the flaviviruses. Our KH tests examine this proposal and reveal that, in contrast with the NS5 gene, the NS3 gene is capable of discriminating between topological hypotheses, making it imperative that NS3 sequences are collected from a larger sample of flaviviruses. The "NS3-like" pattern is supported by a number of other observations; (i) some members of the NKV group, such as PPBV and CIV, are serologically-related to tick-borne viruses [4], and (ii) a typical Asian tick-borne encephalitis strain has been isolated from *Apodemus speciosus*, the natural rodent host of APOIV [36]. More generally, the genus *Flavivirus* has a broad invertebrate range and many flaviviruses have been isolated from arthropods other than their main vectors of virus transmission. For example, YFV has been isolated from ticks [22], SABV has been isolated from both *Anopheles* and from ticks [3], and SLEV and WNV have been isolated from ticks [21]. This may be a reflection of the conservation of genetic characters inherited from a common flaviviral ancestor associated with mosquitoes, although it should be noted that isolation of a virus from an unexpected host could also be due to the chance acquisition of a non-replicating virus in a blood-meal.

Irrespective of which mode of transmission is ancestral in the flaviviruses, it would appear that these viruses have their origins in the Old World. In particular, the earliest evolutionary lineages of the *Aedes*-borne virus clades appear to have an African ancestry since only YFV, the four DENV serotypes and WNV are found in the New World and these appear to be more recent migrations. Second, virtually all of the tick-borne flaviviruses are found in the Old World, with the exception of POWV. Third, of the NKV viruses, only flaviviruses associated with bats (BUKV, CIV, DKV, PPBV and RBV) have been isolated from both the New World and Old World. In contrast, members of the rodent clade (CRV, JUTV, MODV, SVV and SPV) have only been isolated in the New World, with the exception of APOIV. This is in agreement with a single dispersion event from the Old World followed by local infection of rodents, which are less mobile and less likely to play a role in the global dispersion of the flaviviruses in contrast to bats.

If the mosquito-borne flaviviruses do indeed represent the most divergent outgroup, relative to the NKV and tick-borne members of the genus, we would expect to find numerous flaviviruses associated with mosquitoes that fall outside the three main clades, representing earlier lineages. This is exactly the case with the recent discovery of KRV [5, 32]. This flavivirus, found in *Aedes macintoshi* mosquitoes in Kenya, clearly falls with CFAV in all three phylogenies in the current study. Recent theoretical work also suggests there could be a large number of currently unidentified mosquito-borne flaviviruses. Using a phylogenetic method to estimate the level of taxon sampling in a clade, the number of unsampled taxa in the mosquito-borne flavivirus clade is estimated to be approximately 2000 [28]. Since it is clear that the currently known flaviviruses represent only a very small sample of those present in nature, making strong conclusions about the likely absence of a vector as the ancestral transmission mode for the *Flaviviridae* is perhaps premature based on the present data. Exhaustive research aimed at the investigation of further examples of such lineages has not been conducted to date yet holds the key to further clarifying the evolution of the flaviviruses.

Acknowledgements

This work was funded by the Wellcome Trust.

References

1. Billoir F, de Chesse R, Tolou H, Micco P, Gould EA, de Lamballarie X (2000) Phylogeny of the genus *Flavivirus* using complete coding sequences of arthropod-borne viruses and viruses with no known vector. J Gen Virol 81: 781–790
2. Blok J, Gibbs AJ (1995) Molecular systematics of the flaviviruses and their relatives. In: Gibbs A, Calisher CH, Garcia-Arenal F (eds) Molecular basis of virus evolution. Cambridge University Press, Cambridge, England, pp 270–289
3. Butenko AM (1996) Arbovirus circulation in the Republic of Guinea. Med Parazitol (Mosk) 2: 40–45
4. Calisher CH, Karabatsos N, Dalrymple JM, Shope RE, Porterfield JS, Westaway EG, Brandt WE (1989) Antigenic relationships between flaviviruses as determined by cross-neutralisation tests with polyclonal antisera. J Gen Virol 70: 37–43
5. Crabtree MB, Sang RC, Stollar V, Dunster LM, Miller BR (2003) Genetic and phenotypic characterisation of the newly described insect flavivirus, Kamiti River virus. Arch Virol 148: 1095–1118
6. Crochu S, Cook S, Attoui H, Charrel R, De Cheese R, Belhouchet M, Lemasson J-J, de Micco P, de Lamballarie X (2004) Sequence of flavivirus-related RNA viruses persist in DNA form in the genome of *Aedes* spp. mosquitoes. J Gen Virol 85: 1971–1980
7. De Lamballerie X, Crochu S, Billoir F, Neuts J, de Micco P, Holmes EC, Gould EA (2002) Genome sequence analysis of Tamana bat virus and its relationship with the genus Flavivirus. J Gen Virol 83: 2443–2454
8. Domingo E, Holland JJ (1997) RNA virus mutations for fitness and survival. Ann Rev Microbiol 51: 151–178
9. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797
10. Feng D-F, Doolittle RF (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J Mol Evol 25: 351–360
11. Gaunt MW, Sall AA, de Lamballarie X, Falconar AKI, Dzihivanian TI, Gould EA (2001) Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. J Gen Virol 82: 1867–1876
12. Guzman MG, Kouri G (2002) Dengue: an update. Lancet Infect Dis 2: 33–42
13. Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680
14. Holmes EC (1998) Molecular epidemiology and evolution of emerging infectious diseases. Br Med Bull 54: 533–543
15. Huang X, Miller W (1991) A time-efficient, linear-space local similarity algorithm. Adv Appl Math 12: 337–357
16. Jenkins GM, Pagel M, Gould EA, Zanotto PM de A, Holmes EC (2001) Evolution of base composition and codon usage bias in the genus Flavivirus. J Mol Evol 52: 383–390
17. Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol 29: 170–179
18. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB (1998) Phylogeny of the genus Flavivirus. J Virol 72: 72–83
19. Lewis JG, Chang GJ, Lanciotti RS, Kinney RM, Mayer LW, Trent DW (1993) Phylogenetic relationships of dengue-2 viruses. Virology 197: 216–224
20. Marin MS, Zanotto PM de A, Gritsun TS, Gould EA (1995) Phylogeny of TYU, SRE, and CFA virus – different evolutionary rates in the genus *Flavivirus*. Virology 206: 1133–1139

21. Monath TP, Tsai TF (1987) St Louis encephalitis: lessons from the last decade. Am J Trop Med Hyg 37: 40S–59S
22. Monath TP, Heinz FX (1996) Flaviviruses. In: Fields BN (ed) Virology, vol 1. Lipincott-Raven, Philadelphia, pp 961–1034
23. Monath TP (2001) Yellow fever: an update. Lancet Infect Dis 1: 11–20
24. Notredame C, Higgins D, Heringa J (2000) T-coffee: a novel method for multiple sequence alignments. J Mol Biol 302: 205–217
25. Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85: 2444–2448
26. Porterfield JS (1980) Antigenic characteristics and classification of Togaviridae. In: Schlesinger RW (ed) The Togaviruses. Academic Press, New York, pp 13–46
27. Price JL (1978) Isolation of Rio Bravo and a hitherto undescribed agent, Tamana Bat virus, from insectivorous bats in Trinidad, with serological evidence of infection in bats and man. Am J Trop Med Hyg 27: 153–161
28. Pybus OG, Rambaut A, Holmes E, Harvey PH (2002) New inferences from tree shape: numbers of missing taxa and population growth rates. Syst Biol 51: 881–888
29. Reinert (2000) New classification for the composite genus *Aedes* (Diptera: Culicidae: Aedini), elevation of subgenus *Ochlerotatus* to generic rank, reclassification of the other subgenera, and notes on certain subgenera and species. J Am Mosq Control Assoc 16: 175–188
30. Rice CM, Lenches EM, Eddy SR, Shin SJ, Sheets RL, Strauss JH (1985) Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science 229: 726–733
31. Rice CM (1996) Flaviviridae: The Viruses and their Replication. In: Fields BN (ed) Virology, vol 1. Lippincott-Raven, Philadelphia, pp 931–960
32. Sang RC, Gichogo A, Gachoya J, Dunster MD, Ofula V, Hunt AR, Crabtree MB, Miller BR, Dunster LM (2003) Isolation of a new flavivirus related to Cell fusing agent virus (CFAV) from field-collected flood-water *Aedes* mosquitoes from a dambo in central Kenya. Arch Virol 148: 1085–1093
33. Savage HM, Strickman D (2004) The genus and subgenus categories within Culicidae and placement of *Ochlerotatus* as a subgenus of *Aedes*. J Am Mosq Control Assoc 20: 208–214
34. Strimmer K, von Haesler A (1996) Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol Biol Evol 13: 964–969
35. Takashima I, Morita K, Chiba M, Hayasaka D, Sato T, Takesawa C, Igarashi A, Kariwa H, Yoshimatsu K, Ariakawa J, Hashimoto N (1997) A case of tick-borne encephalitis in Japan and isolation of the virus. J Clin Microbiol 35: 1943–1947
36. Takeda T, Ito T, Osada M, Takahashi K, Takashima I (1999) Isolation of tick-borne encephalitis from wild rodents and a seroepizootic survey in Hokkaido, Japan. Am J Trop Med Hyg 60: 287–291
37. Tarry DW, Bernal LSE (1991) Transmission of bovine virus diarrhoea virus by blood feeding flies. Vet Record 128: 82–84
38. Varelas-Wesley L, Calisher CH (1982) Antigenic relationships of flaviviruses with undetermined arthropod-borne status. Am J Trop Med Hyg 31: 1273–1284
39. Wurzel LG, Cable RG, Leiby DA (2002) Can ticks be vectors for hepatitis C virus? N Engl J Med 347: 1724–1725
40. Zanotto PM de A, Gould EA, Gao GF, Harvey PH, Holmes EC (1996) Population dynamics of flaviviruses revealed by molecular phylogenies. Proc Natl Acad Sci USA 93: 548–553

Author's address: Shelley Cook, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK; e-mail: shelley.cook@balliol.oxon.org