Available online at www.sciencedirect.com

ScienceDirect

Virology 361 (2007) 80–92

VIROLOGY

www.elsevier.com/locate/yviro

Genetic characterization of tick-borne flaviviruses: New insights into evolution, pathogenetic determinants and taxonomy

Gilda Grard [a,*], Grégory Moureau [a], Rémi N. Charrel [a], Jean-Jacques Lemasson [b], Jean-Paul Gonzalez [c], Pierre Gallian [d], Tamara S. Gritsun [e], Edward C. Holmes [f], Ernest A. Gould [e], Xavier de Lamballerie [a]

[a] *Unité des Virus Emergents (EA3292, IFR48, IRD UR0178), Faculté de Médecine La Timone, 27 boulevard Jean Moulin, 13005 Marseille, France*
[b] *IRD-UR0178, Conditions et Territoires d'Emergence des Maladies, BP 1386, 18524 Dakar, Senegal*
[c] *IRD-UR0178 Mahidol University, Research Center for Emerging Viral Diseases/Center for Vaccine Development Institute of Sciences, MU, Salaya. 25/25 Phutthamonthon 4, Nakhonpathom 73170, Thailand*
[d] *Etablissement Français du Sang, 149 boulevard Baille, 13005 Marseille, France*
[e] *Centre for Ecology and Hydrology, Mansfield Road, Oxford OX1 3SR, UK*
[f] *Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA*

Received 26 July 2006; returned to author for revision 10 August 2006; accepted 10 September 2006
Available online 13 December 2006

Abstract

Here, we analyze the complete coding sequences of all recognized tick-borne flavivirus species, including *Gadgets Gully*, *Royal Farm* and Karshi virus, seabird-associated flaviviruses, *Kadam virus* and previously uncharacterized isolates of *Kyasanur Forest disease virus* and *Omsk hemorrhagic fever virus*. Significant taxonomic improvements are proposed, e.g. the identification of three major groups (mammalian, seabird and Kadam tick-borne flavivirus groups), the creation of a new species (*Karshi virus*) and the assignment of Tick-borne encephalitis and Louping ill viruses to a unique species (*Tick-borne encephalitis virus*) including four viral types (i.e. Western Tick-borne encephalitis virus, Eastern Tick-borne encephalitis virus, Turkish sheep Tick-borne encephalitis virus and Louping ill Tick-borne encephalitis virus). The analyses also suggest a complex relationship between viruses infecting birds and those infecting mammals. Ticks that feed on both categories of vertebrates may constitute the evolutionary bridge between the three distinct identified lineages.

Keywords: Flavivirus; Tick-borne virus; RNA virus; Emerging virus; Phylogeny; Evolution; Hemorrhagic fever; Encephalitis; Taxonomy

Introduction

Viruses in the genus *Flavivirus* differ from other members of the family *Flaviviridae* in their antigenic, ecological and epidemiological characteristics. For example, most infect both vertebrate and invertebrate species, a feature that is not shared by the members of *Pestivirus* and *Hepacivirus* genera. The genus *Flavivirus* comprises more than 50 recognized species which include a large number (approximately 50%) of human pathogens responsible for biphasic fever, encephalitis or hemorrhagic fever. Dengue hemorrhagic fever, Yellow fever, Japanese encephalitis, West Nile encephalitis and tick-borne encephalitis are examples of (re)emerging flaviviral diseases. The flaviviruses (FVs) also share a complex antigenic relationship and were first divided into twelve serocomplexes, according to cross-neutralization tests with polyclonal antisera (Calisher et al., 1989). The genomic sequence of the prototype Yellow fever virus was first obtained by Rice et al. (1985) and subsequently sequence data for a large number of other flaviviruses have become available. These data have allowed the progressive resolution of phylogenetic relationships that globally correlate with the previous antigenic classification. The flaviviruses form a monophyletic lineage that is currently divided into three main groups: the tick-borne flaviviruses group (TBFV), the mosquito-borne flaviviruses (MBFV) and

* Corresponding author. Fax: +33 491 32 44 95.
E-mail address: gilda.grard@gulliver.fr (G. Grard).

0042-6822/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2006.09.015

the No Known Vector (NKV) flaviviruses group. Further subdivisions are made according to the phylogenetic analysis that generally correlate with the vector responsible for transmission, the host reservoir and the disease association (Gaunt et al., 2001). This illustrates how the adaptation of each virus to specific vertebrate and invertebrate hosts influences virus evolution, dispersal, epidemiology and possibly pathogenesis of flaviviruses. Indeed, phylogenetic data are now a recognized parameter usable for the taxonomic classification of flaviviruses.

The tick-borne flaviviruses currently include twelve recognized species that are divided into two groups, the mammalian tick-borne virus group (M-TBFV) and the seabird tick-borne virus group (S-TBFV). Nevertheless, these viruses share a common ancestor within the genus *Flavivirus* (Thiel et al., 2005). The TBFVs display specific evolutionary characteristics that are largely determined by their modes of transmission. This has important consequences for their antigenic relationships, genetic diversity and geographical distribution (Marin et al., 1995b; Zanotto et al., 1995, 1996). The mammalian tick-borne flavivirus group includes six human and animal pathogens, previously known as the "tick-borne encephalitis (TBE) serocomplex," namely *Louping ill virus* (LIV), *Tick-borne encephalitis virus* (TBEV), *Omsk hemorrhagic fever virus* (OHFV), *Langat virus* (LGTV), *Kyasanur Forest disease virus* (KFDV) and *Powassan virus* (POWV). These are all encephalitic viruses, with the exception of OHFV and KFDV species that cause hemorrhagic fever in humans and have been assigned to biosafety class 4. A closely related hemorrhagic virus that unexpectedly appeared in Saudi Arabia in 1992, identified as Alkhurma hemorrhagic fever virus (AHFV), has been recommended for inclusion as a subtype of KFDV (Charrel et al., 2001). The entire genome sequences of AHFV and OHFV have now been determined (Charrel et al., 2001; Li et al., 2004; Lin et al., 2003) but, to date, only partial sequences of KFDV have been characterized (Kuno et al., 1998; Venugopal et al., 1994). Three other viruses that are not known to be human pathogens are currently included in the M-TBFV group*: Royal Farm virus* (RFV), Karshi virus (KSIV) and *Gadgets Gully virus* (GGYV). Despite the available sequence information, the genetic basis to the different types of disease caused by the mammalian TBFV is not yet understood. The seabird tick-borne virus group includes four species: *Tyuleniy virus* (TYUV), *Meaban virus* (MEAV), *Saumarez Reef virus* (SREV) and *Kadam virus* (KADV).

Herein we report the complete coding sequences of all recognized TBFV species, including new isolates of OHFV and KFDV, and the more distantly related tick-borne viruses that infect birds. We also included in our analysis the sheep encephalomyelitic viruses that were isolated in Turkey (TSEV) and Southern Europe (GGEV in Greece and SSEV in Spain). These new data provided the opportunity to extend current phylogenetic analyses between the TBFVs and to re-examine the taxonomy of the flaviviruses. At the deepest nodes of the evolutionary tree, our analysis suggests a complex relationship between viruses infecting birds and those infecting mammals. Importantly, ticks that feed on and infect both categories of vertebrates may constitute the evolutionary bridge between the distinct identified lineages. In addition, the analyses suggest that the hemorrhagic property cannot as yet be explained by specific amino acids (AA) unique to hemorrhagic viruses and are not inherited phylogenetically.

Results

Sequence information

Complete coding sequence characterization was performed for 12 viruses already recognized as belonging to the tick-borne flaviviruses group: (i) *Meaban virus*, *Saumarez Reef virus*, *Tyuleniy virus and Kadam virus* species, previously assigned to the seabird tick-borne flaviviruses group, (ii) *Gadgets Gully virus*, *Royal Farm virus*, Karshi virus, Spanish sheep encephalomyelitis virus and Turkish sheep encephalitis virus (previously recognized respectively as Spanish and Turkish subtype of *Louping ill virus* species), Greek goat encephalitis virus, *Kyasanur Forest disease virus* and *Omsk hemorrhagic fever virus* within the mammalian tick-borne flavivirus group. With the new data presented herein, genetic information becomes available for all tick-borne virus species recognized by the ICTV (Thiel et al., 2005). ICTV current status and GenBank accession numbers are reported in Table 1.

Genetic organization of complete ORFs

The putative cleavage sites of virus polyproteins were deduced from alignments with other TBFVs and from the analysis of most probable cleavage sites by host signalases using the (−3, −1) rule proposed by von Heijne (1984) and the SignalP 3.0 software (Bendtsen et al., 2004) for the prediction of signal peptides. The results are summarized in Table 2a, and the lengths of complete ORFs and deduced viral proteins are reported in Table 2b.

Based on the length of their polyproteins, mammalian TBFVs form a group with ORFs of 3414 to 3417 AA. The seabird TBFVs have longer ORFs (3421–3422 AA). For TYUV, SREV and MEAV, the difference is due to the non-structural regions, which are 11 to 16 AA longer. KADV, which was originally included in the mammalian TBFV group, and more recently in the seabird TBFV group, has the shortest ORF with 3404 residues. Based on this simple criterion alone, KADV appears to be distinct from both groups. The KADV structural genes VirC and E were 3 to 7, and 2 to 6 AA shorter, respectively, than for the other TBFVs. The GC% content of the TBFV complete coding sequences ranged from 52.1% (GGYV) to 55% (SSEV).

The length of the KFDV polyprotein and the processing pattern are identical to those previously reported for AHFV (Charrel et al., 2001). The length of the different proteins of OHFV is the same as that reported for strain Bogoluvovska (GenBank accession no. AY193805; (Lin et al., 2003)) and Kubrin (GenBank accession no. AY438626 (Li et al., 2004)). The M/E cleavage site appears to be very different from that indicated in the article of Lin et al., but this is due to a typing error in Fig. 1B of Dr. Lin's paper. A complete comparison of

Table 1
Flaviviruses included in genetic analysis

TICK-BORNE FLAVIVIRUSES NEWLY CHARACTERISED						
Virus name	Abbreviation	Strain	Propagation	ICTV8th report status	Suggested modifications	GenBank accession number
Omsk hemorrhagic fever virus	$OHFV^{UVE}$	Bogoluvovska	Suckling mice brain	Mammalian TBF group, OHFV species		AY323489
Kyasanur Forest disease virus	KFDV	It P9605	Vero cells	Mammalian TBF group, KFDV species		AY323490
Spanish sheep encephalomyelitis virus	SSEV	87/2617	SW13 cells	Mammalian TBF group, LIV species, Spanish subtype	See plain text	DQ235152
Turkish sheep encephalitis virus	TSEV	TTE80	SW13 cells	Mammalian TBF group, LIV species, Turkish subtype	Id.	DQ235151
Greek goat encephalitis virus	GGEV	Vergina	SW13 cells	Not included*	Id.	DQ235153
Gadgets Gully virus	GGYV	CSIRO122	SW13 cells	Mammalian TBF group, GGYV species		DQ235145
Kadam virus	KADV	AMP 6640	SW13 cells	Mammalian TBF group, KADV species	Kadam virus group, KADV species	DQ235146
Royal Farm virus	RFV	Eg Art 371	SW13 cells	Mammalian TBF group, RFV species		DQ235149
Karshi virus	KSIV	30517	SW13 cells	Mammalian TBF group, RFV species	Mammalian TBF group, KSIV species	DQ235147
Meaban virus	MEAV	Brest/Ar/T70	SW13 cells	Seabird TBF group, MEAV species		DQ235144
Saumarez Reef virus	SREV	CSIRO 4	SW13 cells	Seabird TBF group, SREV species		DQ235150
Tyuleniy virus	TYUV	6017	SW13 cells	Seabird TBF group, TYUV species		DQ235148

OTHER MAMMALIAN TICK-BORNE FLAVIVIRUSES INCLUDED IN GENETIC ANALYSIS				OTHER FLAVIVIRUSES INCLUDED IN GENETIC ANALYSIS			
Virus name	Abbreviation (strain)	GenBank accession number	ICTV current status	Virus name	Abbreviation	GenBank accession number	ICTV current status
Omsk hemorrhagic fever virus	$OHFV^{Li}$ (Kubrin)	AY438626	OHFV species	*Yellow fever virus*	YFV	AF094612	YFV group, YFV species
Omsk hemorrhagic fever virus	$OHFV^{Lin}$ (Bogoluvovska)	AY193805		*Dengue virus* type 1	DENV-1	U88536	DENV group, DENV species
Langat virus	LGTV	AF253419	LGTV species	*Dengue virus* type 2	DENV-2	AF038403	
				Dengue virus type 3	DENV-3	M93130	
Powassan virus	POWV	L06436	POWV species	*Dengue virus* type 4	DENV-4	M14931	
Deer tick virus*	DTV	NC003218		*Japanese encephalitis virus*	JEV	M18370	JEV group, JEV species
Alkhurma hemorrhagic fever virus*	AHFV	AF331718	KFDV species	*Murray Valley encephalitis virus*	MVEV	AF161266	JEV group, MVEV species
				St. Louis encephalitis virus	SLEV	DQ359217	JEV group, SLEV species
Tick-borne encephalitis virus, European subtype	$TBEV^{EU1}$ (Neudoerfl)	U27495	TBEV species	*West Nile virus*	WNV	M12294	JEV group,WNV species
	$TBEV^{EU2}$ (Hypr)	U39292		Kunjin virus	KUNV	D00246	
Tick-borne encephalitis virus, Far Eastern subtype	$TBEV^{FE}$ (Sofjin)	AB062064		*Rio Bravo virus*	RBV	AF144692	RBV group, RBV species
				Montana myotis leukoencephalitis virus	MMLV	AJ299445	RBV group, MMLV species
Tick-borne encephalitis virus, Siberian subtype	$TBEV^{Sib}$ (Vasilchenko)	AF069066		*Modoc virus*	MODV	AJ242984	MODV group, MODV species
				Apoi virus	APOIV	AF160193	MODV group, APOIV species**
Louping ill virus, British subtype	LIV	Y07863	LIV species	*Cell fusing agent virus*	CFAV	M91671	Tentative species in the genus
				Kamiti river virus	KRV	NC05064	Tentative species in the genus

The names of viruses, abbreviations and classifications are those of the eighth report of the ICTVs, except for the abbreviations SSEV, TSEV and GGEV. Superscript letters with OHFV names indicate the origin of the sequences (UVE = Unité des Virus Emergents; Li = Li et al. (2004); Lin = Lin et al. (2003)). Names of recognized species are italicized. (*) not included in the eighth report of the ICTV. Molecular data suggest that DTV is a subtype of POWV; (**) analysis of distances between complete coding sequences suggests that APOIV belongs to a distinct group (Apoi virus group).

the cleavage sites of both OHFV strains is reported in Table 2a and shows that these sites are similar.

The C-terminal hydrophobic domain (CTHD) of KADV, MEAV, SREV and TYUV was found to be 19 instead of 20 AA in length for all other TBFVs. Predictions were identical using either neural network- or hidden Markov-based model algorithms of the SignalP 3.0 software (Bendtsen et al., 2004). The only exception was GGYV: the AnchC/prM site predicted by the hidden Markov model algorithm was VFISSA/SVRR (which resulted in a 21 AA CTHD), and VFISS/ASVRR using the neural network algorithm. Because it resulted in a 20 AA long CTHD homologous to other mammalian tick-borne flaviviruses (M-TBFVs), the latter prediction was retained in Table 2a.

Analysis of the envelope gene

Several amino acid patterns of interest have been reported previously in flavivirus E genes. They were re-examined including the newly determined sequence data. Results are summarized in Table 3.

1. The 12 cysteine residues that form intramolecular disulfide bonds (Nowak and Wengler, 1987) (AA positions 3, 30, 60, 74, 92, 105, 116, 121, 186, 290, 307 and 338 of LIV E protein) were conserved for all TBFVs.
2. The 3 N-X-T/S potential N-glycosylation sites (Chambers et al., 1990) (positions 154–156, 361–363 and 473–475 of LIV E protein) were conserved for all TBFVs, with the following exceptions: (i) in the case of GGYV, the third motif was missing; (ii) the first and second motifs were missing for SREV and TYUV; (iii) in the case of KADV and MEAV, the first and third motifs were maintained, but the second was missing.
3. The (DSGHD) pentapeptide, previously proposed to be specific for TBFVs (Gao et al., 1993) (positions 320–324 of LIV E protein), was not fully conserved in TBFVs and not specific for the mammalian subgroup. The SGHD signature was specific for mammalian TBFVs while the SQHD signature was specific for seabird TBFVs. KADV exhibited a specific signature (SAHD).
4. The newly identified hexapeptide HDTVVM (positions 323–328 of LIV E protein) was found to be specific for TBFVs and fully conserved in both mammalian and seabird subgroups.
5. The hexapeptide EHLPTA (positions 207–212 of LIV E protein), proposed to be specific for TBFVs (Shiu et al., 1991), was not fully conserved in the mammalian subgroup and constitutes a 4-amino-acid insertion in mammalian viruses as compared with KADV and the seabird subgroup.
6. The sequence of the fusion peptide (Allison et al., 2001; Heinz and Allison, 2003; Roehrig et al., 1989) was previously reported to be DRGWGNHCGXFGKG (positions 98–111 of LIV E protein), with X being a leucine for all TBFVs except for DTV and POWV that contained phenylalanine (Beasley et al., 2001; Kuno et al., 2001; Mandl et al., 1993). This is confirmed for all TBFVs analyzed here except KADV that displays a unique substitution producing the sequence DRGWGNNCGLFGKG.
7. The cluster of 2 or 3 hypervariable AA (positions 232–234 of LIV E protein) previously proposed as a flavivirus genetic marker and shown to identify individual serotypes (Shiu et al., 1991, 1992) still allows the distinction of all the different TBFV species (Table 3).

NS3 helicase core motif

The DEXH core motif of the helicase (where X is a cysteine for CFAV, a serine for Tamana bat virus, and an alanine for all other FVs sequenced to date) is DEGH for the seabird TBFVs, KADV and KSIV.

Genetic determinants for hemorrhagic manifestations

An analysis was made for specific motifs that might be associated with the hemorrhagic manifestations produced following infection with KFDV, AHFV or OHFV: (i) the AKG motif (VirC positions 2–4) was present in KFDV and AHFV but not in OHFV; (ii) the insertion of a basic residue at position VirC:91 was identified in AHFV and KFDV, but not in the OHFV sequence (the presence of a basic residue is common at this position among mosquito-borne flaviviruses); (iii) the AHFV and KFDV specific EHLPKA hexapeptide of the E protein was replaced as reported above by the classical tick-borne specific EHLPTA in the OHFV polyprotein; (iv) the KFDV and AHFV specific EGSK motif related to the non-reactivity of Mab 4.2 with KFDV E protein (Venugopal et al., 1994) was EGTK (E protein positions 277–280) in OHFV (i.e. identical to LGTV and TBEV^{EU} non-hemorrhagic viruses). Finally, OHFV, AHFV and KFDV shared only three specific AA substitutions: position 76 of OHFV E protein and positions 558 and 585 of OHFV NS3 protein. No specific pattern encompassing several amino acids shared by these three hemorrhagic viruses could be identified.

Analysis of genetic distances

Comparison of AA distances between full-length polyproteins is presented in Fig. 1. (i) AA pairwise distances over 0.314 are observed between viruses that belong to different groups. According to this unambiguous cut-off, KADV forms a third group distinct from M-TBFVs and S-TBFVs. (ii) Distances below 0.087 correspond to viruses that belong to the same species and also to the group of M-TBFVs located in the most westerly region of the evolutionary cline which includes TBEV, LIV, SSEV, GGEV and TSEV. As previously demonstrated from the analysis of partial sequences (Charrel et al., 2001), the separation of TBEV and LIV as distinct species cannot be justified on the basis of genetic distance. The genetic distance between the two OHFV strains previously characterized (OHFV^{Li} strain Kubrin and OHFV^{Lin} strain Bogoluvovska) is very low (0.001), but the newly described strain (OHFV^{UVE}) is more divergent with a genetic distance of 0.035. Interestingly, both OHFV^{Lin} and OHFV^{UVE} were supposedly derived from the same strain (Bogoluvovska), but the genetic divergence observed suggests either a mis-identification of one of the

Table 2a
Putative processing of flavivirus polyproteins

Cleavage sites (protease)	VirC/CTHD (VSP)	AnchC/prM (HS)	Pr/M (Furin)	M/E (HS)	E/NS1 (HS)	NS1/NS2A (unknown)
LIV	RGKRR / SVTNW	GMTLA / ATVRK	SRTRR / SVLIP	APVYA / SRCTH	LGVGA / DVGCA	MVVAD / NGELL
SSEV	RGKRR / SVTDW	GMTLA / ATVRK	SRTRR / SVLIP	APVYA / SRCTH	LGVGA / DVGCA	MVVAD / NGELM
GGEV	RGKRR / SATDW	GVVLA / ATVRK	ARTRR / SVLIP	APVYA / SRCTH	LGVGA / DVGCA	MVVAD / NGELL
TSEV	RGKRR / SATDW	GVVLA / ATVRK	ARTRR / SVLIP	APVYA / SRCTH	LGVGA / DVGCA	MVVAD / NGELL
$TBEV^{EU1}$	RGKRR / SATDW	GMTLA / ATVRK	SRTRR / SVLIP	APVYA / SRCTH	LGVGA / DVGCA	MVVAD / NGELL
$TBEV^{EU2}$	RGKRR / SATDW	GMTIA / ATVRK	SRTRR / SVLIP	APVYA / SRCTH	LGVGA / DVGCA	MVVAD / NGELL
$TBEV^{Sib}$	RGKKR / STTDW	GVTLA / ATVRK	SRTRR / SVLIP	APVYA / SRCTH	LGVGA / DVGCA	MVVAD / NGELL
$TBEV^{FE}$	RGKRR / SAVDW	GVTLA / ATVRK	SRTRR / SVLIP	APVYA / SRCTH	LGVGA / DVGGA	MVVAD / NGELL
$OHFV^{Li/Lin}$	RGKRR / STTDW	SIALA / ATVRK	TRSRR / SVLIP	APAYA / SRCTH	LGVGA / DVGCA	MVVAD / NGELL
$OHFV^{UVE}$	RGKRR / STTDW	SITLA / ATVRK	TRSRR / SVLIP	APAYA / SRCTH	LGVGA / DVGCA	MVVAD / NGELL
LGTV	RGSRR / TTIDW	GMCLT / ATVRR	SRSRR / SVLIP	APAYA / SRCTH	LGVGA / DVGCA	MVVAD / NGALL
ALKV	RGKRR / STTGL	TLVIS / ATIRR	GRSRR / SVSIP	APTYA / TRCTH	LGVGA / DMGCA	MVLAD / NGAML
KFDV	RGKRR / STTGL	TLVFS / ATVRR	GRNRR / SVSIP	APTYA / TRCTH	LGVGA / DMGCA	MVLAD / NGAML
DTV	RGRRR / SGVDW	AMAMA / TSIHR	FRGRR / SVVIP	GPVYA / TRCTH	MGVGA / DYGCA	MVMAD / NGAML
POWV	RGRRR / SGVDW	TMAMA / TTIHR	SRGKR / SVVIP	GPVYA / TRCTH	MGVGA / DYGCA	MVVAD / NGALL
RFV	RGKRR / AGEGL	GVALC / AHVRR	SRVTR / SVHIP	APGQA / SRCTH	LGVGA / DVGCS	TVMAA / NEQLT
KSIV	RGKRR / SAVDW	PLCLA / LTMRS	GRSRR / SVNIP	APAYA / SRCVH	LGVGA / DQGCS	TVVAC / NGELM
GGYV	RGNRR / GVRWD	VFISS / ASVRR[1]	SRTRR / SVVIP	APAYA / TKCTH	TGVGA / DVGCA	MVVAE / NGALL
KADV	NGKRK / SAGGG	SVAGA / AVVVQ	RRQTR / SVHIT	APAYA / SRCVH	LGVGA / DIGCA	QVLAF / NDGLV
TUV	GKNRK / GGGAG	PAALG / AVAMT	ARRRR / SVTLN	VPAYA / SRCVH	LGVGA / DYGCA	TVLAM / TDGLE
MEAV	GRNRR / GGGSF	AIIAG / AVLME	KRHRR / ELSIS	VPAYA / SRCVH	LGVGA / DYGCA	TVLAM / TDGFE
SREV	GRNRR / GPGGM	GGCMG / ALVMT	SRARR / SISIT	VPAYA / SKCVH	LGVGA / DYGCA	TVLAM / TDGLE
	:: /	/	* * / .: :	* * / ::*.*	**** / * *::	*:* / . :

Cleavage sites (protease)	NS2A/NS2B (VSP)	NS2B/NS3 (VSP)	NS3/NS4A (VSP)	NS4A/2K (VSP)	2K/NS4B (VSP)	NS4B/NS5 (VSP)
LIV	HRGRR / SFSEP	RSSRR / SDLVY	ASGRR / SFGDV	TGKQR / SSDDN	GLVAA / NEMGF	SGGRR / GGSDG
SSEV	HRGKR / SFSEP	RSSRR / SDLVY	ASGRR / SFGDV	TGKQR / SSDDN	GLVAA / NEMGL	SGGRR / GGSEG
GGEV	RGGRR / SFSEP	RSPRK / SDLVF	ASGRR / SFGDV	AGKQR / SSDDN	GLVAA / NEMGF	SGSRR / GGSEG
TSEV	RGGRR / SFSEP	RSPRK / SDLVF	ASGRR / SFGDV	EGKQR / SSDDN	GLVAA / NEMGF	SGSRR / GGSEG
$TBEV^{EU1}$	HRGRR / SFSEP	RSSRR / SDLVF	ASGRR / SFGDV	AGKQR / SSDDN	GLVAA / NEMGF	SGGRR / GGSEG
$TBEV^{EU2}$	HRGRR / SFSEP	RSSRR / SDLVF	ASGRR / SFGDV	AGKQR / SSDDN	GLVAA / NEMGF	SGGRR / GGSEG
$TBEV^{Sib}$	HRGRR / SFSEP	RSARR / SDLVF	ASGRR / SIGDV	AGKQR / SSDDN	GLVAA / NEMGF	SGSRR / GGAEG
$TBEV^{FE}$	HGRRR / SFSEP	RTARR / SGLVF	ASGRR / SFGDV	AGKQR / SSDDN	GLVAA / NEMGF	SGSRR / GGSEG
$OHFV^{Li/Lin}$	RRDRR / SFSEP	RSARR / SDLVF	ASGRR / SLGDM	AGKQR / SSDDN	GLVAA / NEMGF	SGTRR / GGSEG
$OHFV^{UVE}$	RRERR / SFSEP	RTARR / SDLVF	ASGRR / SLGGM	AGKQR / SSDDN	SLVAA / NEMGF	SGTRR / GGSEG
LGTV	SRGRR / SFNEP	GSPRR / TDLVF	ASGRR / SVGDV	TGKQR / SSDDN	GMVAA / NEMGL	TGTRR / GGSEG
ALKV	RRNRR / SFSEP	SSGRR / SELVF	ASGRR / SVGDV	PGKQR / SSDDN	GLVTA / NEMGM	TGTRR / GGADG
KFDV	RRNRR / SFSEP	GSGRR / SELVF	ASGRR / SVGDV	PGKQR / SSDDN	GLVTA / NEMGM	TGTRR / GGAEG
DTV	GRGRR / SLSEP	SSARR / TDLVF	ASGRR / SAMDI	PGKQR / SGEDN	GLLAA / NELGY	QGARR / GGAEG
POWV	GRGRR / SLSEP	SSTRR / TDLVF	ASGRR / SAVDI	PGKQR / SGEDN	GLVAA / NELGY	QGARR / GGAEG
RFV	RCLRR / SIEEP	FSGKR / GDLVF	ASGRR / SASGI	PGKQR / STDDN	GGVAA / NEMGW	SGTRR / GGGEG
KSIV	HPKRR / SIGEP	TSPRR / GELVF	ASCRR / GASEL	PGKQR / SADDN	GTVAA / NELGW	SEPRR / GGGAG
GGYV	RMERR / SIAEP	QGGRR / GDLVF	ASGRR / SVVNI	VGKQR / SGEDN	GTIAA / NEAGF	LTTRR / GGGEG
KADV	FGKQR / SLEEP	SRNRR / SDLIF	AESRK / SASLV	PGTQR / SGEDT	GAVAA / NELGW	GPTRR / GLTGG
TUV	RVLRR / TVTDT	GSSRR / GDVVW	AEGRR / GVGGG	PGTQR / SFEDN	GSIAA / NEMGW	LRDRK / GNHSR
MEAV	HGNKR / SVSDA	GSERR / SDLVW	AEGRR / GVGRG	PGTQR / SFEDN	GLVAA / NEMGW	ARDRR / GPGST
SREV	RSGRR / SVADA	NDTRR / GDLVW	AEGRR / SIGGG	PGTQR / SFEDN	GIVAA / NEMGW	QRDRR / GPGSA
	:* / :. :.	:: / :::	*. *: / .	*.** / * :*.	. ::* / ** *	*: / *

Table 2b
Amino acid length of the proteins

		VirC	CTHD	Pr	M	E	NS1	NS2A	NS2B	NS3	NS4A	2K	NS4B	NS5	Total length
Mammalian tick-borne flavivirus group	LIV, TBEV OHFV, LGTV	96	20	89	75	496	353	229	131	621	126	23	252	903	3414
	KFDV, ALKV	+1	=	=	=	=	+1	=	=	=	=	=	=	=	3416 (+2)
	POWV, DTV	−2	=	=	=	+1	+1	=	=	+1	=	=	=	=	3415 (+1)
	KSIV	=	=	=	=	=	+1	=	=	−1	=	=	+2	=	3416 (+2)
	RFV	=	=	=	=	=	+1	=	=	=	=	=	+1	=	3417 (+3)
	GGYV	=	=	=	=	+1	+1	=	=	+1	=	=	−1	=	3416 (+2)
Kadam tick-borne flavivirus group	KADV	−6	−1	−1	=	−5	+1	=	=	+1	=	=	=	+2	3404 (−10)
Seabirds tick-borne flavivirus group	TYUV	−2	−1	=	=	−4	=	+5	+1	+1	+1	=	+4	+3	3422 (+8)
	MEAV	−3	−1	=	=	−4	=	+5	+1	+1	+1	=	+6	+3	3421 (+7)
	SREV	−3	−1	=	=	−4	=	+5	+1	+1	+1	=	+5	+3	3422 (+8)

Protein sizes are indicated in AA in the first line that is used as a reference from which the differences in protein length are reported for other viruses. The total length of the polyprotein is reported for each virus, and the difference in number of AA with respect to the first line is indicated in brackets.

strains or different passage histories. (iii) Distances between 0.099 and 0.314 were observed between viruses belonging to different species of a given group. Accordingly, the genetic distance (0.294) between RFV and KSIV (currently assigned to the same species) indicates that these viruses may be assigned to two distinct species.

In the complete E gene, the distribution of AA distances identifies several levels of genetic relationship (Fig. 2). In descending order, they correspond to distances between TBFVs and CFAV, between TBFVs and other FVs, between TBFVs belonging to different groups and between TBFVs belonging to the same group. The robustness of this frequency distribution profile was tested by adding ~100 sequences corresponding to different viral isolates (including a majority of sequences of LIV and TBEV isolates). As expected, the peaks were enlarged but the characteristics of distribution remained the same (result not shown). Further analysis showed that the distances between KADV and other TBFVs fall in the peak corresponding to the inter-groups distances, suggesting that KADV belongs to a third independent group. While low distance values may not confirm that two viruses belong to the same species, elevated values are indicative of two viruses belonging to different species. Accordingly, the distance observed between RFV and KSIV (greater than that between LIV and POWV) indicates that RFV and KSIV should be identified as distinct species.

In the complete NS3 gene, the distribution of AA distances is similar to that observed in the E gene. However, the range of genetic variability is narrower. All observations reported above regarding (i) the belonging of KADV to a distinct group, (ii) the absence of a cut-off for the demarcation of TBEV and LIV species and (iii) the assignment of RFV and KSIV to different species are equally applicable to the NS3 gene.

In the complete NS5 gene, the range of genetic variability is significantly reduced but groups are still clearly defined. The observations regarding the demarcation of species and the taxonomic status of KADV, KSIV and RFV are the same as above.

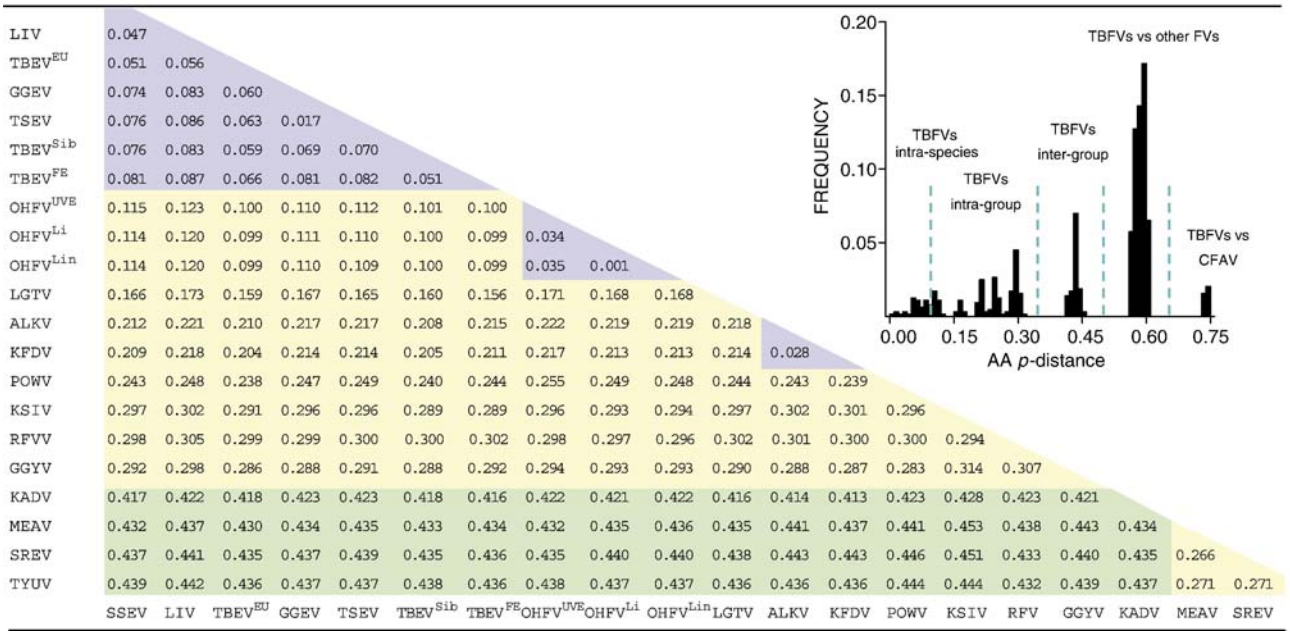
Phylogenetic analyses

Complete coding sequences

A phylogenetic tree produced using complete flavivirus amino acid sequences and the maximum likelihood method is presented in Fig. 3. The general organization of the tree is "NS3-like" i.e. the TBFVs diverged with the NKVs. This is in agreement with the original work by Billoir et al. (2000) and the most recent phylogenetic analysis of the *Flavivirus* genus (Cook and Holmes, 2006).

TYUV, MEAV and SREV form an early diverging lineage (quartet puzzling frequencies ≥99%) inside the TBFV group, in agreement with their taxonomic position in the seabird tick-borne flavivirus group. KADV forms a new evolutionary lineage with quartet puzzling frequencies at 98%. Inside the M-TBFV group, a trifurcation is observed with GGYV in one branch, KSIV and RFV in a second branch and all other M-TBFVs in the third one. In the latter group, the topology globally confirms the clinal distribution previously reported (Zanotto et al., 1995), displaying successively POWV and its close relative DTV, KFDV and its close relative AHFV, LGTV, OHFV and finally the distal group of viruses which includes TBEV, LIV, SSEV, GGEV and TSEV. It is notable that the branching pattern within the latter group is different from a commonly reported evolutionary scheme based on E gene analysis (Charrel et al., 2001; Gould et al., 2001; McGuire et al., 1998). These previous analyses reported that LIV and TSEV have a

Notes to Table 2a:
VirC, mature virion C protein; CTHD, C-terminal hydrophobic domain; AnchC, anchored C protein (mature virion C protein+CHTD); prM, membrane precursor; E, envelope; NS, non-structural protein; VSP, viral serine protease; HS, host signalase. "*" indicates a single fully conserved residue; ":" indicates that one of the strong amino acid groups is conserved; "." indicates that one of the weaker amino acid groups is conserved (group definition according to ClustalX program: www.es.embnet.org/Doc/clustalw/clustax.html).



	SSEV	LIV	TBEV^{EU}	GGEV	TSEV	TBEV^{Sib}	TBEV^{FE}	OHFV^{UVE}	OHFV^{Li}	OHFV^{Lin}	LGTV	ALKV	KFDV	POWV	KSIV	RFV	GGYV	KADV	MEAV	SREV
LIV	0.047																			
TBEV^{EU}	0.051	0.056																		
GGEV	0.074	0.083	0.060																	
TSEV	0.076	0.086	0.063	0.017																
TBEV^{Sib}	0.076	0.083	0.059	0.069	0.070															
TBEV^{FE}	0.081	0.087	0.066	0.081	0.082	0.051														
OHFV^{UVE}	0.115	0.123	0.100	0.110	0.112	0.101	0.100													
OHFV^{Li}	0.114	0.120	0.099	0.111	0.110	0.100	0.099	0.034												
OHFV^{Lin}	0.114	0.120	0.099	0.110	0.109	0.100	0.099	0.035	0.001											
LGTV	0.166	0.173	0.159	0.167	0.165	0.160	0.156	0.171	0.168	0.168										
ALKV	0.212	0.221	0.210	0.217	0.217	0.208	0.215	0.222	0.219	0.219	0.218									
KFDV	0.209	0.218	0.204	0.214	0.214	0.205	0.211	0.217	0.213	0.213	0.214	0.028								
POWV	0.243	0.248	0.238	0.247	0.249	0.240	0.244	0.255	0.249	0.248	0.244	0.243	0.239							
KSIV	0.297	0.302	0.291	0.296	0.296	0.289	0.289	0.296	0.293	0.294	0.297	0.302	0.301	0.296						
RFVV	0.298	0.305	0.299	0.299	0.300	0.300	0.302	0.298	0.297	0.296	0.302	0.301	0.300	0.300	0.294					
GGYV	0.292	0.298	0.286	0.288	0.291	0.288	0.292	0.294	0.293	0.293	0.290	0.288	0.287	0.283	0.314	0.307				
KADV	0.417	0.422	0.418	0.423	0.423	0.418	0.416	0.422	0.421	0.422	0.416	0.414	0.413	0.423	0.428	0.423	0.421			
MEAV	0.432	0.437	0.430	0.434	0.435	0.433	0.434	0.432	0.435	0.436	0.435	0.441	0.437	0.441	0.453	0.438	0.443	0.434		
SREV	0.437	0.441	0.435	0.437	0.439	0.435	0.436	0.435	0.440	0.440	0.438	0.443	0.443	0.446	0.451	0.433	0.440	0.435	0.266	
TYUV	0.439	0.442	0.436	0.437	0.437	0.438	0.436	0.438	0.437	0.437	0.436	0.436	0.436	0.444	0.444	0.432	0.439	0.437	0.271	0.271

Fig. 1. Genetic distances between complete ORF sequences. The upper right diagram displays the distribution of AA *p*-distances (*p*-distances are reported on the *X* axis and the frequency for intervals of 0.01 is reported on the *Y* axis). The lower left matrix shows pairwise distances computed from AA alignments. Intra-species distances are highlighted in blue (see Discussion section about the grouping of LIV and TBEV species), intra-group distances in yellow and inter-group distances in green. OHFV^{UVE}: strain Bogoluvovska, GenBank accession no. AY323489, this study. OHFV^{Li}: strain Kubrin, GenBank accession no. AY438626, Li et al. (2004). OHFV^{Lin}: strain Bogoluvovska, GenBank accession no. AY193805, Lin et al. (2003).

common ancestor and form a sister group to western TBEV. In contrast, our analyses show that GGEV and TSEV have a common ancestor and form a sister group to a group made of west European TBEV and LIV. This topology had been reported by several authors from other E gene studies (Gao et al., 1993, 1997; Marin et al., 1995a,b; Zanotto et al., 1995).

Envelope, NS3 and NS5 genes

Comparative phylogenetic trees were produced using the E, NS3 and NS5 AA sequences, the amino acid *p*-distance method and the Neighbor-Joining algorithm (Fig. 4). The topology at the deepest nodes of all three trees had a low bootstrap support, but the separation of the S-TBFV and the M-TBFV groups was supported by 100% of bootstrap replicates and the general clinal organization of the trees was similar. The topology of the tree constructed from the NS3 gene sequences was extremely similar to that of the tree obtained from complete ORF sequences with only a minor difference observed in the branching order of SREV, MEAV and TYUV. This confirms previous observations that NS3 phylogenies produce a robust prediction of flaviviruses evolution that compares favorably with reconstructions based on complete sequences (Cook and Holmes, 2006; de Lamballerie et al., 2002).

The phylogenetic position of GGEV and TSEV agrees with that observed in complete ORF-based reconstructions using NS3 and NS5 trees, but not E trees where they grouped with LIV and SSEV as discussed above. KADV grouped with the M-TBFV lineage in NS3 and NS5 trees (100% bootstrap support) and with the S-TBFV lineage in the E tree, but in all reconstructions appeared as a third independent evolutionary group.

Discussion

Taxonomic proposal

We examined the impact of new genomic data on flavivirus taxonomy. First, *Kadam virus* was previously assigned to the M-TBFV group (Heinz et al., 2000) and more recently to the S-TBFV group (Thiel et al., 2005). Analyses of genetic distances (Figs. 1 and 2) show that KADV constitutes a third group of tick-borne viruses in addition to the M-TBFV and S-TBFV groups. In addition, (i) phylogenetic analysis assigns KADV to a distinct evolutionary branch with strong bootstrap support (Figs. 3 and 4); (ii) in the group-specific pentapeptide (AA positions 320–324 of LIV E protein), KADV exhibits a specific signature; (iii) in the fusion peptide (positions 98–111 of LIV E protein), KADV has a unique substitution pattern; (iv) the KADV polyprotein is markedly shorter than that of any other TBFVs. In summary, the evidence for assigning KADV to a third group of tick-borne viruses is compelling.

Second, Karshi virus is currently recognized as a variant of the *Royal Farm virus* species. Analysis of AA distances shows that the genetic distance between RFV and KSIV is in the order of magnitude of distances between different species. Accordingly, KSIV should be recognized as a new species within the M-TBFV group (*Karshi virus*).

Third, it has been shown previously that analyses of genetic distances could not robustly discriminate between Tick-borne

Table 3
AA patterns of TBFVs in E gene

		Glycosylation sites (154-156; 361-363; 473-475)			Fusion petide (98-111)	Hexa-peptide EHLPTA (207-212)	Tri-peptide (232-234)	Penta-peptide DSGHD (320-324)	Penta-peptide HDTVVM (323-328)
Mammalian tick-borne virus group	LIV	NET	NPT	NPT	DRGWGNHCGLFGKG	EHLPTA	HDGNPHW	DSGHD	HDTVVM
	SSEV	...	...	...	.....	.....	.E.AQR.	.....	.....
	TSEV	...	...	...	.....	.....	.E.ALG.	.....	.....
	GGEV	...	...	...	.....	.....	.E.ALG.	.....	.....
	TBEV^{EU1}	...	...	...	.....	.....	.E.AQN.	.....	.....
	TBEV^{EU2}	...	...	...	.....	.....	.E.ARN.	.....	.....
	TBEV^{Sib}	...	...	...	.....	.....	.E.AQQ.	.....	.....
	TBEV^{FE}	...	...	...	.....	.....	.E.AQN.	.....	.....
	OHFV	...	...	...	.....	.....	.E.MVG.	.....	.....
	LGTV	...	...	...	.....	.....	...AEA.	.....	.....
	KFDV	..S	..S	...	.....	....K.	.G.AQE.	.....	.....
	ALKV	..S	..S	...	.....	....K.	.E.AHE.	.....	.....
	DTV	...	...	...	.....F....	D...S.	.KDNQD.	.....	.....
	POWV	...	...	...	.....F....	D...S.	.KDNQD.	.....	.....
	KSIV	...	...	.T.	.....	ALH.KV	.HEAEI.	.....	.....
	RFV	.D.	...	.V.	.....	....Q.	.K.ESA.	.....	.....
	GGYV	.D.	...	.MN	.....	S...K.	.V.AET.	E....	.....
Kadam virus group	KADV	.K.	..I	.M.	......N.....	----KV	.G.GDT.	E.A..	.....
Seabird tick-borne virus group	MEAV	..S	..V	.F.	.....	----NV	.GDNP-.	E.Q..	.....
	SREV	.NL	..V	.M.	.....	----DV	.E.TS-.	E.Q..	.....
	TYUV	.VR	..V	.M.	.....	----QV	..ENA-.	G.Q..	.....

AA position with respect to the 1_{st} AA of LIVE protein are indicated in brackets. Dots indicate conserved AA. First column: substitutions that destroy glycolisation sites are highlighted in black. Fourth column: the tri-peptides is highlighted with respect to the adjacent AA. Fifth column: AA that discriminates the three groups is highlighted.

encephalitis viruses and Louping ill viruses (Charrel et al., 2001). Detailed distance and phylogenetic analyses provide new insights into the genetic relationship between the following viruses: (i) the Spanish subtype of LIV is related to the British subtype based on genetic distance and evolutionary clustering; (ii) the Greek goat encephalitis virus (not included in the current

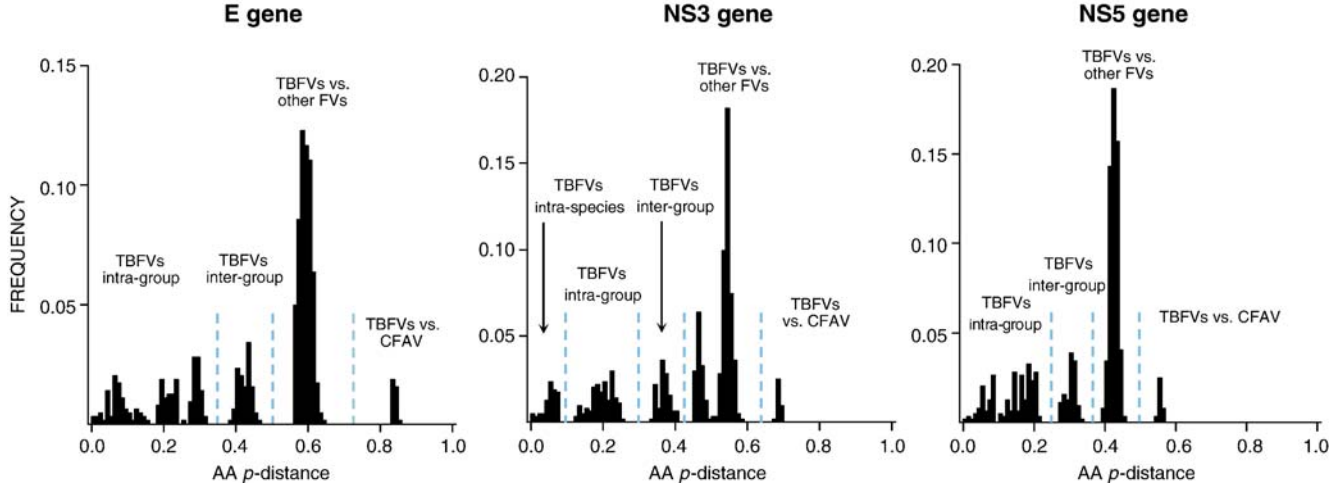


Fig. 2. Distribution of AA pairwise distances in E, NS3 and NS5 genes. The AA *p*-distances are reported on the *X* axis, and the frequency for intervals of 0.01 is reported on the *Y* axis. The grouping of LIV and TBEV species and the cut-off distance for species demarcation in NS3 gene are discussed in the text.

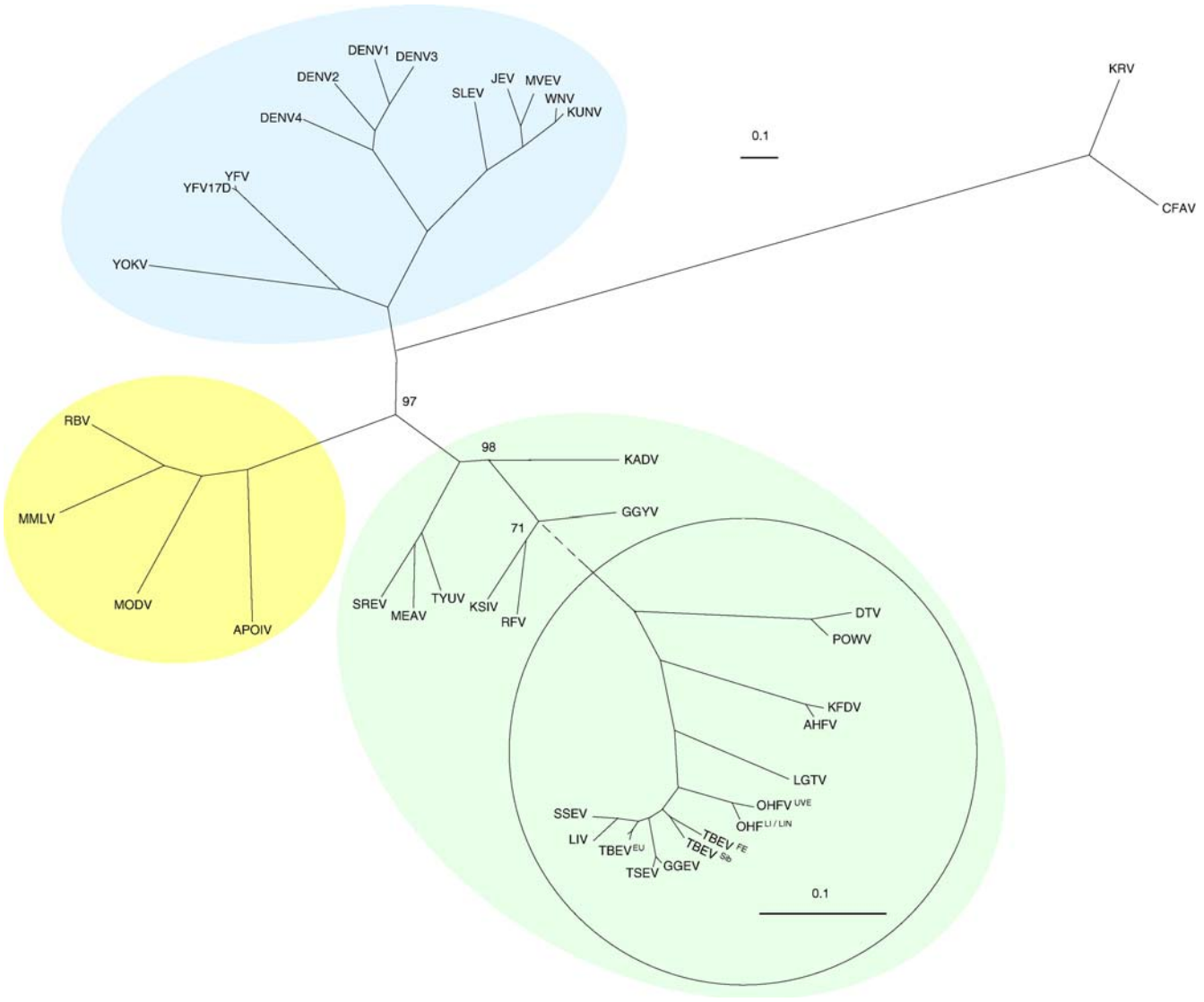


Fig. 3. Phylogenetic analysis based on complete polyprotein sequences. Phylogenetic reconstruction was performed using the maximum likelihood method. All branchings were supported by quartet puzzling frequencies at 99% or 100% except at the forks where a value is indicated. The tick-borne flavivirus group is highlighted in green, the mosquito-borne flavivirus group in blue and the no-known vector flavivirus group in yellow. To improve the legibility of the tree, the distal part of the TBFV branch is presented with a 3.5× magnification.

classification) is closely related to TSEV and it should therefore be considered a strain of this virus; (iii) TSEV, considered previously as a Turkish subtype of LIV, is more closely related, based on genetic distance, to the Tick-borne encephalitis complex viruses (especially the European subtype) than to Louping ill viruses. TSEV forms a branch distinct from the Louping ill viruses. Consequently, despite common biological properties, TSEV is neither a subtype of LIV nor of TBEV. We therefore suggest that the taxonomic organization of this virus group merits a significant revision. We propose that one species could be created, namely *Tick-borne encephalitis virus*, with 4 different types: Louping ill virus, including Spanish, British and Irish subtypes; Western tick-borne encephalitis virus; Eastern tick-borne encephalitis virus including Far Eastern and Siberian subtypes; and Turkish sheep encephalitis virus, including the Greek goat encephalitis virus subtype. This taxonomic proposal is supported by cut-off distances for species demarcation at 0.09 based on complete AA sequences and at 0.063 based on the complete NS3 protein.

Genetic evolution and biogeography

The analyses of genetic distances and tree topologies corroborate previous analyses suggesting that the extant TBFVs have evolved along a geographical cline during the past 2000 years, probably emerging in Africa and gradually dispersing north and westwards through the Asian and European forests and on the moorlands. By contrast, the evolutionary origin of GGYV (southern ocean), RFV (Afghanistan) and KSIV (Uzbekistan) remains poorly understood. Despite their belonging to the M-TBFV group, they appear to have retained the ability to circulate among avian species. They do not display the clinal geographical distribution reported for other M-TBFV and share certain characteristics with the

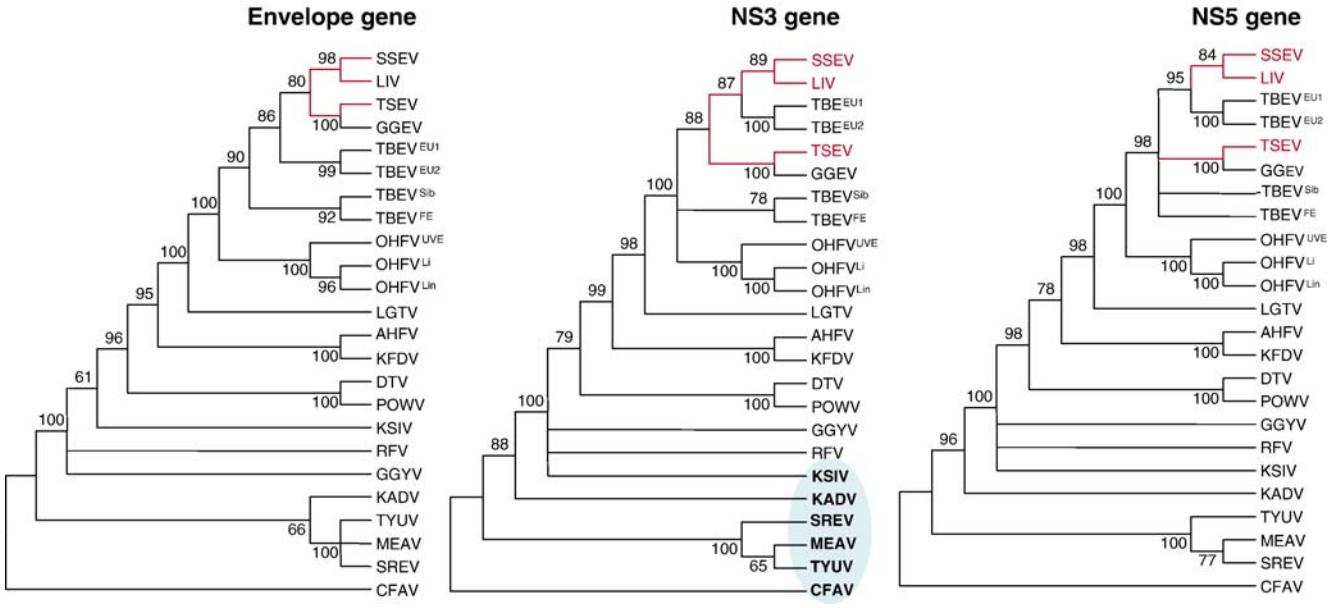


Fig. 4. Phylogenetic analyses in the envelope, NS3 and NS5 complete genes. Phylogenetic reconstruction was performed using the *p*-distance algorithm and the Neighbor-Joining (NJ) method implemented in Mega. Condensed tree are presented in which the length of horizontal lines is not proportional to genetic distances. The % bootstrap values corresponding to 500 replications are indicated with a cut-off value of 60%. Members of the current LIV species are indicated in red. In the NS3 phylogenetic tree, a blue ellipse indicates viruses for which X is not an alanine residue in the helicase DEXH core motif.

seabird-associated flaviviruses, for example, their cladogenetic pattern, their association with soft ticks or birds ticks and the absence of reported associated diseases. They may constitute an evolutionary link between seabird-associated viruses and the more recently emerged mammalian-associated TBFVs, possibly through the infection of ticks, such as certain *Hyalomma* spp., which are associated with avian parasitism during early developmental stages and with mammals during the adult stage (Apanaskevich, 2004; Hoogstraal, 1972).

The seabird tick-borne flaviviruses (*Tyuleniy*, *Saumarez Reef* and *Meaban virus*) are associated with seabird ticks and, with the exception of TYUV, have been isolated only in the Old World. In contrast with the mammalian TBFVs, there is no evidence among seabird TBFVs for a close relationship between genetic evolution and geographic distribution. One can reasonably assume that the seabird TBFVs have been disseminated by independent migratory flights and that their observed genetic evolution reflects adaptation to the different ecological niches they have reached.

Kadam virus is the unique representative of the KADV group. It has been isolated from hard ticks sampled from mammals and has never been associated with seabird ticks. Based on genetic analysis of partial NS5 sequences, it has been suggested that KADV represents an early land based, mammalian-associated, predecessor of the seabird-associated viruses (Gould et al., 2001). However, our data do not preclude that KADV emerged from an ancestor linked to seabirds and could represent the first successful adaptation of a TBFV to mammals and their associated ticks.

A more complete report of the geographical origin, epidemiology and association of the different TBFVs with their respective vectors and hosts is provided as Supplementary data to this study.

NS3 DEXH motif

The analysis of sequences in the NS3 gene revealed an unexpected variability in the highly conserved core motif of the helicase. Up to now, this motif was believed to be DEAH for all flaviviruses except CFAV (DECH) and Tamara bat virus (DESH), which are not arboviruses. The production of sequences for all TBFV species led to the discovery of a DEGH motif for viruses of the seabird and Kadam groups and for KSIV (a member of the M-TBFV group). The mutation does not seem to be linked to the host specificity of the viruses since it is found in both viruses infecting seabirds (TYUV, MEAV and SREV) and mammals (KADV, KSIV) but not in other viruses infecting birds (RFV and GGYV). In phylogenetically related helicase genes (de Lamballerie et al., 2002), the core motif is DECH or DELH in hepaciviruses and the related viruses (GBV-A, GBV-B and GBV-C), DEYH in pestiviruses and DECH, DEFH or DESH in potyviruses. Among these viruses, no precise relationship between the sequence of this motif and specific biological properties has been identified.

Hemorrhagic properties

Among the TBFVs, OHFV, KFDV and AHFV induce hemorrhagic fevers. Examination of the phylogenetic position of OHFV and KFDV/AHFV does not reveal a genetic lineage associated with the ability to produce hemorrhagic disease (Fig. 1b). The clinal distribution of the mammalian TBFVs produces the following alternate sequence of hemorrhagic and non-hemorrhagic viruses: POWV (non-hemorrhagic)–KFDV (hemorrhagic)–LGTV (non-hemorrhagic)–OHFV (hemorrhagic)–TBEV/LIV (non-hemorrhagic). This is also consistent with the observation that viruses responsible for hemorrhagic

fevers exist in distant phylogenetic groups (DENV, YFV) and also with the recent report (Ternovoi et al., 2003) of hemorrhagic viruses in the Far Eastern subtype of TBEV and separately the observation that LIV was isolated from the blood of a laboratory technician suffering from hemorrhagic disease (Cooper et al., 1964).

Previous studies proposed several mutations in the E gene specific for hemorrhagic fever viruses. The analysis of the new OHFV and KFDV sequences confirms the existence of the (T→A) mutation at position E:76. This mutation may be of particular importance (Lin et al., 2003) since the threonine residue is located in a loop that appears to interact with the fusion peptide (Leu 107) (Rey et al., 1995) and possibly with the second E protein in the E protein dimer via a hydrogen bond (Lin et al., 2003). However, the (T→A) mutation also exists in the case of strain D1283 which belongs to the Far Eastern subtype (Hayasaka et al., 2001). D1283 was neuro-pathogenic but was not reported to be hemorrhagic, suggesting that the presence of mutation E:76 is not sufficient to induce hemorrhagic manifestation.

In conclusion, we have proposed significant taxonomic improvements for genus *Flavivirus* such as the identification of three major groups (mammalian, seabird and Kadam tick-borne flavivirus groups), the creation of a new species (*Karshi virus*) and the assignment of Tick-borne encephalitis and Louping ill viruses to a unique species (*Tick-borne encephalitis virus*) including four viral types (i. e. Western Tick-borne encephalitis virus, Eastern Tick-borne encephalitis virus, Turkish sheep Tick-borne encephalitis virus and Louping ill Tick-borne encephalitis virus). In addition, ecological and phylogenetic characteristics of the viruses belonging to the three distinct groups identified suggest that ticks that feed on both mammals and seabirds may constitute the evolutionary bridge between those lineages. Systematic replacement of genes from a non-hemorrhagic virus by the corresponding genes from a closely related hemorrhagic virus may ultimately facilitate identification of the molecular basis for the hemorrhagic characteristic.

Materials and methods

Virus strains and propagation

Abbreviations of virus names are those recommended by the International Committee on Taxonomy of Viruses (ICTV) and are detailed in Table 1. This table also includes the mode of propagation of viruses studied in the current study and their GenBank accession numbers.

Preparation of viral RNAs and cDNAs

OHFV and KFDV

RNA was extracted from either infected mouse brains or infected cell cultures using the RNA NOW™ TC-Kit (Biogentex), resuspended in 50 µl of RNase-free sterile water and stored at −70 °C prior to processing. Reverse transcription (RT) was carried out under standard conditions using the MuMLV Superscript II™ RNase H⁻ Reverse Transcriptase (Life Technologies). Priming on viral RNA was performed using the 3PNC-2R (5′-GCTCAGGGAGAACAAGAACCG-3′) reverse oligonucleotide located in the 3′ non-coding region. The reaction mixture was subsequently treated with DNase-free RNase (Roche Molecular Diagnostics). The resulting non-infectious subgenomic cDNA was received at the Unité des Virus Emergents and further processed.

Other tick-borne flaviviruses

RNA was extracted from either infected mouse brains or infected cell cultures using the QiAmp viral RNA extraction kit (Qiagen). RT was carried out using the CMaster™RTplusPCR System (Eppendorf) with specific primers designed from available data.

Identification of viruses

The identification of all viruses studied (except OHFV) was checked by the amplification and sequencing of a short viral sequence in the NS5 gene, within the region previously characterized (Kuno et al., 1998), using cDNAs prepared as described above and the PF1/PF2 primer set (Crochu et al., 2004). In the case of OHFV, the verification was confirmed by analysis of a region within the envelope (E) gene using specific primers deduced from previously published sequences.

PCR protocols

Characterization of OHFV and KFDV coding sequences

The amplification strategy, primers and sequencing protocols used were the same as reported previously for the genome characterization of AHFV (Charrel et al., 2001) and resulted in the production of 10 (OHFV) and 14 (KFDV) overlapping PCR products that were directly sequenced using the PCR primers and, whenever necessary, additional internal specific primers. Primer sequences and amplification parameters are available upon request to the corresponding author.

PCR amplification in the E region

Primers used were Uni for and Uni rev (positions 1274–1295 and 2230–2252 respectively, of the LIV genome) in the primary PCR, and Uni2 for and Uni2 rev (positions 1298–1323 and 2143–2168 respectively, of the LIV genome) in the nested PCR as described previously (Gaunt and Gould, 2005). PCR amplification was carried out under standard conditions with a polymerization step of 1 min and a hybridization temperature at 50 °C. Sequences previously published were checked using virus-specific primers.

PCR amplification in the NS3 region

A set of degenerate primers enabling PCR amplification in the NS3 region was designed from the alignment of available sequences (Crochu et al., 2004): X1 (5′-YIRTIGGIYTITAYG-GIWWYGG-3′)/X2 (5′-RTTIGCICCCATYTCISHDATRTCIG-

3′) (positions 4913–4935 and 5707–5733 respectively, of the LIV coding sequence) (I: inosine; Y: C/T; R: A/G; H: A/C/T; W: A/T; V: A/C/G; M: A/C). PCR amplification was carried out under standard conditions with a polymerization step of 1 min and a hybridization temperature at 45 °C. Sequences previously published were checked using virus-specific primers.

PCR amplification in the NS5 region

Sequences previously published were checked using virus-specific primers.

Envelope-to-NS3 and NS3-to-NS5 PCR filling

Specific primers in the E, NS3 and NS5 genes were designed from sequences available in the databases and from sequences of the E and NS3 genes. PCR to cover the gaps between E and NS3 sequences on the one hand, and NS3 and NS5 on the other, was performed with the CMaster™RT-plusPCR System (Eppendorf) following manufacturer's recommendations. Sequences were resolved following the Long PCR Product Sequencing (LoPPS) protocol as described elsewhere (Emonet et al., 2006). Briefly, the long PCR products were sheared by ultrasound sonication, termini were repaired using the DNA End repair Kit (Lucigen) and fragments were cloned in pGEMT-vector (Promega) after column purification (Nanosep 100K) and addition of 3′-A overhang with *Taq* DNA polymerase (Invitrogen). 25 to 30 bacterial clones were sequenced with the T7PROM primer, and contigs were constructed from overlapping sequences with the Sequencher 4.5 software (Gene Codes Corporation) using default parameters.

PCR amplification of 5′ and 3′ end of coding sequences

The 5′ ends of the coding sequences were amplified with the CMaster™RTplusPCR System (Eppendorf) following manufacturer's recommendations, using virus-specific reverse primer in the E gene and the primer TB-5′UTR-S (5′-AAAAGACAGCTTAGGAGAACAAGA-3′). The 3′ ends of the coding sequences were amplified using virus-specific forward primer in the NS5 gene and the reverse primer TB-3′ UTR-R (5′-AGAACAAGAACCGCCCCCCC-3′). When necessary, hemi-nested PCR was performed with a second virus-specific primer. Amplified DNA was subsequently cloned with the pGEM-T Vector System (Promega) and sequenced with M13 primers.

Database sequences

Sequences in the NS5 region used for viral identification were as described previously (Kuno et al., 1998). GenBank accession numbers of complete flavivirus coding sequences used for phylogenetic analyses are detailed in Table 1. Database sequences that display a significant identity with the AA sequence of the complete E gene of LIV (strain 369/T2) were identified using BLASTP (protein query–protein database comparison) and subsequently obtained from GenBank.

Evolutionary analysis

Our evolutionary analysis includes the following steps:

(1) Pairwise and multiple alignments of partial or complete AA sequences were generated by the ClustalX (v1.8) program (Thompson et al., 1997) using default parameters. Conserved motifs were used as a control of validity for alignments as previously reported (Billoir et al., 2000). Putative cleavage sites were deduced from AA alignments. Prediction of the cleavage site AnchC/prM by the host signalase was checked using SignalP 3.0 software (Bendtsen et al., 2004). Matrices of AA *p*-distances were calculated with the program MEGA v2.1 (Kumar et al., 2001). The distribution of evolutionary distances upon pairwise comparison was studied.
(2) Phylogenetic trees were estimated using the maximum likelihood method available in the TREE-PUZZLE program (Strimmer and von Haesler, 1996). To determine the best-fit model of AA replacement, the likelihood scores of trees produced by all the six models of AA replacement available in TREE-PUZZLE were compared for the full genome data set, both with equal rates of substitution and with a gamma distribution of rate heterogeneity with a shape parameter (alpha, α) of 1.0. The model that produced the phylogeny with the highest likelihood score for this data set was then used for further analyses. In addition, phylogenetic analyses were conducted using AA alignments, the *p*-distance algorithm and the Neighbor-Joining (NJ) method implemented in Mega. The robustness of branching patterns was tested by 500 bootstrap replications.
(3) Patterns of AA evolutionary change were analyzed, in particular in the envelope and NS3 genes, searching for molecular markers of known biological characteristics.

Acknowledgments

This work was supported in part by the European Community (Scientific program VIZIER, LSHG-CT-2004-511960). The authors are indebted to J-L Camicas for his expertise regarding the biology of arthropods.

References

Allison, S.L., Schalich, J., Stiasny, K., Mandl, C.W., Heinz, F.X., 2001. Mutational evidence for an internal fusion peptide in flavivirus envelope protein E. J. Virol. 75 (9), 4268–4275.

Apanaskevich, D.A., 2004. Host–parasite relationships of the genus Hyalomma Koch, 1844 (Acari, Ixodidae) and their connection with microevolutionary process. Parazitologiia 38 (6), 515–523.

Beasley, D.W., Suderman, M.T., Holbrook, M.R., Barrett, A.D., 2001. Nucleotide sequencing and serological evidence that the recently recognized deer tick virus is a genotype of Powassan virus. Virus Res. 79 (1–2), 81–89.

Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340 (4), 783–795.

Billoir, F., de Chesse, R., Tolou, H., de Micco, P., Gould, E.A., de Lamballerie, X., 2000. Phylogeny of the genus flavivirus using complete coding

sequences of arthropod-borne viruses and viruses with no known vector. J. Gen. Virol. 81 (Pt 3), 781–790.

Calisher, C.H., Karabatsos, N., Dalrymple, J.M., Shope, R.E., Porterfield, J.S., Westaway, E.G., Brandt, W.E., 1989. Antigenic relationships between flaviviruses as determined by cross-neutralization tests with polyclonal antisera. J. Gen. Virol. 70 (Pt 1), 37–43.

Chambers, T.J., Hahn, C.S., Galler, R., Rice, C.M., 1990. Flavivirus genome organization, expression, and replication. Annu. Rev. Microbiol. 44, 649–688.

Charrel, R.N., Zaki, A.M., Attoui, H., Fakeeh, M., Billoir, F., Yousef, A.I., de Chesse, R., De Micco, P., Gould, E.A., de Lamballerie, X., 2001. Complete coding sequence of the Alkhurma virus, a tick-borne flavivirus causing severe hemorrhagic fever in humans in Saudi Arabia. Biochem. Biophys. Res. Commun. 287 (2), 455–461.

Cook, S., Holmes, E.C., 2006. A multigene analysis of the phylogenetic relationships among the flaviviruses (Family: Flaviviridae) and the evolution of vector transmission. Arch. Virol. 151 (2), 309–325.

Cooper, W.C., Green, I.J., Fresh, J.W., 1964. Laboratory Infection with Louping-Ill Virus: a case study. Br. Med. J. 5425, 1627–1630.

Crochu, S., Cook, S., Attoui, H., Charrel, R.N., De Chesse, R., Belhouchet, M., Lemasson, J.J., De Micco, P., De Lamballerie, X., 2004. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes. J. Gen. Virol. 85 (Pt 7), 1971–1980.

de Lamballerie, X., Crochu, S., Billoir, F., Neyts, J., de Micco, P., Holmes, E.C., Gould, E.A., 2002. Genome sequence analysis of Tamana bat virus and its relationship with the genus Flavivirus. J. Gen. Virol. 83 (Pt 10), 2443–2454.

Emonet, S., Grard, G., Brisbarre, N., Moureau, G., Temmam, S., Charrel, R., de Lamballerie, X., 2006. LoPPS: a long PCR product sequencing method for rapid characterisation of long amplicons. Biochem. Biophys. Res. Commun. 344 (4), 1080–1085.

Gao, G.F., Hussain, M.H., Reid, H.W., Gould, E.A., 1993. Classification of a new member of the TBE flavivirus subgroup by its immunological, pathogenetic and molecular characteristics: identification of subgroup-specific pentapeptides. Virus Res. 30 (2), 129–144.

Gao, G.F., Zanotto, P.M., Holmes, E.C., Reid, H.W., Gould, E.A., 1997. Molecular variation, evolution and geographical distribution of Louping ill virus. Acta Virol. 41 (5), 259–268.

Gaunt, M.W., Gould, E.A., 2005. Rapid subgroup identification of the flaviviruses using degenerate primer E-gene RT-PCR and site specific restriction enzyme analysis. J. Virol. Methods 128 (1–2), 113–127.

Gaunt, M.W., Sall, A.A., de Lamballerie, X., Falconar, A.K., Dzhivanian, T.I., Gould, E.A., 2001. Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. J. Gen. Virol. 82, 1867–1876.

Gould, E.A., de Lamballerie, X., Zanotto, P.M., Holmes, E.C., 2001. Evolution, epidemiology, and dispersal of flaviviruses revealed by molecular phylogenies. Adv. Virus Res. 57, 71–103.

Hayasaka, D., Ivanov, L., Leonova, G.N., Goto, A., Yoshii, K., Mizutani, T., Kariwa, H., Takashima, I., 2001. Distribution and characterization of tick-borne encephalitis viruses from Siberia and Far-Eastern Asia. J. Gen. Virol. 82 (Pt 6), 1319–1328.

Heinz, F.X., Allison, S.L., 2003. Flavivirus structure and membrane fusion. Adv. Virus Res. 59, 63–97.

Heinz, F.X., Collett, M.S., Purcell, R.H., Gould, E.A., Howard, C.R., Houghton, M., Moormann, R.J.M., Rice, C.M., Thiel, H.-J., 2000. Family Flaviviridae. In: van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R., Wickner, R.B. (Eds.), Virus Taxonomy—Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, pp. 859–878.

Hoogstraal, H., 1972. Birds as tick hosts and as reservoirs and disseminators of tickborne infectious agents. Wiad. Parazytol. 18 (4), 703–706.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17 (12), 1244–1245.

Kuno, G., Chang, G.J., Tsuchiya, K.R., Karabatsos, N., Cropp, C.B., 1998. Phylogeny of the genus Flavivirus. J. Virol. 72 (1), 73–83.

Kuno, G., Artsob, H., Karabatsos, N., Tsuchiya, K.R., Chang, G.J., 2001. Genomic sequencing of deer tick virus and phylogeny of powassan-related viruses of North America. Am. J. Trop. Med. Hyg. 65 (5), 671–676.

Li, L., Rollin, P.E., Nichol, S.T., Shope, R.E., Barrett, A.D., Holbrook, M.R., 2004. Molecular determinants of antigenicity of two subtypes of the tick-borne flavivirus Omsk haemorrhagic fever virus. J. Gen. Virol. 85 (Pt 6), 1619–1624.

Lin, D., Li, L., Dick, D., Shope, R.E., Feldmann, H., Barrett, A.D., Holbrook, M.R., 2003. Analysis of the complete genome of the tick-borne flavivirus Omsk hemorrhagic fever virus. Virology 313 (1), 81–90.

Mandl, C.W., Holzmann, H., Kunz, C., Heinz, F.X., 1993. Complete genomic sequence of Powassan virus: evaluation of genetic elements in tick-borne versus mosquito-borne flaviviruses. Virology 194 (1), 173–184.

Marin, M.S., McKenzie, J., Gao, G.F., Reid, H.W., Antoniadis, A., Gould, E.A., 1995a. The virus causing encephalomyelitis in sheep in Spain: a new member of the tick-borne encephalitis group. Res. Vet. Sci. 58 (1), 11–13.

Marin, M.S., Zanotto, P.M., Gritsun, T.S., Gould, E.A., 1995b. Phylogeny of TYU, SRE, and CFA virus: different evolutionary rates in the genus Flavivirus. Virology 206 (2), 1133–1139.

McGuire, K., Holmes, E.C., Gao, G.F., Reid, H.W., Gould, E.A., 1998. Tracing the origins of louping ill virus by molecular phylogenetic analysis. J. Gen. Virol. 79 (Pt 5), 981–988.

Nowak, T., Wengler, G., 1987. Analysis of disulfides present in the membrane proteins of the West Nile flavivirus. Virology 156 (1), 127–137.

Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., Harrison, S.C., 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. Nature 375 (6529), 291–298.

Rice, C.M., Lenches, E.M., Eddy, S.R., Shin, S.J., Sheets, R.L., Strauss, J.H., 1985. Nucleotide sequence of Yellow fever virus: implications for flavivirus gene expression and evolution. Science 229 (4715), 726–733.

Roehrig, J.T., Hunt, A.R., Johnson, A.J., Hawkes, R.A., 1989. Synthetic peptides derived from the deduced amino acid sequence of the E-glycoprotein of Murray Valley encephalitis virus elicit antiviral antibody. Virology 171 (1), 49–60.

Shiu, S.Y., Ayres, M.D., Gould, E.A., 1991. Genomic sequence of the structural proteins of louping ill virus: comparative analysis with tick-borne encephalitis virus. Virology 180 (1), 411–415.

Shiu, S.Y., Jiang, W.R., Porterfield, J.S., Gould, E.A., 1992. Envelope protein sequences of dengue virus isolates TH-36 and TH-Sman, and identification of a type-specific genetic marker for dengue and tick-borne flaviviruses. J. Gen. Virol. 73 (Pt 1), 207–212.

Strimmer, K., von Haesler, A., 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Ternovoi, V.A., Kurzhukov, G.P., Sokolov, Y.V., Ivanov, G.Y., Ivanisenko, V.A., Loktev, A.V., Ryder, R.W., Netesov, S.V., Loktev, V.B., 2003. Tick-borne encephalitis with hemorrhagic syndrome, Novosibirsk region, Russia, 1999. Emerg. Infect. Dis. 9 (6), 743–746.

Thiel, H.-J., Collett, M.S., Gould, E.A., Heinz, F.X., Meyers, G., Purcell, R.H., Rice, C.M., Houghton, M., 2005. Flaviviridae. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), Virus Taxonomy—Eighth Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, pp. 981–998.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Venugopal, K., Gritsun, T., Lashkevich, V.A., Gould, E.A., 1994. Analysis of the structural protein gene sequence shows Kyasanur Forest disease virus as a distinct member in the tick-borne encephalitis virus serocomplex. J. Gen. Virol. 75 (Pt 1), 227–232.

von Heijne, G., 1984. How signal sequences maintain cleavage specificity. J. Mol. Biol. 173 (2), 243–251.

Zanotto, P.M., Gao, G.F., Gritsun, T., Marin, M.S., Jiang, W.R., Venugopal, K., Reid, H.W., Gould, E.A., 1995. An arbovirus cline across the northern hemisphere. Virology 210 (1), 152–159.

Zanotto, P.M., Gould, E.A., Gao, G.F., Harvey, P.H., Holmes, E.C., 1996. Population dynamics of flaviviruses revealed by molecular phylogenies. Proc. Natl. Acad. Sci. U. S. A. 93 (2), 548–553.