New Insights into Flavivirus Evolution, Taxonomy and Biogeographic History, Extended by Analysis of Canonical and Alternative Coding Sequences

Gregory Moureau[1]*, Shelley Cook[2], Philippe Lemey[3], Antoine Nougairede[1], Naomi L. Forrester[4], Maxim Khasnatinov[5], Remi N. Charrel[1], Andrew E. Firth[6], Ernest A. Gould[1], Xavier de Lamballerie[1]

1 Aix Marseille Université, IRD French Institute of Research for Development, EHESP French School of Public Health, EPV UMR_D 190 Emergence des Pathologies Virales, Marseille, France, 2 Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom, 3 Department of Microbiology and Immunology, Rega Institute, KU Leuven, Minderbroedersstraat 10, 3000 Leuven, Belgium, 4 Institute for Human Infections and Immunity and Department of Pathology, University of Texas Medical Branch, Galveston, TX 77555, United States of America, 5 Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Crowmarsh, Gifford, Wallingford, Oxfordshire, OX10, United Kingdom, 6 Division of Virology, Department of Pathology, University of Cambridge, Cambridge CB2 1QP, United Kingdom

* gregory.moureau@univ-amu.fr

Abstract

To generate the most diverse phylogenetic dataset for the flaviviruses to date, we determined the genomic sequences and phylogenetic relationships of 14 flaviviruses, of which 10 are primarily associated with *Culex* spp. mosquitoes. We analyze these data, in conjunction with a comprehensive collection of flavivirus genomes, to characterize flavivirus evolutionary and biogeographic history in unprecedented detail and breadth. Based on the presumed introduction of yellow fever virus into the Americas via the transatlantic slave trade, we extrapolated a timescale for a relevant subset of flaviviruses whose evolutionary history, shows that different *Culex*-spp. associated flaviviruses have been introduced from the Old World to the New World on at least five separate occasions, with 2 different sets of factors likely to have contributed to the dispersal of the different viruses. We also discuss the significance of programmed ribosomal frameshifting in a central region of the polyprotein open reading frame in some mosquito-associated flaviviruses.

INTRODUCTION

The flaviviruses constitute a fascinating group of diverse arboviruses that exhibit uniquely clear correlations between phylogenetic relationships and virus-vector-host interactions [1–4]. The genus *Flavivirus* includes an unusually large number of taxonomically recognised species (more than 50 at the present time, of which more than 40 are human pathogens) with a global

distribution. The genus also includes a large and increasing number of unclassified or "tentative" species. Pathogenic mosquito- and/or tick-borne flaviviruses cause a variety of clinical diseases in a wide range of vertebrate species. These disease syndromes include mild/severe febrile illness, "flu-like" syndromes with a rash, or in other cases severe encephalitis or haemorrhagic disease. Dengue with / without warning signs and severe dengue, is the most devastating arboviral disease in tropical and, increasingly, sub-tropical areas of the world (300 to 400 million cases each year) [5]. Yellow fever virus and Japanese encephalitis virus also considerably contribute to the human flavivirus disease burden. Other flavivirus diseases, including West Nile encephalitis, Usutu encephalitis, Zika fever, Bagaza encephalitis and duck egg drop syndrome, are recognised as "emerging diseases".

Flavivirus ecological networks are varied, complex, and poorly understood. Importantly, most of the natural pathogens are transmitted by arthropods (*i.e.* they are "arboviruses"). However, flaviviruses with no known vector (NKV), or that infect only insects (*i.e.*, insect-specific flaviviruses—ISFVs), have also been identified. This remarkable diversity is associated with broad genetic variability, complex mechanisms of pathogenesis and intriguing virus/vector/host associations.

The taxonomy of the flaviviruses is constantly being updated to reflect newly-identified viruses and advances in analytical methods. Interestingly, the first mammalian viruses to be identified included 4 arboviruses, three of which were flaviviruses: louping ill virus, yellow fever virus and dengue virus [6]. This has had a deep and long-lasting influence on the development of virological research and more specifically, on taxonomy [7] and phylogeography.

With the publication of many new complete genomic flavivirus sequences [3,8] and the data generated in the current study, it is now timely and appropriate to re-examine the phylogenetic relationships in the context of flavivirus vector-host relationships, evolution and biogeographical characteristics.

The concept that the phylogenetic relationships of the tick-borne flaviviruses (TBFV) may correlate with their epidemiology, disease association and biogeography was first proposed in 1996 with the publication of the clinal evolution concept of the tick-borne encephalitic flaviviruses [9]. These relationships were then corroborated and extended by the inclusion of the mosquito-borne and non-vectored flaviviruses [1] but analyses were still based on a limited number of recognised flaviviruses and only partial gene sequence data (44 species based on the envelope gene). Subsequent analyses [3,8,10,11], using more extensive datasets improved our understanding of these virus-vector-host relationships in the context of their evolution and dispersal [2,12]. Additional flaviviruses including Lammi virus [13], N'goye virus [14], Alkhumra haemorrhagic fever virus [15], Usutu virus [16], New Mapoon virus [17], and Marisma mosquito virus [18] have subsequently been discovered. Together with the increasing numbers of documented species and strains of ISFV, that do not appear to be arboviruses [19–23], these discoveries have shed new light on our perception of the evolution and taxonomy of this complex genus.

Based on the flavivirus arthropod vectors and vertebrate hosts, current phylogenies recognise three major groups in addition to the ISFVs [1,3,8,10,13]: the tick-borne, mosquito-borne, and no known-vector flaviviruses (TBFV, MBFV and NKV respectively). The TBFV are subdivided into pathogenic flaviviruses primarily associated with *Ixodes* spp., and apathogenic flaviviruses associated with *Ornithodorus* spp. ticks that primarily feed on or parasitize, seabirds. The mosquito-borne flaviviruses (MBFV) are sub-divided into those primarily associated with *Culex* spp. (ornithophilic) mosquitoes and those primarily associated with *Aedes* spp. (mammalophilic) mosquitoes. In contrast, the viruses with no known vectors (NKV) are divided into those primarily associated either with rodents or bats. These specific lineage NKVs are defined in the text as NKV^{SL}. In addition, the three viruses (Sokoluk, Entebbe bat and Yokose virus)

are exceptional in that they diverged with the MBFV but arthropod vectors have not been associated with these viruses. In other words, they diverged within the MBFV—*Aedes* spp.-associated clade but appear to have lost this mosquito association [10]. These NKV are defined in the text as NKV-like.

For further clarity, we will also distinguish ISFVs which are very divergent from the other members of the genus *Flavivirus* by referring to them as ISFV^{SL}. Finally, we will distinguish recently discovered viruses such as LAMV, which fall phylogenetically within the MBFV group but appear to infect only insects, as ISF-like.

The genus *Flavivirus* also includes 2 highly divergent genetic lineages not associated with any recognized group in the ninth report of the International Committee on Taxonomy of Viruses (ICTV) [7], represented by Tamana bat virus [TABV] [11] and Ngoye virus [14]. Recently, another highly divergent lineage, Mogiana tick virus [MGTV], was isolated in Brazil [24].

Whilst early interpretations of the evolutionary and taxonomic relationships within the genus *Flavivirus* have proved informative, they were restricted to some extent by the lack of complete genomic sequence data, and robust analytical methods. As a result, it has not yet proven possible to resolve the issue that phylogenies based on complete genomes and, separately, the NS3 gene show different branching characteristics from those based on the NS5 gene, despite a lack of evidence for recombination within the respective data sets [10,11,25]. In this manuscript we report the genomic sequences of fourteen flaviviruses, for which previously only limited data were available. Based on these more comprehensive datasets, we attempt to resolve hitherto unanswered questions relating to the flaviviruses, and explore frame shift characteristics. Additionally, we identify viruses that appear to have been introduced from the Old World into the New World, estimate the likely times prior to the present that these introductions occurred, and discuss factors that probably contributed to the global dispersal of these viruses.

MATERIALS AND METHODS

Viruses

MBFVs included in the study were all subcultured at least once in C6/36 cells. Batu cave virus [BCV], Jutiapa virus [JUTV], Phnom Penh bat virus [PPBV] and Sokuluk virus [SOKV] were amplified in the mammalian cell-line BHK21 and Sitiawan virus [STWV] was amplified in Vero cell cultures. STWV virus was kindly provided by Dr. Yuji Kono as inactivated nucleic acid in RNA-Now lysis buffer.

Nucleic acid preparation

Viral RNA was either extracted using the BioRobot EZ1 (Viral RNA Mini kit: Qiagen) or RNA-Now (Biogentex) using the manufacturer's recommendations. Reverse transcription was carried out using Taqman Reverse transcription reagents (Applied Biosystems) under standard conditions with random hexamers as primers.

Polymerase chain reaction (PCR) in the conserved region of the genome

PCR was targeted at the E, NS3 and NS5 gene-conserved regions using consensus degenerate primers [8,26–28]. Sequences for the NS3 gene region were obtained using NS3-FS (5'-GGIGTIYTICAYACIATGTGGCAYGTIAC-3')/NS3-FR (5'-TKICKICCIAYICKICCICK-ICKYTGIGCNGY-3') primers in first round PCR, followed by nested PCR using X1 (5'-YIRTIGGIYTITAYGGIWWYGG-3')/X2 (5'-RTTIGCICCCATYTCISHDATRTCIGT-3') primers, with standard conditions and a hybridisation temperature of 45°C.

Sequencing strategy

Specific primers were designed from the 3 conserved region sequences defined above and long-range PCR was conducted to complete the sequencing of gap between the E to NS3 genes, and NS3 to NS5 genes, using the cMaster RTplusPCR system (Eppendorf) [3]. A long PCR product sequencing protocol (LoPPs) was employed to sequence amplicons [29,30].

PCR amplification of the 5' and 3' end of coding sequences

The 5'-terminal region of the genome sequence was obtained using an E-gene specific reverse primer and a forward degenerate primer in the 5'UTR designed using an alignment of available 5' UTRs of *Culex*-spp. associated flavivirus sequences (5'-CULEX-S1: 5'-AGwiGTTCryCT-GyGTGArCT-3'; position 1–21 of the Japanese encephalitis virus genome). Semi-nested PCR with a second inner virus-specific primer was also used when necessary. The 3'-terminal region of the genome was obtained using a similar strategy with virus-specific forward primers in the NS5 region and a reverse degenerate primer in the 3'UTR (3'UTR-MOS: 5'-GGTCTCCWMTAACCTCTAG-3').

Sequencing was conducted with the primers used for amplification, or with M13 primers after cloning in a pCR2 cloning vector (Invitrogen).

Complete polyprotein open-reading frame (ORF) sequences, excluding the partial untranslated regions (UTR) regions obtained by this protocol were used for further analyses.
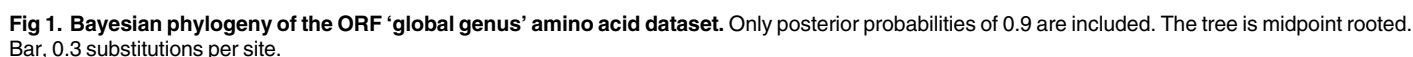
Next Generation Sequencing (NGS)

Resequencing of seven of the eleven new complete polyprotein ORF sequences (AROAV, CPCV, ITV, KOUV, NTAV, TMUV and YAOV) was performed using the Ion PGM Sequencer (Life Technologies SAS, Saint Aubin, France) [31] and a random reverse transcription-amplification protocol. Reads, of minimum length 30 nucleotides, were trimmed using CLC Genomic Workbench 6.5 (QIAGEN Company), with a minimum of 99% quality per base and mapped to reference sequences previously obtained by the Sanger method. Parameters were set such that each accepted read had to map to the reference sequence for at least 50% of its length, with a minimum of 80% identity to the reference.

BCV, JUTV and PPBV sequences were obtained using the same NGS method and *de novo* assembly.

Sequence analysis

Sequences were refined using Sequencher 4.8 (Gene Codes, Ann Arbor, MI) and combined with other flavivirus sequences available in the Genbank database, to obtain a dataset including a representative of at least one sequence for each species available as full polyprotein ORF for the genus *Flavivirus*. Genbank accession numbers of sequences used for the analysis are noted after each virus abbreviation on Fig. 1 (tree). Complete polyprotein ORF amino acid alignments were generated using both Clustal W2 [32,33] and MUSCLE [34] available at the EMBL server (http://www.ebi.ac.uk/Tools/) and refined manually, for comparison. Nucleotide alignments were then deduced using amino acid (AA) alignments as a template using the TranAlign software available via the EMBOSS server (http://emboss.bioinformatics.nl/cgi-bin/emboss/tranalign). The effect of removing regions of ambiguous alignment via the GBlocks algorithm [35] using less stringent parameters was also investigated.

Amino acid phylogenetic trees were reconstructed using Markov chain Monte Carlo (MCMC) analysis implemented in MrBayes v3.1.2 [36]. The analysis was performed using the WAG substitution model with gamma distributed rate variation among sites and using default

Fig 1. Bayesian phylogeny of the ORF 'global genus' amino acid dataset. Only posterior probabilities of 0.9 are included. The tree is midpoint rooted. Bar, 0.3 substitutions per site.

doi:10.1371/journal.pone.0117849.g001

priors. Five independent Markov chains were run for 10 million generations, with the first 10% of samples discarded as burn-in. Stationarity was confirmed based on effective sample sizes >400 using Tracer v1.4.1 [37]. A maximum clade credibility tree was summarised using TreeAnnotator which annotates all nodes with posterior probability support values. A range of nucleotide analyses was also conducted (both including and excluding third codon positions).

Secondary RNA structures were investigated using the RNAfold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) and pknotsRG [38]. Synonymous site conservation was analysed as described previously [39].

Bayesian inference of a time-measured evolutionary history

A second data set of polyprotein ORF flaviviruses was compiled by excluding the highly divergent sequences of all ISFVs and including additional sequences of YFV that were isolated from either the New World or the Old World. An amino acid alignment was produced using Clustal W2 and regions of ambiguous alignment were removed using the GBlocks algorithm with standard parameters.

A time-measured evolutionary history was inferred using MCMC analysis implemented in BEAST [40]. Based on model testing using Prottest, we used the LG substitution model [41] with gamma distributed rate variation among sites in conjunction with a relaxed uncorrelated lognormal molecular clock model [42] and a Bayesian skygrid tree-prior [43].

Time-calibration of the evolutionary history was based on the recognised slave-trade introduction of YFV to Brazil [12,44,45]. Specifically, based on estimates from www.slavevoyages.org, we constrained the common ancestor of the American YFVs to have existed before 1860, and the divergence from the West-African YFVs to have occurred after 1561. The divergence of the South American YFV from the African YFV strains represents the upper boundary for the introduction. Therefore, this node is assumed to be younger than 449 years. The common ancestor of all South American YFV strains represents the lower boundary for the slave trade introduction, and this node is therefore assumed to be older than 150 years. Given the long evolutionary time-scale and the fact that we performed our analysis at the amino acid level to allow the estimation of relatively deeper divergence times, instead of resorting to potentially saturated nucleotides, and to avoid a potential disconnect between short-term and long-term evolutionary rates [46], we did not consider sampling time differences when estimating the timed history.

We also incorporated a two-state discrete diffusion model for the Old World and New World locations in our analyses and jointly estimated the ancestral geographical states with the evolutionary history [47]. The BEAST analysis was run for 20 million generations and diagnosized using Tracer. Trees were summarized using TreeAnnotator and visualized using FigTree (Fig. 2).

Biogeographical data

Available ecological data, geographical dispersal patterns, host association and pathogenicity for all mosquito-borne flaviviruses were retrieved from the CRORA viral database (Centre collaborateur OMS de Reference et de Recherche sur les Arbovirus, Institut Pasteur de Dakar, Africa http://www.pasteur.fr/recherche/banques/CRORA/) and the International Catalogue of Arboviruses [48]. Three viruses tentatively assigned to the genus *Flavivirus*, TABV [11], Ngoye virus [14] and Mogiana tick [24] virus were not included in analyses because they are genetically too divergent to be incorporated without ambiguities using currently-available methods [10].
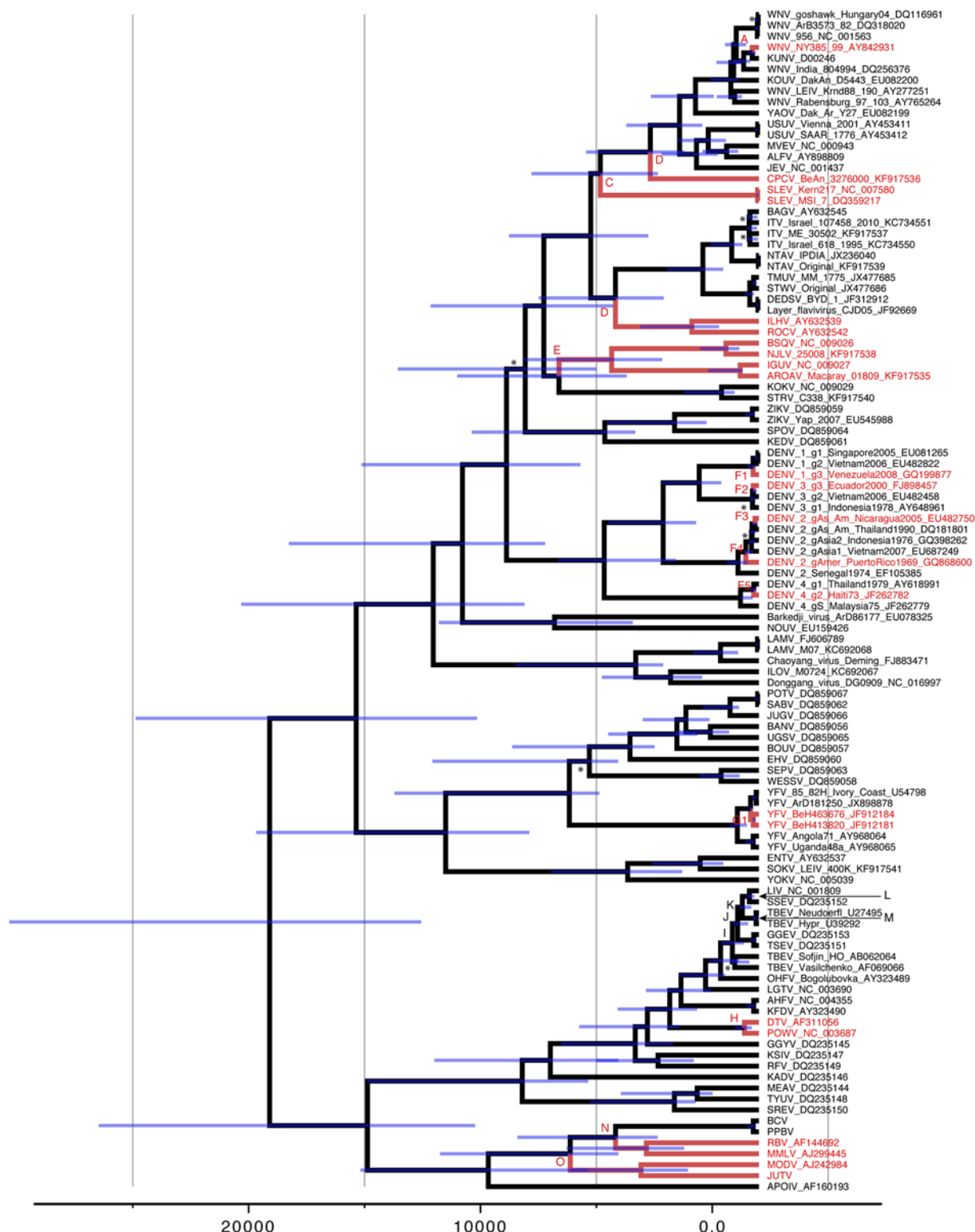
Fig 2. Maximum clade credibility tree summarized from the Bayesian molecular clock analysis. All nodes attained maximal probability support except for those indicated by a * (posterior probability < 0.95). The node age uncertainty is indicated using 95% highest posterior density (HPD) interval bars in blue. Old World and New World ancestral states are indicated by black and red branches/nodes respectively. Nodes of interest are listed A to N as referred in the text.

doi:10.1371/journal.pone.0117849.g002

RESULTS

Using our sequencing strategy, we obtained complete coding sequence for 10 *Culex*-spp. associated flaviviruses (Aroa virus [AROAV], Naranjal virus [NJLV], Cacipacore virus [CPCV], Koutango virus [KOUV], Yaounde virus [YAOV], Stratford virus [STRV], Israel turkey meningoencephalomyelitis virus [ITV], Ntaya virus [NTAV], Tembusu virus [TMUV] and STWV. Additionally, the remaining non-vectored viruses (NKV-like) in the MBFV group, viz., SOKV and 3 NKV[SL] flaviviruses BCV, JUTV and PPBV were sequenced. With the addition of these new sequences, and by taking into account other recently-published data [49–52], we have produced the most comprehensive review and phylogeny of flavivirus full polyprotein ORF sequences and phylogeographic information to date.

Phylogenetic analysis

Fig. 1 presents a phylogenetic tree based on all the sequence data yielding the highest posterior probabilities out of all analyses performed. The tree is based on amino acid sequences aligned using ClustalW2. There are high levels of support at all nodes (with a posterior probability $\geq$ to 0.98 for all nodes, with the exception of one node intra WNV lineage 2 at 0.6 and one node intra DENV-3 strains at 0,5). Clades suggested by this analysis correspond to the recognised virus-vector-host groups defined above and previously [1].

TBFVs

The phylogenetic relationships of the TBFVs agree with previous studies [3] and are not elaborated upon further in this manuscript.

NKVs

The inclusion of the complete polyprotein ORF sequence for SOKV corroborates and extends previous suggested groupings for the three NKV-like bat-associated viruses, namely Entebbe bat virus [ENTV], Yokose virus [YOKV] and SOKV which cluster within the *Aedes*-associated mosquito-borne virus clade. These three NKV-like viruses represent the ICTV "Entebbe bat virus group" [7] and form a basal lineage for the YFV group and the Edge Hill virus [EHV] group. In addition, our analysis confirms that this NKV-like group of viruses currently sits in the clade that contains *Aedes* spp. MBFVs with a robust branching pattern supported by high posterior probabilities.

Sokoluk virus was isolated in 1970 from *Vespertilio pipistrellus* bats in Kyrgyzstan [53]. According to the 9th report of the ICTV [7] SOKV is a member of the Entebbe bat virus group, and appears to be a strain of the species ENTV. *Entebbe bat virus* is an African virus isolated from the insectivorous migratory bat species *Tadarida* (Chaerephon) *limbata*. It therefore seems reasonable to propose that SOKV probably emerged in Africa and dispersed to Asia via transmission across areas of overlapping bat habitats. Importantly, ENTV and SOKV have been shown to replicate in C6/36 mosquito cells *in vitro* [54] whereas the recognised NKV[SL] that form a genetically distant clade (Rio bravo [RBV], Montana myotis leukoencephalitis virus [MMLV], etc.) do not replicate in mosquito cells [55]. This observation together with the phylogenetic position of these viruses in the MBFV clade is consistent with the idea that the transmission or maintenance cycles of YOKV, ENTV and SOKV (or ancestral representatives of these viruses), may have involved mosquitoes in the past.

The addition of the complete ORF sequences of BCV, JUTV and PPBV corroborates and extends our knowledge regarding the poorly-described NKV[SL] group. Apoi virus (a Japanese virus isolated from *Apodemus* spp. and *Clethrionomys* spp.) constitutes a separate group which

appears to have diverged earlier from the other viruses, which are distributed into two groups that include viruses isolated from rodents and bats respectively [25]. Bat-associated viruses constitute 2 sub-groups including MMLV and RBV, PPBV and BCV, respectively. Rodent-associated viruses are represented by Modoc virus and JUTV in our analysis. The other representatives of rodent and bat NKVs were not included in this analysis due to the absence of complete ORF sequences.

MBFVs

Using partial genome sequence-based phylogenies, the MBFVs were previously divided into 2 major groups reflecting their vector associations (*i.e.*, *Aedes*- and *Culex*-associated groups), the principal vertebrate host and associated diseases [1]. Our analysis, based on complete ORF sequence data, confirmed this division of MBFVs into 2 major epidemiologically distinct vector groups, *i.e.*, those primarily associated either with *Culex* spp. or *Aedes* spp mosquitoes. However, additional ISFV-"like" MBFVs became available for phylogenetic analysis. Lammi virus [LAMV] was isolated from *A. cinereus* mosquitoes in Finland [13], Ilomantsie virus [ILOV] was isolated from *Ochlerotatus* mosquitoes in Finland (Huhtamo et *al.*, submitted paper), Marisma mosquito virus [MMV] from *O. caspius* in Spain [18], Donggang virus [DGV] from *Aedes* mosquitoes in China (Unpublished data, GenBank acc. Number: NC_016997), Chaoyang virus [CHAOV] from *A. vexans nipponii* in Korea [56] and in China (Article published in Chinese, GenBank acc. Number: FJ883471), Nounané virus [NOUV] from *Uranotaenia* spp. mosquitoes in Ivory Coast [57] and Barkedji virus [BJV] from *Culex perexiguus* in Israel [58] and Senegal (Unpublished data, GenBank acc. Number: EU078325). The mosquito vector was originally unknown for CHAOV (China) and BJV (Senegal) but these viruses have also been recently described in other locations [58]. Thus, based on this more recent evidence the probable vectors for CHAOV and BJV are *Aedes* and *Culex* species respectively.

These viruses form two distinct groups (Fig. 1) that diverge from the two previously recognised groups *Aedes* spp.-associated and *Culex* spp.-associated MBFV. Moreover, the branching pattern of the tree implies that they emerged after the *Aedes* spp.-associated virus lineage but before the *Culex* spp.-associated virus lineage. Currently, they sit in the major clade of *Aedes*-associated viruses and represent an integral part of the evolutionary continuum amongst the MBFVs. These two groups of Eurasian viruses, containing LAMV, ILOV, DGV and CHAOV and the African virus group containing BJV and NOUV, appear to replicate only in mosquito cells, possibly with transitory replication in vertebrate cell lines [13,57]. This is compatible with 2 independent events during which these MBFVs lost their ability to infect vertebrate cells. Recently, Nanay virus was isolated from *Culex ocossa* in Peru [59]. This virus was only partially sequenced in the E and NS5 gene but it seems to be closely related to NOUV, leaving the question of a potentially more widespread dispersal of this new ISFV-like virus in the NW. This is not surprising, in view of other papers which reported the widespread distribution of ISFV and the huge extent of undersampling.

Aedes-associated flaviviruses (Fig. 1) include the dengue virus [DENV] group, the yellow fever virus [YFV] group, the EHV group, and the Kedougou virus [KEDV] group. Results from our analyses are in accordance with the most recently published studies [8].

The complete ORF sequences of 10 *Culex* spp.-associated flaviviruses were determined and included in the phylogenetic tree (Fig. 1). Thus, all viruses currently known to fall within this group have now been characterized.

CPCV

CPCV is a bird-associated virus, isolated in Brazil from the blood of the black-faced antbird. This virus has never been found in mosquitoes. However, this could reflect insufficient

sampling of field materials. In the most recent phylogenies (Fig. 1), CPCV always roots the JEV group with a posterior probability of 1. Recently, it was isolated in Brazil from a human presenting clinically with leptospirosis and/or yellow fever-like illness [60].

YAOV, KOUV and WNV

YAOV and KOUV represent ancestral lineages of WNV in the phylogenetic tree (Fig. 1). YAOV has been isolated in Africa in the Cameroon, Central African Republic, Congo, Senegal and Ivory Coast from birds, mammals and both *Culex* and *Aedes* mosquito species and has never been identified as a human pathogen [61]. In contrast, KOUV was isolated in Senegal from rodents (*Tatera kempi* and *Mastomys sp.*) and also from a human following a laboratory infection [48].

The sequences of two viruses isolated in Eastern Europe and Russia, namely Rabensburg virus (RABV) and Krasnodar virus (KRDV) respectively, from a pool of *Culex pipiens* mosquitoes and *Dermacentor marginatus* ticks, were also included in the phylogenetic analysis because they are related to, but show significant divergence, from WNV [62,63]. In addition, the Indian lineage WNV strain India 804994 isolated from a human, was included in the analysis [64].

The phylogeny (Fig. 1) supports previous suggestions [65] that WNV has an African origin, as the African YAOV roots the WNV group.

MVEV and ALFV

Other members within the JEV serocomplex including JEV, Usutu virus [USUV], Murray Valley encephalitis virus [MVEV] and Alfuy virus [ALFV], form a strongly supported sister group to these viruses, all sharing a common ancestor with CPCV. Notably, the 2 Australian viruses MVEV and ALFV share the same ecological niche and are considered to be a single species [7], with ALFV being a strain of MVEV. In common with other JEV serocomplex viruses, MVEV and ALFV have been isolated birds and mosquitoes. MVEV causes hundreds of human cases of encephalitis annually in Australia. In contrast, there is only one unconfirmed case of mild polyarticular disease (in 1987) doe to ALFV [48]. Previous studies in laboratory animals showed that ALFV is less neuroinvasive than MVEV following peripheral challenge [66].

ITV and BAGV

Within the recognized Ntaya virus group [7], we have determined the polyprotein ORF sequences of NTAV, ITV, TMUV and STWV. ITV and Bagaza virus [BAGV] are bird-associated viruses that cause encephalitis in poultry and wild birds. Both of these viruses appear to have their evolutionary origins in Africa, although ITV is a frequent cause of avian disease in Israel and BAGV was recently identified as the aetiological agent of bird fatalities in birds in southern Spain [67,68]. BAGV was also isolated in India and human exposure was implied by detection of BAGV neutralizing antibodies in 15% of the human population [69]. No human exposures have been reported for ITV.

During the preparation of this manuscript, the sequences of 5 strains of ITV were reported [70]. The authors suggested that ITV and BAGV should be considered a single species with 2 different clades representing "old isolates" and "recent isolates". The sequence presented in our analysis is an isolate from 1959, included in the clade of "old isolates" (Fig. 1) and is 6% divergent at the nucleotide level from all other ITV and BAGV sequences.

TMUV, STWV and DEDSV

TMUV, STWV and the recently described duck egg drop syndrome virus [DEDSV] [51] and layer flavivirus (unpublished data, Genbank acc nb.: JF926699) are also closely related southern Asian strains, mainly isolated in Thailand, China and Korea, but they are phylogenetically

distinct from ITV and BAGV [71–73]. They also cause severe disease pathologies in domestic birds (chicken, duck, etc.) but have not been associated with human disease.

NTAV

NTAV was originally isolated in Uganda. During the preparation of this manuscript, the sequence of another strain of Ntaya virus was reported [74]. The strain provisionally designated IPD/A was collected in the Cameroon region of Africa in 1966. This virus differs by only 0.08% in nucleotide sequence when compared with the original strain that was sequenced in our study (Fig. 1). Positive human serological evidence in many regions of Africa confirms that NTAV does infect humans [74].

AROAV group

With determination of the ORF sequences of AROAV and NJLV, the sequences of all viruses recognised to date within the AROAV group are now known. Both AROAV and NJLV were isolated from sentinel hamsters in South America (Venezuela and Ecuador respectively). NJLV has also been isolated from *Culex* spp., mosquitoes.

Bussuquara virus [BSQV] and Iguape virus [IGUV] are 2 other representatives of this group and have been respectively isolated in sentinel monkeys and sentinel mice, in Brazil. As shown in Fig. 1, NJLV and BSQV form a sister group, as do AROAV and IGUV. These 4 viruses have all been isolated from mammals and each has its own ecological and biological niche. With the exception of BSQV which causes symptoms in humans that include fever, headache and arthralgia, other members are not known to be human pathogens.

STRV and KOKV

With the complete ORF determination of STRV in this study, all of the recognised viruses known to date in the KOKV group have now been sequenced. Interestingly, STRV has only been isolated from *Aedes vexans* mosquitoes whereas KOKV was also isolated from *Culex annulirostris* [48], both in Australia and Papua New Guinea. STRV is not known to cause human pathology, whereas serological evidence in humans has been reported for KOKV and occasionally it is responsible for acute polyarthritic disease (3 cases) with fever, headache and lethargy [75]. Additionally, the KOKV group also appears to include 2 new members designated TS5273 and New Mapoon virus [NMV] (CY1014). New Mapoon virus is included with complete polyprotein ORF sequence in Fig. 1 [17,76].

ISFV^{SL}

ISFV^{SL}

During the past decade, many new ISFV^{SL} have been isolated and their sequence data are consistent with the concept that they should be classified as a fourth major group of flaviviruses. Moreover it was recently observed that the ISFV^{SL} could be sub-divided into 2 sub-groups: *Stegomyia* (*Aedes*) associated viruses and *Culex* associated viruses based on all inferred phylogenies. With the addition of recent discoveries such as Nakiwogo virus (NAKV) isolated from *Mansonia* [20] and Palm Creek virus (PCV) isolated from *Coquillettidia* [77] the ISFV^{SL} group is becoming increasingly complex. Indeed the phylogeny now shows a potential third and/or fourth sub-group that includes NAKV and PCV. This is not surprising considering that the likely mosquito vectors of the two viruses, *Coquillettidia* and *Mansonia* spp., are considered by morphological data to be sister groups (Harbach & Kitching, 1998). Also, these viruses are significantly undersampled and more vector distribution studies and virus discovery would help to clarify their phylogenetic status.

A timed evolutionary perspective for Old to New World introductions

From a phylogeographic point of view, the evolutionary origins and dispersal patterns of many of the flaviviruses can be tentatively deduced by considering their association between phylogenetic clustering and geographic location, knowledge of historical anthropological patterns, host/vector associations and estimated times from the present, of divergence from a common ancestor. For example, all the *Culex*-spp. associated virus clades are rooted by Old World *Aedes*-spp. associated viruses.

To provide formal support for these observations, we performed a separate Bayesian phylogenetic analysis under a relaxed molecular clock model. A summary of this analysis is represented by the maximum clade credibility tree, which has strong branching support for most of the clades; all nodes had high posterior probabilities (> 0.95) except those labelled with a star (∗). Time-calibration for the tree, was based on the principle that YFV was introduced from the Old World into the New World during the slave-trade [12,44,45]. Although this provides only a single calibration point in the evolutionary history, it allows us to position other relative divergence times (Fig. 2). In this tree, viruses included in the highly divergent ISFV groups were removed to avoid the extrapolation of divergence times too far into the past, based on a relatively recent calibration.

In Fig. 2, the black nodes and branches indicate that the ancestral viral lineage was inferred to have existed in the Old World (OW) as opposed to the red nodes and branches for the New World (NW) virus lineages. Mean divergence time estimates and 95% credibility intervals (translucent light blue bars) are shown for each node in the tree. The phylogenetic estimate identifies, with high confidence, 11 independent nodes that represent introductions of mosquito-borne virus lineages (identified as A to E, F1–5 and G1–2) from the OW to the NW (Fig. 2). Table 1 summarizes the credible intervals for each node (or 95% HPD interval) and the median estimate.

Estimated times of divergence that belong to the "current time", i.e., the past 500 hundred years were authenticated against historical facts relating to the slave trade and introduction of yellow fever virus into the Americas [78].

Similar estimates for divergence between OW ancestral DENV and NW strains (150 to 450 years ago) are also consistent with the concept of human introduction of these viruses into the NW via the Slave trade and other commercial exchanges (nodes F1 to F5—Fig. 2).

The divergence between Far eastern/Siberian TBEV and European TBEV is estimated to have occurred 1087 [1610–649] years ago (node I—Fig. 2)[79] and, the divergence between European TBEV and LIV/SSEV occurred about 572 [844–328] years ago (node K—Fig. 2). The estimates for LIV are consistent with historical records of the dispersal patterns of this virus in the British Isles [80].

The first outbreak of WNV in North America occurred in August 1999 in New York (http://www.cdc.gov/mmwr/preview/mmwrhtml/mm4838a1.htm). The virus was most likely introduced from North Africa [81]. In the OW, the eastward dispersal of WNV, probably via migratory birds and/or shipping, seems to have occurred many years earlier leading to the appearance of Kunjin virus (KUNV) in Australia. Fig. 2 (node A) shows that KUNV diverged from WNV approximately 277 [475–137] years ago. This corresponds quite closely with the dates of the early emigrants from the British Isles to Australia. The shipping routes included stopovers in Africa, during which KUNV, or an ancestral lineage, could have gained access to the ships via infected mosquitoes. The virus could then have been carried from Africa to Australia in these mosquitoes on the ships that transported the British emigrants.

In the Old World, Powassan virus (POWV) is found in Russia [82] and since the ancestral lineages of this virus are also OW viruses, it is reasonable to assume that POWV originated in

Table 1.

Virus	Interval low	Nodes of divergence with the common ancestor	Interval high	Fig. 2 nodes
WNV NY99 and KUNV	137	277	475	A
CPCV	3217	5239	7453	B
SLEV	4356	6928	9806	C
ILHV group	4107	6801	9496	D
AROAV group	5703	9351	13013	E
DENV-1 gen3 NW / OW	157	321	515	F1
DENV-3 gen3 NW / OW	89	186	311	F2
DENV-2 gen As-Am NW / OW	74	148	250	F3
DENV-2 gen Am NW / OW	245	450	664	F4
DENV-4 gen2 NW / OW	115	225	377	F5
YFV split OW / NW	281	385	450	G1
YFV NW	150	199	277	G2
DTV and POWV	308	601	1025	H
FE and Sib TBEV group and Europe TBEV group/LIV	649	1087	1610	I
GGEV/TSEV group and rest of Europe TBEV/LIV	465	783	1133	J
Europe TBEV and LIV	328	572	844	K
LIV and SSEV	195	369	569	L
TBEV Neud and TBEV Hypr	57	141	254	M
Bat's NKV split NW /OW	4370	7421	10414	N
Rodent's NKV split OW /NW	6065	9580	13750	O

doi:10.1371/journal.pone.0117849.t001

the OW [12,83]. However, POWV is also found in the NW in Canada and the United States of America, although no other related tick-borne encephalitic viruses have been found in the New World. The ancestral lineage of POWV in the NW has diverged to produce 2 lineages [79,84] namely the current POWV lineage and a variant that adapted to deer ticks, currently referred to as deer tick virus (DTV). This is depicted in Fig. 2 as node H, where the ancestral lineage is displayed in black on the assumption of an OW origin.

There has been considerable discussion concerning the significance of these observations [79] and the hypothesis that POW was introduced into North America using the Beringian land bridge that connected Asia and North America between 15,000–11,000 years ago has recently been used to estimate modelling the temporal origin and evolution of flaviviruses (see the discussion section)[85].

Based on the phylogenetic and phylogeographic evidence, we deduce that mosquito-borne flaviviruses, corresponding to those identified at nodes A to E, F1–5 and G1–2 (Fig. 2), originated in the OW and were introduced into the New World 11 times. This does not preclude the possibility that in some cases ancestral African lineages may have dispersed to the NW before emerging as the currently recognised viruses. According to the TMCRA, these 11 events can be divided into two categories, those that have an estimated TMRCA that overlaps with the period of commercial trading by ships crossing the Atlantic Ocean, approximately during the past 500 years (YFV, node G1–2; DENV 1–4, node F1–5; WNV, node A—Fig. 2) and those for which the estimated TMRCA pre-dates this period of trading (AROAV group, node E; ILHV group, node D; SLEV, node C; CPCV, node B—Fig. 2).

On the other hand, the TMRCA estimates for the other 6 NW lineages suggest that these viruses potentially could have been introduced to the NW more than one thousand years before the initiation of historic commercial and slave trading across the Atlantic Ocean (AROAV group (5700–13000; Fig. 2 node E); ILHV group (4100–9500; Fig. 2 node D); SLEV

(4400–9800; Fig. 2 node C); CPCV (3200–7500; Fig. 2 node B); NW bat NKV^{SL} (4400–10400; node N) and NW rodent NKV^{SL} (6100–13800; node O). These estimate are less accurate, due to the relatively short term calibration date we have used for YFV as reference, and could be thousands or even more years in the past. All apparent virus migrations during this period are referred to as having occurred "before the slave trade".

Node E of Fig. 2 displays a TMRCA of 9400 years, which represents the common ancestor of the NW Aroa virus (AROAV) and Kokobera virus (KOKV) groups, i.e. groups of viruses that dispersed respectively westward to the NW and eastward to Australia, in the OW.

A similar pattern was observed at node D (TMRCA 6800 years before the present). The ILHV group emerged presumably in Africa and dispersed westward to the NW. On the other hand, the BAGV group remained in Africa, eventually emerging in Europe and the Middle East, whereas the TMUV related ancestral lineages emerged and dispersed eastward into Asia (node D—Fig. 2).

On the other hand, the JEV/USUV/ALFV/MVEV group shares a common ancestor with a TMRCA of about 2000–3000 years before the present, implying that an ancestral lineage dispersed eastwards out of Africa and appears to have dispersed to south East Asia before emerging as JEV and then dispersing widely, throughout Asia [86].

Nodes N and O in Fig. 2 identify the no known vector NKV^{SL} viruses. In common with the arboviruses, there are distinct lineages in the OW and NW (coloured black and red respectively). The TMRCA of the OW NKV, Apoi virus (APOIV), which was isolated from rodents in Japan, pre-dates all the other recognised NKV^{SL}, and the MRCA of this lineage diverged to produce descendant NKV and arthropod-borne virus lineages. This leaves open the possibility that the association of flaviviruses with tick and mosquito vectors may have been an acquired trait from a non-vectored ancestral virus [25]. Nodes N and O indicate that the NKVs were introduced to the NW on at least two independent occasions. Based on the TMRCA predictions, the NKV diverged over a period of 4000 to 14,000 years ago. It has been suggested (Varelas-Wesley & Calisher, 1982) that these NKVs could have been introduced into the Americas during the Miocene/Pliocene period, possibly by migrating rodents and/or bats. This idea is consistent with the TMRCA predictions presented in Fig. 2. Nevertheless, the alternative possibility that these NW viruses emerged in the OW and were introduced more recently via rodents into the NW, following the development of trading via ships between the OW and the NW should not be ruled out. Two African viruses Dakar Bat and Bukulasa bat virus are not included in this analysis as only partial sequences for these viruses are currently available.

A search for sequence elements associated with ribosomal-1 frameshifting

Many viruses harbour sequences that induce a proportion of translating ribosomes to shift-1 nt and continue translating in the new reading frame to produce a 'transframe' fusion protein [87]. Where functionally utilized, this is referred to as programmed-1 ribosomal frameshifting (-1 PRF). The eukaryotic-1 frameshift site typically consists of a 'slippery' heptanucleotide sequence fitting the consensus motif X_XXY_YYZ, where XXX represents any three identical nucleotides; YYY represents AAA or UUU; Z represents A, C or U; and underscores separate zero-frame codons. In the tandem slippage model, the P-site anticodon re-pairs from XXY to XXX, whereas the A-site anticodon re-pairs from YYZ to YYY, thus allowing for perfect repairing except at the wobble position. Certain deviations from the canonical XXX of the slippery site are tolerated in the P-site, including UCC in some members of the JEV serogroup, GGU in cardioviruses and some luteoviruses, GUU in equine arteritis virus (family *Arteriviridae*), and GGA in many ISFVs besides insect nidoviruses of the family *Mesoniviridae* and some

umbraviruses and dianthoviruses. The efficiency of frameshifting depends on the identity of the slippery site nucleotides but is typically less than 1% in the absence of additional stimulatory elements. Thus, most known instances of eukaryotic-1 frameshifting are stimulated (typically to a level between 1% and 50%) by the presence of a 3' stable RNA secondary structure, such as a pseudoknot or stem-loop, that is separated from the slippery heptanucleotide by a 'spacer' region of 5–9 nt.

Sequence elements associated with-1 PRF were previously described in the JEV serogroup viruses for JEV, WNV, USUV, MVEV and ALFV [88,89](Fig. 3a). In these viruses, -1 PRF occurs (with an estimated efficiency of 20–50%) when ribosomes are positioned on the codons encoding the 8th and 9th amino acids of NS2A. When PRF occurs, ribosomes translate a 43-codon ORF in the-1 reading frame relative to the polyprotein ORF and then terminate. The resulting 52 amino acid 'transframe' polypeptide is not cleaved at the NS1|NS2A cleavage site and thus frameshifting results in the production of a C-terminally extended version of NS1, known as NS1'. Our study extends the range of viruses that possess this frameshift site to include CPCV, YAOV and KOUV. The overlapping ORF displayed a constant length (43 codons) for all viruses within the JEV serogroup, with the exception of SLEV which lacks the frameshift site.

A nucleotide alignment of the putative stimulatory elements responsible for the-1 PRF is presented in Fig. 3b. The slippery heptanucleotide Y_CCU_UUU (highlighted in orange on Fig. 3b) is absolutely conserved for CPCV, YAOV and KOUV as with other members of the JEV group except for SLEV. Predicted 3' RNA stimulatory elements are highlighted in Fig. 3b. KOUV retains the canonical stable 3' pseudoknot structure that stimulates-1 PRF in other JEV serogroup flaviviruses (Fig. 3a) [89]. In YAOV, however, a stem-loop structure was predicted instead of a pseudoknot (Fig. 3c). The CPCV sequence also lacks the potential to form the canonical JEV serogroup pseudoknot. In this case, however, several possible alternative structures could be predicted and, in the absence of experimental analysis or comparative genomic inference, it remains unclear which structure (if any) might be functionally relevant for CPCV.

Analysis of nucleotide conservation at synonymous sites within alignments of related flavivirus sequences has previously been successful in identifying additional coding ORFs overlapping internal regions of the polyprotein ORF and accessed via-1 PRF [88,90]. Due to the sequence constraints imposed by simultaneous coding in two overlapping reading frames, besides maintaining functional frameshift-stimulatory elements, such sites (if functionally important and phylogenetically conserved) are associated with greatly increased nucleotide conservation at synonymous sites in the polyprotein reading frame relative to the genome average. To investigate the potential presence of frameshifting in other flaviviruses, we constructed sequence alignments of selected flavivirus clades, and analyzed conservation at synonymous sites as described previously [39]. A selection of these analyses is presented in Figs. 4 and 5.

Most flaviviruses exhibit enhanced synonymous site conservation at the 5' end of the polyprotein ORF. This is presumably associated with non-coding elements such as functional RNA structures involved in replication and/or translation enhancement [91–96]. Several flavivirus clades also exhibit localized regions of statistically significantly enhanced synonymous site conservation in internal regions of the polyprotein ORF, notably the JEV serogroup, the ISFVs, the CHAOV-LAMV-DGV-ILOV clade (Fig. 4), and the WESSV-SEPV clade (Fig. 5).

In the ISFVs (CxFV-QBV clade shown in Fig. 4), enhanced synonymous site conservation is apparent in the NS2A/2B-encoding region; this corresponds to a long overlapping ORF that is accessed via-1 PRF [90]. Many more ISFV sequences are now available, and the overlapping ORF is conserved in all except for the cell culture adapted original isolate of CFAV (GenBank acc. M91671) [23]. The overlapping ORF ranges from 270 to 293 codons in CxFV, QBV, PCV, NAKV and *Cx theileri* flavivirus, and 253 to 257 codons in KRV, CFAV, HANKV, AEFV
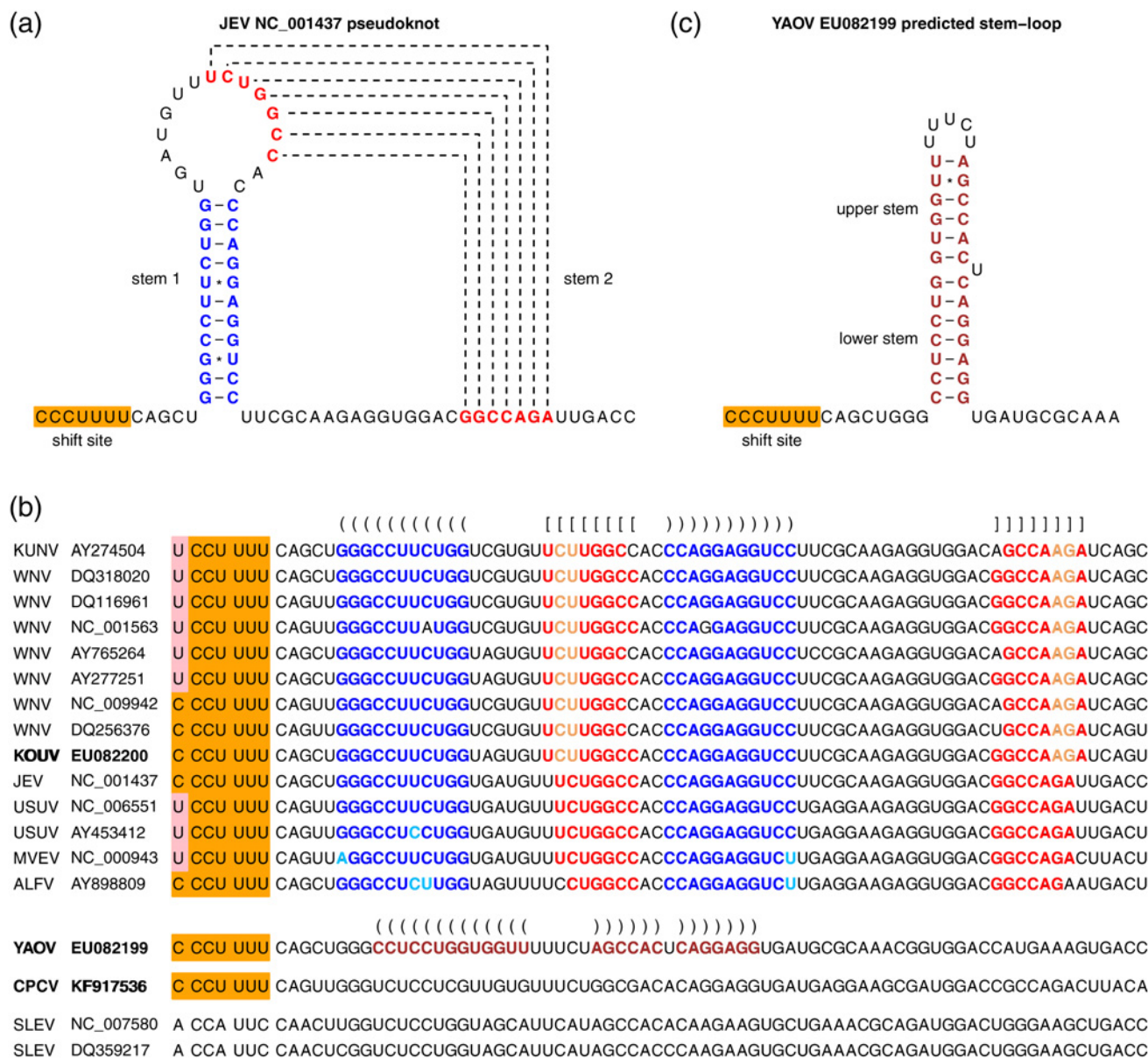
Fig 3. Frameshift stimulatory elements in viruses of the JEV serogroup. (a) Previously identified frameshift site (Y_CCU_UUU; Y = C or U; orange) and 3'-adjacent stable pseudoknot structure responsible for stimulating-1 PRF in the NS2A-encoding region of JEV and related flaviviruses. Stems 1 and 2 of the pseudoknot are indicated in blue and red respectively. (b) The shift site and pseudoknot are preserved in the newly sequenced KOUV but not in YAOV or CPCV. Substitutions that preserve the base-pairings in stem 1 (blue) or stem 2 (red) of the pseudoknot are indicated in pale blue and orange respectively. In YAOV, a simple stem-loop (brown) was predicted at an appropriate spacing from the shift site to act as a stimulator of-1 PRF. CPCV maintains the shift site but multiple possible 3'-proximal structures (not shown) could be predicted. SLEV sequences lack a suitable shift site at this genomic location. (c) Predicted frameshift stimulatory elements (shift site and 3'-adjacent stem-loop) in YAOV.

doi:10.1371/journal.pone.0117849.g003

and *Ochlerotatus caspius* flavivirus. The frameshift site generally appears to be G_GAU_UUY (Y = U or C), with notable exceptions appearing to be G_UUU_UUU in NAKV and A_AAU_UUU_UUC (potential tandem shift sites) in PCV.

In the CHAOV-LAMV-DGV-ILOV clade enhanced synonymous site conservation is apparent in the region encoding NS2B. This is associated with a conserved G_GAU_UUU
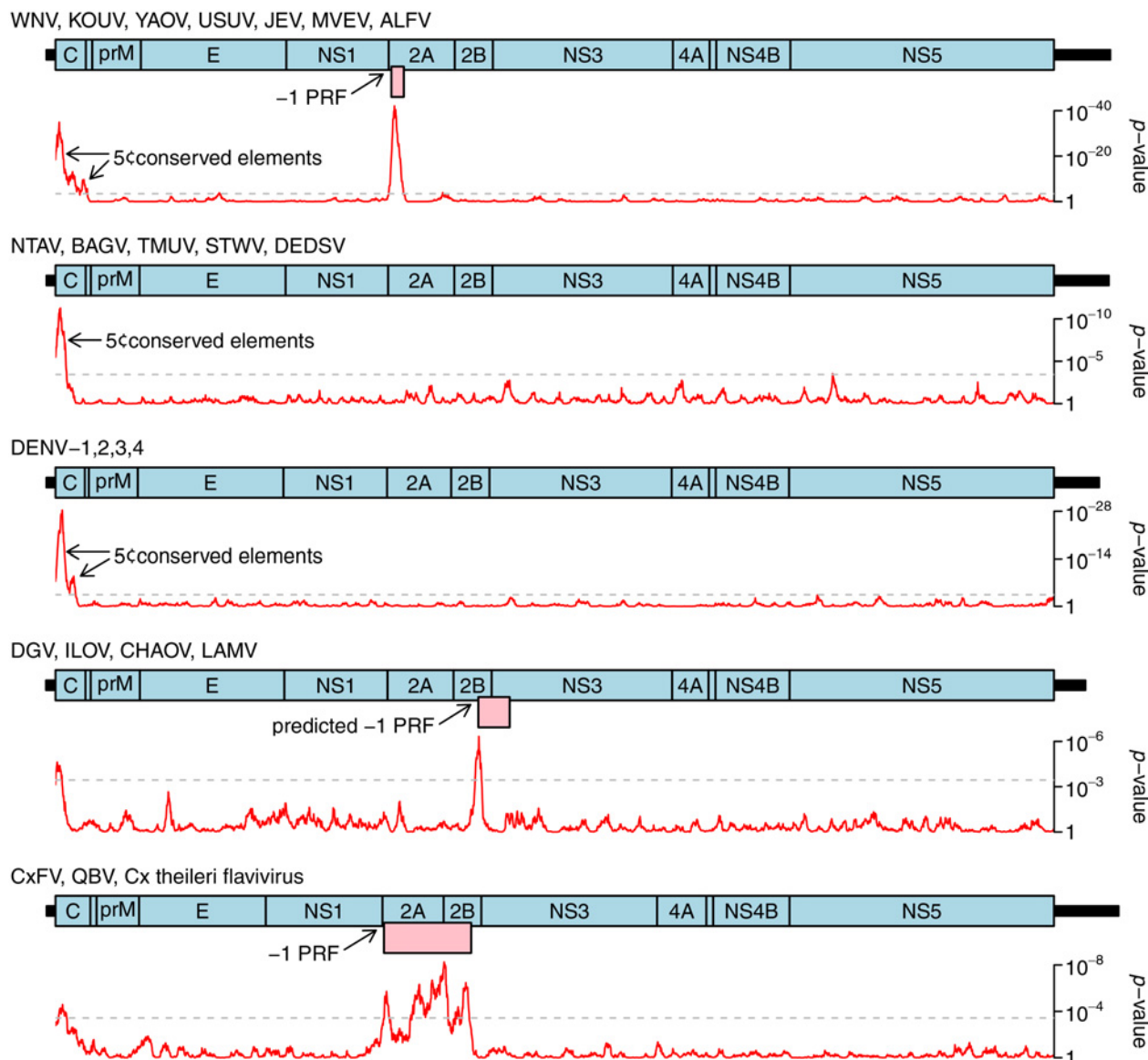
Fig 4. Synonymous site conservation analysis for selected flavivirus clades. Alignments of 249 JEV serogroup, 49 NTAV/TMUV clade, 89 DENV, 6 DGV/LAMV clade and 29 CxFV/QBV clade polyprotein ORF sequences were analyzed for synonymous site variability as decribed previously (Firth et al., 2011 PMID 21525127). The accession numbers of all sequences used in the analysis are available on request. Red lines indicate the probability (*p*-value) of obtaining not more than the observed number of synonymous substitutions, in a 25-codon sliding window, under a null model of neutral evolution at synonymous sites. Dashed grey lines indicate an approximate 5% false positive threshold after correcting for multiple tests (i.e. ~136 x 25-codon windows in the ~3400-codon polyprotein ORF). Statistically significant peaks in synonymous site conservation are indicative of overlapping functional elements, either coding or non-coding. Genome maps are shown for each clade. UTR lengths may be uncertain for less well-studied clades. Known and predicted overlapping ORFs accessed via-1 PRF are shown in pink. The predicted overlapping ORF in the DGV/LAMV clade is much shorter in DGV than in other members of the clade; the long form of the ORF is indicated. Note that *p*-values can not be directly compared between different clades because the statistical significance (i.e. p-value) of observed reductions in synonymous site variabilty depends on the diversity of the specific sequence alignment being analyzed.

doi:10.1371/journal.pone.0117849.g004

slippery heptanucleotide and a 3'-adjacent predicted stem-loop structure that stimulates-1 PRF in dual reporter assays and whose functionality is supported by compensatory substitutions (i.e. paired substitutions that preserve the predicted base-pairings) [52,90]. Here, the putative frameshift ORF has variable length (107 codons in LAMV and CHAOV, 71 codons in ILOV,
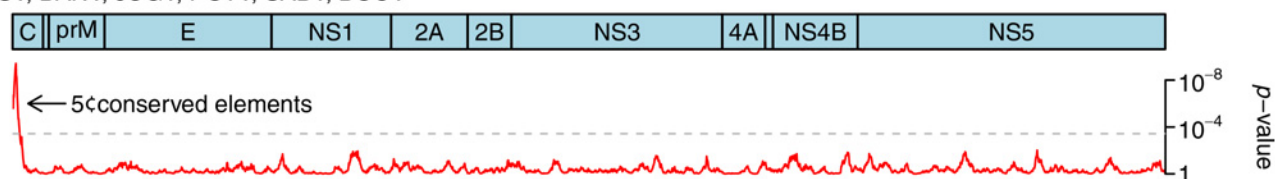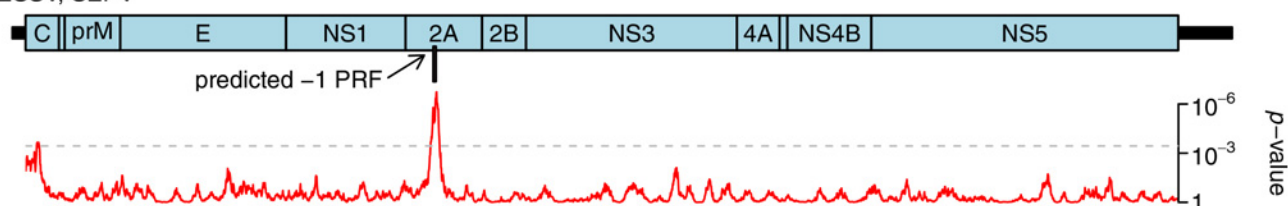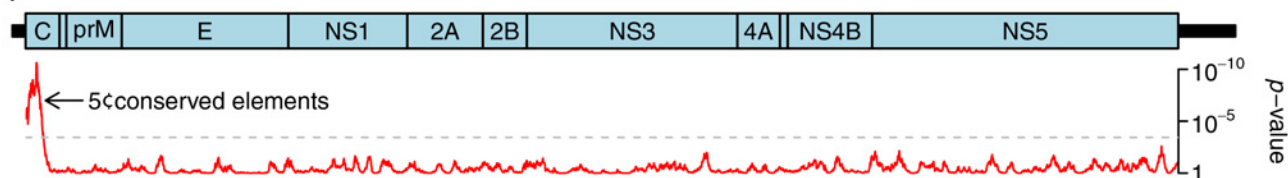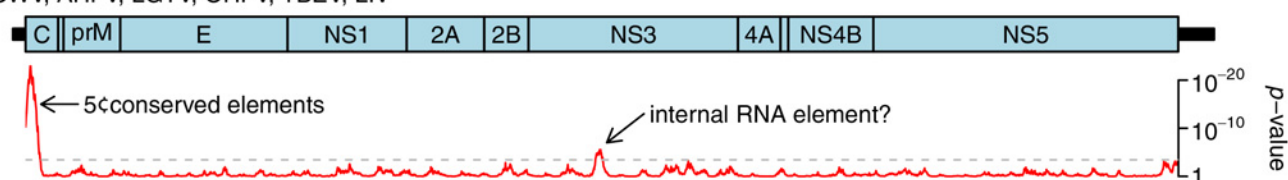
Fig 5. Synonymous site conservation analysis for selected flavivirus clades. Alignments of 6 UGSV/BOUV clade, 6 WESSV/SEPV clade, 56 YFV, and 144 POWV/TBEV clade polyprotein ORF sequences were analyzed for synonymous site variability as decribed previously (Firth et al., 2011 PMID 21525127). The accession numbers of all sequences used in the analysis are available on request. Red lines indicate the probability (*p*-value) of obtaining not more than the observed number of synonymous substitutions, in a 25-codon sliding window, under a null model of neutral evolution at synonymous sites. Dashed grey lines indicate an approximate 5% false positive threshold after correcting for multiple tests (i.e. ~136 x 25-codon windows in the ~3400-codon polyprotein ORF). Statistically significant peaks in synonymous site conservation are indicative of overlapping functional elements, either coding or non-coding. Genome maps are shown for each clade. UTR lengths may be uncertain for less well-studied clades and are omitted for the UGSV/BOUV clade due to lack of sequence data. Known and predicted overlapping ORFs accessed via-1 PRF are shown in pink. Note that *p*-values can not be directly compared between different clades because the statistical significance (i.e. *p*-value) of observed reductions in synonymous site variabilty depends on the diversity of the specific sequence alignment being analyzed.

doi:10.1371/journal.pone.0117849.g005

but only 6 codons in DGV)[52], perhaps explaining why the conservation peak does not extend throughout the 107-codon ORF annotated on the genome map in Fig. 4.

A dramatic and statistically significant peak in synonymous site conservation in an internal region was observed in one other clade of flaviviruses—the WESSV-SEPV clade (Fig. 5). This could represent an overlapping non-coding RNA element or yet another-1 PRF site. Inspection of the sequences corresponding to the conservation peak revealed a conserved slippery hepta-nucleotide, G_GUU_UUU [the same shift site that is utilized for-1 PRF in cardioviruses and some species of luteovirus, besides both-1 and-2 PRF in porcine reproductive and respiratory syndrome arterivirus [97–99]], and the potential for a 3'-adjacent stem-loop to form at the ap-propriate spacing potentially to act as a stimulator of-1 PRF (Fig. 6). In this case the overlap-ping ORF has just 7 codons, with the termination codon embedded within the predicted stem-
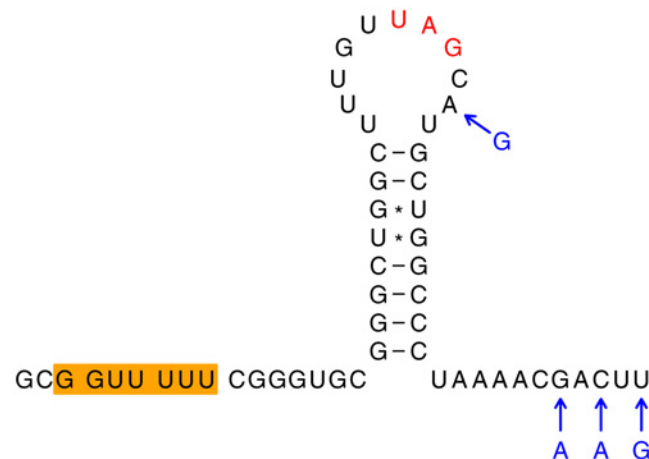
Fig 6. Predicted frameshift stimulatory elements in WESSV and SEPV. Frameshifting is predicted to occur on a conserved G_GUU_UUU heptanucleotide (orange) in the NS2A-encoding region, stimulated by a 3'-adjacent stem-loop structure. The-1 frame stop codon is indicated in red. The WESSV RefSeq NC_012735 is shown; nucleotide differences in the SEPV RefSeq NC_008719 are indicated in blue.

doi:10.1371/journal.pone.0117849.g006

loop structure. Frameshifting here would result in a greatly truncated version of the NS2A protein with a distinct 7 amino acid C-terminal end encoded by the short overlapping ORF.

A statistically significant peak in synonymous site conservation was also observed in the NS3-encoding region of TBEV and related flaviviruses (Fig. 5). However this was not thought to represent a PRF site, first, because the degree of conservation was relatively modest compared to what is typically observed at a PRF site, and, second, because we were unable to identify a canonical PRF-compatible shift site at this genomic location in these viruses. Instead, it appears to correspond to a small conserved RNA stem-loop structure.

Internal frameshift sites and/or overlapping ORFs were not predicted using these methods for other flavivirus groups such as SLEV, YFV and DENV. It should be noted, however, that the synonymous site conservation analysis would not necessarily detect cases of frameshifting where there is not an overlapping ORF subject to purifying selection and where the frameshift stimulatory elements comprise just a few codons (e.g. primary sequence rather than secondary structure stimulators) and/or are located in the nascent peptide rather than the RNA sequence. Also these analyses do not yet provide statistically useful results for all flavivirus clades (e.g. where there is insufficient sequence divergence within a clade and too much sequence divergence between phylogenetically adjacent clades to observe the effects of purifying selection at synonymous sites).

DISCUSSION

As phylogenetic methods have developed and more complete genome (ORF) sequence data have become available, increasingly robust analyses and interpretations of the flaviviruses have become possible. For example, we now know that the African virus, Kadam virus (KADV), sits at the root of the TBEV lineages and KADV diverged from a common ancestor of the seabird tick-associated viruses (Fig. 1).

Our phylogeny strongly supports the NS3 branching pattern of divergence defined previously [10,11,25] in which the TBFVs and NKVs share a common ancestor. The analysis confirms previous suggestions that the *Culex*-associated flaviviruses evolved from ancestral *Aedes*-associated viruses [8,12]. However, the recent discovery of several new viruses included in the ISFV-like group, now reveals the possibility of two potentially new virus groups, viz., the

LAMV group and the NOUV group. They are geographically very widely dispersed but the mechanisms for their dispersal are unknown. Currently exclusively OW viruses found in Africa, Europe or Asia have been characterised. However, with the addition of the newly isolated Nanay virus in Peru [59], which is closely related to NOUV, the ISFV-like group now seems to extend to the NW and this will need to be re-visited when genomic sequences of NW isolates become available.

These recently-discovered apparently non-pathogenic mosquito-associated viruses (LAMV and NOUV group) are phylogenetically closely-related to human pathogenic viruses, such as DENV, YFV, WNV, JEV. However, they do not appear to replicate in vertebrate cells [13,52,57] a property which to date has precluded the possibility of isolating such viruses from field material by intracerebral inoculation of newborn mice. Possibly, they represent independent lineages of viruses that, like the NKV, have lost a particular genetic trait not required for their long-term survival.

Within the MBFVs, three NKV-like viruses, ENTV, SOKV and YOKV, have not been associated with any known vector. Nevertheless, they are genetically closely related to the YFV and EHV groups, have only been found in the Old World, and are associated with migratory bats. YOKV and SOKV were both isolated in Asia but appear to have their ancestral roots in Africa, as do many of the related viruses in the EHV group. Thus, flaviviruses seem able to lose or acquire a variety of genetic traits as illustrated by acquisition of the requirement for a tick or mosquito vector and therefore becoming arboviruses, loss of ability to infect vertebrate cells, exemplified by LAMV and NOUV, loss of the requirement for vector transmission illustrated by the ENTV group and originally proposed by Mattingly in 1960 [100] and supported by Kuno and Chang [101], inability to infect vertebrate cells, as illustrated by the ISFVs and acquisition of a frameshift as typified by the JEV serogroup viruses (discussed in detail later).

No known vector flaviviruses specific lineage (NKV^{SL})

In contrast with the MBFV and TBFV groups, the NKV^{SL} group remains poorly described. Based on the phylogenetic and TMRCA data, viruses within this group have been introduced to the NW on at least two separate occasions and these introductions may have occurred thousands of years ago. Estimations for the time of divergence between these OW and NW NKV^{SL} ranged between 4000 to 14000 years ago (Fig. 2 nodes N and O). Despite the recognised large errors in TMRCA estimates, they all pre-date, by thousands of years, the recognised commercial trading period across the Atlantic Ocean that commenced 400–500 years ago. Therefore, viruses could have dispersed to the NW from the OW thousands of years before transoceanic trading was taking place. The possible mechanisms of dispersal remain a mystery. It is now established that OW human populations (presumably of both Asian and European origin) became established in the Americas during the period of TMRCA estimates, and possibly even before these estimated times. The question arises, during such migrations is it likely that infected animals could have been transported (e.g., rodents). The arrival of viruses in the NW could alternatively have occurred via the gradual dispersal of OW rodents and migratory bats. On the other hand, these transoceanic dispersions might have occurred after the development of trading between the OW and the NW.

Insect-specific flaviviruses (ISFVs)

Until about ten years ago, only one ISFV, cell fusing agent virus (CFAV), was recognised. Currently, at least 9 genetically distinguishable ISFVs have been isolated and their sequences determined. This group of mosquito- and potentially sandfly-borne viruses is an extremely genetically diverse group with a divergence of up to 61% at the amino acid level. This compares

with the genetic distances between the three current genera in the family *Flaviviridae*. More-over, they infect only invertebrate hosts and DNA forms of ISFV genomes are generated during infection of cell cultures [26]. Integrated DNA sequences have been identified in mosquito ge-nomes and isolated DNA forms have been detected in field samples [18,26,102,103]. Currently, there is no clear indication of the biological significance of these DNA forms amongst ISFVs. We therefore propose that the distinct differences of the ISFVs justify their inclusion as a sepa-rate genus in the family *Flaviviridae*. These viruses are also interesting because similar ISFVs have been isolated from mosquito species that inhabit different ecological niches, raising the question how might this occur? Possible mechanisms of virus transfer between mosquito spe-cies include the diffusion of viruses between larvae at sites shared by different mosquito species, or biting midges could become infected when taking blood meals from infected mosquitoes [104–108] and they could then act as vectors of the virus if they subsequently feed on different mosquito species.

A perspective on Old World to New World flavivirus introductions

It is recognised that YFV and DENV were transported frequently to the Americas from Africa on the ships that transported slaves across the Atlantic Ocean during the centuries of slave trading [44,45,78,109–111]. Based on this assumption, we estimated divergence times and an-cestral OW/NW relationships to provide a deeper understanding of the evolution and dispersal patterns of flaviviruses. Since virus evolution characterised by lineage-specific and non-con-stant substitution rates (Pettersson 2014), it would be unwise to assert that a single calibration point can produce precise estimates for both short/recent times, and long/deep times. Not sur-prisingly, our estimates, based on a calibration point from the recent historical period and in-volving mosquito-borne viruses, differ significantly from those of Pettersson and Omar—whose calibration point was derived from a more distant historical time period and involved a tick-borne virus. An important consequence of the latter calibration is the hypothetical origin of mosquito-borne flavivirus diversification during the last glaciation period, whilst the former calibration suggests that the major part of mosquito-borne flavivirus diversification which in-volves human epidemiology, occurred after the end of the most recent major Ice Age [12].

Our choice was to use an 'independently confirmed' historical hypothesis [44] and to ex-clude from analysis highly divergent viruses for which genomic and biological information re-mains scarce (insect specific viruses, Tamana bat virus, Ngoye virus, Mogiana tick virus...). Therefore, based on these assumptions our estimates are more accurate for the recent historical period than for the deepest nodes of the trees, and for mosquito-borne viruses than for tick-borne or no-known vector flaviviruses.

Our results demonstrate that such movement of arboviruses from Africa to the Americas is not unique to YFV and DENV. The ancestral history presented in Fig. 2 provides clear indica-tions that multiple introductions of viruses have occurred from the OW to the NW during rela-tively recent millenia. All of the *Culex*-associated viruses that circulate in Europe, Asia, Australia and/or the Americas, appear to have their evolutionary roots in Africa. Given the OW diversity of flaviviruses and the comparatively lower number of NW viruses, this analysis attempts to formalise a parsimonious argument for an 'Out-of-Africa' history. Unless there is a matching, but largely unsampled flavivirus diversity in the NW, this remains the most plausible interpretation of their phylogenetic distribution.

Given that we rely on a single, relatively recent calibration to estimate old divergence times, we do not consider our estimates to be a precise historical record of flavivirus evolution, but it does allow us to distinguish 2 major temporal periods during which flaviviruses may have been introduced from the OW into the NW. All historical and phylogenetic data are consistent with

the concept that there was a recent period of 400 to 500 years during which African viruses such as YFV and DENV were transported on slave and commercial ships across the Atlantic Ocean to the Americas, [12,44,45,78,111,112] which is why we base our estimates on this for the divergence time estimation. This "recent evolutionary period" includes viruses that have gradually emerged during the past few decades viruses (e.g., WNV, HIV, SARS, CoV, HepC virus, USUV, MERS CoV) and are still being dispersed via human movement and commercial transportation.

Based on published estimates for the times of divergence in the tree [9,44,113,114] it also appears that these introductions to the New World may have occurred gradually over the period of time corresponding to the movement of humans from Africa to the New World during the four or five hundred years of the Slave trade. Support for the concept that DENV can be included with YFV also comes from the known presence of the American genotype dengue viruses in South America that have African ancestral lineages. These viruses were presumably also transported to the Americas over the same range of time as the introduction of yellow fever viruses [12] and even more recently as slave trading was finally abolished [115]. It has been recorded that febrile syndromes and even haemorrhagic fever, clinically compatible with dengue fever, were diagnosed clinically in southern parts of North America over a long period of time [111,116,117]. The most recent and persuasive example of an OW introduction to the NW was the spectacular appearance of West Nile virus (WNV) in New York in August 1999 [118]. Since the first cases of WNV encephalitis in birds and humans were discovered in the area of the Bronx zoo and relatively close to a major international airport, it is considered possible that WNV was inadvertently introduced into North America either via infected birds or mosquitoes transported to the New York area by air transport [12,65,119].

Fig. 2 also identifies 4 examples of other flaviviruses (CPCV—node B; SLEV—node C; ILHV group—node C; AROAV group—node E) that have been introduced from the OW to the NW. However, the TMRCA for these introduced viruses predates the period of slave trading, by thousands of years and will be referred to as "ancient period" viruses.

If we consider the possibility of a more ancient introduction of viruses, i.e. between 4000 to 14000 years ago, we have to consider the possibility that ancestral viruses emerged in the OW and were then introduced into the NW many years before the slave trade period, possibly by birds, rodents, bats, arthropods and/or humans. The tree shows that introductions to the NW were multiple and independent. The first introduction is represented in Fig. 2—node E for the AROAV group. Except for BSQV, these viruses have only been isolated from rodents. The TMRCA for this node is 5700 to 13000 years ago. An ancestral virus present in the OW could have been introduced into the NW via rodents or migratory birds and then adapted to other species such as rodents. Other predicted introductions are presented in Fig. 2 nodes-C, B and D for ILHV/ROCV, SLEV and CPCV. These viruses are bird associated and may therefore have been introduced via migratory birds or bats. However, if birds were a major cause of virus introduction into the NW, it seems surprising that viruses such as WNV have only been successfully introduced once, as suggested by the numerous phylogenetic analyses of many North America WNV isolates [81,120,121]. Additionally, Nanay virus isolated in *Culex* mosquitoes in Peru might represent another example of potential virus introduction from the OW to the NW.

Recent discoveries of viruses in the ISF group such as Culex flavivirus [20,21,122,123] show that *Culex* spp.-associated viruses contain members from both the OW and the NW. It has been proposed that in the past, several different insect-specific flaviviruses have been introduced independently into Latin America, from the OW, rather than a single virus having been introduced with subsequent divergence to generate the different viruses found in the Americas [49,124].

The NKV[SL] Apoi virus, an OW virus isolated in Japan from *Apodemus* mice (*Muridae* family, *Murinae* sub-family) roots all other NKV[SL]. MODV and JUTV were the first recognised rodent NKV[SL] to be introduced to the NW ([Fig. 2](#)—node O) about 9500 years ago. MODV was isolated in North America from *Peromyscus* mice (*Muridae* family, *Sigmodontinae* sub-family) whereas JUTV was isolated in South America. These viruses were possibly introduced into the NW via rodents, when the OW and NW land masses in the northern hemisphere were joined by ice.

Rio Bravo virus (RBV) and Montana myotis leukoencephalitis virus (MMLV, represent a second introduction of NKVs into the NW ([Fig. 2](#)—node N). Their ancestral lineage appears to have diverged from an ancestral lineage of the OW bat NKVs.

Studies on Arenaviruses and Hantaviruses originally led to the development of the concept that viruses have co-evolved with their rodent hosts over time-scales of millions of years during which the viruses were assumed to have been transported during the gradual introduction of rodents from the OW into the NW [125–127]. However more recent studies do not appear to support this concept since for hantaviruses, estimates of virus divergence are in the order of thousands of years, as they exhibit short-term substitution rates of 10^{-2} to 10^{-4} substitutions/site/year [128,129]. It is hence more likely that Hantaviruses adapted relatively recently to their rodent hosts.

In the case of rodent-associated flaviviruses, and particularly NKV[SL], only limited genetic data are available and due to presumed undersampling of viruses, to date there is no clear indication of the exact period of introduction of viruses to the rodent species in the NW. However, by analogy with the hantaviruses, and taking into account the TMRCA presented in [Fig. 2](#), it seems unlikely that flaviviruses have coevolved with their rodent hosts over millions of years.

Although the uncertainty in the TMRCA is quite significant, there are clearly two periods of emergence, *i.e.* an "ancient period" in the order of magnitude of thousands of years ago, and a more "recent" period in the order of decades to a few hundred years ago. We note that it is important to be aware of the fact that estimates for the "ancient period" including estimates for NK[SL], are based on a single calibration event of an *Aedes* vectored virus (YFV) during the "recent period". This is almost certainly not an accurate estimation of the evolutionary dynamics that occurred during the "ancient period", which in the case of the NKV[SL] are not vectored by arthropods.

Frameshifting in the MBFVs and ISFVs

It was previously reported that members of the JEV serocomplex express a transframe fusion protein, NS1', via-1 PRF [89]. The frameshift-stimulatory elements—a 'slippery' heptanucleotide sequence and the potential to form a stable 3'-adjacent pseudoknot structure—are conserved in all recognised members of the JEV serocomplex with the exception of SLEV [88]. However, the sequences of CPCV, YAOV and KOUV had not been determined when this conclusion was reached. We now show that the frameshift site is also conserved in these newly sequenced viruses. Thus, the ability to produce NS1' via PRF appears to have been acquired as a genetic trait after the branching point which separates the New World virus SLEV ([Fig. 2](#)—node C) from another New World virus (CPCV) and the remaining Old World JEV serocomplex viruses ([Fig. 2](#)—node B).

The CPCV sequence also lacks the potential to form the canonical JEV-serogroup 3' RNA stable pseudoknot that is associated with highly efficient-1 PRF (20–50%; [89]); thus it is possible that frameshifting in CPCV is relatively inefficient.

The significance of the NS1' protein is not yet fully understood but it has been reported that it plays a role in viral neuroinvasiveness and reduced neurovirulence [89,130,131].

Frameshifting at a very similar genomic location occurs in the ISFVs but results in the translation of a much longer overlapping ORF [90]. Potential frameshift sites have also been bioinformatically predicted for NOUV, CHAOV, LAMV and KEDV (Firth *et al.*, 2010); ILOV and DGV [52]; and SEPV and WESSV (see above). These potential frameshift sites are also located within the genomic region encoding NS2A/NS2B. Thus, the acquisition of PRF at internal regions of the polyprotein ORF appears to be a common (though not ubiquitous) theme of flavivirus evolution. Aside from the ability to produce new functional proteins, where PRF is efficient, it may also play a role in downregulating production of the 3'-encoded replicative proteins and more quickly recycling the host cell translational machinery for increased production of the 5'-encoded structural proteins.

In conclusion, with the addition of 14 new flavivirus ORF sequences, the estimation of phylogenies via Bayesian methods, plus biogeographic and bioinformatic considerations, we have identified 11 likely introductions of mosquito-borne flaviviruses from the OW to the NW over two separate temporal periods. We have also demonstrated that similar introductions have occurred eastwards from Africa to Australia, again over two distinct time periods. Clearly there have been far more introductory movements of flaviviruses from one part of the world to another than we have described here. However, more data for a larger number of samples will be required before we can draw specific and more detailed conclusions. Finally, in the context of flavivirus evolution and dispersal, we have extended our current understanding of frameshifting amongst the flaviviruses.

Author Contributions

Conceived and designed the experiments: GM AN SC PL AEF. Performed the experiments: GM NLF MK SC PL AEF. Analyzed the data: GM AN SC PL AEF EAG XDL. Contributed reagents/materials/analysis tools: GM RNC XDL. Wrote the paper: GM SC NLF MK AEF PL EAG XDL.

REFERENCES

1. Gaunt MW, Sall AA, de Lamballerie X, Falconar AK, Dzhivanian TI, et al. (2001) Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. J Gen Virol 82: 1867–1876. PMID: 11457992

2. Gould EA, Solomon T (2008) Pathogenic flaviviruses. Lancet 371: 500–509. doi: 10.1016/S0140-6736(08)60238-X PMID: 18262042

3. Grard G, Moureau G, Charrel RN, Lemasson JJ, Gonzalez JP, et al. (2007) Genetic characterization of tick-borne flaviviruses: new insights into evolution, pathogenetic determinants and taxonomy. Virology 361: 80–92. PMID: 17169393

4. Zanotto PM, Gao GF, Gritsun T, Marin MS, Jiang WR, et al. (1995) An arbovirus cline across the northern hemisphere. Virology 210: 152–159. PMID: 7793067

5. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, et al. (2013) The global distribution and burden of dengue. Nature. doi: 10.1038/nature12879 PMID: 24390349

6. Calisher CH, Gould EA (2003) Taxonomy of the virus family Flaviviridae. Adv Virus Res 59: 1–19. PMID: 14696325

7. Pletnev A, Gould E, Heinz FX, Meyers G, Thiel H-J, et al. (2011) Flaviviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, editors. Virus Taxonomy—Ninth Report of the International Committee on Taxonomy of Viruses. Oxford: Elsevier. pp. 1003–1020. doi: 10.1016/j.bjid.2014.12.003 PMID: 25636714

8. Grard G, Moureau G, Charrel RN, Holmes EC, Gould EA, et al. (2010) Genomics and evolution of Aedes-borne flaviviruses. J Gen Virol 91: 87–94. doi: 10.1099/vir.0.014506-0 PMID: 19741066

9. Zanotto PM, Gould EA, Gao GF, Harvey PH, Holmes EC (1996) Population dynamics of flaviviruses revealed by molecular phylogenies. Proceedings of the National Academy of Sciences 93: 548–553. PMID: 8570593

10. Cook S, Holmes EC (2005) A multigene analysis of the phylogenetic relationships among the flaviviruses (Family: Flaviviridae) and the evolution of vector transmission. Arch Virol. PMID: 16380812

11. de Lamballerie X, Crochu S, Billoir F, Neyts J, de Micco P, et al. (2002) Genome sequence analysis of Tamana bat virus and its relationship with the genus Flavivirus. J Gen Virol 83: 2443–2454. PMID: 12237426

12. Gould EA, de Lamballerie X, Zanotto PM, Holmes EC (2003) Origins, evolution, and vector/host coadaptations within the genus Flavivirus. Adv Virus Res 59: 277–314. PMID: 14696332

13. Huhtamo E, Putkuri N, Kurkela S, Manni T, Vaheri A, et al. (2009) Characterization of a novel flavivirus from mosquitoes in northern europe that is related to mosquito-borne flaviviruses of the tropics. J Virol 83: 9532–9540. doi: 10.1128/JVI.00529-09 PMID: 19570865

14. Grard G, Lemasson JJ, Sylla M, Dubot A, Cook S, et al. (2006) Ngoye virus: a novel evolutionary lineage within the genus Flavivirus. J Gen Virol 87: 3273–3277. PMID: 17030860

15. Charrel RN, Zaki AM, Fakeeh M, Yousef AI, de Chesse R, et al. (2005) Low diversity of Alkhurma hemorrhagic fever virus, Saudi Arabia, 1994–1999. Emerg Infect Dis 11: 683–688. PMID: 15890119

16. Bakonyi T, Gould EA, Kolodziejek J, Weissenbock H, Nowotny N (2004) Complete genome analysis and molecular characterization of Usutu virus that emerged in Austria in 2001: comparison with the South African strain SAAR-1776 and other flaviviruses. Virology 328: 301–310. PMID: 15464850

17. Nisbet DJ, Lee KJ, van den Hurk AF, Johansen CA, Kuno G, et al. (2005) Identification of new flaviviruses in the Kokobera virus complex. J Gen Virol 86: 121–124. PMID: 15604438

18. Vazquez A, Sanchez-Seco MP, Palacios G, Molero F, Reyes N, et al. (2011) Novel Flaviviruses Detected in Different Species of Mosquitoes in Spain. Vector Borne Zoonotic Dis. doi: 10.1089/vbz.2011.0736 PMID: 22141733

19. Cook S, Bennett SN, Holmes EC, De Chesse R, Moureau G, et al. (2006) Isolation of a new strain of the flavivirus cell fusing agent virus in a natural mosquito population from Puerto Rico. J Gen Virol 87: 735–748. PMID: 16528021

20. Cook S, Moureau G, Harbach RE, Mukwaya L, Goodger K, et al. (2009) Isolation of a novel species of flavivirus and a new strain of Culex flavivirus (Flaviviridae) from a natural mosquito population in Uganda. J Gen Virol 90: 2669–2678. doi: 10.1099/vir.0.014183-0 PMID: 19656970

21. Kim DY, Guzman H, Bueno R Jr, Dennett JA, Auguste AJ, et al. (2009) Characterization of Culex Flavivirus (Flaviviridae) strains isolated from mosquitoes in the United States and Trinidad. Virology 386: 154–159. doi: 10.1016/j.virol.2008.12.034 PMID: 19193389

22. Sang RC, Gichogo A, Gachoya J, Dunster MD, Ofula V, et al. (2003) Isolation of a new flavivirus related to cell fusing agent virus (CFAV) from field-collected flood-water Aedes mosquitoes sampled from a dambo in central Kenya. Arch Virol 148: 1085–1093. PMID: 12756616

23. Stollar V, Thomas VL (1975) An agent in the Aedes aegypti cell line (Peleg) which causes fusion of Aedes albopictus cells. Virology 64: 367–377. PMID: 806166

24. Maruyama SR, Castro-Jorge LA, Ribeiro JM, Gardinassi LG, Garcia GR, et al. (2014) Characterisation of divergent flavivirus NS3 and NS5 protein sequences detected in Rhipicephalus microplus ticks from Brazil. Memorias do Instituto Oswaldo Cruz 109: 38–50. doi: 10.1590/0074-0276130166 PMID: 24626302

25. Billoir F, de Chesse R, Tolou H, de Micco P, Gould EA, et al. (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. PMID: 10675416

26. Crochu S, Cook S, Attoui H, Charrel RN, De Chesse R, et al. (2004) Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. J Gen Virol 85: 1971–1980. PMID: 15218182

27. Gaunt MW, Gould EA (2005) Rapid subgroup identification of the flaviviruses using degenerate primer E-gene RT-PCR and site specific restriction enzyme analysis. J Virol Methods 128: 113–127. PMID: 15927275

28. Moureau G, Temmam S, Gonzalez JP, Charrel RN, Grard G, et al. (2007) A real-time RT-PCR method for the universal detection and identification of flaviviruses. Vector Borne Zoonotic Dis 7: 467–477. PMID: 18020965

29. Emonet S, Grard G, Brisbarre N, Moureau G, Temmam S, et al. (2006) LoPPS: a long PCR product sequencing method for rapid characterisation of long amplicons. Biochem Biophys Res Commun 344: 1080–1085. PMID: 16643852

30. Emonet SF, Grard G, Brisbarre NM, Moureau GN, Temmam S, et al. (2007) Long PCR Product Sequencing (LoPPS): a shotgun-based approach to sequence long PCR products. Nat Protoc 2: 340–346. PMID: 17406595

31. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, et al. (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475: 348–352. doi: 10.1038/nature10242 PMID: 21776081

32. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948. PMID: 17846036

33. 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 research 22: 4673–4680. PMID: 7984417

34. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32: 1792–1797. PMID: 15034147

35. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56: 564–577. PMID: 17654362

36. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755. PMID: 11524383

37. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214. PMID: 17996036

38. Reeder J, Steffen P, Giegerich R (2007) pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. Nucleic acids research 35: W320–324. PMID: 17478505

39. Firth AE, Wills NM, Gesteland RF, Atkins JF (2011) Stimulation of stop codon readthrough: frequent presence of an extended 3' RNA structural element. Nucleic acids research 39: 6679–6691. doi: 10.1093/nar/gkr224 PMID: 21525127

40. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29: 1969–1973. doi: 10.1093/molbev/mss075 PMID: 22367748

41. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21: 2104–2105. PMID: 15647292

42. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLoS biology 4: e88. PMID: 16683862

43. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, et al. (2013) Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol Biol Evol 30: 713–724. doi: 10.1093/molbev/mss265 PMID: 23180580

44. Bryant JE, Holmes EC, Barrett AD (2007) Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. PLoS Pathog 3: e75. PMID: 17511518

45. Tabachnick WJ (1991) Evolutionary genetics and arthropod-borne diseases: The yellow fever mosquito, Aedes aegypti. Am J Entomol 37:14–24.

46. Worobey M, Telfer P, Souquiere S, Hunter M, Coleman CA, et al. (2010) Island biogeography reveals the deep history of SIV. Science 329: 1487. doi: 10.1126/science.1193550 PMID: 20847261

47. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. PLoS computational biology 5: e1000520. doi: 10.1371/journal.pcbi.1000520 PMID: 19779555

48. Karabatsos N (1985) International Catalogue of Arboviruses; Karabatsos N, editor. San Antonio, Texas: The American Society of Tropical Medecine and Hygiene.

49. Cook S, Moreau G, Kitchen A, Gould EA, de Lamballerie X, et al. (2012) Molecular evolution of the insect-specific flaviviruses. J Gen Virol 93: 223–234. doi: 10.1099/vir.0.036525-0 PMID: 22012464

50. Huhtamo E, Moreau G, Cook S, Julkunen O, Putkuri N, et al. (2012) Novel insect-specific flavivirus isolated from northern Europe. Virology 433: 471–478. doi: 10.1016/j.virol.2012.08.038 PMID: 22999256

51. Liu P, Lu H, Li S, Moreau G, Deng YQ, et al. (2012) Genomic and antigenic characterization of the newly emerging Chinese duck egg-drop syndrome flavivirus: genomic comparison with Tembusu and Sitiawan viruses. J Gen Virol 93: 2158–2170. doi: 10.1099/vir.0.043554-0 PMID: 22764316

52. Huhtamo E, Cook S, Moreau G, Uzcátegui NY, Sironen T, et al. (2014) Novel flaviviruses from mosquitoes: Mosquito-specific evolutionary lineages within the phylogenetic group of mosquito-borne flaviviruses. Virology: In press.

53. Lvov DK, Tsyrkin YM, Karas FR, Timopheev EM, Gromashevski VL, et al. (1973) "Sokuluk" virus, a new group B arbovirus isolated from Vespertilio pipistrellus Schreber, 1775, bat in the Kirghiz S.S.R. Arch Gesamte Virusforsch 41: 170–174. PMID: 4727779

54. Varelas-Wesley I, Calisher CH (1982) Antigenic relationships of flaviviruses with undetermined arthropod-borne status. Am J Trop Med Hyg 31: 1273–1284. PMID: 6293325

55. Kuno G (2007) Host range specificity of flaviviruses: correlation with in vitro replication. J Med Entomol 44: 93–101. PMID: 17294926

56. Lee JS, Grubaugh ND, Kondig JP, Turell MJ, Kim HC, et al. (2013) Isolation and genomic characterization of Chaoyang virus strain ROK144 from Aedes vexans nipponii from the Republic of Korea. Virology 435: 220–224. doi: 10.1016/j.virol.2012.10.020 PMID: 23127596

57. Junglen S, Kopp A, Kurth A, Pauli G, Ellerbrok H, et al. (2009) A new flavivirus and a new vector: characterization of a novel flavivirus isolated from uranotaenia mosquitoes from a tropical rain forest. J Virol 83: 4462–4468. doi: 10.1128/JVI.00014-09 PMID: 19224998

58. Kolodziejek J, Pachler K, Bin H, Mendelson E, Shulman L, et al. (2013) Barkedji virus, a novel mosquito-borne flavivirus identified in Culex perexiguus mosquitoes, Israel, 2011. J Gen Virol doi: 10.1099/vir.0.060822-0 PMID: 24362963

59. Evangelista J, Cruz C, Guevara C, Astete H, Carey C, et al. (2013) Characterization of a novel flavivirus isolated from Culex (Melanoconion) ocossa mosquitoes from Iquitos, Peru. J Gen Virol 94: 1266–1272. doi: 10.1099/vir.0.050575-0 PMID: 23515021

60. Batista WC, Tavares Gda S, Vieira DS, Honda ER, Pereira SS, et al. (2011) Notification of the first isolation of Cacipacore virus in a human in the State of Rondonia, Brazil. Rev Soc Bras Med Trop 44: 528–530. PMID: 21860907

61. Williams RAJ, Vázquez A, Asante I, Bonney K, Odoom S, et al. (2012) Yaoundé-like virus in resident wild bird, Ghana. African Journal of Microbiology Research 6: 1966–1969.

62. Hubalek Z, Halouzka J, Juricova Z (1999) West Nile fever in Czechland. Emerg Infect Dis 5: 594–595. PMID: 10458974

63. Lvov DK, Butenko AM, Gromashevsky VL, Kovtunov AI, Prilipov AG, et al. (2004) West Nile virus and other zoonotic viruses in Russia: examples of emerging-reemerging situations. Arch Virol Suppl: 85–96. PMID: 15119764

64. Bondre VP, Jadi RS, Mishra AC, Yergolkar PN, Arankalle VA (2007) West Nile virus isolates from India: evidence for a distinct genetic lineage. J Gen Virol 88: 875–884. PMID: 17325360

65. May FJ, Davis CT, Tesh RB, Barrett AD (2011) Phylogeography of West Nile virus: from the cradle of evolution in Africa to Eurasia, Australia, and the Americas. J Virol 85: 2964–2974. doi: 10.1128/JVI.01963-10 PMID: 21159871

66. May FJ, Lobigs M, Lee E, Gendle DJ, Mackenzie JS, et al. (2006) Biological, antigenic and phylogenetic characterization of the flavivirus Alfuy. J Gen Virol 87: 329–337. PMID: 16432019

67. Aguero M, Fernandez-Pinero J, Buitrago D, Sanchez A, Elizalde M, et al. (2011) Bagaza virus in partridges and pheasants, Spain, 2010. Emerg Infect Dis 17: 1498–1501. doi: 10.3201/eid1708.110077 PMID: 21801633

68. Gamino V, Gutierrez-Guzman AV, Fernandez-de-Mera IG, Ortiz JA, Duran-Martin M, et al. (2012) Natural Bagaza virus infection in game birds in southern Spain. Veterinary research 43: 65. doi: 10.1186/1297-9716-43-65 PMID: 22966904

69. Bondre VP, Sapkal GN, Yergolkar PN, Fulmali PV, Sankararaman V, et al. (2009) Genetic characterization of Bagaza virus (BAGV) isolated in India and evidence of anti-BAGV antibodies in sera collected from encephalitis patients. J Gen Virol 90: 2644–2649. doi: 10.1099/vir.0.012336-0 PMID: 19570951

70. Fernandez-Pinero J, Davidson II, Elizalde M, Perk S, Khinich Y, et al. (2014) Bagaza virus and Israel turkey meningoencephalomyelitis virus are a single virus species. J Gen Virol. doi: 10.1099/vir.0.000037 PMID: 25537376

71. Pandey BD, Karabatsos N, Cropp B, Tagaki M, Tsuda Y, et al. (1999) Identification of a flavivirus isolated from mosquitos in Chiang Mai Thailand. Southeast Asian J Trop Med Public Health 30: 161–165. PMID: 10695806

72. Kono Y, Tsukamoto K, Abd Hamid M, Darus A, Lian TC, et al. (2000) Encephalitis and retarded growth of chicks caused by Sitiawan virus, a new isolate belonging to the genus Flavivirus. Am J Trop Med Hyg 63: 94–101. PMID: 11358004

73. Su J, Li S, Hu X, Yu X, Wang Y, et al. (2011) Duck egg-drop syndrome caused by BYD virus, a new Tembusu-related flavivirus. PLoS One 6: e18106. doi: 10.1371/journal.pone.0018106 PMID: 21455312

74. Dilcher M, Sall AA, Hufert FT, Weidmann M (2013) Full-length genome sequence of Ntaya virus. Virus Genes 46: 162–164. doi: 10.1007/s11262-012-0825-7 PMID: 22996737

75. Boughton CR, Hawkes RA, Naim HM (1986) Illness caused by a Kokobera-like virus in south-eastern Australia. The Medical journal of Australia 145: 90–92. PMID: 3016489

76. May FJ, Clark DC, Pham K, Diviney SM, Williams DT, et al. (2013) Genetic divergence among members of the Kokobera group of flaviviruses supports their separation into distinct species. J Gen Virol 94: 1462–1467. doi: 10.1099/vir.0.049940-0 PMID: 23426358

77. Hobson-Peters J, Yam AW, Lu JW, Setoh YX, May FJ, et al. (2013) A new insect-specific flavivirus from northern Australia suppresses replication of West Nile virus and Murray Valley encephalitis virus in co-infected mosquito cells. PLoS One 8: e56534. doi: 10.1371/journal.pone.0056534 PMID: 23460804

78. Bloom KJ (1993) The Mississippi Valley's Great Yellow Fever Epidemic of 1878. In: Press LSU, editor. Medical History LA: BatonRouge. pp. 1–290. doi: 10.1111/gwat.12306 PMID: 25510437

79. Heinze DM, Gould EA, Forrester NL (2012) Revisiting the clinal concept of evolution and dispersal for the tick-borne flaviviruses by using phylogenetic and biogeographic analyses. J Virol 86: 8663–8671. doi: 10.1128/JVI.01013-12 PMID: 22674986

80. McGuire K, Holmes EC, Gao GF, Reid HW, Gould EA (1998) Tracing the origins of louping ill virus by molecular phylogenetic analysis. J Gen Virol 79 (Pt 5): 981–988.

81. Mann BR, McMullen AR, Swetnam DM, Barrett AD (2013) Molecular epidemiology and evolution of West Nile virus in North America. International journal of environmental research and public health 10: 5111–5129. doi: 10.3390/ijerph10105111 PMID: 24135819

82. Leonova GN, Kondratov IG, Ternovoi VA, Romanova EV, Protopopova EV, et al. (2009) Characterization of Powassan viruses from Far Eastern Russia. Arch Virol 154: 811–820. doi: 10.1007/s00705-009-0376-y PMID: 19363586

83. Gritsun TS, Nuttall PA, Gould EA (2003) Tick-borne flaviviruses. Adv Virus Res 61: 317–371. PMID: 14714436

84. Telford SR 3rd, Armstrong PM, Katavolos P, Foppa I, Garcia AS, et al. (1997) A new tick-borne encephalitis-like virus infecting New England deer ticks, Ixodes dammini. Emerg Infect Dis 3: 165–170. PMID: 9204297

85. Pettersson JH, Fiz-Palacios O (2014) Dating the origin of the genus Flavivirus in the light of Beringian biogeography. J Gen Virol. doi: 10.1099/vir.0.000037 PMID: 25537376

86. Gao X, Liu H, Wang H, Fu S, Guo Z, et al. (2013) Southernmost Asia is the source of Japanese encephalitis virus (genotype 1) diversity from which the viruses disperse and evolve throughout Asia. PLoS Negl Trop Dis 7: e2459. doi: 10.1371/journal.pntd.0002459 PMID: 24069502

87. Firth AE, Brierley I (2012) Non-canonical translation in RNA viruses. J Gen Virol 93: 1385–1409. doi: 10.1099/vir.0.042499-0 PMID: 22535777

88. Firth AE, Atkins JF (2009) A conserved predicted pseudoknot in the NS2A-encoding sequence of West Nile and Japanese encephalitis flaviviruses suggests NS1' may derive from ribosomal frameshifting. Virol J 6: 14. doi: 10.1186/1743-422X-6-14 PMID: 19196463

89. Melian EB, Hinzman E, Nagasaki T, Firth AE, Wills NM, et al. (2010) NS1' of flaviviruses in the Japanese encephalitis virus serogroup is a product of ribosomal frameshifting and plays a role in viral neuroinvasiveness. J Virol 84: 1641–1647. doi: 10.1128/JVI.01979-09 PMID: 19906906

90. Firth AE, Blitvich BJ, Wills NM, Miller CL, Atkins JF (2010) Evidence for ribosomal frameshifting and a novel overlapping gene in the genomes of insect-specific flaviviruses. Virology 399: 153–166. doi: 10.1016/j.virol.2009.12.033 PMID: 20097399

91. Clyde K, Harris E (2006) RNA secondary structure in the coding region of dengue virus type 2 directs translation start codon selection and is required for viral replication. J Virol 80: 2170–2182. PMID: 16474125

92. Clyde K, Barrera J, Harris E (2008) The capsid-coding region hairpin element (cHP) is a critical determinant of dengue virus and West Nile virus RNA synthesis. Virology 379: 314–323. doi: 10.1016/j.virol.2008.06.034 PMID: 18676000

93. Groat-Carmona AM, Orozco S, Friebe P, Payne A, Kramer L, et al. (2012) A novel coding-region RNA element modulates infectious dengue virus particle production in both mammalian and mosquito cells and regulates viral replication in Aedes aegypti mosquitoes. Virology 432: 511–526. doi: 10.1016/j.virol.2012.06.028 PMID: 22840606

94. Liu Y, Wimmer E, Paul AV (2009) Cis-acting RNA elements in human and animal plus-strand RNA viruses. Biochimica et biophysica acta 1789: 495–517. doi: 10.1016/j.bbagrm.2009.09.007 PMID: 19781674

95. Tuplin A, Evans DJ, Buckley A, Jones IM, Gould EA, et al. (2011) Replication enhancer elements within the open reading frame of tick-borne encephalitis virus and their evolution within the Flavivirus genus. Nucleic acids research 39: 7034–7048. doi: 10.1093/nar/gkr237 PMID: 21622960

96. Khromykh AA, Meka H, Guyatt KJ, Westaway EG (2001) Essential role of cyclization sequences in flavivirus RNA replication. J Virol 75: 6719–6728. PMID: 11413342

97. Domier LL, McCoppin NK, Larsen RC, D'Arcy CJ (2002) Nucleotide sequence shows that Bean leafroll virus has a Luteovirus-like genome organization. Journal of General Virology 83: 1791–1798. PMID: 12075101

98. Loughran G, Firth AE, Atkins JF (2011) Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus genome. Proc Natl Acad Sci U S A 108: E1111–1119. doi: 10.1073/pnas.1102932108 PMID: 22025686

99. Fang Y, Treffers EE, Li Y, Tas A, Sun Z, et al. (2012) Efficient-2 frameshifting by mammalian ribosomes to synthesize an additional arterivirus protein. Proc Natl Acad Sci U S A 109: E2920–2928. doi: 10.1073/pnas.1211145109 PMID: 23043113

100. Mattingly PF (1960) Ecological aspects of the evolution of mosquito-borne virus diseases. Transactions of the Royal Society of Tropical Medicine & Hygiene 54: 97–112. doi: 10.1093/trstmh/tru112 PMID: 25059525

101. Kuno G, Chang GJ (2005) Biological transmission of arboviruses: reexamination of and new insights into components, mechanisms, and unique traits as well as their evolutionary trends. Clin Microbiol Rev 18: 608–637. PMID: 16223950

102. Roiz D, Vazquez A, Seco MP, Tenorio A, Rizzoli A (2009) Detection of novel insect flavivirus sequences integrated in Aedes albopictus (Diptera: Culicidae) in Northern Italy. Virol J 6: 93. doi: 10.1186/1743-422X-6-93 PMID: 19575816

103. Sanchez-Seco MP, Vazquez A, Collao X, Hernandez L, Aranda C, et al. (2009) Surveillance of arboviruses in Spanish wetlands: detection of new flavi- and phleboviruses. Vector Borne Zoonotic Dis 10: 203–206.

104. Edwards FW (1922) On Some Malayan and Other Species of Culicoides, with a Note on the Genus Lasiohelea. Bulletin of entomological research 13: 161–167.

105. Laird M (1946) A ceratopogonine midge (Culicoides anophelis Edwards, 1922) sucking engorged blood from a mosquito (Armigeres lacuum Edwards, 1922) at Palmalmal, New Britain. Transactions and Proceedings of the Royal Society of New Zealand 76: 158–161.

106. Chu Feng I (1959) FOUR SPECIES OF THE GENUS CULICOIDES (DIPTERA, HELEIDAE) FROM HAINAN ISLAND, SOUTH CHINA. ACTA ENTOMOLOGICA SINICA 9: 161–166.

107. Chhilar JS, Chaudhry S (2011) First Report of a Biting Midge Culicoides Anophelis Parasitzing Mosquito Anopheles stephensi from North-Western India.

108. Reddy CVS, Hafeez M (2011) Studies on certain aspects of prevalence of Culicoides species.

109. Gould EA, Coutard B, Malet H, Morin B, Jamal S, Weaver SC, Get al. (2009) Understanding the alphaviruses: Recent research on important emerging pathogens and progress towards their control. Antiviral Research. doi: 10.1016/j.antiviral.2009.12.007 PMID: 20035788

110. Gould EA, de Lamballerie X, Zanotto PM, Holmes EC (2001) Evolution, epidemiology, and dispersal of flaviviruses revealed by molecular phylogenies. Adv Virus Res 57: 71–103. PMID: 11680389

111. Gubler DJ (1987) Dengue and dengue hemorrhagic fever in the Americas. Puerto Rico health sciences journal 6: 107–111. PMID: 3313490

112. Strode GK (1951) Yellow Fever. New York: McGraw-Hill.

113. Kramer LD, Chandler LJ (2001) Phylogenetic analysis of the envelope gene of St. Louis encephalitis virus. Arch Virol 146: 2341–2355. PMID: 11811684

114. Gould EA, Moss SR, Turner SL (2004) Evolution and dispersal of encephalitic flaviviruses. Arch Virol Suppl: 65–84. PMID: 15119763

115. Allicock OM, Lemey P, Tatem AJ, Pybus OG, Bennett SN, et al. (2012) Phylogeography and population dynamics of dengue viruses in the Americas. Mol Biol Evol 29: 1533–1543. doi: 10.1093/molbev/msr320 PMID: 22319149

116. Gubler DJ (1997) The emergence of dengue/dengue haemorrhagic fever as a global public health problem. In: Saluzzo JF, Dodet B, editors. Factors in the Emergence of Arbovirus Diseases. Paris: Elsevier. pp. 83–92.

117. Gubler DJ (2006) Dengue/dengue haemorrhagic fever: history and current status. Novartis Foundation Symposium 277: 3–16; discussion 16–22, 71–13, 251–253. PMID: 17319151

118. Briese T, Jia XY, Huang C, Grady LJ, Lipkin WI (1999) Identification of a Kunjin/West Nile-like flavivirus in brains of patients with New York encephalitis. Lancet 354: 1261–1262. PMID: 10520637

119. Lanciotti RS, Roehrig JT, Deubel V, Smith J, Parker M, et al. (1999) Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. Science 286: 2333–2337. PMID: 10600742

120. Reisen WK (2013) Ecology of West Nile virus in North America. Viruses 5: 2079–2105. doi: 10.3390/v5092079 PMID: 24008376

121. Roehrig JT, Layton M, Smith P, Campbell GL, Nasci R, et al. (2002) The emergence of West Nile virus in North America: ecology, epidemiology, and surveillance. Curr Top Microbiol Immunol 267: 223–240. PMID: 12082991

122.   Huanyu W, Haiyan W, Shihong F, Guifang L, Hong L, et al. (2012) Isolation and identification of a distinct strain of Culex Flavivirus from mosquitoes collected in Mainland China. Virol J 9: 73. doi: 10.1186/1743-422X-9-73 PMID: 22452813

123.   Morales-Betoulle ME, Monzon Pineda ML, Sosa SM, Panella N, Lopez MR, et al. (2008) Culex flavivirus isolates from mosquitoes in Guatemala. J Med Entomol 45: 1187–1190. PMID: 19058647

124.   Harbach R, Kitching IJ (1998) Phylogeny and classification of the Culicidae (Diptera). Systematic Entomology 23: 327–370.

125.   Charrel RN, de Lamballerie X, Fulhorst CF (2001) The Whitewater Arroyo virus: natural evidence for genetic recombination among Tacaribe serocomplex viruses (family Arenaviridae). Virology 283: 161–166. PMID: 11336541

126.   Engel SR, Hogan KM, Taylor JF, Davis SK (1998) Molecular systematics and paleobiogeography of the South American sigmodontine rodents. Mol Biol Evol 15: 35–49. PMID: 9491603

127.   Hughes AL, Friedman R (2000) Evolutionary diversification of protein-coding genes of hantaviruses. Mol Biol Evol 17: 1558–1568. PMID: 11018161

128.   Ramsden C, Holmes EC, Charleston MA (2009) Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. Mol Biol Evol 26: 143–153. doi: 10.1093/molbev/msn234 PMID: 18922760

129.   Ramsden C, Melo FL, Figueiredo LM, Holmes EC, Zanotto PM (2008) High rates of molecular evolution in hantaviruses. Mol Biol Evol 25: 1488–1492. doi: 10.1093/molbev/msn093 PMID: 18417484

130.   Sun J, Yu Y, Deubel V (2012) Japanese encephalitis virus NS1' protein depends on pseudoknot secondary structure and is cleaved by caspase during virus infection and cell apoptosis. Microbes and infection / Institut Pasteur 14: 930–940. doi: 10.1016/j.micinf.2012.03.007 PMID: 22504173

131.   Ye Q, Li XF, Zhao H, Li SH, Deng YQ, et al. (2012) A single nucleotide mutation in NS2A of Japanese encephalitis-live vaccine virus (SA14–14–2) ablates NS1' formation and contributes to attenuation. J Gen Virol 93: 1959–1964. doi: 10.1099/vir.0.043844-0 PMID: 22739060