Arch Virol (2003) 148: 1095–1118
DOI 10.1007/s00705-003-0019-7

Archives of Virology
Printed in Austria

Genetic and phenotypic characterization of the newly described insect flavivirus, Kamiti River virus

M. B. Crabtree[1], **R. C. Sang**[2], **V. Stollar**[3], **L. M. Dunster**[2], and **B. R. Miller**[1]

[1]Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Fort Collins, Colorado, U.S.A.
[2]Centre for Virus Research, Kenya Medical Research Institute, Nairobi, Kenya
[3]Department of Molecular Genetics and Microbiology, UMDNJ-Robert Wood Johnson Medical School, Piscataway, New Jersey, U.S.A.

Received October 23, 2003; accepted January 6, 2003
Published online April 9, 2003

Summary. We have described in the accompanying paper by Sang, et al., ([57], Arch Virol 2003: 1085–1093) the isolation and identification of a new flavivirus, Kamiti River virus (KRV), from *Ae. macintoshi* mosquitoes that were collected as larvae and pupae from flooded dambos in Central Province, Kenya. Among known flaviviruses, KRV was shown to be most similar to, but genetically and phenotypically distinct from, Cell fusing agent virus (CFAV). KRV was provisionally identified as an insect-only flavivirus that fails to replicate in vertebrate cells or in mice. We report here the further characterization of KRV. Growth in cell culture was compared to that of CFAV; although growth kinetics were similar, KRV did not cause the cell fusion that is characteristic of CFAV infection. The KRV genome was found to be 11,375 nucleotides in length, containing a single open reading frame encoding 10 viral proteins. Likely polyprotein cleavage sites were identified, which were most similar to those of CFAV and were comparable to those of other flaviviruses. Sequence identity with other flaviviruses was low; maximum identity was with CFAV. Possible terminal secondary structures for the 5′ and 3′ non-coding regions (NCR) were similar to those predicted for other flaviviruses. Whereas CFAV was isolated from insect cells in the laboratory, the isolation of KRV demonstrates the presence of an insect-only flavivirus in nature and raises questions regarding potential interactions between this virus and other mosquito-borne viruses in competent vector populations. Additionally, this virus will be an important tool in future studies to determine markers associated with flavivirus host specificity.

Introduction

Kamiti River virus (KRV) is an insect-only virus that was recently isolated in Kenya (Sang et al., Arch Virol 2003: 1085–1093 [57]). Based on preliminary sequence information, KRV was provisionally determined to be a flavivirus and was found to be most closely related to the only other known insect-only flavivirus, Cell fusing agent virus (CFAV). The genus *Flavivirus*, family *Flaviviridae*, contains a large number of single-stranded, positive-sense RNA viruses, many of which are associated with human disease. Most known flaviviruses are arthropod-borne, being transmitted to vertebrate hosts by a mosquito or tick, although there are members of the genus that have a vertebrate host but no known arthropod vector. CFAV has no known vertebrate host; the virus replicates in *Aedes albopictus* mosquitoes and in cultured mosquito cells but does not replicate in cultured vertebrate cells [23, 46].

CFAV was isolated from laboratory-cultured *Ae. aegypti* mosquito cells and has, until recently, been the only known insect-only flavivirus [46]. There has been no immunologic relationship demonstrated between CFAV and other known flaviviruses [23]. However, the genome organization of CFAV is similar to that of other flaviviruses and although sequence identity is low between the structural proteins of CFAV and other flaviviruses, there is considerable sequence identity in the nonstructural (NS) 3 and NS5 proteins [8]. Phylogenetic studies of the genus *Flavivirus* suggest that CFAV may represent a basal lineage of the genus and that it diverged from the other flaviviruses before the separation of the mosquito and tick-borne groups [8, 28, 32].

KRV is the first insect-only flavivirus to be isolated from nature; the virus was isolated from field-collected, immature mosquitoes collected from a flooded dambo in Central Province, Kenya (Sang et al. [57]). The discovery of this virus raises questions regarding whether the presence in mosquitoes of KRV can interfere with superinfection by other arboviruses and, if so, what effect this interference might have on the cycle of arbovirus transmission. The isolation of KRV also presents an opportunity to explore the nature of flavivirus host specificity and to further characterize the evolutionary relationships among viruses within the genus.

We report here the genetic and phenotypic characterization of KRV. KRV growth in cell culture was found to be similar to that of CFAV, although KRV does not cause the formation of massive syncytia characteristic of mosquito cell infection with CFAV [46]. The complete KRV genome was sequenced and nucleotide (nt) and deduced amino acid (aa) sequences were analyzed and compared with those of other flaviviruses. The KRV genome was found to be single-stranded, positive sense RNA, 11,375 nucleotides (nt) in length, containing a single, long open reading frame (ORF) that codes for three structural and seven non-structural proteins. Sequence identity with other flaviviruses was very low, particularly at the nucleotide level, although most conserved flavivirus sequence motifs were also conserved in KRV. Maximum identity at both nt and aa levels was with CFAV. Highest identity was seen between amino acid sequences of the NS3 and NS5

proteins. Likely polyprotein cleavage sites were identified, which, except for the anchored capsid/virion capsid (anchC/virC) site, were comparable to those of other flaviviruses and, with the exception of the premembrane/membrane (pr/M) cleavage site, were most similar to those of CFAV. Possible terminal secondary structures for the 5′ and 3′ non-coding regions (NCR) were similar to those predicted for other flaviviruses and potential cyclization sequences in the 5′ and 3′ NCRs were identified.

Materials and methods

Viruses and cells

Two isolates of KRV, SR-75 and SR-82, were isolated from *Ae. macintoshi* mosquitoes as previously described (Sang et al., Arch Virol 2003: 1085–1093 [57]). CFAV, (H-9-1, passage 10) was obtained from Victor Stollar, UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ [46]. Cells used in this study include *Ae. albopictus* C6/36-ATCC (American Type Culture Collection, Manassas, VA), C6/36-H (Hawaii strain), AP-61 (*Ae. pseudoscutellaris*), and *Ae. aegypti* mosquito cells, and BHK-21 (baby hamster kidney), clone 13, Vero (African green monkey kidney), and LLC-MK_2 (Rhesus monkey kidney) mammalian cells. Mosquito and vertebrate cells were maintained at 28 °C and 37 °C, respectively. Purified dengue virus type 2 (DENV-2), strain 16681 (purified from infected LLC-MK_2 cells), was generously provided by Dr. Richard Kinney, Centers for Disease Control and Prevention, Fort Collins, CO [55].

Characterization in cell culture

Growth characteristics of the KRV SR-75 and SR-82 virus isolates and CFAV in C6/36-ATCC and C6/36-H cells were compared. C6/36 cells were infected at a multiplicity of infection (m.o.i.) of 0.1 and samples were removed at two-day intervals for 10 days. Virus titers were assayed on C6/36 cell monolayers in six-well plates using a double-overlay method in media without antifungal additives (fungizone) with SeaPlaque low-melting temperature agarose (BioWhittaker, Rockland, MD) [34]. Second overlays containing neutral red were added at four days or seven days for CFAV and KRV viruses, respectively. Growth of the KRV SR-75 virus isolate was also assayed in AP-61 and *Ae. aegypti* mosquito cells.

Virus purification and analysis of viral proteins

C6/36 cells were infected with virus at a m.o.i. of 10 for virus purification. Cell culture supernatant was harvested 3 or 5 days after infection with CFAV and KRV, respectively, when there was significant evidence of cytopathic effects (CPE) in the cell monolayer. Virus was purified by using the method of Obijeski et al. [37]. Purified viral protein was quantitated with the BCA Protein Assay kit (Pierce Chemical Co., Rockford, IL) and subjected to polyacrylamide gel electrophoresis (PAGE) on a NuPAGE Novex 4–12% Bis–Tris gel with MES buffer (Invitrogen/Life Technologies, Baltimore, MD). Protein samples were reduced by heating at 70 °C for 10 minutes in LDS Sample Buffer with NuPage Reducing Agent (Invitrogen/Life Technologies) prior to electrophoresis. The glycosylation status of viral proteins was analyzed by peptide:N-glycosidase F (PNGaseF) deglycosylation prior to electrophoresis using the GlycoShift kit (Oxford GlycoSystems, Rosedale, NY). Following electrophoresis gels were stained with the SilverQuest Silver Staining kit (Invitrogen/Life Technologies) and air-dried. Expected molecular weights (mw) of the viral proteins were predicted on the basis of their aa

sequences and potential for glycosylation and compared with mw estimates determined by analysis of stained PAGE gels using a GS710 Calibrated Imaging Densitometer and Quantity One quantitation software (Bio Rad Laboratories, Hercules, CA).

Nucleic acid sequencing

Viruses were amplified in C6/36 cells and viral RNA was extracted from cell culture supernatant using the QIAamp Viral RNA Mini-Kit (Qiagen, Valencia, CA). RT-PCR was conducted using the Titan One Tube RT-PCR System (Roche, Indianapolis, IN); amplified products were purified by agarose gel electrophoresis followed by extraction of DNA fragments using the QiaQuick Gel Extraction Kit (Qiagen). DNA was eluted in 50 μl of 10 mM Tris-HCL, pH 8.5, and stored at −20 °C. Purified DNA fragments were sequenced using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit with Amplitaq DNA Polymerase FS (PE Applied Biosystems, Foster City, CA) and analyzed using a Model 377 PRISM automated DNA sequencer (PE Applied Biosystems). Both strands of DNA fragments were sequenced.

The full-length genomes of the KRV SR-75 and SR-82 virus isolates were sequenced as described above beginning with fragments amplified by the FU2/CFD3 flavivirus universal primer pair (NS5 gene region) [28]. Sequence obtained from this amplified fragment was used to design primers to amplify and determine the sequence of adjacent fragments, which were then used to design further primers, and so on. The viral 5′ and 3′ terminal sequences were determined by using the 5′3′ RACE Kit (Roche).

Genome characterization

The nt sequence of the KRV SR-75 isolate genome was analyzed for open reading frames (ORF) using the EditSeq module of Lasergene (DNASTAR, Inc., Madison, WI) and translated into aa sequence. This deduced aa sequence was analyzed for protein cleavage sites by comparison with other flaviviruses and published signalase-type cleavage consensus sequences to determine the boundaries of each of the proteins in the polyprotein sequence [41, 49]. The aa sequences were also analyzed for the presence of potential N-linked glycosylation (N-GLY) sites and cysteine (CYS) residues; codon usage patterns were determined by using the Codon Frequency program in the Wisconsin Package, version 10.2 [13]. The nt sequences of the 5′ and 3′ noncoding regions (NCR) were analyzed for possible secondary structures (RNA mfold, ver. 3.0, fold temperature = 28 °C, 5% suboptimality, http://bioinfo.math.rpi.edu/~mfold/rna/form1.cgi) and repeat sequences (GeneQuest module, Lasergene ver. 5.03, DNASTAR, Inc., Madison, WI) [33, 56]. Additionally, to investigate possible genome cyclization structures formed by interactions between the 5′ and 3′ NCRs, sequences from the 5′ 200 nt and 3′ 400 nt of the genome were joined with a 50 nt poly A spacer and analyzed by RNA *mfold* as described above.

Comparison with other flaviviruses

The aa sequences of the individual viral proteins of the KRV SR-75 isolate were compared in a pairwise fashion to those of other flaviviruses including CFAV (M91671); DENV-2, strain 16681 (U87411); *Yellow fever virus* (YFV), Asibi strain [20]; *Japanese encephalitis virus* (JEV), strain SA-14 (U14163); *Tick-borne encephalitis virus* (TBEV), Vasilchenko strain (M27157); and *Modoc virus* (MODV), (AJ242984) using the GAP program in the Wisconsin Package. In addition, aa alignments of the premembrane/envelope (prM/E) protein region and the NS5 protein from the KRV SR-75 and SR-82 isolates, the six flaviviruses listed above and *Langat virus* (LGTV), strain TP21 (AF253419), and *Apoi virus* (APOIV) (AF160193) were conducted using the PILEUP program in the Wisconsin Package. The prM/E and NS5 alignments were then analyzed using the maximum parsimony method with 500 bootstrap

replicates in the program PAUP, version 4.08b, to produce phylogenetic trees showing the relationships between the viruses [47]. Trees were presented as unrooted phylograms in order to avoid assumptions of ancestry. Comparisons of hydrophobicity plots of aa sequences from KRV, CFAV, YFV, and TBEV were conducted using the PepPlot program in the Wisconsin Package.

Results and discussion

Growth of KRV in cell culture

The SR-75 and SR-82 isolates of KRV replicated in both C6/36-H and C6/36-ATCC cells and caused CPE, including cell rounding, detachment from flask, and death, in C6/36-H cells, but did not cause the formation of syncytia as does CFAV (Fig. 1). CFAV caused CPE, including massive syncytium formation, in both C6/36 strains. Results of 10-day growth curves in C6/36-H cells are shown in Fig. 2. The SR-75 and SR-82 KRV isolates replicated to higher titers in C6/36-H cells than did CFAV; maximum titers of 8.3 and 8.4 $\log_{10}$ plaque forming units (PFU)/ml were achieved for SR-75 and SR-82, respectively, compared to 7.4 $\log_{10}$ PFU/ml for CFAV. Comparable titers were observed in C6/36-ATCC cells. Although the time required for replication to maximum titer was similar (4–6 days post infection), onset of CPE was more rapid in cells infected with CFAV, becoming apparent by 48 hours post infection. CPE in C6/36-H cells infected with the KRV isolates was

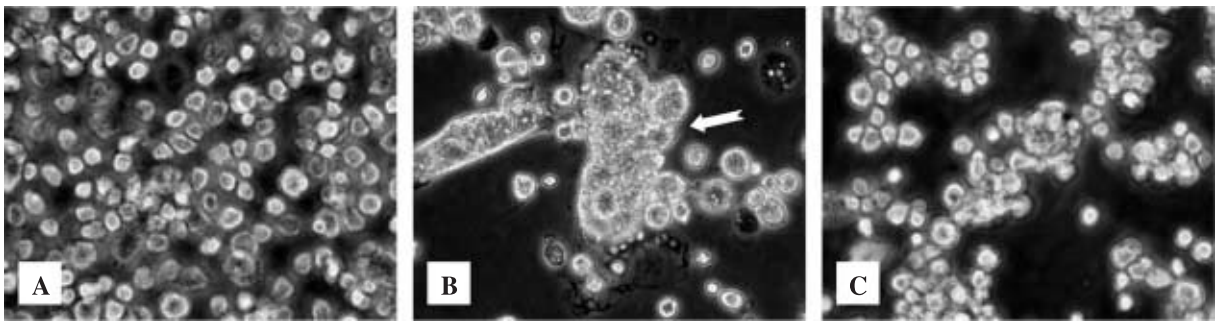

Fig. 1. C6/36-H cells. **A**, mock-infected with virus-free media, day 7; **B**, CFAV-infected, day 5, arrow indicates syncitium; **C**, KRV-infected, day 7

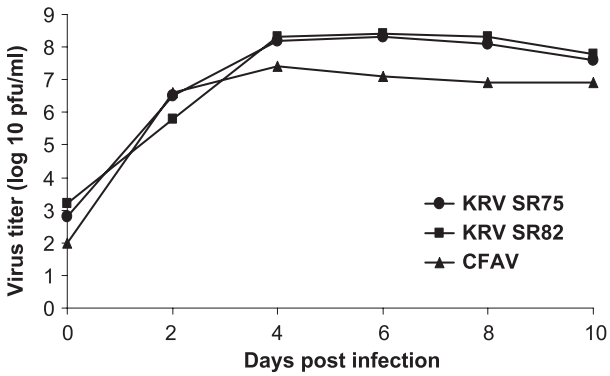

Fig. 2. Ten-day growth curves in C6/36-H cells. KRV, Kamiti River virus; CFAV, Cell fusing agent virus

not observed until approximately 6–7 days post infection. Plaque morphology of the KRV isolates in both C6/36 cell strains was similar to that observed with CFAV: diffuse plaques of 2–3 mm in diameter were observed six days following infection with CFAV and 10 days following infection with the KRV isolates. The KRV isolates also replicated in AP-61 (6.5 $\log_{10}$ PFU/ml) and *Ae. aegypti* (5.9 $\log_{10}$ PFU/ml) cells in culture, however, KRV did not cause CPE in either of these mosquito cell types.

Sequence analysis of the KRV genome

The complete genomes of both of the SR-75 and SR-82 KRV isolates were sequenced. The full-length genome was found to be 11,375 nucleotides in length (GenBank accession no. AY149904 and AY149905, for SR-75 and SR-82, respectively), which is significantly longer than any other flavivirus genome, and contained a single, long ORF of 10,071 nucleotides. The nt sequence of the ORF was translated into aa sequence; the polyprotein sequence was 3,357 aa in length. The nt and aa sequences of the two KRV isolates were found to be nearly identical, with 55 nucleic acid differences and one aa difference (NS2B aa position 8, SR-75-Ile to SR-82-Thr) between them (99.5 and 99.9% sequence identity, respectively).

Based on comparison of the KRV polyprotein aa sequence with other flaviviruses the organization of the viral genome was found to be the same: 5′ non-coding region (5′NCR), followed by three structural proteins, capsid (C), premembrane/membrane (pr/M) and envelope (E), seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5), and a 3′ non-coding region (3′NCR).

The deduced aa sequence was analyzed, using known protein cleavage consensus patterns and comparison with other flaviviruses, to determine the sites of proteolytic cleavage of the polyprotein into the 10 flavivirus proteins. The proposed cleavage sites are listed and described in Table 1. The first AUG initiation codon in the nt sequence begins at position 22. However, this codon exists in an unfavorable context for initiation, due to the presence of a T residue in both positions −3 upstream from the AUG and +4 downstream, and is quickly followed by two termination codons [27]. The second initiation codon starts at nt position 97. This codon is in-frame with the polyprotein ORF and sits in good context for initiation with a G residue in position −3 preceded by a C in position −4. Although two subsequent AUG codons exist in the first 260 nt of the KRV sequence, both are situated in weaker context for initiation and aa sequence alignments between KRV, CFAV and other flaviviruses suggest that the AUG-97 is the start of the coding region for the polyprotein as well as for the C protein.

It is likely that the cleavage pattern of the KRV polyprotein into the 10 viral proteins is similar to that of the other flaviviruses [8, 10, 41]. The suggested C/prM, prM/E, E/NS1, and NS4A/NS4B cleavage sites conform to predicted sequence patterns for signalase cleavage while the NS1/NS2A cleavage site conforms only partially to a signalase sequence; the −3 and −1 residues conform yet the upstream hydrophobic sequence usually seen with signalase cleavage is absent [49].

Table 1. Summary of suggested cleavage sites of the KRV polyprotein

	KRV		CFAV[a]	Other flaviviruses
	Cleavage site	Cleavage type		
anchC/virionC	RLEKQR↓SGPNL	after QR	after dibasic residues	after dibasic residues
C/prM	LGLCYG↓EMLRY	signalase	signalase[b]	signalase
pr/M	QVRRRR↓APQPQ	after dibasic residues	?[c]	after dibasic residues (lumen)
prM/E	WNVVKA↓SSIEP	signalase	signalase[b]	signalase
E/NS1	VRSVSA↓DVGCG	signalase	signalase	signalase
NS1/NS2A	YGKAHA↓CSDFR	novel signalase[d]	novel signalase	novel signalase
NS2A/NS2B	WAAERA↓QQPTI	after RA	after KA	after dibasic residues KR or RR
NS2B/NS3	LSEQNR↓SDDLL	after NR	after NR	after dibasic residues or after QR
NS3/NS4A	EWDTRK↓LSIEF	after dibasic residues	after dibasic residues	after dibasic residues
NS4A/NS4B	VCGVLA↓WEMRL	signalase	signalase	signalase
NS4B/NS5	FNQFRA↓LEKST	after RA	after RA	after dibasic residues KR or RR

[a] Suggested cleavages unless otherwise noted [8]
[b] Confirmed by amino acid sequencing [8]
[c] CFAV results indicate inefficient or absent cleavage of pr/M [8]
[d] Novel signalase = satisfies (−1, −3) rule for signalase cleavage, but not the requirement for an upstream hydrophobic sequence

Six flavivirus polyprotein cleavages (anchC/virC, pr/M, NS2A/NS2B, NS2B/NS3, NS3/NS4A, and NS4B/NS5) occur following a pair of basic aa residues (RR or KR) that precede a short side chain aa residue (G or S). These cleavages have been shown to be mediated by the serine protease domain of the viral NS3 in conjunction with NS2B [11, 15]. The suggested cleavage of CFAV varies at all but two of these sites.

In the KRV C protein, as in other flaviviruses, several dibasic aa pairs are present in the region 20–30 residues upstream from the carboxy terminus where the anchC/virC cleavage site is located. None of these dibasic pairs, however, is followed by a short side chain residue nor do they immediately precede the highly hydrophobic membrane anchor region as seen in other flaviviruses. Evidence from aa alignments with other flaviviruses along with hydrophobicity analysis suggests that the most likely KRV anchC/virC cleavage site follows a QR pair (neutral/basic) preceding a S residue that is immediately followed by a hydrophobic region of 19 aa that would constitute the hydrophobic membrane anchor of the protein. Polyacrylamide gel electrophoresis analysis of KRV structural proteins resulted in a strongly-staining band that corresponded to the expected mw of anchC (∼15 kDa) and a weakly-staining band that correlated with the expected mw of the virC protein (∼13 kDa) (Fig. 3). These results suggested that anchC is the major form present in the mature KRV virion and that anchC to virC cleavage

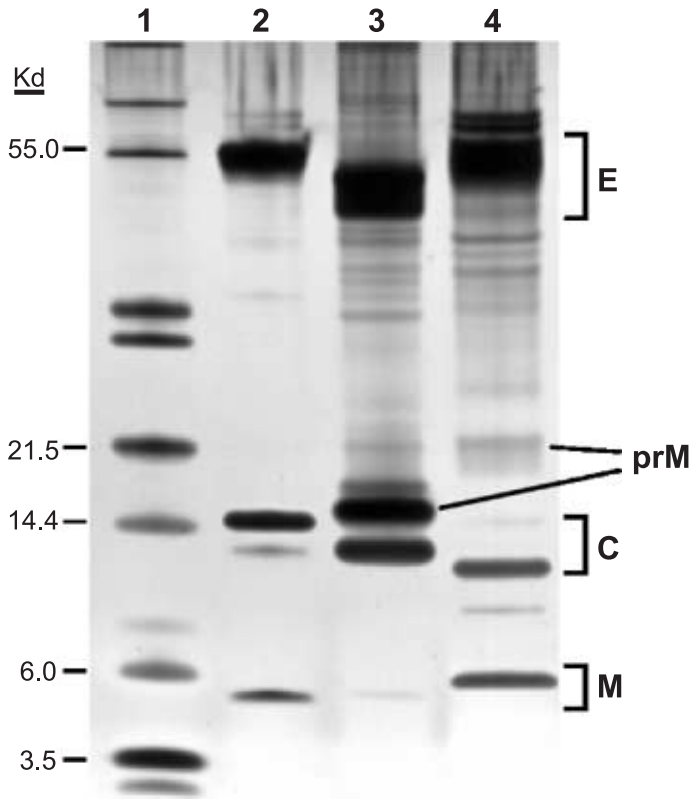

Fig. 3. Polyacrylamide gel electrophoresis of purified viral proteins. *1*, molecular weight marker (Magic Mark, Invitrogen/Life Technologies); *2*, Kamiti River virus; *3*, Cell fusing agent virus; *4*, Dengue virus type 2

occurs inefficiently in KRV. This may be due to the lack of a dibasic aa pair at the putative cleavage site, although QR as well as other monobasic aa pairs have been shown to be flaviviral serine protease cleavage sites at other locations in the polyprotein of other flaviviruses (see NS2A/NS2B and NS2B/NS3 below) [10]. Further research will be necessary to elucidate the factors affecting the efficiency of the KRV anchC/virC cleavage and any effect it may have on the virus.

The potential KRV pr/M cleavage motif (V-R-R-R-R ↓ A) is similar to that found in most flaviviruses (R-R-S/T-R-R ↓ S/A), with substitution of a V for R at the −5 position and an R for S/T at the −3 position. The CFAV aa sequence in this region (K-R-E-K-R ↓ S) is identical to the DENV-2 motif with the exception of K for R at the −5 position; the negatively charged E at position −3 is seen only in the dengue viruses and CFAV. Recent mutational analysis of the DENV-2 pr/M cleavage motif indicates that cleavage is enhanced by changing the DENV-2 −3 position from E to S, as is found in most other flaviviruses, suggesting that an E residue at this position may reduce cleavage efficiency (J. T. Roehrig, personal communication). Amino-terminal sequencing of CFAV structural proteins suggests that processing of prM to M does not occur or occurs very inefficiently [8].

This was confirmed by PAGE analysis of purified CFAV that showed a strongly staining band of approximately 16 kDa that correlated with the expected mw of the CFAV prM protein (Fig. 3). A very faint band consistent with the expected mw of the CFAV M protein (∼5 kDa) was also observed. The pr/M cleavage motif found in KRV is more similar to that of most other flaviviruses than to CFAV. PAGE analysis of KRV structural proteins resulted in a strong M band (mw ∼5 kDa) (Fig. 3) with a faint band representing a protein of the expected mw for prM (∼16 kDa) that was only observed when a very high concentration of viral protein was electrophoresed (data not shown). These results suggest that, as with other flaviviruses, KRV prM is efficiently cleaved to M; in this respect it differs significantly from CFAV.

The proposed NS2A/NS2B cleavage in KRV follows RA, rather than a dibasic aa pair (KR or RR) as in most other flaviviruses; NS2A/NS2B cleavage in CFAV follows KA. Similarly, cleavage at this site in APOIV follows a pair containing a single basic aa, QR. KRV NS2B/NS3 cleavage follows a single basic aa (NR), as in CFAV, while in most other flaviviruses it follows a dibasic aa pair or QR or QK, as in DENV-2 and APOIV, respectively. NS3/NS4A cleavage in KRV follows a pair of basic aa residues as seen in CFAV and other flaviviruses. NS4B/NS5 cleavage in KRV occurs after RA with one basic aa residue; cleavage at this site in CFAV also follows RA whereas in other flaviviruses it occurs following a pair of basic residues (KR or RR).

Codon usage in the KRV polyprotein aa sequence was non random, as has been demonstrated for other flaviviruses including YFV and CFAV. The Gly codon GGA is used in KRV more than twice as often as GGG, GGU or GGC. Similarly, the GUG codon for Val is used more than three times as often as the GUA codon. The frequency of these codons are similarly nonrandom in YFV and CFAV [8, 40]. CUC is the most frequently used codon for Leu and is used more than five times as often as the UUA codon. CUC is also the most frequently used Leu codon in CFAV; in YFV the CUG codon for Leu has the highest frequency.

Biases in dinucleotide frequency have also been demonstrated in flavivirus coding region sequences, including a deficiency of CpG doublets and an excess of UpG doublets compared to expected levels predicted from base composition [24]. This bias has been observed in flaviviruses from the no-known-vector, mosquito/vertebrate, and tick/vertebrate groups and is similar to what has been documented for vertebrate DNA [4, 24]. For these three virus groups, mean ratios of actual/expected CpG are 0.32 ± 0.04, 0.47 ± 0.09 and 0.55 ± 0.03 and UpG are 1.48 ± 0.01, 1.42 ± 0.04 and 1.45 ± 0.02, respectively. Whereas variation is observed in dinucleotide frequencies between these viruses, there is no covariance with arthropod association [24]. In contrast, the CpG and UpG doublet frequencies in KRV and CFAV are closer to levels predicted from the nt composition, as seen in invertebrate DNA. In these viruses, ratios of actual/expected CpG were 0.77 and 0.82, and actual/expected UpG ratios were 1.33 and 1.24, respectively. Therefore, it would appear that flaviviruses that replicate in vertebrate hosts have, to some extent, evolved to include dinucleotide frequencies that are representative of vertebrate cell types and that dinucleotide frequencies characteristic of

invertebrate cells are retained only in viruses that replicate exclusively in invertebrate hosts.

Comparison of KRV proteins with other flaviviruses

Amino acid sequences of the 10 KRV proteins were individually aligned in a pairwise fashion with those of a representative group of six other flaviviruses. Table 2 lists the aa length of each protein encoded by KRV, CFAV, DENV-2, YFV, JEV, TBEV and MODV along with the percent sequence identity between the KRV proteins and those of the other flaviviruses listed. For all 10 proteins, sequence identity between KRV and CFAV proteins (35.4–81.3% identity) was higher than between KRV proteins and those of other flaviviruses (14.8–47.6%). This was especially notable with the nonstructural proteins. For example, whereas KRV NS5 showed 81.2% sequence identity to CFAV NS5, it showed only 45.7–47.6% identity to NS5 of the other flaviviruses in the comparison. Overall, the greatest identity was observed for the NS5 and NS3 protein sequences. As expected, in comparing KRV and CFAV, sequence identity was much greater for the nonstructural proteins (58.1–81.3%) than for the structural proteins (35.4–39.3%).

Phylogenetic analysis was conducted on alignments of prM-E and NS5 aa sequences from KRV, CFAV and other flaviviruses. These regions were selected in order to provide a comparative analysis of both highly conserved nonstructural and less conserved structural sequences. In accordance with the pairwise sequence

Table 2. Length and percent identity comparisons between protein sequences from KRV and other flaviviruses[a]

Vector/host:	Mosquito			Mosquito/vertebrate						Tick/vertebrate		NKV[b]/rodent	
	KRV[c]	CFAV		DENV-2		YFV		JEV		TBEV		MODV	
Protein	aa[d]	aa	%[e]	aa	%	aa	%	aa	%	aa	%	aa	%
C	143	136	35.4	114	25.2	121	25.2	127	25.2	112	27.7	110	30.1
prM	143	142	39.3	166	20.7	164	25.0	167	24.3	168	29.1	162	22.1
E	432	427	36.2	495	19.8	493	22.3	500	21.9	496	22.6	482	23.4
NS1	390	390	73.6	352	23.8	409	27.3	415	27.0	352	29.0	353	25.1
NS2A	232	232	68.1	218	20.7	167	21.8	227	23.3	230	28.0	221	20.1
NS2B	124	124	58.1	130	21.9	130	23.9	131	14.8	131	15.0	132	22.5
NS3	577	577	81.3	618	39.6	623	36.0	619	38.3	621	36.8	618	37.9
NS4A	168	168	58.9	150	23.0	287	27.4	149	25.7	149	24.0	144	24.6
NS4B	261	258	67.2	248	18.3	250	17.5	255	21.5	252	20.2	254	17.1
NS5	887	887	81.2	935	47.6	905	46.3	905	46.8	903	46.1	898	45.7

[a]Pairwise alignments were constructed by using the GAP program, Wisconsin Package, Version 10.2
[b]No-known-vector
[c]KRV isolate SR-75
[d]Amino acid sequence length
[e]Percent identity value, gaps excluded, calculated based on alignments

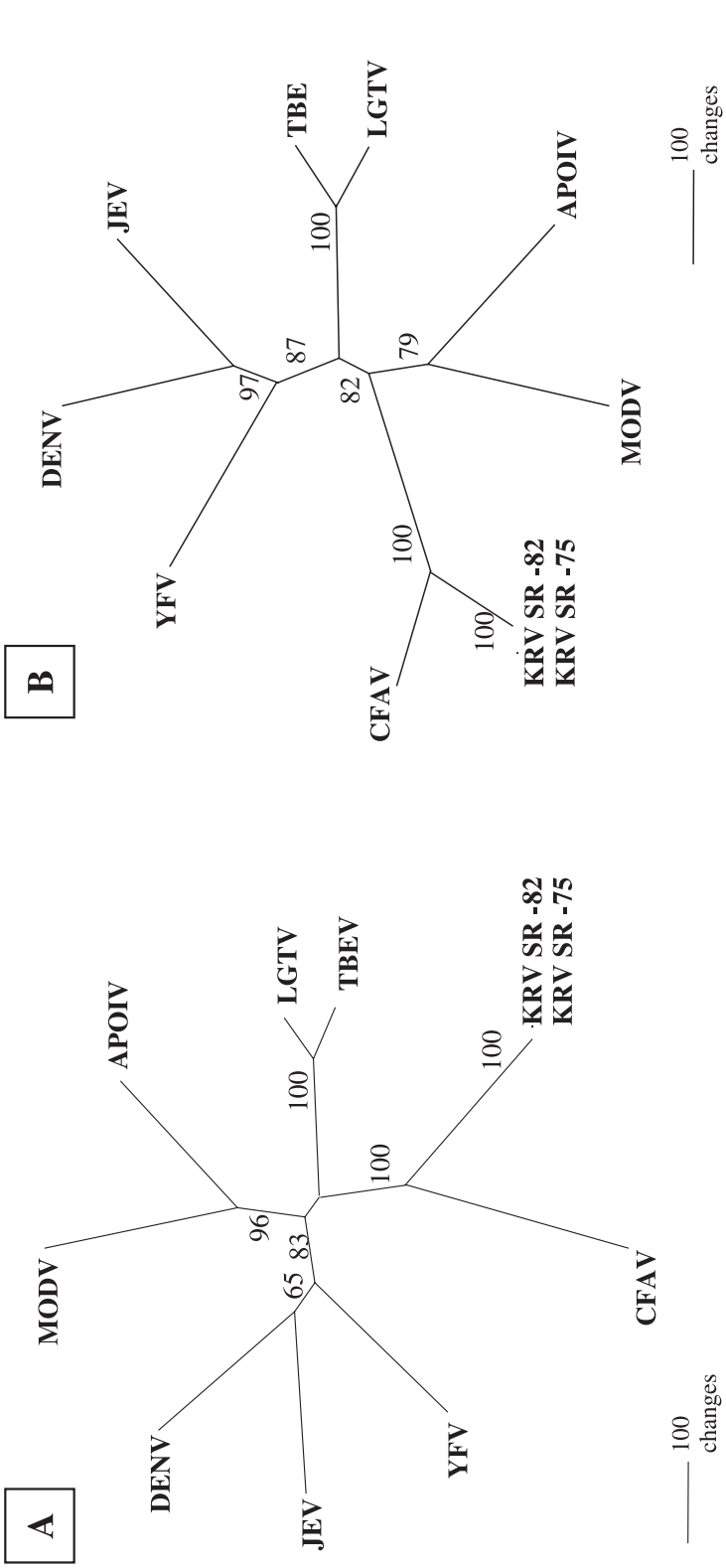

Fig. 4. Unrooted phylogenetic trees generated by maximum parsimony analysis (PAUP) of aligned (**A**) prM and E aa sequences and (**B**) NS5 aa sequences. KRV, Kamiti River virus; CFAV, Cell fusing agent virus; APOIV, *Apoi virus*; MODV, *Modoc virus*; LGTV, *Langat virus*; TBEV, *Tick-borne encephalitis virus*; DENV, *Dengue virus*; JEV, *Japanese encephalitis virus*; YFV, *Yellow fever virus*. Flavivirus species groups represented include the Mosquito (CFAV and KRV), No-Known-Vector (APOIV and MODV), Mosquito/Vertebrate (JEV, DENV and YFV) and Tick/Vertebrate (TBEV and LGTV). Numbers indicate bootstrap percentages out of 500 replicates

identity comparisons, results of phylogenetic analyses showed the SR-75 and SR-82 KRV isolates to be most closely related to CFAV (Fig. 4). As in previous studies of flavivirus phylogeny, clustering of viruses with respect to vector and host relationships was observed in both trees with the mosquito, no-known-vector, tick/vertebrate and mosquito/vertebrate viruses grouped in separate clades [3, 8, 28, 32]. Our analysis of NS5 sequences showed KRV and CFAV to be more closely related to the no-known-vector flaviviruses whereas the prM-E analysis showed a closer relationship with the tick/vertebrate flaviviruses. The average number of informative changes between the KRV isolates and the no-known-vector and tick/vertebrate species on the NS5 tree was 594 and 600, respectively (data not shown). The number of changes between KRV and these species on the prM-E tree averaged 708 and 593, respectively, reflecting the lower degree of sequence conservation in this structural gene region (data not shown). It has been hypothesized that arthropod-borne viruses may have evolved from insect viruses [44]. Previous phylogenetic studies of flavivirus E protein sequences have indicated that CFAV may represent a basal lineage of the genus, having diverged from the other flaviviruses before the separation of the mosquito and tick-borne groups [8, 32]. In our analysis CFAV and KRV were the most divergent taxa on both the prM-E and NS5 trees. The inclusion of KRV in future studies will therefore present an additional opportunity for evolutionary analysis of the genus.

Comparisons of hydrophobicity plots based on aa sequences from KRV, CFAV, YFV and TBEV showed that although sequence identity was low, structural homology based on predicted hydrophobicity/hydrophilicity of individual proteins was potentially high (data not shown). Conserved regions of hydrophobicity were observed in KRV proteins as in other flaviviruses, including C-terminal hydrophobic domains in the structural proteins, C, prM and E.

Structural proteins

Results of PAGE analysis of the viral structural proteins are shown in Fig. 3. Electrophoresis of purified KRV resulted in three strongly-staining bands with estimated mw that corresponded to mw expected for the E (∼54 kDa), anchC (∼15 kDa) and M (∼5 kDa) proteins; a weakly staining band migrating faster than the anchC band corresponded to the virC protein (∼13 kDa) (see discussion of cleavage sites above). Four strongly staining bands were observed for CFAV, including a doublet of approximately 51/49 kDa corresponding to the E protein (expected mw ∼46–57 kDa depending on glycosylation), and bands of 16 and 13 kDa consistent with the expected mw of the prM and C proteins, respectively. For the DENV-2 sample, three strongly staining bands of approximately 54, 11 and 6 kDa were observed that correlated with expected mw for the E, C and M proteins, respectively. A weaker band at approximately 22 kDa was also observed that most likely was uncleaved DENV-2 prM protein.

The KRV C protein contains 143 aa, compared to 136 aa in CFAV and 110–121 aa in other flaviviruses (Table 2). As observed for other flaviviruses the C protein is highly basic with K and R residues making up 22.3% of the protein

sequence, concentrated at the N and C termini of the protein. The KRV C protein contains one potential N-GLY site at aa 54–56; however, treatment with PNGaseF did not result in a change in C protein mobility on PAGE analysis (data not shown) therefore it is unlikely that this site is utilized. Interestingly, as discussed above, PAGE results suggested that the KRV C protein is inefficiently processed from anchC to virC and therefore anchC is the major form present in the mature KRV virion (Fig. 3). The CFAV anchC is cleaved to virC; an approximately 13 kDa band in PAGE analysis correlated with the expected size of the CFAV virC protein (Fig. 3).

The prM protein of KRV is 143 aa long and, unlike most other flaviviruses including CFAV, does not contain any potential N-GLY sites. The CFAV prM contains two potential sites for N-linked glycosylation. The mw of the strongly staining CFAV prM band (~16 kDa) in PAGE analysis correlated with the expected mw of the prM protein without carbohydrate (Fig. 3); however, treatment with PNGase F caused a mobility shift in the CFAV prM band (~14 kDa) that correlated with the loss of a single carbohydrate unit suggesting that one of the N-GLY sites is utilized. A weaker band seen on PAGE analysis of untreated protein that migrated slightly slower than the prM band (~17 kDa) was also observed and may represent a doubly glycosylated form of the prM protein (Fig. 3). Further research will be necessary to determine why the estimated mw of the protein is lower than expected; however, one possible explanation is that additional post-translational processing occurs at the C-terminus that reduces the final mw of the protein. Most flavivirus prM proteins contain six Cys residues, all of which are located in the pr portion of the protein, are highly conserved and are involved in disulfide bridges. The KRV and CFAV prM proteins contain only five Cys residues. In alignments with other flavivirus prM aa sequences the positions of four of the KRV Cys residues appeared to be conserved with those of other flaviviruses and all five are conserved with those of CFAV. The fifth Cys residue in KRV and CFAV, which did not align with the position of a Cys residue in other flavivirus prM sequences, is located in the M portion of the protein within 30 aa of the C terminus; TBEV and MODV each contain seven Cys residues with the additional residue located near the C terminus.

The E proteins in KRV (432 aa) and CFAV (427 aa) were found to be 50–70 aa shorter than the E protein of other flaviviruses (Table 2). There are 15 Cys residues in KRV E compared to a highly conserved 12 Cys in most other flaviviruses and 14 Cys in CFAV. Most flavivirus E proteins contain 1–3 potential sites for N-linked glycosylation; KRV and CFAV contain 6 and 5 such sites, respectively, although E protein mw estimation by PAGE analysis suggested that only two or three of these sites are utilized in the mature E protein (Fig. 3). Results of PAGE analysis showed the CFAV E protein as a doublet (~49 and 51 kDa) of approximately equal staining intensities; a possible explanation for this observation is that two species of the E protein, glycosylated at different sites, existed in the virus preparation. The KRV E protein shares little aa sequence identity with other flaviviruses (19.8–36.2%, Table 2). However, analysis of an alignment of E aa sequences from KRV, CFAV, JEV, DENV-2, YFV,

TBEV and MODV suggested considerable homology with regard to structural features.

Crystallographic examination of the TBEV E protein has revealed a structure for the flavivirus E protein that includes three domains [39]. Domain I, which comprises three noncontiguous segments of the protein sequence, contains four highly conserved Cys residues that form two disulfide bridges. Only the two most C-terminal of these are conserved in KRV E, thus there is only one disulfide bridge; the disulfide bridge close to the amino terminus of E is absent in KRV. With the exception of YFV, most flaviviruses contain one potential N-GLY site located in the second segment of domain I; KRV domain I contains two potential N-GLY sites, one of which is located near the C-terminus of the first domain I segment; the other comprises the last three aa of the domain. CFAV contains one potential N-GLY site in domain I which falls in close proximity to the position of the N-GLY site of other flaviviruses in the alignment.

Domain II, the dimerization domain, includes two noncontiguous sequence segments, the first of which contains the proposed 14 aa fusion peptide of the protein [42]. This sequence has been shown to be highly conserved among flaviviruses, with the exception of CFAV in which only eight of the 14 aa are conserved [8]. KRV E contains the same eight conserved residues and differs from CFAV at two of the six nonconserved residues. It has been suggested that differences observed in the fusion peptide sequence might account for the high degree of fusion caused by CFAV in mosquito cells. If this is the case, the increased fusion characteristic may be attributed to the fusion peptide residues that differ between KRV and CFAV, since KRV has not been observed to cause fusion. The first segment of domain II also contains six Cys residues that pair to form three disulfide bridges, contributing greatly to the folded structure of this region of the protein. These Cys residues are conserved among all flaviviruses, including KRV and CFAV, confirming the strict requirement for the structural integrity of this region. In contrast to these domain II structural and sequence homologies between KRV E and other flavivirus E proteins, this domain of the KRV and CFAV E contains two potential N-GLY sites whereas among other flaviviruses DENV-2 contains one site and the other flaviviruses have none. The position of the first E domain II potential N-GLY site is not conserved between KRV and CFAV whereas the position of the second site is conserved. It is not yet known whether either or both of these sites are utilized in KRV although, based on the estimated KRV E protein mw from PAGE analysis, only approximately two of the six potential N-GLY sites in the complete E protein are utilized.

Domain III of the flavivirus E protein contains residues that have been implicated as determinants of host range, tropism and virulence and therefore may be involved in receptor recognition and/or cell attachment of the virus [9]. Mutations affecting these characteristics have been localized to a specific region of domain III, the FG loop [22, 31]. Sequence in this region differs significantly between the tick-borne TBEV and the mosquito-borne flaviviruses which contain four additional aa residues; MODV in the no-known-vector group of flaviviruses does not contain the additional residues. This suggests that the additional residues in the

FG loop may be involved in mosquito cell tropism of the viruses. Although there was very little sequence identity observed in this domain between KRV and CFAV and other flaviviruses, the E protein alignment suggests that KRV and CFAV do include additional residues in this region compared to TBEV, and, furthermore, the sequence of these additional residues contains a potential N-GLY site that is conserved between KRV and CFAV. KRV and CFAV contain two potential N-GLY sites in this domain; the position of the first N-GLY site, however, is not conserved between them. Among the other flaviviruses, potential domain III N-GLY sites occur in YFV, TBEV and MODV, however each of these sites contains a Pro residue making it unlikely that they would be utilized [1]. Two conserved Cys residues are found in domain III, both of which are also conserved in KRV and CFAV. However, KRV and CFAV contain an additional four and three Cys residues in this domain, respectively, one of which falls in the FG loop region adjacent to the KRV/CFAV-conserved N-GLY site. An additional Cys residue occurs in KRV and CFAV closely following domain III in the C terminal region of the alignment making a total of seven or six Cys, respectively, in this 87 aa section of the protein. The presence of these additional Cys residues suggests that the folded structure of this domain differs significantly from that of other flaviviruses, a feature that may be responsible for the restricted host range of these two viruses.

Nonstructural proteins

The function of the flavivirus NS1 protein is not yet fully understood, however it has been implicated in RNA replication [29, 30, 35, 36]. NS1 exists in cell-associated, cell-surface and extracellular forms, as dimers and multimers, and has been shown to elicit a strong immune response [16, 19, 43, 54]. The KRV NS1 protein contains 390 aa and, like CFAV, includes 12 Cys residues and four potential N-GLY sites, the locations of which are conserved between the two viruses. Other flavivirus NS1 proteins contain 12 highly conserved Cys residues and 2–3 potential N-GLY sites. The positions of 10 of the 12 Cys residues in KRV are conserved with those of other flaviviruses; positions of the potential N-GLY sites are not conserved.

The flavivirus NS3 protein functions as a protease in processing of the viral polyprotein, as a helicase in RNA replication, and as a mRNA triphosphatase which may be involved in modification of the 5′ end of the viral genome [10, 53]. A number of sequence motifs have been identified that are associated with these functions and are conserved among the flaviviruses including KRV [2, 17]. The N-terminal one third of the protein contains motifs associated with the serine protease activity including the His_{55}, Asp_{79} and Ser_{138} catalytic triad of the protease active site and the Gly_{136}-X-Ser-Gly-X-Pro that surrounds Ser_{138}. A motif proposed to form part of the protease substrate binding pocket is defined in most flaviviruses by the sequences Asp_{132}-Tyr/Phe/Leu and Gly_{150}-Leu-Tyr-Gly-Asn-Gly, however KRV and CFAV contain Phe in place of Leu_{151} and Asn_{154}. Seven sequence elements associated with helicase activity have been identified in the flavivirus NS3 protein; portions of each are conserved including specifically the

segment I Gly_{192}-X-Gly-Lys-Thr/Ser and segment II Asp_{278}-Glu-Ala-His which are proposed to be the A and B sites of the helicase NTP-binding motif [18]. In KRV, as in CFAV, the segment I site is conserved and the segment II site contains a Cys residue in place of the Ala_{280}. Lastly, RNA triphosphatase activity has been demonstrated for the flavivirus NS3 and a seven-residue motif, the 5′-terminus recognition element, has been identified near the C-terminus of the protein which may function in positioning the 5′-terminal triphosphate of the substrate RNA into the active center of the enzyme [53]. This motif is loosely conserved among most flaviviruses and is conserved at only one residue (Pro_{561}) in KRV.

NS5 is the largest and most conserved of the flavivirus proteins and is believed to function as an RNA-dependent RNA polymerase (RdRP) [10]. The KRV NS5 contains 887 aa and shares 81.2% sequence identity with CFAV NS5 and approximately 46% identity with other flaviviruses. The invariant Gly-Asp-Asp motif found in RdRP enzymes is located at position 648–650 in the KRV protein. Additionally, sequences in the N-terminal portion of the flavivirus NS5 are homologous to methyltransferase enzymes; specific sequence regions have been identified including a putative S-adenosyl-methionine-utilizing methyltransferase binding motif that is highly conserved among flaviviruses including KRV [14, 26].

Noncoding regions

The 5′NCR of KRV is 96 nt in length, compared to the CFAV 5′NCR which is 113 nt and other flaviviruses which average 100 nt. The 5′NCR is not conserved between flaviviruses and there is only 58.6% sequence identity between the 5′NCR of KRV and CFAV, with highest identity in the 5′ terminal region. Secondary structures have been predicted for the 5′NCR of several flaviviruses including CFAV [7, 8]. Each of these structures consists of a stem-loop (SL) structure that includes a side loop. A possible structure for the terminal region of the KRV 5′NCR is shown in Fig. 5A. The first large SL includes nt 4–78 ($\Delta G = -19.3$ Kcal) and was present in 11 of 15 predicted structures. This structure is larger than that predicted for CFAV and does not include a large side loop, however the largest loop in the KRV structure includes the sequence AUUUU (nt 52–56) which is also present on the large side loop of the CFAV SL (nt 50–54). As seen in other flaviviruses, a second SL including nt 79–113 ($\Delta G = -12.5$ Kcal) was present in 13 of 15 predicted structures and included the proposed initiation codon (arrow) in the terminal loop (Fig. 5A).

The flavivirus 3′NCR exhibits a great deal of both size and sequence heterogeneity, however there is a high degree of structural homology. Stem-loop secondary structures have been predicted for the 3′ terminal ∼100 nt of most of the known flaviviruses and similarities have been identified that are specific to viruses within the mosquito/vertebrate and tick/vertebrate groups [6, 38]. Studies have demonstrated the binding of viral NS3 and NS5 replicase components as well as host proteins to these terminal SL structures suggesting their involvement in the replication process [5, 12, 48].

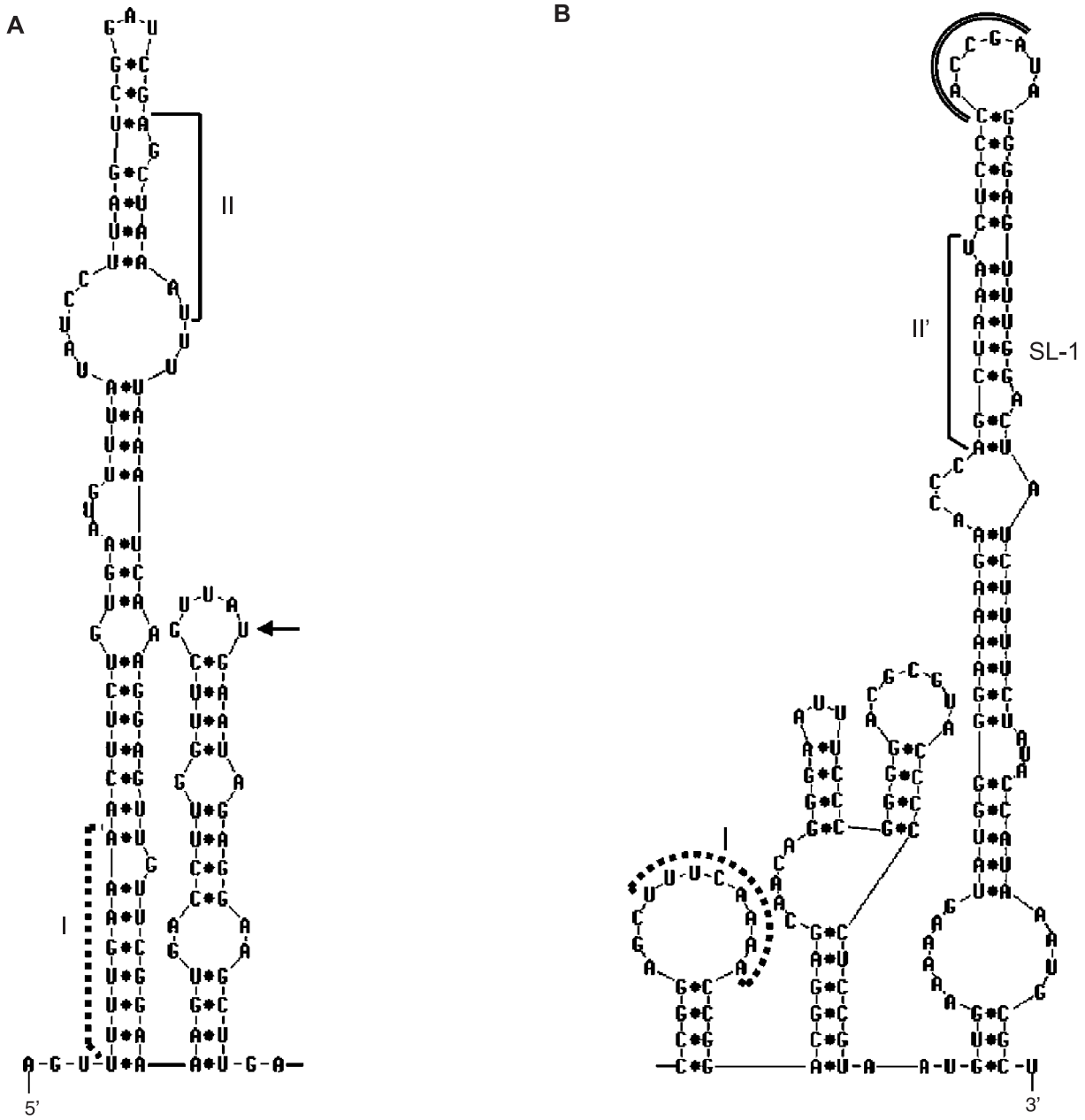

Fig. 5. Predicted secondary structures for 5′ and 3′ terminal regions of KRV. **A**, nucleotides 1–115 of the 5′NCR, arrow indicates AUG codon at 5′ end of the open reading frame. **B**, nucleotides 11227–11375 of the 3′NCR, double line marks conserved loop. Complementary sequence (CS) motif I indicated by dashed lines; 5′ CS component of motif II and 3′ reverse complement motif II′ indicated by solid lines

The 3′NCR of KRV is 1208 nt long which is significantly longer than that of CFAV (556 nt) or other flaviviruses (400–700 nt). There is no significant sequence identity between the 3′NCR of KRV and that of other flaviviruses, with the

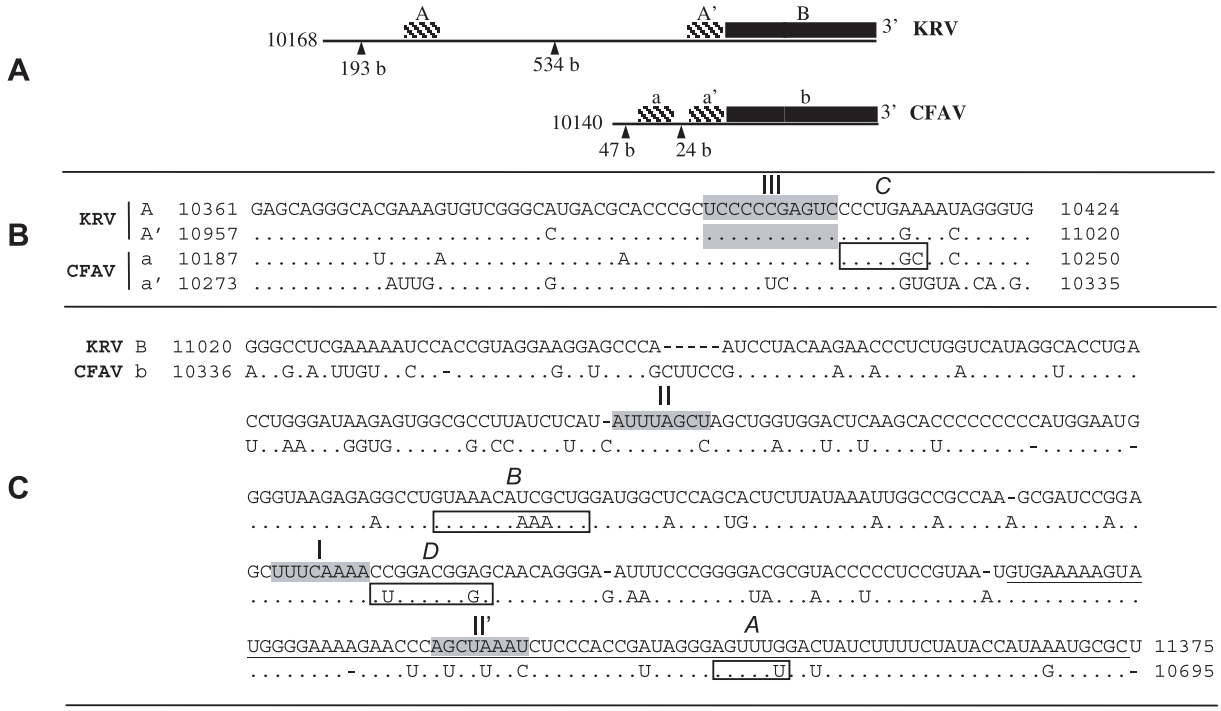

Fig. 6. **A**, diagram of the KRV and CFAV 3′NCRs, diagonal slashed lines indicate position of conserved repeat sequences, solid lines indicate 3′ terminal 'core' region. **B**, alignment of conserved repeat sequences. **C**, alignment of 3′ terminal 'core' region; sequence of nt included in the 3′ terminal stem-loop are underlined. Shaded sequences show positions of KRV 5′–3′ complementary sequence motifs labeled I, II, II′ and III; boxed sequences show positions of CFAV 5′–3′ complementary sequence motifs labeled in italics *A*, *B*, *C* and *D*

exception of CFAV. The 3′NCR of both KRV and CFAV contains a ∼64 nt, highly conserved, imperfect direct repeat motif. The repeat sequence occurs from nt 10361–10424 (A) and nt 10957–11020 (A′) in KRV (Fig. 6A and 6B) and from nt 10187–10250 (a) and nt 10273–10335 (a′) in CFAV (Fig. 6A and 6B). The length of the sequence between the repeat motifs is quite different in the two viruses (KRV = 534 nt, CFAV = 24 nt). This long sequence between the two repeat motifs in the KRV 3′NCR, which is likely the result of an insertion/deletion event facilitated by the adjacent repeat motifs, can account in large part for the size difference between the KRV genome and the genomes of CFAV and the other flaviviruses. A BLAST search conducted with the KRV 534 nt sequence between the repeats resulted in no significant matches. The region of the 3′NCR that lies upstream from the first repeat sequence also varies; this sequence is 193 nt in KRV and 47 nt in CFAV and there is no significant homology between the sequences. Conversely, the sequence that is downstream from the second repeat sequence (B and b, ∼356 nt) (Fig. 6A and 6C) exhibits approximately 80% identity between the two viruses. These observations are in agreement with previous studies of other flaviviruses that found the central portion of the 3′NCR to be highly variable

and to frequently contain repeat sequences and the 3′ terminal region to be most conserved [50]. Researchers studying the 3′NCR of YFV and DENV suggest that differences and repeats in the variable region may be associated with replication in mosquito hosts [45, 51].

As in other flaviviruses, the terminal 84 nt of the KRV genome (nt 11291–11375) forms a stable SL (SL-1) structure ($\Delta G = -29.5$) that occurs in 10 of 12 fold structures predicted for the terminal 500 nt of the genome (Fig. 5B). The structure of SL-1 contains four loops and is similar to the SL predicted for the terminal region of CFAV [8]. The terminal loop includes the sequence CACCGA (CACCGU in CFAV) that is homologous to the CACAG(A/U) sequence located in the outermost loop of the 3′ terminal SLs of other flaviviruses [52]. Additionally, in 9 of the 12 predicted KRV 3′NCR structures, a pair of smaller SLs are formed upstream that include nt 11227–11288 ($\Delta G = -21.5$) (Fig. 5B).

In CFAV, as in several other flaviviruses, complementary sequences (CS) between the 5′NCR/C region and the 3′NCR (CFAV: CS-A through D, Fig. 6B and 6C) have been identified [8, 21, 25]; these sequence motifs are not conserved in KRV. Three complementary sequence motifs (I–III) were identified in the KRV 5′ and 3′NCRs and their sequences and locations in the genome are listed in Table 3. Motif I (nt 4–11 complementary to nt 11234–11241) is located in the lower stem of the predicted 5′NCR SL structure and its complement is in the large loop of the 5′-most small SL structure predicted for the 3′NCR terminus (Figs. 5A, 5B and 6C). In an alignment of 3′NCR sequences this motif is located adjacent to the position of the CFAV-CS D motif. Motif II (nt 46–53 complementary to nt 11117–11124) is located in the upper stem of the 5′NCR SL and terminates in the large upper loop that contains similarity to the CFAV 5′NCR SL side loop sequence (Fig. 5A). The 3′ component of motif II is located 258 nt upstream from the 3′ terminus of the genome (nt 11117–11124) away from the terminal structures (area not included in Fig. 5B, location shown in Fig. 6C). A reverse complement of this 3′ component (motif II′) is also found downstream at nt 11318–11325, which is located in the upper stem of the predicted terminal SL-1 structure (Figs. 5B and 6C). The 5′NCR II and 3′NCR II′ AGCUAAAU motifs are homologous to a region of a conserved sequence (CS1, AG/CCAUAUU. . .) that has been identified in the 3′NCR of other

Table 3. List of KRV 5′NCR/3′NCR complementary sequence motifs

Motif designation	5′ location[a]	5′ sequence[b]	3′ location	3′ sequence
I	4–11	UUUUGAAA	11234–11241	UUUCAAAA
II	46–53	AGCUAAAU	11117–11124	AUUUAGCU[c]
III	166–177	GACTCGGC[d]GGGA	10398–10408 10994–11004	UCCCCCGAGUC

[a]Nucleotide location in full-length genome
[b]Sequence in 5′–3′ direction
[c]Sequence also indirectly repeated downstream in 3′NCR at position 11318–11325
[d]Base not paired between 5′ and 3′ motifs

flaviviruses and may be involved in cyclization of the genome [21]. These multiple repeats may have a regulatory function in viral replication; competition between several complementary motifs may determine which genome structure occurs at various times during the process. The 5′ component of motif III (5′ nt 166–177) is located in the C gene and is complementary to two 3′ motif III components (nt 10398–10408 and nt 10994–11004) which are located in the conserved repeat sequences A and A′ (nt 173 is not paired between the 5′ and 3′ motifs) (Fig. 6B). The 3′ motif III in the A repeat aligns adjacent to the position of CFAV-CS C (Fig. 6B).

Previous studies have suggested that complementary sequences between the 5′ and 3′ regions of flavivirus genomes may be involved in circularization of the genome during replication [21, 25]. Although specific CS sequences differ between the mosquito-borne, tick-borne and no-known-vector flavivirus groups, 5′–3′ complementary circularization sequences have been identified for each group and studies have shown that mutations in these regions deleteriously affect replication [25]. RNA folding analysis of KRV sequence constructs that included both 5′ and 3′ sequences separated by a poly-A 'stuffer' sequence suggested two basic structures for pairing between the ends of the viral genome (structures not shown). The first of these predicted structures maintained the 3′ terminal SL structures but involved unfolding of the 5′ SL and included pairing of the 5′ and 3′ CS components of both motif I and motif II. The second structure maintained both the 5′ and 3′ terminal SL structures and included pairing between the motif III 5′ and 3′ CS components. The 5′ motif III is located downstream of the initiation codon in the C gene, approximately the same position as the 5′ circularization sequences identified for other flaviviruses. The 3′ motif III, however, is located in the two repeat motifs, A and A′ (Fig. 6A and 6B), much farther upstream (−382 and −978) than the 3′ circularization sequences of other flaviviruses, which are located just upstream of the terminal SL structure. It is unclear at this time whether either or both of these structures actually plays a role in KRV replication.

Conclusions and future directions

We have shown that KRV, although highly divergent in many ways from other flaviviruses, also shows a great degree of homology in terms of components associated with important viral functions. The International Committee on Taxonomy of Viruses (ICTV) lists characteristics common among species in the genus *Flavivirus* including virion properties, nucleic acid and protein characteristics, genome organization, replication strategy, and antigenic properties. Also listed are biological properties including host range, transmission mode and vector relationships, geographic distribution and association with disease. We have demonstrated the significant similarity of KRV to other known flaviviruses in terms of genome size and organization, na and aa sequence identity, and homology in certain structural and functional components of the viral proteins. Similarity to CFAV in host range and apparent lack of associated disease has also been demonstrated. Based on this evidence we recommend placement of KRV in the genus *Flavivirus*.

The comparison of KRV with other viruses in the genus presents a unique opportunity to study the association of function with sequence and structure in divergent, yet related, species. Additionally, the availability of two insect flaviviruses, KRV and CFAV, will present the opportunity to investigate components associated with flavivirus host specificity and to conduct further studies on the evolution of the genus.

It has been observed that while CFAV usually causes massive syncytium formation in C6/36 cells, it can also infect these cells in a persistent manner [46]. The persistently-infected cultures do not show large amounts of syncytium formation, continue to produce low levels of the virus and are resistant to subsequent CFAV superinfection. Similarly, we observed that although KRV did cause CPE in the C6/36-H cells, it did not cause CPE in the C6/36-ATCC cells; neither did it cause any apparent pathology in the mosquitoes from which it was isolated. Therefore, it will be important to determine if the presence of KRV in mosquitoes has an effect on superinfection with other arboviruses.

CFAV has not yet been found in nature. However, the cell line from which it was isolated was derived from mosquito embryos and it has been suggested that the virus may have been present in these original embryos. KRV was isolated from mosquitoes that had been collected in the field as larvae or pupae, then reared to adults, pooled and assayed for the presence of virus. This suggests that KRV may be maintained in nature via transovarial transmission from one mosquito generation to the next. Alternatively, the virus may be transmitted by larval ingestion of infected mosquitoes in the breeding environment. Future studies will be conducted to determine the mode of transmission of the virus.

References

1. Bause E (1983) Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. Biochem J 209: 331–336
2. Bazan JF, Fletterick RJ (1989) Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. Virology 171: 637–639
3. Billoir F, de Chesse R, Tolou H, de Micco P, Gould EA, 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 9: 2339
4. Bird AP (1986) CpG-rich islands and the function of DNA methylation. Nature 321: 209–213
5. Blackwell JL, Brinton MA (1995) BHK cell proteins that bind to the 3′ stem-loop structure of the West Nile virus genome RNA. J Virol 69: 5650–5658
6. Brinton MA, Fernandez AV, Dispoto JH (1986) The 3′-nucleotides of flavivirus genomic RNA form a conserved secondary structure. Virology 153: 113–121
7. Brinton MA, Dispoto JH (1988) Sequence and secondary structure analysis of the 5′-terminal region of flavivirus genome RNA. Virology 162: 290–299
8. Cammisa-Parks H, Cisar LA, Kane A, Stollar V (1992) The complete nucleotide sequence of cell fusing agent (CFA): homology between the nonstructural proteins encoded by CFA and the nonstructural proteins encoded by arthropod-borne flaviviruses. Virology 189: 511–524

9. Cecilia D, Gould EA (1991) Nucleotide changes responsible for loss of neuroinvasiveness in Japanese encephalitis virus neutralization-resistant mutants. Virology 181: 70–77
10. Chambers TJ, Hahn CS, Galler R, Rice CM (1990) Flavivirus genome organization, expression, and replication. Annu Rev Microbiol 44: 649–688
11. Chambers TJ, Grakoui A, Rice CM (1991) Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. J Virol 65: 6042–6050
12. Chen CJ, Kuo MD, Chien LJ, Hsu SL, Wang YM, Lin JH (1997) RNA-protein interactions: involvement of NS3, NS5, and 3′ noncoding regions of Japanese encephalitis virus genomic RNA. J Virol 71: 3466–3473
13. Devereux J (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395
14. Egloff MP, Benarroch D, Selisko B, Romette JL, Canard B (2002) An RNA cap (nucleoside-2′-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. EMBO J 21: 2757–2768
15. Falgout B, Pethel M, Zhang YM, Lai CJ (1991) Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. J Virol 65: 2467–2475
16. Flamand M, Megret F, Mathieu M, Lepault J, Rey FA, Deubel V (1999) Dengue virus type 1 nonstructural glycoprotein NS1 is secreted from mammalian cells as a soluble hexamer in a glycosylation-dependent fashion. J Virol 73: 6104–6110
17. Gorbalenya AE, Donchenko AP, Koonin EV, Blinov VM (1989) N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases. Nucleic Acids Res 17: 3889–3897
18. Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM (1989) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res 17: 4713–4730
19. Gould EA, Buckley A, Barrett AD, Cammack N (1986) Neutralizing (54 K) and non-neutralizing (54 K and 48 K) monoclonal antibodies against structural and non-structural yellow fever virus proteins confer immunity in mice. J Gen Virol 67: 591–595
20. Hahn CS (1987) Comparison of the virulent Asibi strain of yellow fever virus with the 17D vaccine strain derived from it. Proc Natl Acad Sci USA 84: 2019–2023
21. Hahn CS, Hahn YS, Rice CM, Lee E, Dalgarno L, Strauss EG, Strauss JH (1987) Conserved elements in the 3′ untranslated region of flavivirus RNAs and potential cyclization sequences. J Mol Biol 198: 33–41
22. Holzmann H, Heinz FX, Mandl CW, Guirakhoo F, Kunz C (1990) A single amino acid substitution in envelope protein E of tick-borne encephalitis virus leads to attenuation in the mouse model. J Virol 64: 5156–5159
23. Igarashi A, Harrap KA, Casals J, Stollar V (1976) Morphological, biochemical, and serological studies on a viral agent (CFA) which replicates in and causes fusion of *Aedes albopictus* (Singh) cells. Virology 74: 174–187
24. Jenkins GM, Pagel M, Gould EA, de A. Zanotto PM, Holmes EC (2001) Evolution of base composition and codon usage bias in the genus flavivirus. J Mol Evol 52: 383–390
25. Khromykh AA, Meka H, Guyatt KJ, Westaway EG (2001) Essential role of cyclization sequences in flavivirus RNA replication. J Virol 75: 6719–6728
26. Koonin EV (1993) Computer-assisted identification of a putative methyltransferase domain in NS5 protein of flaviviruses and lambda 2 protein of reovirus. J Gen Virol 74: 733–740
27. Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44: 283–292

28. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB (1998) Phylogeny of the genus *Flavivirus*. J Virol 72: 73–83
29. Lindenbach BD, Rice CM (1997) trans-Complementation of yellow fever virus NS1 reveals a role in early RNA replication. J Virol 71: 9608–9617
30. Lindenbach BD, Rice CM (1999) Genetic interaction of flavivirus nonstructural proteins NS1 and NS4A as a determinant of replicase function. J Virol 73: 4611–4621
31. Lobigs M, Usha R, Nestorowicz A, Marshall ID, Weir RC, Dalgarno L (1990) Host cell selection of Murray Valley encephalitis virus variants altered at an RGD sequence in the envelope protein and in mouse virulence. Virology 176: 587–595
32. Marin MS, Zanotto PM, Gritsun TS, Gould EA (1995) Phylogeny of TYU, SRE, and CFA virus: different evolutionary rates in the genus Flavivirus. Virology 206: 1133–1139
33. Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288: 911–940
34. Miller BR, Mitchell CJ, Ballinger ME (1989) Replication, tissue tropisms and transmission of yellow fever virus in *Aedes albopictus*. Trans R Soc Trop Med Hyg 83: 252–255
35. Muylaert IR, Chambers TJ, Galler R, Rice CM (1996) Mutagenesis of the N-linked glycosylation sites of the yellow fever virus NS1 protein: effects on virus replication and mouse neurovirulence. Virology 222: 159–168
36. Muylaert IR, Galler R, Rice CM (1997) Genetic analysis of the yellow fever virus NS1 protein: identification of a temperature-sensitive mutation which blocks RNA accumulation. J Virol 71: 291–298
37. Obijeski JF, Bishop DH, Murphy FA, Palmer EL (1976) Structural proteins of La Crosse virus. J Virol 19: 985–997
38. Proutski V, Gould EA, Holmes EC (1997) Secondary structure of the 3′ untranslated region of flaviviruses: similarities and differences. Nucleic Acids Res 25: 1194–1202
39. Rey FA, Heinz FX, Mandl C, Kunz C, Harrison SC (1995) The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. Nature 375: 291–298
40. Rice CM, Lenches EM, Eddy SR, Shin S, Sheets RL, Strauss JH (1985) Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science 229: 726–733
41. Rice CM, Strauss JH (1990) Production of flavivirus polypeptides by proteolytic processing. Seminars in Virology 1: 357–367
42. Roehrig JT, Hunt AR, Johnson AJ, Hawkes RA (1989) Synthetic peptides derived from the deduced amino acid sequence of the E-glycoprotein of Murray Valley encephalitis virus elicit antiviral antibody. Virology 171: 49–60
43. Schlesinger JJ, Brandriss MW, Cropp CB, Monath TP (1986) Protection against yellow fever in monkeys by immunization with yellow fever virus nonstructural protein NS1. J Virol 60: 1153–1155
44. Schlesinger RW (1971) New opportunities in biological research offered by arthropod cell cultures. I some speculations on the possible role of arthropods in the evolution of arboviruses. Curr Top Microbiol Immunol 55: 241–245
45. Shurtleff AC, Beasley DW, Chen JJ, Ni H, Suderman MT, Wang H, Xu R, Wang E, Weaver SC, Watts DM, Russell KL, Barrett AD (2001) Genetic variation in the 3′ non-coding region of dengue viruses. Virology 281: 75–87
46. 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
47. Swofford DL (1998) PAUP*: phylogenetic analysis using parsimony (*and other methods), ver. 4.08b. Sinauer Associates, Sunderland, MA

48. Ta M, Vrati S (2000) Mov34 protein from mouse brain interacts with the 3′ noncoding region of Japanese encephalitis virus. J Virol 74: 5108–5115
49. von Heijne G (1984) How signal sequences maintain cleavage specificity. J Mol Biol 173: 243–251
50. Wallner G, Mandl CW, Kunz C, Heinz FX (1995) The flavivirus 3′-noncoding region: extensive size heterogeneity independent of evolutionary relationships among strains of tick-borne encephalitis virus. Virology 213: 169–178
51. Wang E, Weaver SC, Shope RE, Tesh RB, Watts DM, Barrett AD (1996) Genetic variation in yellow fever virus: duplication in the 3′ noncoding region of strains from Africa. Virology 225: 274–281
52. Wengler G, Castle E (1986) Analysis of structural properties which possibly are characteristic for the 3′-terminal sequence of the genome RNA of flaviviruses. J Gen Virol 67: 1183–1188
53. Wengler G, Wengler G (1993) The NS 3 nonstructural protein of flaviviruses contains an RNA triphosphatase activity. Virology 197: 265–273
54. Winkler G, Maxwell SE, Ruemmler C, Stollar V (1989) Newly synthesized dengue-2 virus nonstructural protein NS1 is a soluble protein but becomes partially hydrophobic and membrane-associated after dimerization. Virology 171: 302–305
55. Yoksan S, Bhamarapravati N, Halstead SB (1986) Dengue virus vaccine development: study on biological markers of uncloned dengue 1–4 viruses serially passaged in primary kidney cells. In: Arbovirus Research in Australia, Proceedings of the Fourth Symposium. CSIRO/QIMR, Brisbane, pp 35–38
56. Zuker M, Mathews DH, Turner DH (1999) Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide. In: Barciszewski J, Clark BFC (eds) RNA Biochemistry and Biotechnology. Kluwer Academic Publishers NATO ASI Series, pp 11–43
57. Sang RC, Gichogo A, Gachoya J, Dunster MD, Ofula V, Hunt AR, Crabtree MB, Miller BR, Dunster LM (2003) Isolation of a new flavivirus related to cell fusing agent virus (CFAV) from field-collected flood-water *Aedes* mosquitoes sampled from a dambo in central Kenya. Arch Virol 148: 1085–1093

Author's address: Mary B. Crabtree, Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, P.O. Box 2087, Fort Collins, Colorado 80522, U.S.A.; e-mail: meb3@cdc.gov