Characterisation of divergent flavivirus NS3 and NS5 protein sequences detected in *Rhipicephalus microplus* ticks from Brazil

Sandra Regina Maruyama[1], Luiza Antunes Castro-Jorge[1], José Marcos Chaves Ribeiro[2],
Luiz Gustavo Gardinassi[1], Gustavo Rocha Garcia[1], Lucinda Giampietro Brandão[1,3],
Aline Rezende Rodrigues[4,5], Marcos Ituo Okada[3,6], Emiliana Pereira Abrão[1], Beatriz Rossetti Ferreira[7],
Benedito Antonio Lopes da Fonseca[1], Isabel Kinney Ferreira de Miranda-Santos[1]/+

[1]Faculdade de Medicina de Ribeirão Preto [7]Escola de Enfermagem de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP,
Brasil [2]National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD, USA [3]Faculdade de Tecnologia,
Araçatuba, SP, Brasil [4]Vallée SA, São Paulo, SP, Brasil [5]Laticínios Vale dos Buritis, Buritis, MG, Brasil
[6]Vetmais Medicina Animal, Campinas, SP, Brasil

Transcripts similar to those that encode the nonstructural (NS) proteins NS3 and NS5 from flaviviruses were found in a salivary gland (SG) complementary DNA (cDNA) library from the cattle tick Rhipicephalus microplus. *Tick extracts were cultured with cells to enable the isolation of viruses capable of replicating in cultured invertebrate and vertebrate cells. Deep sequencing of the viral RNA isolated from culture supernatants provided the complete coding sequences for the NS3 and NS5 proteins and their molecular characterisation confirmed similarity with the NS3 and NS5 sequences from other flaviviruses. Despite this similarity, phylogenetic analyses revealed that this potentially novel virus may be a highly divergent member of the genus Flavivirus. Interestingly, we detected the divergent NS3 and NS5 sequences in ticks collected from several dairy farms widely distributed throughout three regions of Brazil. This is the first report of flavivirus-like transcripts in* R. microplus *ticks. This novel virus is a potential arbovirus because it replicated in arthropod and mammalian cells; furthermore, it was detected in a cDNA library from tick SGs and therefore may be present in tick saliva. It is important to determine whether and by what means this potential virus is transmissible and to monitor the virus as a potential emerging tick-borne zoonotic pathogen.*

Key words: *Rhipicephalus microplus* - tick - NS3 - NS5 - flavivirus

Ticks transmit a greater variety of infectious agents to humans and other animal species than does any other blood-feeding arthropod (Jongejan & Uilenberg 2004). The pathogens transmitted by ticks include bacteria, protozoa and viruses, with members of the Flaviviridae family among the most common tick-borne viruses. The Flaviviridae family is composed of three genera: Pestivirus, which contains the species responsible for zoonotic infections (ICTV 1995), Hepacivirus, which contains the human hepatitis C virus and (tentatively) the GB viruses (Ferron et al. 2005) and Flavivirus, which contains over 70 species, most of which are arthropod-borne (arboviruses) (Thiel et al. 2005). The tick-borne viral diseases caused by viruses of the Flavivirus genus are mainly infections of the central nervous system characterised by severe encephalitis in humans. Important and complex environmental interactions involving animal reservoirs of ticks and tick-borne viruses may influence the incidence of these viral diseases (Norman et al. 1999, Laurenson et al. 2003, Cope et al. 2004). In addition to the tick-borne group, the Flavivirus genus contains a mosquito-borne group and an arthropod group with no known vector (NKV) that infects vertebrate hosts, but does not have an identified arthropod vector (Kuno et al. 1998). The Tamana bat virus (TABV) and the cell fusing agent virus are classified as tentative species in the Flavivirus genus (Thiel et al. 2005) because they are considered highly divergent from the flaviviruses.

While analysing a salivary gland (SG) transcriptome from females of the cattle tick *Rhipicephalus microplus*, we identified transcripts with significant similarity to various viral genes. A few of these transcripts were similar to the nonstructural (NS)3 and NS5 genes of viruses in the Flavivirus genus. The Flavivirus NS proteins NS3 and NS5 perform the enzymatic activities necessary for RNA capping and genome replication (Bollati et al. 2010). Although virus-like particles have been found in the SGs of *R. microplus* (Megaw 1978), there have been no reports of Flaviviridae viruses in this tick species.

To ascertain whether the transcripts found in the *R. microplus* SG transcriptome were indeed from an RNA virus, we performed a molecular assay [polymerase chain reaction (PCR)] using DNA and RNA [reverse

transcribed into complementary DNA (cDNA)] from a different tick sample. The results confirmed that the transcripts were not a product of viral integration into the tick genome. The virus was then isolated from tick extracts in cell culture. The virus samples from the culture supernatants were PCR-positive for the same transcripts detected in the SGs of *R. microplus,* similar to the NS3 and NS5 sequences, suggesting that a potentially novel virus had been isolated. We refer to this possible novel virus as the Mogiana tick virus (MGTV) because of the region where it was isolated.

Viral RNA from the culture supernatant of MGTV-infected Vero cells was deep sequenced and we obtained the complete sequences of NS3 and NS5 from the novel virus. We conducted a comparative molecular analysis of these sequences with those from other Flaviviridae viruses and the results revealed that these sequences were highly divergent from those of other members of the Flavivirus genus. In addition, we detected these divergent NS3 and NS5 sequences in field samples of ticks, which suggests a wide geographical distribution of the virus in Brazil. These findings highlight the importance of studying the presence of arboviruses in *R. microplus* and the associations of these viruses with vectors and vertebrate hosts. These results also increase awareness of possible emerging zoonoses and/or tick-borne viral diseases in bovines.

MATERIALS AND METHODS

cDNA library and bioinformatics analysis - Adult female *R. microplus* ticks were collected from naturally infested cattle and the SGs were immediately dissected from 30 ticks. An *R. microplus* SG library was used and its construction has been described in detail by Maruyama et al. (2010). The bioinformatics analysis of the transcriptome data was performed as previously described (Ribeiro et al. 2006), with some modifications. The alignments were performed using the CLUSTALW program (Thompson et al. 1994) and BioEdit (Hall 1999) sequence alignment editing software. The phylogenetic associations were determined using the neighbour joining (NJ) or maximum likelihood (ML) methods (MEGA 4.0) (Tamura et al. 2007) and the node support of each clade was evaluated using a bootstrap analysis (1,000 replicates). The hydropathy profiles of the proteins were obtained using the web-based tool ProtScale (web.expasy.org/protscale/) from the ExPASy Bioinformatics Resource Portal (Artimo et al. 2012) with the Kyte and Doolittle (1982) scale option and a window size of nine amino acids (aa). The hydropathy plot were constructed based on the alignment of NS3 and NS5 transcripts found in *R. microplus* with NS3 and NS5 transcripts from dengue virus type 2 (DENV-2). Because the transcripts had different lengths the score was zero for gaps. The codon usage adaptation index (CAI) and base composition values were obtained using tools provided by the CAIcal server (genomes.urv.es/CAIcal/) (Puigbò et al. 2008a). The codon usage tables for Flavivirus spp, *R. microplus*, *Bos taurus* and *Homo sapiens* were obtained from the Codon Usage Database (kazusa.or.jp/codon/), which is compiled from the GenBank DNA sequence database

(Kazusa DNA Research Institute).The NS3 and NS5 protein sequences of Flaviviridae viruses used in this work were obtained from National Center for Biotechnology Information (NCBI) RefSeq collection (Supplementary data, Table SI).

Viral isolation - The virus was isolated from ticks collected from Holstein bulls on a farm in Ribeirão Preto, state of São Paulo (SP), Brazil (21º13'09.28''S 48º31'34.91''W). Pools of 20 ticks were crushed in liquid nitrogen, resuspended in 1 mL of sterile phosphate-buffered saline (PBS) (pH 7.0) containing 10% foetal bovine serum (Invitrogen, San Diego, CA, USA), penicillin (500 IU/mL) and streptomycin (500 mg/mL) (Invitrogen) and centrifuged at 2,500 *g* for 5 min. The supernatant was collected and stored at -70ºC until use. Prior to the inoculation of the culture medium, the supernatant was filtered in a 0.22-μm filter.

Viral isolation was performed using both arthropod and mammalian cells. African green monkey kidney (Vero) and baby hamster kidney (BHK)-21 cells were grown at 37ºC under 5% CO_2 in minimum essential medium (MEM) (Invitrogen) supplemented with 10% inactivated foetal calf serum (FCS) (Invitrogen), 1% L-glutamine (Cultilab, Campinas, SP, Brazil) and 1% antibiotic-antimycotic (Invitrogen). *Boophilus microplus* cattle tick (BME 26) cells were obtained from Dr Ulrike Munderloh in the Department of Entomology at the University of Minnesota. *Aedes albopictus* mosquito (C6/36) cells were grown at 30ºC in Leibovitz's medium (L-15) (Invitrogen) supplemented with 10% FCS and 1% antibiotic-antimycotic.

A 30-μL aliquot of each tick pool filtrate was diluted in 310 μL of PBS (pH 7.0) with 2% antibiotic-antimycotic (Invitrogen). This solution was inoculated into paired wells of six-well plates (Corning, Corning, NY, USA) containing semiconfluent monolayers of C6/36, BME 26, Vero and BHK cells. The filtrate was allowed to adsorb for 60 min at 37ºC (Vero and BHK cells) or 30ºC (C6/36 and BME 26). To prevent the detection of residual, nonreplicating viruses from the inoculum after the adsorption of the virus from the tick extracts to the cell monolayers, the plates were thoroughly washed with sterile saline to remove the inoculum before the maintenance medium was added. The cells were incubated for seven days at 37ºC in a 5% CO2 atmosphere (Vero and BHK) or at 30ºC (C6/36 and BME 26) with maintenance medium (for Vero and BHK cells: MEM with 2% FCS, 1% L-glutamine and 2% antibiotic-antimycotic; for C6/36 and BME 26 cells: L-15 supplemented with 2% FCS and 2% antibiotic-antimycotic). The plates were observed daily for any cytopathic effects (CPE) and infection was confirmed by the molecular detection of the virus. RNA was extracted from the cell culture supernatants of the first and second passages using a QIAamp Viral RNA kit (Qiagen, Valencia, CA, USA) and reverse-transcription(RT)-PCR was performed for molecular detection using primers for the NS3 and NS5 coding sequences of MGTV. Additionally, the first passage virus was replicated in Vero cells. Cells were infected as described above and the supernatant was collected 1 h, 3 h and 6 h after infection and on days one,

three, five, seven and nine after infection. The viral load (VL) was determined by amplifying the NS5 region herein described. A KAPA SYBR® FAST Universal Kit was used the NS5 region through real-time PCR, using the KAPA SYBR (KAPA Biosystems, Woburn, MA, USA) under the amplification conditions recommended by the manufacturer. The reactions were performed in a 7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). A standard curve was generated by cloning the 401 region using the InsTAclone PCR Cloning Kit (Thermo Scientific) and the final value was expressed as MGTV RNA copies/mL calculated as copies/mL.

Deep sequencing of viral RNA - Purification of viruses by sucrose cushion - Cultures of Vero cells (2 x 10⁶) were infected with tick pools as described. The culture supernatants were collected six days post-infection. The cellular debris was removed by centrifugation at 10,000 *g* for 30 min at 4ºC. The supernatants were layered onto a 20% sucrose cushion prepared in TNM buffer [10 nM Tris HCl (pH 7.5), 5 mM $MgCl_2$ and 150 mM NaCl]. The solution was centrifuged in an AH-629 Thermo Scientific rotor at 27,000 rpm for 9 h at 4ºC. The viral pellet was resuspended in 0.3 mL of cold PBS and used for RNA extraction.

RNA preparation - A 0.15-mL aliquot of purified virus was treated with 100 units of RNAse I and four units of DNAse I at 37ºC for 1 h to remove nucleic acids that were not protected by viral capsids. After incubation, 2 μL of dithiothreitol and 2 μL of RNAse inhibitor were added and RNA was extracted using a QIAamp viral RNA mini kit (Qiagen). The viral RNA was sequenced by an external company.

Deep sequencing using the Illumina MiSeq platform - Sequence data were generated by the High-Throughput Sequencing and Genotyping Unit at the University of Illinois in Urbana-Champaign. The deep sequencing was performed according to the manufacturer's (Illumina, San Diego, CA) instructions. Briefly, after a sample quality control analysis, the RNA was converted into cDNA using random hexamers and was then nebulised, adaptored and quantitated. The paired-end reads [150 nucleotides (nt) in length] were sequenced in one lane of a MiSeq instrument for 2 x 150 cycles and were further analysed using Casava 1.8.2. A total of 11,307,270 Illumina paired-end reads were obtained.

Bioinformatics analysis of the short reads - The workflow for the bioinformatic analysis consisted of rounds of mapping steps to reference genomes to filter out of unwanted sequences. The TopHat2 (Kim et al. 2013) and Bowtie2 (Langmead & Salzberg 2012) aligners were used with the default options to remove the following: sequences that were identical to host cell sequences [Rhesus monkey *Macaca mulatta* genome assembly MMUL_1, downloaded from Ensembl (ensembl.org/index.html)], then to sequences from bacterial contaminants [*Mycoplasma hyorhinis* SK6 genome assembly GCA_000313635.1, downloaded from EnsemblBacteria (bacteria.ensembl.org/index.html)] and ultimately to unwanted viral sequences [artificial reference

genome built from DNA viruses, retrovirus genomes and sequences from endogenous viruses in primates, all downloaded from RefSeq-NCBI (ncbi.nlm.nih.gov/refseq/)]. After the successive purging steps, 6,512,646 paired-end reads were used as input for *de novo* assembly with the program SOAPdenovo-Trans (Luo et al. 2012). Variable k-mer values ranging from k = 13 to k = 31 (in steps of 2) and from k = 33 to k = 73 (in steps of 4) and the parameters -d 3 -D 5 -L 150 -u -e 3 for every selected k-mer size were employed. All the resulting assembled sequences larger than 150 nt were combined into a FASTA file and submitted as input to an assembly pipeline consisting of BLASTN (Altschul et al. 1997) and CAP3 (Huang & Madan 1999) iterations, as previously described (Karim et al. 2011). The assembled contigs were BLASTed against the NCBI nonredundant (NR) protein database Swiss Prot (O'Donovan et al. 2002), the Conserved Domains Database (CDD) (Marchler-Bauer et al. 2002), a customised protein database containing only viral proteins from the NR database and a second customised database containing only Flavivirus proteins from RefSeq-NCBI. The data from the bioinformatics analysis are presented in Supplementary data (Table SII), a hyperlinked Excel spreadsheet.

Collection of the ticks - Ticks were collected in March 2006 from 19 livestock farms located in different regions of Brazil (Supplementary data, Table SIII) in a partnership with Valleé SA [Montes Claros, state Minas Gerais (MG), Brazil]. All the samples were placed in RNAlater (Invitrogen) and stored at -70ºC for later nucleic acid extraction.

DNA and total RNA extraction - Pools of ticks (organised according to locality and life stage) (Supplementary data, Table SIII) were crushed and resuspended in PBS for simultaneous DNA and RNA isolation. DNA was extracted from an aliquot of the homogenate using the QIAamp DNA kit (Qiagen). The remaining homogenate was mixed with Trizol (Invitrogen) and total RNA was isolated from it using the SV Total RNA Isolation Kit (Promega, Madison, WI, USA). cDNA was synthesised from the total RNA from the tick samples with the ImProm-II™ Reverse Transcription System kit (Promega). All the procedures were performed according to the manufacturers' instructions.

Molecular detection - PCR was performed to detect the divergent NS3 and NS5 fragments in genomic DNA (gDNA) and cDNA. The contig sequences were used as templates for primer design. Previously described actin primers (de la Fuente et al. 2008) were used as reaction controls. The reactions were performed using the primers listed in Supplementary data (Table SIV).

Accessions - All the sequences obtained from the ticks and culture cells have been deposited in GenBank (NCBI). The accessions include the following: (i) the partial sequences of isolates from tick pools (extract and culture) are JQ289026-JQ289041, as detected in ticks, (ii) the Illumina reads are the short reads archive SRA 055953, (iii) the complete sequences of NS3 and NS5 are JX390985 (protein_id: "AGL39759") and JX390986

(protein_id: "AGL39760"), respectively, and (iv) the partial sequences from the field samples of ticks and the cDNA library are HS586608-HS586670.

RESULTS AND DISCUSSION

Similarities to the NS NS3 and NS5 proteins in the tick cDNA library - The cDNA library from the SGs of female *R. microplus* generated 1152 expressed sequence tags (ESTs) clustered into 533 contigs (Maruyama et al. 2010), which were compared to the NCBI NR protein database using BLASTX. The annotation of the contigs revealed that five contigs aligned with NS flaviviral proteins (Table I) spanning seven ESTs. NS3 and NS5 proteins from three different Flavivirus species (TABV, Apoi virus and Kamiti River virus) represented the best matches in the NR protein database, as indicated by low e-values (except for contig 319, with an e-value of 0.002).

To evaluate the presence of this putative Flavivirus member in a different tick sample and to confirm that our sequences were not derived from a viral insertion into the tick genome, as has been reported for the mosquito *Ae. albopictus* (Crochu et al. 2004), we performed molecular detection assays in different samples of *R. microplus* ticks. These samples were collected from two Brazilian farms in Ribeirão Preto and Araguari (MG); these cities are located in the Mogiana macroregion (between SP and MG). We designed primer pairs for , sequences 317, 401 and 2743, which produced amplicons of 258 bp, 281 bp and 248 bp, respectively. A protocol for extracting gDNA and total RNA from the same tick sample was developed to facilitate viral detection through PCR with gDNA or cDNA as a template. The 317, 401 and 2743 primers amplified one conserved motif from the nucleotidases domain of NS3 (motif VI: QRRGRVGR) (Wu et al. 2005), one conserved motif from the methyltransferase domain of NS5 (motif IV: DTLLFDGGE) (Egloff et al. 2002) and one conserved motif from the RNA-dependent RNA polymerase domain (RdRp) of NS5 (motif B: SG_VVTYALNTL) (Poch et al. 1989), respectively. PCR using each set of primers (317, 401, 2743 and actin as a positive control) was used to amplify the gDNA and cDNA samples. The agarose gel electrophoresis analysis revealed that no products were obtained using the RNAse-treated gDNA as a template for the detection of fragments of the 317 (NS3), 401 (NS5) and 2743 (NS5) sequences, except in the positive control reactions using actin primers. The PCR-negative results for the gDNA template indicated that these sequences were not derived from the tick genome. In contrast, when cDNA (RNA) was used as a template, all nine tick samples were PCR-positive for the 317 primer set detection, whereas seven tick samples were PCR-positive for the 401 and 2743 primer set detection (Supplementary data, Fig. S1). These results indicated that the viral-like transcripts observed in the tick cDNA library were genuinely derived from RNA (cDNA) and were likely from an RNA virus.

Viral isolation in cultured cells and deep sequencing of RNA in viral capsids - To confirm the presence of the presumed Flavivirus found in the tick samples, viral isolation was performed in mammalian and arthropod cells. Five pools of ticks (4 from females and 1 from larvae) from the Ribeirão Preto farm, all of which were PCR-positive for the 317, 401 and 2743 sequences, were used to inoculate the cell cultures. No CPE was observed in the mosquito (C6/36), tick (BME 26) or hamster (BHK) cell lines; a CPE was observed only in the Vero cells (data not shown). Upon primary isolation, the CPE was minimal and developed slowly (for 5 days after the inoculation), resulting in the rounding and shrinkage of the cells, which became refractile and detached from the plate surface. After the second passage, the CPE disappeared; therefore, we could not titre it with a plaque-forming assay. The CPE observed during the initial passage alone

TABLE I

Contigs of *Rhipicephalus microplus* ticks salivary glands transcriptome
that presented matches with nonstructural (NS) proteins NS3 and NS5 of flaviviruses

| Contigs[a] | Number of ESTs | Best mach to NR database | | Virus |
		BLASTX	e-value	
317	2 (gi 317017979; gi 317017980)	Putative NS protein NS3 (gi 27735332)	5e-009	TABV
319	1 (gi 317017985)	Putative NS protein NS3 (gi 27735332)	0.002	TABV
400	2 (gi 317017981; gi 317017982)	NS protein NS5 (gi 27697405)	3e-006	APOIV
401	1 (gi 317017983)	NS5 protein (gi 37695589)	2e-005	KRV
2,743	1 (gi 317017984)	Putative NS protein NS5 (gi 27735336)	1e-010	TABV

a: these five contigs are part of a transcriptome analysis that generated more than 3,000 contigs; APOIV: Apoi virus; KRV: Kamiti virus; TABV: Tamana bat virus.

might have been due to toxicity in the inoculum; however, no CPE was observed in the C6/36, BME 26 or BHK cells inoculated under the same conditions.

The molecular detection of the virus was performed using supernatants from the four cell lines and primer sets 317, 401 and 2743. Each supernatant was inoculated with a different pool of ticks. The cell cultures were inoculated with the tick extracts, the cell monolayers were thoroughly washed after 1 h of incubation to remove any residue from the inoculum and maintenance medium was added. Seven days later, the culture supernatants were subjected to PCR. Sequences 317, 401 and 2,743 were successfully amplified from all the cell cultures inoculated with female tick pools, but not from the cell cultures inoculated with unfed larvae (UL) pools (Supplementary data, Fig. S2 displays the results from primer set 401). The culture supernatants from the second passage of the four cell lines remained PCR-positive for primer set 401 (Supplementary data, Fig. S3 shows the results from the Vero cells). Thus, the positive PCR results suggested that the virus replicated in the mammalian and arthropod cell lines. The viral strains isolated from the four tick pools were sequenced using primer sets 317 and 401 and the results confirmed that these strains represented the same fragments previously identified in the *R. microplus* SG cDNA library (data not shown).

The VL from the first passage was further evaluated in Vero cells and the viral titre was highest on the seventh day, at 1.7×10^4 copies/mL (data not shown). This result differs from findings from other flaviviruses, which exhibit the highest VL between the third and fifth days after infection (Bonaldo et al. 2007, Orlinger et al. 2011). The low observed VL may have been responsible for the absence of a CPE in the second passage. Because this was a field isolate, it may have required more passages to increase the titre.

To obtain longer sequences than those of tick cDNA library contigs 307, 401 and 2,743, we first performed PCR using the cDNA produced from viral RNA isolated from the tick extracts and culture supernatants described. A previously described set of universal Flavivirus primers was employed (Gaunt & Gould 2005, Maher-Sturgess et al. 2008). None of the amplifications with the universal primers produced amplicons (data not shown), indicating that the genome of the isolated virus may be highly divergent from those of classical flaviviruses. We then purified the virus from an infected Vero cell supernatant (2nd passage). Following RNAse and DNAse treatment to degrade the nucleic acids that were not protected by viral capsids, the viral RNA was extracted and the molecular detection of the divergent NS5 sequence was confirmed (Supplementary data, Fig. S3). This viral RNA was used for deep sequencing on an Illumina platform. The MiSeq run generated 11,898,134 paired-end reads, which were purged of reads with sequence identity to the cellular host bacterial contamination (*M. hyorhinis*), DNA viruses, retroviruses or endogenous viruses of primates. The purged paired-end read files were then assembled using the SOAPdenovo-Trans program. An automatic annotation was performed using a customised bioinformatics workflow on the BLASTX protein database. The RNA-seq dataset, which included over 6,500 contigs, was plotted in a hyperlinked Excel spreadsheet (Supplementary data, Table SII). We searched this dataset for significant BLAST results against the Flavivirus protein database that were identical to the tick cDNA library contigs (317, 401 and 2,743) described in Table I.

Despite the use of high-throughput sequencing, the assembled reads did not reveal the full-length genome sequence of the potentially novel virus. Although steps to remove background had been performed, most of the assembled reads were assigned to bacteria and miscellaneous endogenous virus sequences. This contamination, together with the low MGTV VL in the cell cultures, appears to have hindered the sequencing of transcripts expressed by MGTV at a lower frequency. The deep sequencing results revealed that the majority of the reads from our viral RNA sample from cultured MGTV-infected cell supernatants may have been derived from an endogenous primate retrovirus present in Vero cells. The chemical induction of endogenous retroviruses in Vero cells has been reported (Ma et al. 2011). Furthermore, the mobilisation of endogenous retroviruses in mice following infection with an exogenous retrovirus has occurred (Evans et al. 2009). Because we did not investigate whether viral RNA was present in supernatants of uninfected Vero cells (control), we could not determine whether these endogenous viral-like particles were constitutive or induced by MGTV infection.

Only 0.025% of the reads exhibited identity with contig sequences 401 and 317. Although viral RNA sequences from MGTV were underrepresented, we found two RNA-seq contigs, 1,961 and 2,579, with greater than 90% nt identity with the tick cDNA library contigs (Supplementary data, Table SII, highlighted rows). Both contigs exhibited significant similarity to the TABV polyprotein. RNA-seq contigs 1,961 (2,993 nt) and 2,579 (2,721 nt) were similar to the Flavivirus NS5 and NS3 NS proteins, respectively. Several contigs exhibited intermediate similarity with other Flavivirus structural and NS proteins (Supplementary data, Fig. S6). However, because most of these contigs represented short consensus sequences, we were unable to build a proper scaffold genome for MGTV. Therefore, RNA-seq contigs 1,961 and 2,579 were the only ones confirmed as MGTV-derived sequences. Although the complete genomic sequence could not be identified through deep sequencing, the complete NS3 and NS5 sequences from MGTV were obtained. Because NS3 and NS5 are the two largest and most conserved Flavivirus proteins (Chambers et al. 1990), we performed a comparative sequence analysis between these two MGTV proteins and NS3 and NS5 from Flaviviridae viruses to confirm our findings concerning the presence of Flavivirus-like transcripts in *R. microplus* ticks and to investigate the relationship between this presumed flavivirus and other members of the Flaviviridae family.

Molecular characterisation of the NS3 and NS5 protein sequences found in R. microplus ticks - The NS3 and NS5 proteins are the two main components of the flaviviral replication machinery. To obtain the complete NS3 and NS5 protein sequences from MGTV, open

reading frames +1 and -2 of the nt consensus sequences of RNA-seq contigs 1,961 and 2,579, respectively, were translated into aa sequences and aligned with different flaviviral NS3 and NS5 protein sequences [the Flavivirus sequences used are listed in the Supplementary data (Table SI)]. The alignment results were used to obtain the complete protein sequences for NS3 (554 aa, GenBank accession JX390985) and NS5 (866 aa, GenBank accession JX390986) from the potentially novel virus. Both proteins were shorter than known proteins from other flaviviruses, except for NS5 from TABV (831 aa).

Conserved motifs - To analyse sequence conservation in motifs of the MGTV NS3 and NS5 proteins, we performed sequence alignments using flavivirus sequences from one representative virus for each group vector and highlighted the conserved motifs. The N-terminal region of the flavivirus NS3 protein encodes a viral serine protease (Gorbalenya et al. 1989, Falgout et al. 1991), whereas the C-terminal portion encodes a helicase (Lain et al. 1989). The alignment of the flavivirus NS3 protein sequences revealed that residues in important regions for these functions were conserved in NS3 from MGTV (Fig. 1). The catalytic triad of histidine, aspartate

and serine residues required for serine protease activity (Polgar 2005) was conserved in the MGTV sequence at positions 45, 69 and 126, respectively. Together with a glycine residue at position 124 [Fig. 1A, Supplementary data (Fig. S7)], these four residues provide an electrostatic environment for the active site of the enzyme. The flavivirus helicase/NTPase catalyses the unwinding of the RNA strand (Dumont et al. 2006) to facilitate the initiation of viral replication and contains seven conserved motifs (Gorbalenya & Koonin 1993). Compared with the set of motifs described for the yellow fever virus (Wu et al. 2005), all seven motifs were well conserved in the corresponding novel sequence (Fig. 1B). Of note, motifs I and VI were the most strongly conserved. The complete sequence alignment of the NS3 proteins displayed in Fig. 1 can be viewed in the Supplementary data (Fig. S4).

The flaviviral NS5 protein contains a C-terminal (RdRp), whereas the N-terminus possesses the methyltransferase activity implicated in capping the 5'-end of the flaviviral RNA genome (Egloff et al. 2002, Davidson 2009). The alignment of the flavivirus NS5 protein sequences revealed that NS5 from MGTV contained conserved motifs (Fig. 2). The NS5 N-terminus also con-
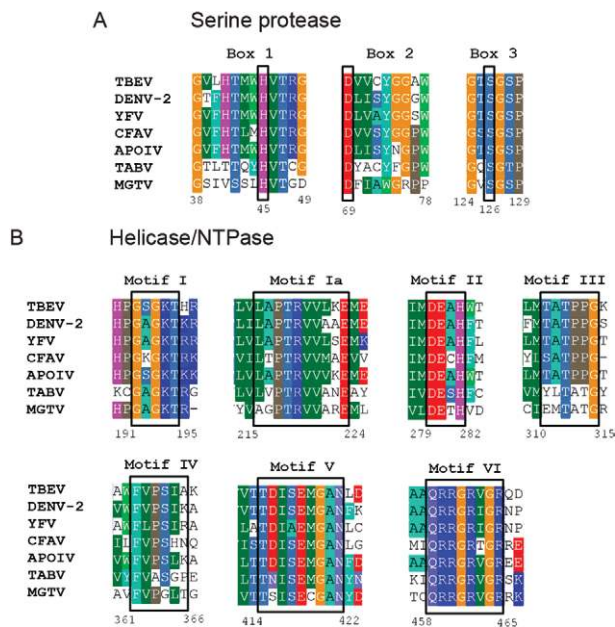

Fig. 1: conserved motifs in nonstructural (NS)3 of Mogiana tick virus (MGTV). Multiple sequence alignment of NS3 from MGTV and representative of tick-borne encephalitis virus (TBEV), mosquito-borne [dengue virus type 2 (DENV-2) and yellow fever virus (YFV)], insect-only [cell fusing agent virus (CFAV)] and not known vector [Apoi virus (APOIV) and Tamana bat virus (TABV)] flavivirus groups. For better visualisation, only blocks of conserved regions along the alignment are displayed. Motifs are delimited accordingly. A: N-terminal regions of NS3 contain a serine protease, amino acids (aa) of the catalytic triad of which are highlighted; B: C-terminus portion of NS3 contain a helicase/nucleotidases (NTPases) where seven conserved motifs (I, Ia and II-VI) are found. Numbers at the bottom of the alignments refer to MGTV sequence. The threshold for shading colours of aa similarity was 50%.
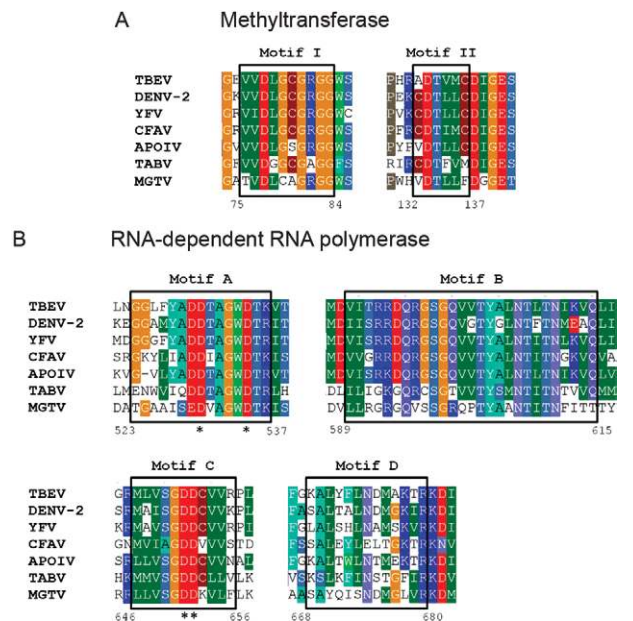

Fig. 2: conserved motifs in nonstructural (NS)5 of Mogiana tick virus (MGTV). Multiple sequence alignment of NS5 from MGTV and representative of tick-borne encephalitis virus (TBEV), mosquito-borne [dengue virus type 2 (DENV-2) and yellow fever virus (YFV)], insect-only [cell fusing agent virus (CFAV)] and no known vector [Apoi virus (APOIV) and Tamana bat virus (TABV)] flavivirus groups. To better visualisation, only blocks of conserved regions along the alignment are displayed. Motifs are accordingly delimited. A: N-terminal region of NS5 presents methyltransferase activity; two conserved motifs are found (I and II); B: C-terminus portion of NS5 contains an RNA-dependent RNA polymerase; four conserved motifs (A-D) are found; asterisk indicates the conserved aspartate residues important for enzyme activity. Numbers at the bottom of the alignments refer to MGTV sequences. The threshold for shading colours of amino acid similarity was 50%.

tains an RNA guanylyltransferase, which uses GTP as a substrate and its activity is stimulated by the NS3 protein (Issur et al. 2009). The N-terminal methyltransferase domain displayed two conserved motifs, I and II, involved in S-adenosyl methionine binding (Koonin 1993). The novel (MGTV) NS5 sequence exhibited approximately 70% identity with residues from motifs I and II of the methyltransferase region (Fig. 2A). In addition to capping the viral genome, NS5 is involved in the synthesis of RNA through the RdRp activity at the C-terminus; thus, it plays a key role in viral replication. The RdRp domain includes four conserved motifs: A, B, C and D (Poch et al. 1989). Of these, the D motif appears to be the least conserved among the analysed flaviviruses; the RdRp from the MGTV NS5 sequence showed conserved residues mainly in motifs A and C (Fig. 2B). Remarkably, the four strictly conserved aspartate residues, D-X_4-D and G-D-D (which are located in motifs A and C, respectively) were also strictly conserved in the MGTV NS5 sequence at positions 530-535 and 650-652, respectively (Fig. 2B, indicated by an asterisk). The complete sequence alignment of the NS5 proteins displayed in Fig. 2 can be viewed in the Supplementary data (Fig. S5).

Phylogenetic analysis - To verify the phylogenetic associations between MGTV and other Flaviviridae members, we performed ML analyses of the NS3 and NS5 sequences. In addition to the Flavivirus genus (including the divergent TABV), members of the Pestivirus and Hepacivirus genera were included in the analyses [the sequences IDs are listed in Supplementary data (Table SI)]. The phylograms of NS3 and NS5 indicated that MGTV is positioned closer to the Flavivirus genus than to the other genera of Flaviviridae (Fig. 3), which was supported by significant bootstrap values. However, MGTV is highly divergent from the other flaviviruses and is distantly related to all Flavivirus species described to date, both vectored and non-vectored, including TABV.

The phylogeny of the Flavivirus genus has been studied extensively (Kuno et al. 1998, Billoir et al. 2000, Gaunt et al. 2001) and different patterns of phylogenetic positions have been observed, depending on the gene analysed. All these studies used the divergent insect-only viruses and did not consider the divergent TABV virus. We included TABV in our analyses due to the BLASTX results, which showed that the best hit was a TABV polyprotein [Table I, Supplementary data (Table SII)]. A more recent comprehensive phylogenetic analysis of NS3, NS5 and entire flaviviruses genomes sequences using the ML method revealed that the NS3 and complete genome trees exhibited the same phylogenetic topology in the flavivirus groups (Cook & Holmes 2006). Because the NS3 tree appears to reflect information from relationships between complete flavivirus genome sequences, the association of MGTV with the Flavivirus genus is also likely reliable (Fig. 3A) despite the high divergence of this novel virus. The complete genomic sequence of MGTV will be critical for clarifying the taxonomic organisation of the Flaviviridae family.

Hydropathy profile - The aa chemical properties of a protein are reflected in its hydrophobicity/hydrophilicity profile. The MGTV NS3 and NS5 proteins exhibited a hydrophobic content of 33% (Supplementary data, Table SV). Because the flaviviral NS3 and NS5 proteins are hydrophilic (Chambers et al. 1990), we compared the hydropathy profiles of NS3 and NS5 from MGTV with the profiles of the corresponding proteins from a classical flavivirus, DENV-2. Both NS3 and NS5 from MGTV displayed distributions of hydrophilic residues (Fig. 4) similar to those of DENV-2. Together with the observed similarities in the conserved motifs (Figs 1, 2), this result highlights the physiochemical similarities between the MGTV NS3 and NS5 proteins and those that form the flaviviral replication complex.

Codon usage - RNA viruses evolve quickly; however, particular genomic sites may be more conserved due to purifying selection. The resulting genome reflects these opposing pressures and specific sites may reflect more conservative forces in the form of signatures that may be identified by an analysis of their base composition. We explored the NS3 and NS5 nt sequences from MGTV and from other flaviviruses by analysing their base composition. We used three different approaches [Nc (effective number of codons), overall GC (G + C bases composition) and GC3 (G + C content of third-base codon position) and CAI] that have been applied to analyse codon bias; these methods are explained in the Supplementary data (Fig. S7) together with details of the results of the analy-
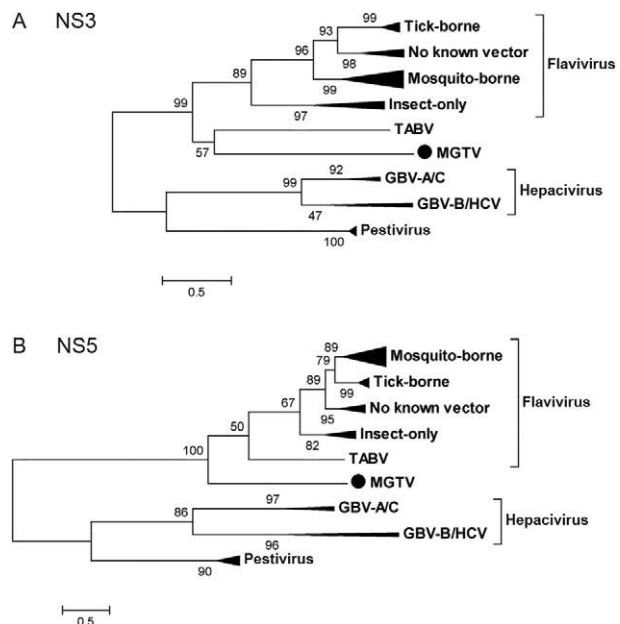

Fig. 3: phylogenetic analysis of nonstructural (NS)3 and NS5 sequences of Mogiana tick virus (MGTV) virus within the Flaviviridae family. Maximum likelihood analyses were performed with 1,000 bootstrap replicates. The evolutionary distances were computed using the JTT matrix-based method. A: NS3 tree; B: NS5 tree. Virus groups were condensed for better visualisation and all viruses used in trees construction are listed in Supplementary data (Table SIII). The bar at the bottom indicates 50% amino acid divergence. GBV: GB virus; HCV: hepatitis C virus; TABV: Tamana bat virus (TABV).

ses. All the codon usage computations were performed using the CAIcal server (Puigbò et al. 2008a). The CAI calculation for NS3 and NS5 from the potentially novel virus indicated similarity with flaviviral codon usage.
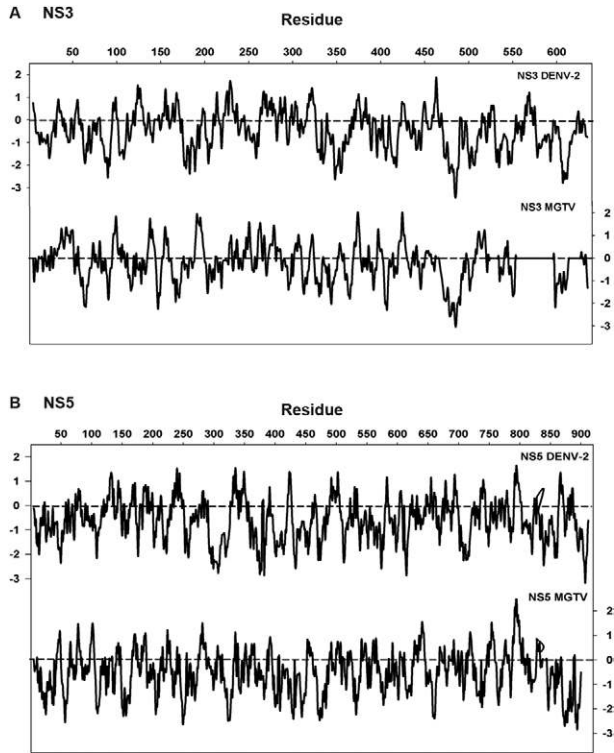

A NS3

B NS5

Fig. 4: hydropathy profile. Hydropathy values were determinate by Kyte and Doolittle scale using ProstScale web-based tool. Plot of non-structural (NS)3 (A) and NS5 (B) of Mogiana tick virus (MGTV). Both are compared with dengue virus type 2 (DENV-2) proteins. Negative scores mean a hydrophilic residue. Dashed line is the upper threshold for hydrophilicity.

Because the CAI value (e-CAI) indicates a codon bias towards a reference set, we used the reference set from the codon usage table of Flavivirus spp to determine the e-CAI of the sequences of MGTV NS3 and NS5 and of the flaviviruses addressed in this study. A normalised e-CAI of ≥ 1.0 indicates that the observed CAI is equal to or greater than the expected value (eCAI); such a result could be interpreted as a codon usage adaptation towards the Flavivirus genus. The normalised CAI_{NS3} and CAI_{NS5} values for MGTV were 1.00247 and 1.00732, respectively, suggesting that both genes indeed exhibit a codon usage adaptation towards the Flavivirus genus (Fig. 5). Interestingly, the e-CAI of both MGTV genes showed greater accordance with the Flavivirus codon usage patterns than some well-described flaviviruses.

The circulation of RNA arboviruses in nature, such as mosquito-borne and tick-borne Flavivirus species, rely on their capacity of replication in a range of hosts (vertebrates and invertebrates). Some experimental studies of the evolution of RNA arboviruses have shown that the host alternation cycles constrain the adaptation and evolution of these viruses towards a single host, corroborating the theory that an adaptation selected by one host involves a fitness trade-off for the other (Weaver et al. 1999, Coffey et al. 2008). Analyses of the dinucleotide compositions and codon usage patterns of Flaviviridae viruses and their hosts have suggested that host-induced pressure shapes the viral codon usage pattern (Lobo et al. 2009): the NKV mosquito-borne and tick-borne groups displayed dinucleotide usage patterns that were more closely related to the vertebrate genomic signature.

We used the normalised CAI data to explore the relationship between the codon usage patterns of MGTV and its potential invertebrate host, the cattle tick *R. microplus*. Furthermore, we extrapolated this analysis to virtual vertebrate (bovine and human) hosts because we had observed the replication of MGTV in mammalian cells (Vero and BHK) and arthropod cells (BME 26 and
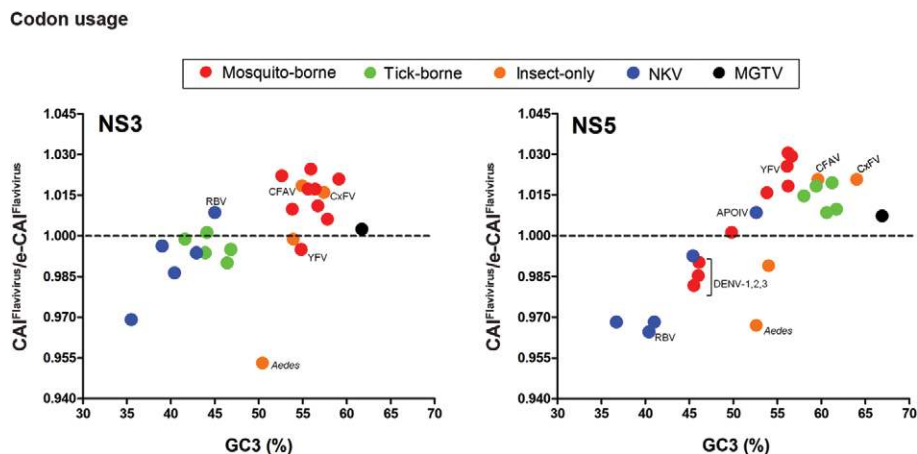

Fig. 5: codon usage pattern in Mogiana tick virus (MGTV) sequences. Codon bias in MGTV sequences were analysed with CAIcal Server. Codon usage adaptation index (CAI) calculation used Flavivirus sp. codon usage table as reference. The values were normalised with the expected CAI value (eCAI) for codon usage pattern in Flavivirus, calculated with input sequences. The dashed line indicates the threshold for which equal or higher values are interpreted as a codon usage adaptation towards Flavivirus codon usage pattern. APOIV: Apoi virus; CFAV: cell fusing agent virus; CxFV: *Culex* flavivirus; DENV: dengue virus; NKV: no known vector; NS: nonstructural; RBV: Rio Bravo virus; YFV: yellow fever virus.

C6/36), which suggested that the virus was a true arbovirus. The CAI and eCAI values of the NS3 and NS5 genes were calculated using the codon usage tables for *R. microplus*, *B. taurus* and *H. sapiens* and normalised CAI values were obtained for each condition (Table II). The three normalised CAI values, referring to the hosts, fluctuated for NS3 and were more stable for NS5. However, all three values were lower than 1.0, indicating a constrained codon usage adaptation towards a single host (neither a tick nor a vertebrate). This finding was theoretically consistent with what it would be expect for arboviruses. In conclusion, the codon usage analyses revealed characteristics of MGTV codon usage that are similar to those of flaviviruses.

Molecular detection of the NS3 and NS5 sequences from MGTV in field samples of ticks - The cattle tick *R. microplus* is broadly distributed in midwestern, southeastern and southern Brazil (Estrada-Pena 1999), where it causes heavy infestations in dairy cattle and is the main factor limiting animal health and production. We obtained a collection of tick samples from 19 dairy farms located in these three regions and searched for potentially novel viruses in these field samples through the molecular detection of partial sequences of NS3 (using primer set 317) and NS5 (using primer set 401). Different stages of the *R. microplus* life cycle were evaluated and both DNA and RNA from identical tick samples were used as templates for the reactions. Tick actin primers were used for positive control experiments. PCR amplifications using each set of primers (317, 401 and actin) were performed for each of the 81 gDNA and 81 cDNA samples. The agarose gel electrophoresis analysis revealed that there were no products of the PCR amplification of the gDNA templates, except from the reactions using actin primers (data not shown), as previously observed (Supplementary data, Fig. S1).

Interestingly, most of the analysed farms provided tick samples that were PCR-positive for the MGTV NS3 and/or NS5 sequences (Fig. 6). The analysed tick samples were collected from a large area in Brazil that includes the main cattle-producing regions (Fig. 6A). Most of the sampled farms received positive results for at least one tick life stage (Fig. 6B). The MGTV NS3 and NS5 fragments were detected in 42 samples, representing 51.8% of the total tick pools. Notably, the NS3 and/

or NS5 fragments were also detected in most of the UL samples. Because this is an unfed life stage, it is possible that the vertical transmission of this virus occurs in nature. The only tick samples in which the NS3 and NS5 fragments were not detected were from farms located in Presidente Prudente (SP) and Ribas do Rio Pardo, state of Mato Grosso do Sul (MS). Interestingly, although these regions are separated by the Paraná River, they are joined by a highway along which there are several abattoirs that receive carcasses from MS.

The positive samples (Fig. 6B) were sequenced to confirm that they corresponded to the same fragments found in the *R. microplus* SG cDNA library. The aa compositions of these samples were identical to those of contigs 317 and 401 (data not shown). However, we observed synonymous substitutions along the sequences, a common feature of RNA viruses due to the high mutation rates and consequently rapid evolution of RNA genomes (Holland et al. 1982). A high level of genetic diversity has been reported in populations of West Nile virus, a mosquito-borne flavivirus; this diversity confers a fitness benefit to mosquitoes (Jerzak et al. 2005, Fitzpatrick et al. 2010). To examine the diversity among our samples, phylogenetic trees were constructed for sequences 317 (NS3-derived) and 401 (NS5-derived) (Fig. 7). The sequences were distributed along numerous branches, demonstrating high sequence variability (nt substitutions) in the NS3 (Fig. 7A) and NS5 trees (Fig. 7B).

To confirm the origin of the samples, we highlighted several branches composed of samples from the same farm. Branch I of the NS3 tree (Fig. 7A) was formed by samples from farm Q (Santa Vitória,MG, Southeast Region) and branch I of the NS5 tree (Fig. 7B) was formed by samples from farm M [Piracanjuba, state of Goiás (GO), Central-West Region]. In both phylogenetic trees, branch II comprised samples from farm N (Itauçu, GO), whereas branches III and IV of the NS3 tree were formed by samples from farms T (Água Clara, MS) and H (South Region), respectively (Fig. 7A). Samples from farm X (Ribeirão Preto, Southeast Region) were grouped in both phylogenetic trees, on branches V and III of the NS3 and NS5 trees, respectively.

Final remarks - The original viral sequences described herein were detected in the SGs of female *R. microplus* ticks and in UL, suggesting a potential viral

TABLE II
Codon usage pattern observed for nonstructural (NS) genes in Mogiana tick virus and its relation with potential hosts

Gene	Length (bp)	Cattle tick			Bovine			Human		
		CAI_{Rm}	$eCAI_{Rm}$*	$CAI_{Rm}/eCAI_{Rm}$	CAI_{Bt}	$eCAI_{Bt}$*	$CAI_{Bt}/eCAI_{Bt}$	CAI_{Hs}	$eCAI_{Hs}$*	$CAI_{Hs}/eCAI_{Hs}$
NS3	1,662	0.740	0.776	0.95361	0.687	0.745	0.92215	0.729	0.781	0.93342
NS5	2,598	0.757	0.776	0.97551	0.724	0.745	0.97181	0.764	0.781	0.97823

codon usage tables of cattle tick, bovine and human were used as reference set for codon usage adaptation index (CAI) computation (asterisk means p = 0.05). eCAI: expected CAI value; *Rm*: *Rhipicephalus microplus*; *Bt*: *Bos taurus*; *Hs*: *Homo sapiens*.

transmission through saliva and the maintenance of a viral reservoir in ticks by vertical transmission. More pools of larvae should be analysed to confirm this possible mechanism of transmission because we could not isolate the virus from the UL, only from the female ticks. The vertical transmission of arboviruses in mosquitoes has been reported for mosquito-borne Flavivirus viruses, such as DENV (Rosen et al. 1983). This phenomenon has also been described in ixodid ticks; the best known case is the transovarial transmission of tick-borne encephalitis virus (Danielova et al. 2002). An interesting survival strategy of arboviruses in nature involves vertical transmission in the arthropod host so that the viruses are maintained in the environment even under adverse conditions. *R. microplus* is a monoxenic tick, i.e., it spends its entire cycle on a single host. Thus, it is essential that the UL are vertically infected to complete the parasite's life cycle, as is the case for Babesia protozoa transmission by *Rhipicephalus* spp (Guglielmone 1995).

Brazil has the largest number of commercial cattle worldwide. Other species of ticks also affect cattle, such as *Amblyomma* spp, which are heteroxenic ticks that parasitise many different hosts, including humans. In a study concerning the potential transmission of tick-borne pathogens in Brazil, a Flavivirus was isolated from *Amblyomma cajennense* ticks (Figueiredo et al. 1999); however, this virus has not been further characterised. Therefore, increasing awareness about the possibilities of emerging zoonoses is important because other tick species, such as those of the *Amblyomma* genus, can infest cattle and then parasitise humans, occasionally infecting them with a new pathogen.

Our findings suggest that the potential novel virus (MGTV) may be an arbovirus belonging to Flavivirus genus because MGTV was able to infect both vertebrate and invertebrate cells and because viral sequences were found in tick SGs. The molecular characterisation of NS3 and NS5, the largest and most highly conserved flaviviral proteins, showed that MGTV is more closely related to the Flavivirus genus, although it is a highly divergent member of this genus. The complete genome sequence of this novel virus will be necessary to define its correct taxonomic position. Unexpectedly, the results of the deep sequencing of viral RNA purified from cul-
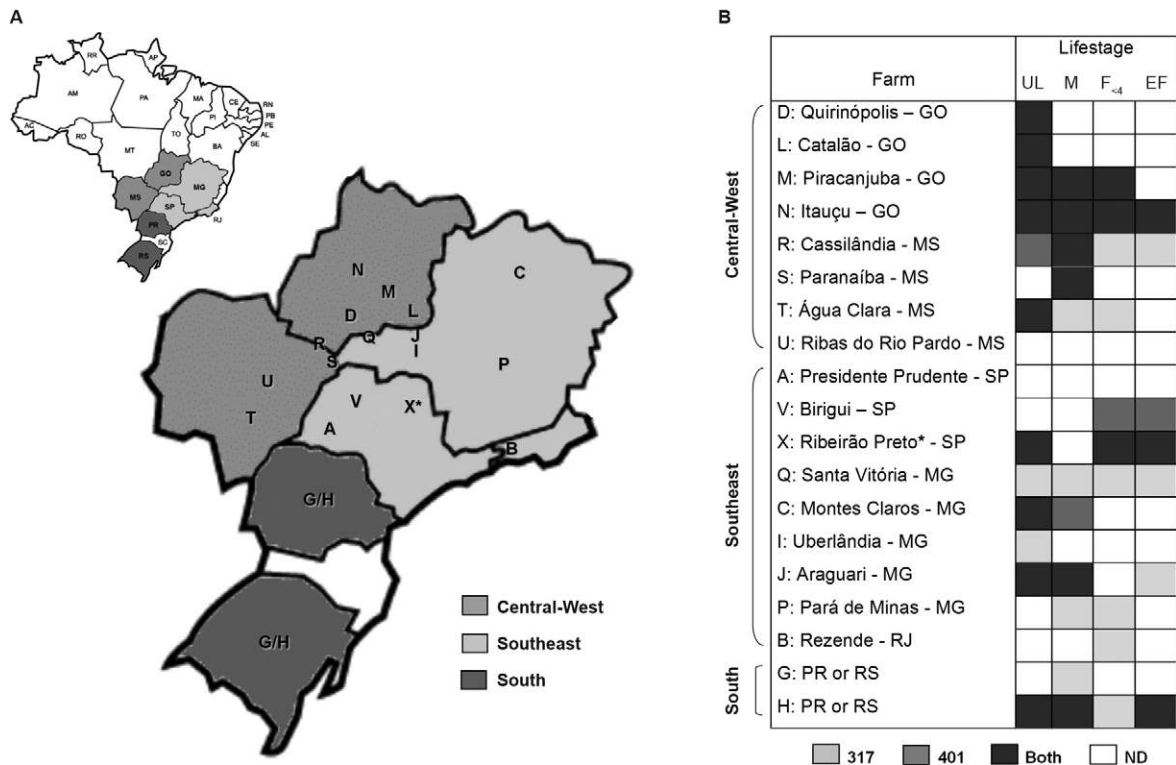

Fig. 6: molecular detection of viral transcripts in total RNA of tick samples collected in farms from seven Brazilian states. Detections were done through reverse-transcription-polymerase chain reaction using primers that target the 317 contig sequence [nonstructural (NS)3 fragment] or primers that target the 401 contig sequence (NS5 fragment). A: tick samples were collected from seven Brazilian states: Mato Grosso do Sul (MS) and Goiás (GO) (Central-West Region), São Paulo (SP), Minas Gerais (MG) and Rio de Janeiro (RJ) (Southeast Region), Paraná (PR) and Rio Grande do Sul (RS) (Southeast Region); B: panel for molecular detection linking farms to results for primers 317 (light grey), 401 (medium grey), both (dark grey) or no detection (ND) using any of the primers. Life stages of ticks: EF: engorged female (F); $F_{<4}$: F less than 4 mm (before the rapid engorgement phase of feeding); M: male; UL: unfed larvae; *: samples from farm X (Ribeirão Preto, SP) were the only ones that presented positive detection when 2,743 primers were tested. Other Brazilian states: AC: Acre; AL: Alagoas; AP: Amapá; AM: Amazonas; BA: Bahia; CE: Ceará; DF: Distrito Federal; ES: Espírito Santo; MA: Maranhão; MT: Mato Grosso; PA: Pará; PB: Paraíba; PE: Pernambuco; PI: Piauí; RN: Rio Grande do Norte; RO: Rondônia; RR: Roraima; SC: Santa Catarina; SE: Sergipe; TO: Tocantins;
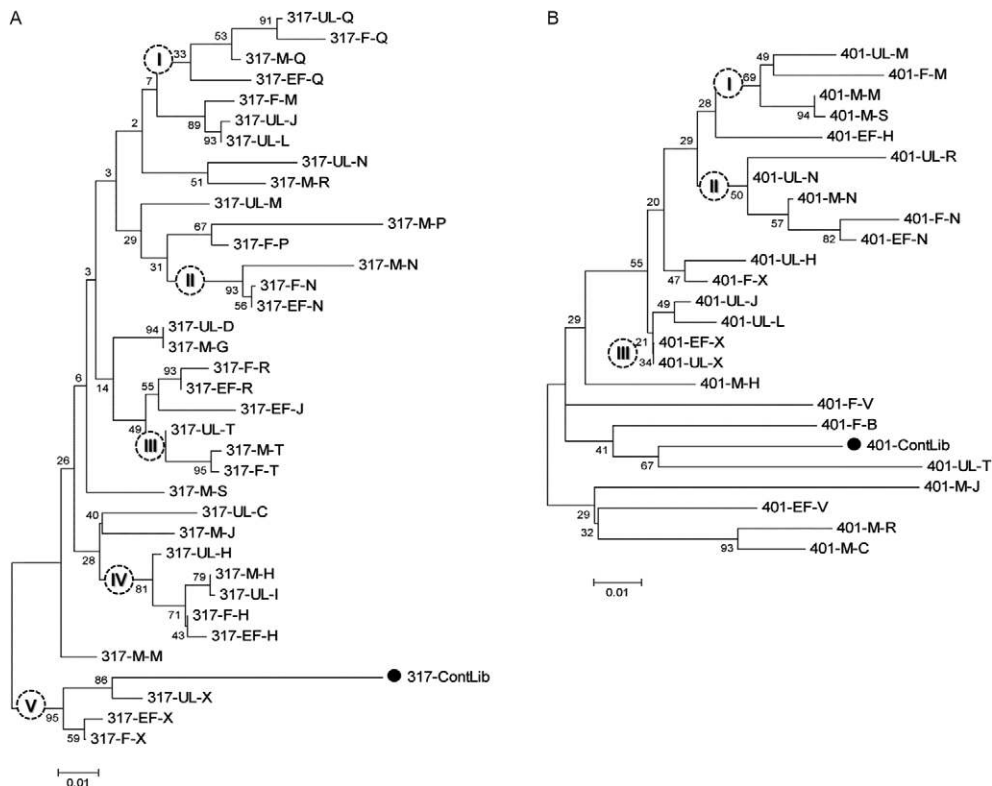
Fig. 7: phylogenetic trees of nonstructural (NS)3/NS5 sequences from positive tick samples. Sequences are named according to detection primers (317 or 401) followed by tick life stage [UL: unfed larvae; EF: engorged female (F) less than 4 mm; M: male] and the farm alphabetic label (see Table I for corresponding location). A: 36 sequences positive for detection using primers 317 (NS3 fragment) plus original contig sequence from the complementary DNA **(cDNA) library (named ContLib highlighted in closed circle); B: 21 sequences positive for detection using prim**ers 401 (NS5 fragment) plus original contig sequence from the cDNA library (named ContLib highlighted in closed circle); I-V: branches composed of samples from same farm. Neighbour-joining analysis was performed with 1,000 bootstrap replicates. The bar at the bottom indicates 1% nucleotide substitution.

tured infected cells did not provide the complete genome sequence of MGTV. This failure may have been caused by a technical artefact during the isolation and purification of the virus from the Vero cell culture, during the subsequent processing of the samples to obtain the viral RNA, or during the freezing and thawing prior to sequencing. Alternatively, this difficulty in obtaining the full genome sequence may reflect in inherent characteristics of this novel virus, such as the low VL observed. This virus is being further passaged to increase the VL.

The PCR assays as molecular detection of viral RNA fragments for molecular detection of viral RNA fragments for assigning viral infection has been debated. For example, Telis (2012) criticised a study by Bingham et al. (2012) that described the presence of the eastern equine encephalomyelitis virus in snakes using quantitative RT-PCR. Despite this controversy, our findings should be publicised due to their importance for Public Health authorities, who should be alerted to the possibility of emerging zoonoses and/or tick-borne viral diseases. In addition to the complete genome sequence, the potential transmission cycle and pathogenicity of MGTV have yet to be defined. Further research to characterise MGTV genotypically and phenotypically is important.

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res 25*: 3389-3402.

Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res 40*: W597-W603.

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*: 781-790.

Bingham AM, Graham SP, Burkett-Cadena ND, White GS, Unnasch TR 2012. Detection of eastern equine encephalomyelitis virus RNA in North American snakes. *Am J Trop Med Hyg 87*: 1140-1144.

Bollati M, Alvarez K, Assenberg R, Baronti C, Canard B, Cook S, Coutard B, Decroly E, de Lamballerie X, Gould EA, Grard G,

Grimes JM, Hilgenfeld R, Jansson AM, Malet H, Mancini EJ, Mastrangelo E, Mattevi A, Milani M, Moureau G, Neyts J, Owens RJ, Ren J, Selisko B, Speroni S, Steuber H, Stuart DI, Unge T, Bolognesi M 2010. Structure and functionality in flavivirus NS-proteins: perspectives for drug design. *Antiviral Res 87*: 125-148.

Bonaldo MC, Mello SM, Trindade GF, Rangel AA, Duarte AS, Oliveira PJ, Freire MS, Kubelka CF, Galler R 2007. Construction and characterization of recombinant flaviviruses bearing insertions between E and NS1 genes. *Virol J 4*: 115.

Carbone A, Zinovyev A, Képès F 2003. Codon adaptation index as a measure of dominating codon bias. *Bioinformatics 19*: 2005-2015.

Chambers TJ, Hahn CS, Galler R, Rice CM 1990. Flavivirus genome organization, expression and replication. *Annu Rev Microbiol 44*: 649-688.

Coffey LL, Vasilakis N, Brault AC, Powers AM, Tripet F, Weaver SC 2008. Arbovirus evolution in vivo is constrained by host alternation. *Proc Natl Acad Sci USA 105*: 6970-6975.

Cook S, Holmes EC 2006. A multigene analysis of the phylogenetic relationships among the flaviviruses (family Flaviviridae) and the evolution of vector transmission. *Arch Virol 151*: 309-325.

Cope DR, Iason GR, Gordon IJ 2004. Disease reservoirs in complex systems: a comment on recent work by Laurenson et al. *J Anim Ecol 73*: 807-810.

Crochu S, Cook S, Attoui H, Charrel RN, de Chesse R, Belhouchet M, Lemasson JJ, 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*: 1971-1980.

Danielova V, Holubova J, Pejcoch M, Daniel M 2002. Potential significance of transovarial transmission in the circulation of tick-borne encephalitis virus. *Folia Parasitol (Praha) 49*: 323-325.

Davidson AD 2009. Chapter 2. New insights into flavivirus nonstructural protein 5. *Adv Virus Res 74*: 41-101.

de la Fuente J, Maritz-Olivier C, Naranjo V, Ayoubi P, Nijhof AM, Almazan C, Canales M, de la Lastra JMP, Galindo RC, Blouin EF, Gortazar C, Jongejan F, Kocan KM 2008. Evidence of the role of tick subolesin in gene expression. *BMC Genomics 9*: e372.

Dumont S, Cheng W, Serebrov V, Beran RK, Tinoco Jr I, Pyle AM, Bustamante C 2006. RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. *Nature 439*: 105-108.

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.

Estrada-Pena A 1999. Geostatistics and remote sensing using NOAA-AVHRR satellite imagery as predictive tools in tick distribution and habitat suitability estimations for *Boophilus microplus* (Acari: Ixodidae) in South America. National oceanographic and atmosphere administration-advanced very high resolution radiometer. *Vet Parasitol 81*: 73-82.

Evans LH, Alamgir AS, Owens N, Weber N, Virtaneva K, Barbian K, Babar A, Malik F, Rosenke K 2009. Mobilization of endogenous retroviruses in mice after infection with an exogenous retrovirus. *J Virol 83*: 2429-2435.

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.

Ferron F, Bussetta C, Dutartre H, Canard B 2005. The modeled structure of the RNA dependent RNA polymerase of GBV-C virus suggests a role for motif E in Flaviviridae RNA polymerases. *BMC Bioinformatics 6*: e255.

Figueiredo LT, Badra SJ, Pereira LE, Szabo MP 1999. Report on ticks collected in the Southeast and Mid-West regions of Brazil: analyzing the potential transmission of tick-borne pathogens to man. *Rev Soc Bras Med Trop 32*: 613-619.

Fitzpatrick KA, Deardorff ER, Pesko K, Brackney DE, Zhang B, Bedrick E, Shi PY, Ebel GD 2010. Population variation of West Nile virus confers a host-specific fitness benefit in mosquitoes. *Virology 404*: 89-95.

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

Gaunt MW, Sall AA, de Lamballerie X, Falconar AK, Dzhivanian TI, Gould EA 2001. Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. *J Gen Virol 82*: 1867-1876.

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.

Gorbalenya AE, Koonin EV 1993. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol 3*: 419-429.

Guglielmone AA 1995. Epidemiology of babesiosis and anaplasmosis in South and Central America. *Vet Parasitol 57*: 109-119.

Hall TA 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser 41*: 95-98.

Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, VandePol S 1982. Rapid evolution of RNA genomes. *Science 215*: 1577-1585.

Huang X, Madan A 1999. CAP3: a DNA sequence assembly program. *Genome Res 9*: 868-877.

ICTV - International Committee on Taxonomy of Viruses 1995. Virus taxonomy: classification and nomenclature of viruses. Sixth report of the International Committee on Taxonomy of Viruses. In FA Murphy, CM Fauquet, DHL Bishop, SA Ghabrial, AW Jarvis, GP Martelli, MA Mayo, MD Summers (eds.), *Virus taxonomy: classification and nomenclature of viruses. Sixth report of the International Committee on Taxonomy of Viruses*, Spring-Verlag, New York, 586 pp.

Issur M, Geiss BJ, Bougie I, Picard-Jean F, Despins S, Mayette J, Hobdey SE, Bisaillon M 2009. The flavivirus NS5 protein is a true RNA guanylyltransferase that catalyzes a two-step reaction to form the RNA cap structure. *RNA 15*: 2340-2350.

Jerzak G, Bernard KA, Kramer LD, Ebel GD 2005. Genetic variation in West Nile virus from naturally infected mosquitoes and birds suggests quasispecies structure and strong purifying selection. *J Gen Virol 86*: 2175-2183.

Jongejan F, Uilenberg G 2004. The global importance of ticks. *Parasitology 129* (Suppl.): S3-S14.

Karim S, Singh P, Ribeiro JM 2011. A deep insight into the sialotranscriptome of the Gulf Coast tick, *Amblyomma maculatum*. *PLoS ONE 6*: e28525.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol 14*: R36.

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.

Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB 1998. Phylogeny of the genus Flavivirus. *J Virol 72*: 73-83.

Kyte J, Doolittle RF 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol 157*: 105-132.

Lain S, Riechmann JL, Martin MT, Garcia JA 1989. Homologous potyvirus and flavivirus proteins belonging to a superfamily of helicase-like proteins. *Gene 82*: 357-362.

Langmead B, Salzberg SL 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods 9*: 357-359.

Laurenson MK, Norman RA, Gilbert L, Reid HW, Hudson PJ 2003. Identifying disease reservoirs in complex systems: mountain hares as reservoirs of ticks and louping-ill virus, pathogens of red grouse. *J Anim Ecol 72*: 177-185.

Lobo FP, Mota BE, Pena SD, Azevedo V, Macedo AM, Tauch A, Machado CR, Franco GR 2009. Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PLoS ONE 4*: e6282.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience 1*: 18.

Ma H, Ma Y, Ma W, Williams DK, Galvin TA, Khan AS 2011. Chemical induction of endogenous retrovirus particles from the vero cell line of African green monkeys. *J Virol 85*: 6579-6588.

Maher-Sturgess SL, Forrester NL, Wayper PJ, Gould EA, Hall RA, Barnard RT, Gibbs MJ 2008. Universal primers that amplify RNA from all three flavivirus subgroups. *Virol J 5*: 16.

Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res 30*: 281-283.

Maruyama SR, Anatriello E, Anderson JM, Ribeiro JM, Brandão LG, Valenzuela JG, Ferreira BR, Garcia GR, Szabo MP, Patel S, Bishop R, de Miranda-Santos IK 2010. The expression of genes coding for distinct types of glycine-rich proteins varies according to the biology of three metastriate ticks, *Rhipicephalus (Boophilus) microplus*, *Rhipicephalus sanguineus* and *Amblyomma cajennense*. *BMC Genomics 11*: e363.

Megaw MW 1978. Virus-like particles pathogenic to salivary glands of the tick *Boophilus microplus*. *Nature 271*: 483-484.

Norman R, Bowers RG, Begon M, Hudson PJ 1999. Persistence of tick-horne virus in the presence of multiple host species: tick reservoirs and parasite mediated competition. *J Theor Biol 200*: 111-118.

O'Donovan C, Martin MJ, Gattiker A, Gasteiger E, Bairoch A, Apweiler R 2002. High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief Bioinform 3*: 275-284.

Orlinger KK, Hofmeister Y, Fritz R, Holzer GW, Falkner FG, Unger B, Loew-Baselli A, Poellabauer EM, Ehrlich HJ, Barrett PN, Kreil TR 2011. A tick-borne encephalitis virus vaccine based on the European prototype strain induces broadly reactive cross-neutralizing antibodies in humans. *J Infect Dis 203*: 1556-1564.

Poch O, Sauvaget I, Delarue M, Tordo N 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J 8*: 3867-3874.

Polgar L 2005. The catalytic triad of serine peptidases. *Cell Mol Life Sci 62*: 2161-2172.

Puigbò P, Bravo IG, Garcia-Vallvé S 2008a. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct 3*: 38.

Puigbò P, Bravo IG, Garcia-Vallvé S 2008b. e-CAI: a novel server to estimate an expected value of codon adaptation index (eCAI). *BMC Bioinformatics 9*: 65.

Ribeiro JM, Alarcon-Chaidez F, Francischetti IM, Mans BJ, Mather TN, Valenzuela JG, Wikel SK 2006. An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochem Mol Biol 36*: 111-129.

Rosen L, Shroyer DA, Tesh RB, Freier JE, Lien JC 1983. Transovarial transmission of dengue viruses by mosquitoes: *Aedes albopictus* and *Aedes aegypti*. *Am J Trop Med Hyg 32*: 1108-1119.

Schubert AM, Putonti C 2010. Evolution of the sequence composition of flaviviruses. *Infect Genet Evol 10*: 129-136.

Sharp PM, Li WH 1987. The codon adaptation index-a measure of directional synonymous codon usage bias and its potential applications. *Nucleic Acids Res 15*: 1281-1295.

Tamura K, Dudley J, Nei M, Kumar S 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol 24*: 1596-1599.

Telis G 2012. Vipers go viral. Available from: news.sciencemag.org/sciencenow/2012/10/vipers-go-viral.html.

Thiel HJ, Gould EA, Heinz FX, Houghton M, Meyers G, Purcell RH, Rice CM 2005. Family Flaviviridae. In CM Fauquet, MA Mayo, J Maniloff, U Desselberger, LA Ball (eds.), *Virus taxonomy, VIII report of the International Committee on Taxonomy of Viruses*, Academic Press, San Diego, p. 981-998.

Thompson JD, Higgins DG, Gibson TJ 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*: 4673-4680.

Wan XF, Xu D, Kleinhofs A, Zhou J 2004. Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evol Biol 4*: 19.

Weaver SC, Brault AC, Kang W, Holland JJ 1999. Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. *J Virol 73*: 4316-4326.

Wright F 1990. The 'effective number of codons' used in a gene. *Gene 87*: 23-29.

Wu J, Bera AK, Kuhn RJ, Smith JL 2005. Structure of the Flavivirus helicase: implications for catalytic activity, protein interactions and proteolytic processing. *J Virol 79*: 10268-10277.
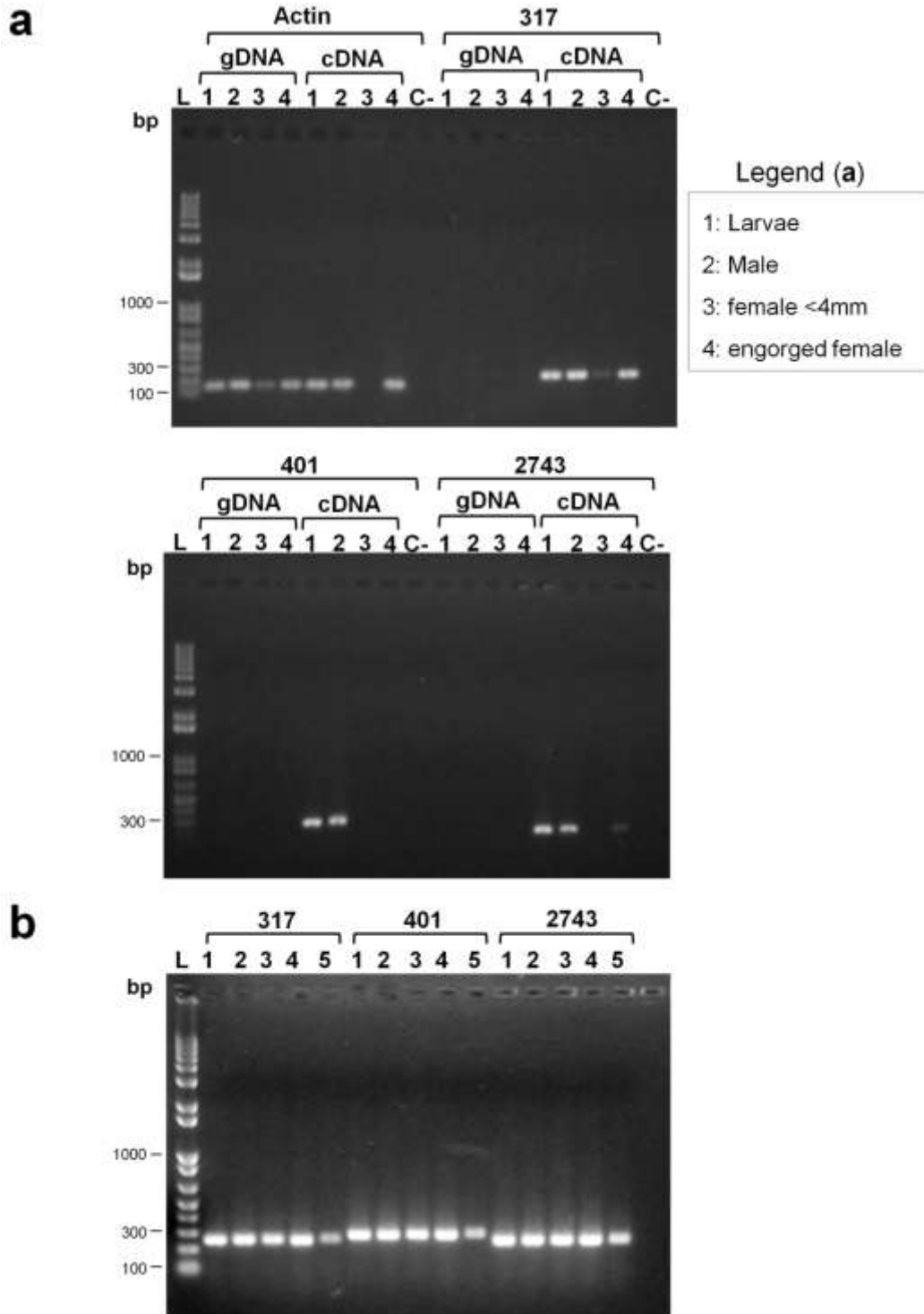
Fig. S1: molecular detection of Mogiana tick virus (MGTV) partial sequences in tick samples from different farms located in Mogiana region (Line: 1 Kb plus DNA ladder; C-: negative control). A: tick samples from a farm located in Araguari, state of Minas Gerais (1: larvae; 2: male; 3: female < 4 mm; 4: engorged female); B: tick samples from a farm located in Ribeirão Preto, state of São Paulo. Primers: actin, 317 [nonstructural (NS)3], 401 and 2,743 (NS5). Genomic DNA (gDNA) and RNA [represented by complementary DNA (cDNA)] from ticks were analysed. [1: pool of tick 1; 2: pool of tick 2; 3: pool of tick 4; 4: pool of tick 8 (all from females); 5: pool from larvae].
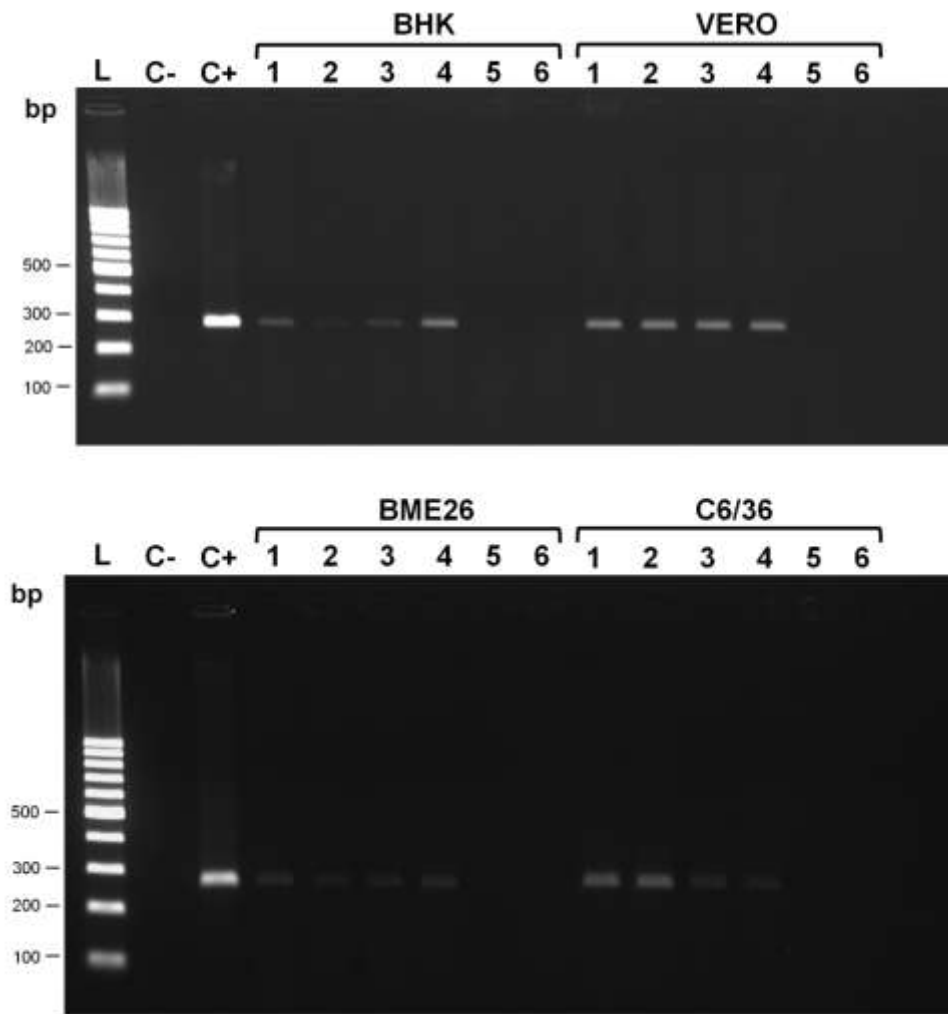
Fig. S2: detection of Mogiana tick virus in mammalian and arthropod cell cultures. Two percent agarose gel electrophoresis demonstrating the products generated by nonstructural 5 amplification (primer set 401-5-65 and 401-3-349) in baby hamster kidney (BHK), Vero (monkey), *Boophilus microplus* cattle tick (BME) 26 and C6/36 (mosquito) cell lines. Lane L: 100 bp DNA ladder; C-: reverse-transcription-polymerase chain reaction (RT-PCR) negative control; C+: RT-PCR positive control; 1: isolates from pool 1; 2: isolates from pool 2; 3: isolates from pool 4; 4: isolates from pool 8; 5: isolates from larval pool; 6: mock infected cells.
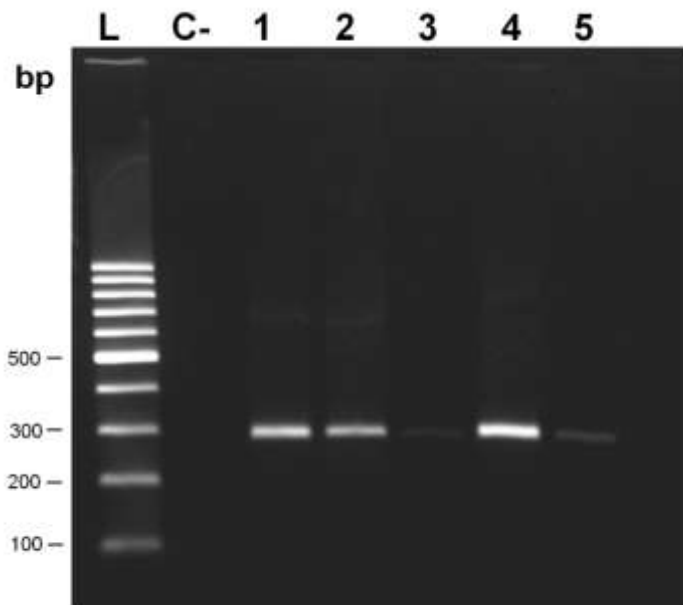
Fig. S3: molecular detection of purified Mogiana tick virus after ultracentrifugation of culture supernatant from second passage in Vero cells. Two percent agarose gel electrophoresis demonstrating the amplicons generated by nonstructural 5 amplification (primer set 401-5-65 and 401-3-349) in virus purified from Vero cells. Lane L: 100 bp DNA ladder; C-: reverse-transcription-polymerase chain reaction (RT-PCR) negative control; 1: purified pool 4 isolate; 2: purified pool 4 isolate treated with RNAse and DNAse (which was used in deep sequencing); 3: pool 4 isolate in culture supernatant before ultracentrifugation; 4: purified pool 1 isolate treated with RNAse and DNAse; 5: pool 1 isolate from culture supernatant before ultracentrifugation.

Fig. S4: multiple sequence alignment of nonstructural 3 from Mogiana tick virus and representative of tick-borne encephalitis virus (TBEV), mosquito-borne [dengue virus type 2 (DENV-2) and yellow fever virus (YFV)], insect-only [cell fusing agent virus (CFAV)] and not known vector [Apoi virus (APOIV) and Tamana bat virus (TABV)] flavivirus groups. Conserved residues and motifs are indicated by narrows and curly brackets, respectively.

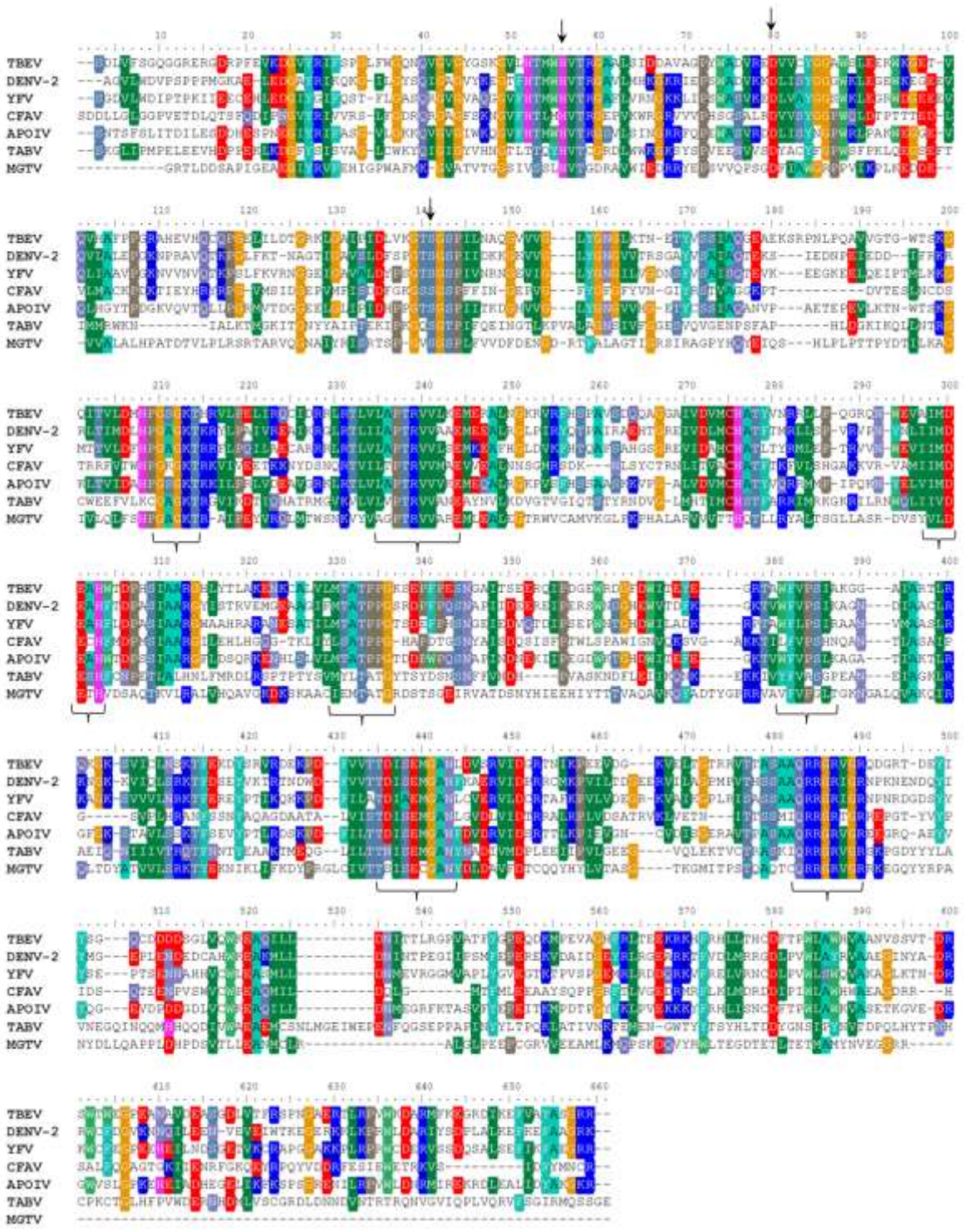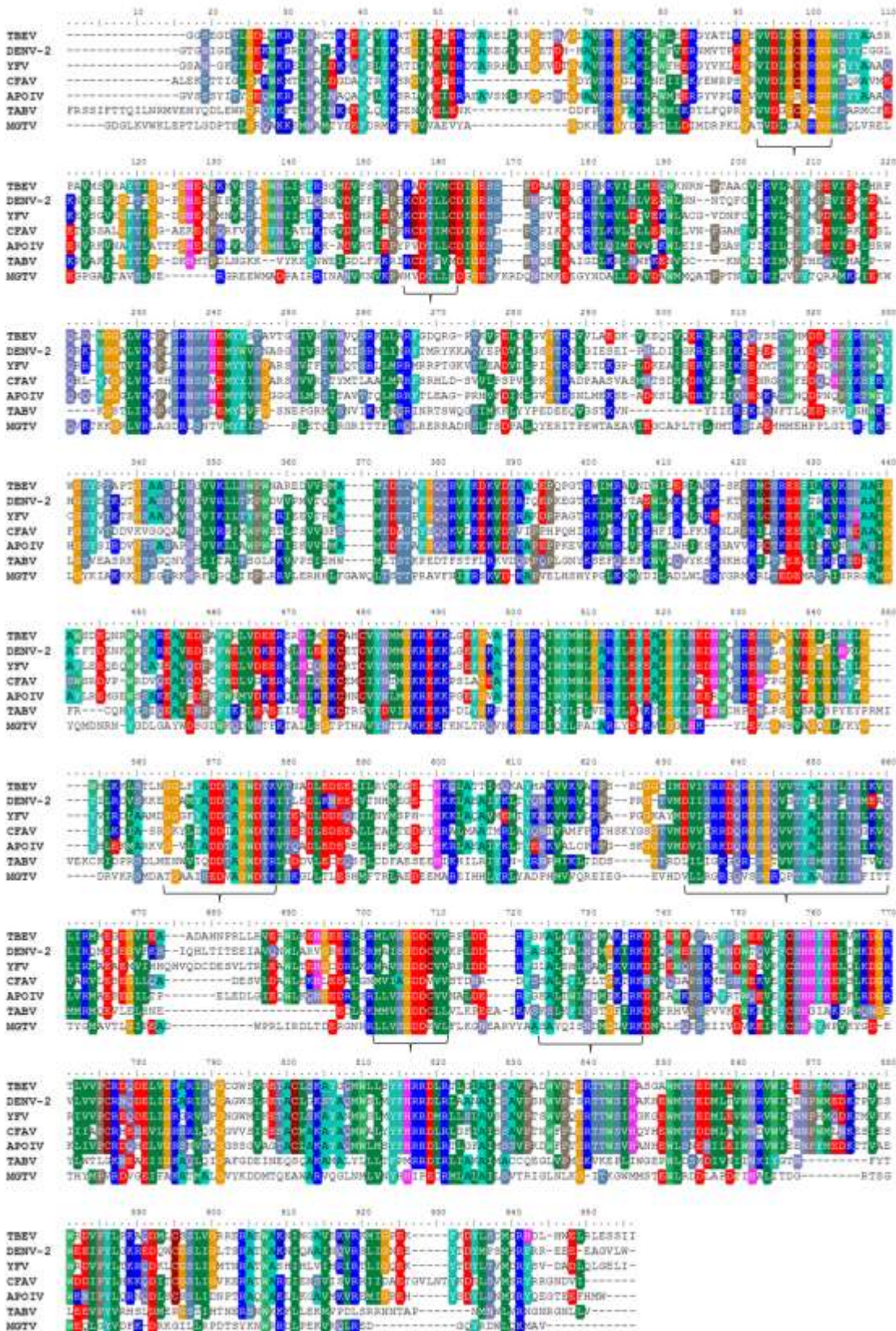Fig. S5: multiple sequence alignment of nonstructural 5 from Mogiana tick virus and representative of tick-borne encephalitis virus (TBEV), mosquito-borne [dengue virus type 2 (DENV-2) and yellow fever virus (YFV)], insect-only [cell fusing agent virus (CFAV)] and not known vector [Apoi virus (APOIV) and Tamana bat virus (TABV)] flavivirus groups. Conserved motifs are indicated by curly brackets.

TABLE SI

Viral sequences from National Center for Biotechnology Information (NCBI)-RefSeq used in this study

Virus	Abbreviation	Vector	RefSeq polyprotein	RefSeq NS3	RefSeq NS5
Louping ill virus	LIV	T	NC_001809.1	NP_740726.1	NP_740729.1
Langat virus	LGTV	T	NC_003690.1	NP_740299.1	NP_740302.1
Powassan virus	PWV	T	NC_003687.1	NP_775520.1	NP_775524.1
Alkhurma virus	ALKV	T	NC_004355.1	NP_775474.1	NP_775478.1
Tick-borne encephalitis virus	TBEV	T	NC_001672.1	NP_775507.1	NP_775511.1
Japanese encephalitis virus	JEV	M	NC_001437.1	NP_775670.1	NP_775674.1
West Nile virus	WNV	M	NC_009942.1	YP_001527884.1	YP_001527887.1
Dengue virus 2	DENV-2	M	NC_001474.2	NP_739587.2	NP_739590.2
Dengue virus 4	DENV-4	M	NC_002640.1	NP_740321.1	NP_740325.1
Dengue virus 3	DENV-3	M	NC_001475.2	YP_001531172.2	YP_001531176.2
Dengue virus 1	DENV-1	M	NC_001477.1	NP_722463.1	NP_722465.1
Usutu virus	USUV	M	NC_006551.1	YP_164814.1	YP_164818.1
Murray Valley encephalitis virus	MVEV	M	NC_000943.1	NP_722535.1	NP_722539.1
Yellow fever virus	YFV	M	NC_002031.1	NP_776005.1	NP_776009.1
Aedes flavivirus	-	I	NC_012932.1	YP_003084129.1	YP_003084132.1
Kamiti River virus	KRV	I	NC_005064.1	NP_937777.1	NP_937780.1
Culex flavivirus	CxFV	I	NC_008604.2	YP_006470615.1	YP_006470619.1
Cell fusing agent virus	CFAV	I	NC_001564.1	NP_776044.1	NP_776048.1
Montana myotis leukoencephalitis virus	MMLV	NKV	NC_004119.1	NP_775649.1	NP_775653.1
Modoc virus	MODV	NKV	NC_003635.1	NP_740264.1	NP_740267.1
Apoi virus	APOIV	NKV	NC_003676.1	NP_775684.1	NP_775688.1
Rio Bravo virus	RBV	NKV	NC_003675.1	NP_776076.1	NP_776080.1
Tamana bat virus	TABV	NKV	NC_003996.1	NP_776031.1	NP_776035.1
Pestivirus Giraffe-1[a]	-	-	NC_003678.	NP_777527.1	NP_777531.1-NP_777532.1
Border disease virus X818[a]	BDV	-	NC_003679.1	NP_777540.1	NP_777544.1-NP_777545.1
Classical swine fever virus[a]	CSFV	-	NC_002657.1	NP_777501.1	NP_777505.1-NP_777506.1
Bovine viral diarrhoea virus genotype 2[a]	BVDV-2	-	NC_002032.1	NP_777488.1	NP_777492.1-NP_777493.1
Bovine viral diarrhoea virus 1[a]	BVDV-1	-	NC_001461.1	NP_776266.1	NP_776270.1-NP_776271.1
Hepatitis C virus genotype 1	HCV	-	NC_004102.1	NP_803144.1	NP_751927.1-NP_751928.1
GB virus A[a]	GBV-A	-	-	NP_803213.1	NP_803216.1-NP_803217.1
GB virus B[a]	GBV-B	-	-	NP_757357.1	NP_757360.1-NP_757361.1
GB virus C[a]	GBV-C	-	-	NP_803205.1	NP_803208.1-NP_803209.1

a: sequences used in phylogenetic analysis; I: insect-only; M: mosquito; NKV: non known vector; T: tick.

TABLE SII

Deep sequencing data of viral RNA isolated from Mogiana tick virus-infected Vero cell culture

(Available from: fmrp.usp.br/imsantos at data download link. Password for file: a01b02c03)

➤ AnchC
    Contig 2841, Contig 3667

> Contig 2841-blastx

==> gi|123205972|ref|YP_001008348.1| polyprotein* [St. Louisencephalitis virus]
    Length = 3430

 Score = 23.5 bits (49), Expect = 0.59
 Identities = 13/33 (39%), Positives = 16/33 (48%)
 Frame = -3

Query: 119 LAHSVVAARPLLPTRTLQGSLAHGVVPVRHLLA 21
    L V PL + + GSL G PVR +LA
Sbjct: 17 LKRGVSRVNPLTGLKRILGSLLDGRGPVRFILA 49

*capsid protein C from St. Louis encephalitis virus, region 6..106

>Contig2841_sequence_frame-3
WGHAPFWPLGALQDP<mark>LAHSVVAARPLLPTRTLQGSLAHGVVPVRHLLA</mark>AKDSAG

>Contig3667-blastx

==> gi|20178609|ref|NP_620044.1| polyprotein [Rio Bravo virus]
    Length = 3379

 Score = 27.3 bits (59), Expect = 0.18
 Identities = 13/28 (46%), Positives = 17/28 (60%)
 Frame = +2

Query: 167 EVCFPLPDLRSRSSLQVSRNVWGWYLSP 250
      E FP+ L ++QVS+N GW LSP
Sbjct: 92 ESLFPIMFLTGLMAMQVSQNGDGWLLSP 119

*anchored core protein C from Rio Bravo virus, region 1..102

>Contig3667_sequence_frame+2
AYSGCENESREVPLFGPTWPRALKRGGCRSWWGKPNSQADQSRVCGVGVHDRLVA
<mark>EVCFPLPDLRSRSSLQVSRNVWGWYLSP</mark>TPRFAVELXXXXXXXXXXXXXXPTLYST
RIDLTVWLIPRVNGRGQSTSCQYSL

preM(region_name="Flavi_propep", "Flavi_M")

    Contig 547, Contig 4248, Contig 3735, Contig 3663

>Contig547-blastx

==> gi|27697395|ref|NP_775678.1| PreM protein [Apoi virus]
    Length = 161

 Score = 26.9 bits (58), Expect = 0.12
 Identities = 18/65 (27%), Positives = 31/65 (47%), Gaps = 2/65 (3%)
 Frame = +2

Query: 32 KCQKSKTLFVLSLMHKKYSICVNCRHRAVQQISRTSPSC--ITETAHP*TTSDSPHSQLL
205
     +C+++ T ++L + + ++C R V+ + T P+C T T T D P S L
Sbjct: 40 ECEETMTYPCITLAATEEPVDLDCFCRDVKNVMVTYPTCKRNTRTRRDVTIQDHPPSVTL
99

Query: 206 ATSSL 220
       SL
Sbjct: 100 TKPSL 104

>Contig547_sequence_frame+2_ region_Flavi_propep
//QRFRIK`KCQKSKTLFVLSLMHKKYSICVNCRHRAVQQISRTSPSCITETAHP.TTSDSPHSQLLATSSLFSDS`
LWNPLT//

>Contig4248-blastx

==> gi|226377836|ref|YP_002790882.1| polyprotein* [Kedougou virus]
    Length = 3408

 Score = 27.7 bits (60), Expect = 0.50
 Identities = 12/40 (30%), Positives = 21/40 (52%)
 Frame = +1

Query: 637 PQRGPRAGALSPKDEPVLTTPRSRWPPNRAPTSHIQRNEQ 756
    P+R R+ +L P E L T W R+ +H+++ E+
Sbjct: 202 PRRSRRSVSLPPHTEKKLETRHESWLETRSYLAHLEKTER 241

*preM protein from Kedougou virus, region 121..282

>Contig4248_sequence_frame+1_ region_Flavi_M
DRQLSQWDSTVGIMAFWQKKTKKVTWATSYFLPVERHPHTPRCQRRPGNEPTQHLVSSXXXXXXX
XXSTPARDTHTPHLTPNT.GSRGMAGSPRRTGHPVRSVQGYA`PQRGPRAGALSPKDEPVLTTPRSR`
`WPPNRAPTSHIQRNEQ`VPAAHGAQAPRRAQDEPTVQAWGRHQPREASEALPSNGHXEVTTQAHQE
FPHYAGFRSRDRGSGIVAQTSTDPMEMRAWSLHPTSLHGYSILLPSHTSPHSHVPSPFPPLGSSNQTS
LSLANIPLIYLLPRTKAHLQLT

>Contig3735-blastx

==> gi|11528014|ref|NP_041724.2| unnamed protein product* [West Nile virus]
     Length = 3430
 Score = 26.9 bits (58), Expect = 0.58
 Identities = 17/57 (29%), Positives = 27/57 (47%)
 Frame = -2

Query: 219 VASRTGWMQSEVLSALSMFADVFLYKLFLETFNCHAHNSTSYNRELTGVFGFTWMNV
49
   VA+ GWM +FA + L +FNC ++ + L GV G TW+++
Sbjct: 260 VAAVIGWMLGSNTMQRVVFAILLLLVAPAYSFNCLGMSNRDF---LEGVSGATWVDL 313

*preM protein from West Nile virus, region 124..290

>Contig3735_sequence_frame-2_ region_Flavi_M
//LGQEWGCPAFNNLDALNEDSVDNILCKYALSARVSKAQGAGGSSVLDRCLWQGQLCGCQSECD
QEASVPLRWPAFLLHVA.AEGGRCWGTCSCSCFLIGNSRFLXXXXXXXXXXXXXXXXXXXXCVFFSAFR
IYTTWHLA<mark>VASRTGWMQSEVLSALSMFADVFLYKLFLETFNCHAHNSTSYNRELTGVFGFTWMNV</mark>

>Contig3663-blastx

==> gi|20178609|ref|NP_620044.1| polyprotein* [Rio Bravo virus]
     Length = 3379

 Score = 27.3 bits (59), Expect = 0.18
 Identities = 13/28 (46%), Positives = 17/28 (60%)
 Frame = +2

Query: 167 EVCFPLPDLRSRSSLQVSRNVWGWYLSP 250
       E FP+ L ++QVS+N GW LSP
Sbjct: 92 ESLFPIMFLTGLMAMQVSQNGDGWLLSP 119

*preM protein from Rio Bravo virus, region 103..262

>Contig3663_sequence_frame+2
AYSGCENESREVPLFGPTWPRALKRGGCRSWWGKPNSQADQSRVCGVGVHDRLVA<mark>EVCFPLPDL
RSRSSLQVSRNVWGWYLSP</mark>TPXXXXXXXSSHPLLSTDEPSSPTLYSTRIDLTVWLIPRVNGRGQSTSC
QYSL.GQGRGPTNCTNLSRVIRGILAQSDHLQRETGPYRGWRGLPSPREAGLLARA//

➢   Envelope (region_name="Flavi_glycoprot", "Flavi_glycop_C", "flavi_E_stem")
      Contig 553, Contig 561, Contig 4362

>Contig553-blastx

==> gi|25121895|ref|NP_740271.1| envelope protein [Louping ill virus]
     Length = 496

 Score = 28.5 bits (62), Expect = 0.018
 Identities = 18/51 (35%), Positives = 24/51 (47%), Gaps = 7/51 (13%)
 Frame = -2

Query: 189 ARAERTGPSGAMRRQSVGS-----SGDEGWCQGPGRRGKGSM--CLKAAME 58
       AR GP+ + +G+ D GW G GKGS+ C+KAA E

Sbjct: 72 ARCPTMGPAVLTEERQIGTVCKRDQSDRGWGNHCGLFGKGSIVACVKAACE 122

>Contig553_sequence_frame-2_ Flavi_glycoprot
ARAERTGPSGAMRRQSVGSSGDEGWCQGPGRRGKGSMCLKAAMELAVPAVKSGLQRTWTATRQ

>Contig561-blastx
==> gi|380877199|ref|YP_005352889.1| unnamed protein product* [Donggang virus]
    Length = 3444

 Score = 27.7 bits (60), Expect = 0.93
 Identities = 18/67 (26%), Positives = 26/67 (38%), Gaps = 6/67 (8%)
 Frame = +1

Query: 1717 YPADLRARFVTEEAIWEWDPGATWTVAEMDYMATILHAVIFSL------AVYTMLTDSGE 1878
        Y AD+ ++ + W D WT D+ + V FS +VYT+ G
Sbjct: 499
YTADMSGKWWLVKRDWYHDIALPWTAPSADFWHDMDRLVEFSTPHATKQSVYTLGDQEGA 558

Query: 1879 TDRVLGD 1899
        LGD
Sbjct: 559 MSTALGD 565

*Envelope protein from Donggang virus, region 299..797

>Contig561_sequence_frame+1_region_ Flavi_glycoprot
SGLPGFSVEGGHAQPVPPRSRKTTPTGREGWIPDWMTIAGDYFKRLQGKANVMGEGFDIILDEARD
LFPSTKEFALPGLIRQYLERRDVPLDSAFFLDVFLPLVLMALAITTNRWTRMALAVGVYLTGYYYV
AIMLAATSVVSLAFKAFAPREHKRGDVVIENGRLSVLGLTVVAVAATVGIHYYQPSPSLIFAVASLA
GFAAILMALPTIQYHGATDVAKMMIAVLLFAGVIYIASSFDVDKDQLYLFVKTGTPTYHIPRGEREN
AVKMDQVYNLARYYARYPADLRARFVTEEAIWEWDPGATWTVAEMDYMATILHAVIFSLAVYT
MLTDSGETDRVLGDIKMRILEKIKTTELFGDVTMAKITLAVSRVEWLVVIAGNLLHTYFALGLPGV
AMEILLAAGLMVPTYHLWSRTFRVMSYLRGTNGYRQPAEGLLPAPTIRSTQTQYTAGLATIVGIAA
VFVNLYYYLTGSNPYYVAHSVLVFAGTFVMVQDNDQLNGYPLLMLMAYTFNSPSLALIGSIYKKC
LGRVLWSRTT//

>Contig4362-blastx

==> gi|24418982|ref|NP_722531.1| envelope protein E [Murray Valle encephalitis virus]
    Length = 501

 Score = 28.5 bits (62), Expect = 0.55
 Identities = 29/106 (27%), Positives = 40/106 (37%), Gaps = 1/106 (0%)
 Frame = -3

Query: 2684
GSAVGACRKGFECLVRTSSGTEACPLGATAWNHSLPGETRGSEESRDHALLGRQFHQYVL 2505
        GS++G K F ++ + A LG TAW+ G S H + G F
Sbjct: 401 GSSIG---KAFSTTLKGAQRLAA--LGDTAWDFGSVGGVFNSIGKAVHQVFGGAFRTLFG
455

Query: 2504 WKEWILP-VMGVVLTPSWWPPPVPDKLTQRGSSAANGTVATLANRV 2370
        WI P ++G +L W DK A G + LA V
Sbjct: 456 GMSWISPGLLGALLL--WMGVNARDKSIALAFLATGGVLLFLATNV 499

>Contig4362_sequence_frame-3_region_ flavi_E_stem

APFLWRLPLPPALSSEAAGSAVGACRKGFECLVRTSSGTEACPLGATAWNHSLPGETRGSEESRDHALLGRQFHQYVLWKEWILPVMGVVLTPSWWPPPVPDKLTQRGSSAANGTVATLANRVTSVTWPPGAS//

NS1
Contig 3959, Contig 1327

>Contig3959-blastx

==> gi|119952255|ref|YP_950477.1| polyprotein* [Entebbe bat virus]
     Length = 3411

 Score = 28.9 bits (63), Expect = 0.17
 Identities = 37/131 (28%), Positives = 56/131 (42%), Gaps = 22/131 (16%)
 Frame = -1

Query: 901 RICPLSTQNLAH----TKQR----LSPGRPAKHKCSYKAEKLEGVGLCSSGTSVSQ--ED 752
            R+CPLS Q LA T +R L+ +H+ + E L +G +S D
Sbjct: 809 RLCPLSPQELASIIQATSERGACGLNSVDELEHRMWKEIEDEVNHVLDENGIDLSMVVGD
868

Query: 751 P-GHYRVQGKGF--------LGSRTWNKNPFPDPVEYESAKGRFPRRTEH-CPTTP--WY 608
            P G YR G F G +TW K F + VE ++ R ++ CP + W
Sbjct: 869 PMGVYRRGGMSFSNATRELSYGWKTWGKT-FYNAVERKNHSFIIDSRDQNECPDSQRVWN 927

Query: 607 GYVITYVGMGM 575
            +++ GMG+
Sbjct: 928 SFILEEFGMGL 938

*NS1 from Entebbe bat virus, region 778..1132

>Contig3959_sequence_frame-1
//VFGHRIWKRKRWFWAGWLPA<mark>RICPLSTQNLAHTKQRLSPGRPAKHKCSYKAEKLEGVGLCSSGT
SVSQEDPGHYRVQGKGFLGSRTWNKNPFPDPVEYESAKGRFPRRTEHCPTTPWYGYVITYVGMG</mark>M
LLPIPR//

>Contig1327-blastx

==> gi|27669991|ref|NP_775646.1| non-structural protein NS1 [Montana myotis leukoencephalitis virus]
     Length = 354

 Score = 27.3 bits (59), Expect = 0.39
 Identities = 18/52 (34%), Positives = 25/52 (48%)
 Frame = -3

Query: 626 HTLESLVSWNIFYSSIDHMASKGWLILNTVIGDPASYLGWVAGTPQPSWVLW 471
           HT+E L W I ++ + M LIL +G PAS L V G + + W
Sbjct: 226 HTVECL--WPITHTLGNRMVLDSKLILPKEMGGPASILNMVEGYSEQNKCPW 275

>Contig1327_sequence_frame-3
//KKAYNCMMPCSFQLLSCGSFFVLLKCFPRRERCFSCIDEEPGFQMVRNSA<mark>HTLESLVSWNIFYSSID
HMASKGWLILNTVIGDPASYLGWVAGTPQPSWVLW</mark>GGVAELWQPLPAEQQGWGKLIPFPLVLTR//

➢ NS2
Contig 3162 (NS2A), Contig 2431 (NS2B), Contig 3173 (NS2B)

>Contig3162-blastx

==> gi|27735365|ref|NP_776074.1| non-structural protein NS2a [Rio Bravo virus]
    Length = 229

 Score = 24.6 bits (52), Expect = 0.27
 Identities = 12/26 (46%), Positives = 17/26 (65%)
 Frame = -2

Query: 111 SVSSRVMISLWREHSQLTFVAVLKIL 34
        S+S + M S W + +QLT + LKIL
Sbjct: 180 SISPKFMQSDWIQKTQLTILGGLKIL 205

>Contig3162_sequence_frame-2
//HHNEVAVSGTSVSSRVMISLWREHSQLTFVAVLKILVFCSCLLFIMA

>Contig2431-blastx

==> gi|254688384|ref|YP_003084128.1| putative NS2B protein [*Aedes* flavivirus]
    Length = 125

 Score = 23.9 bits (50), Expect = 0.45
 Identities = 17/68 (25%), Positives = 29/68 (42%), Gaps = 12/68 (17%)
 Frame = -2

Query: 185 IVVLTLASLLYSPSAGVLGAVVILTVSFLPRG------------DLRGRALDDSAPIGEA 42
        + ++T++ + LY A V + L+ +P G DLRG D+ IG+
Sbjct: 9 LALVTIIAFLYMDQANVTMELEFLSTGDVPDGIALEEDEGGNFRDLRGTYSDEGITIGQD 68

Query: 41 EGIYRVFE 18
         G ++ E
Sbjct: 69 MGSAQIPE 76

>Contig2431_sequence_frame-2
TRNYLLTAIVVLTLASLLYSPSAGVLGAVVILTVSFLPRGDLRGRALDDSAPIGEAEGIYRVFEHIGP
W

>Contig3173-blastx

==> gi|226377838|ref|YP_002790883.1| polyprotein* [Bagaza virus]
    Length = 3426

 Score = 24.3 bits (51), Expect = 0.54
 Identities = 17/64 (26%), Positives = 30/64 (46%), Gaps = 3/64 (4%)
 Frame = +3

Query: 126 TGSNR---LKLDQSGSNRLKLDQNITDRFSPAQTGLIRVLTGSNQARPAQTGFDWLKPDS 296

      TGS++ +++D G+ +L DQ + +TGLI + P G W+ +

Sbjct: 1438 TGSSQRYDVEIDCDGNMKLMNDQGVPFSIWALRTGLILASAYNPYILPVTLGAYWM--TT 1495

Query: 297 HRPR 308

      H P+

Sbjct: 1496 HSPK 1499

*NS2B from Bagaza virus, region 1373..1500

>Contig3173_sequence_frame+3

DQLKPFLTGSNRTQSRPDRLKPVLNGSNRTQTGIDRLKQEQTGSNRLKLDQSGSNRLKLDQNITDRFSPAQTGLIRVLTGSNQARPAQTGFDWLKPDSHRPRP

➢ NS4
Contig 2370 (NS4A), Contig 2651 (NS4B)

>Contig2370-blastx

==> gi|159024817|ref|NP_739588.2| Nonstructural protein NS4A [dengue virus 2]
    Length = 127

 Score = 24.3 bits (51), Expect = 0.35
 Identities = 9/26 (34%), Positives = 16/26 (61%)
 Frame = +2

>Contig2370_sequence_frame+2
Query: 23 FLLQFIGNGQHSIGKCVCLSVSSILW 100

     FL+ G G+ ++G C ++ S +LW

Sbjct: 71 FLMSGRGIGKMTLGMCCIITASILLW 96

>Contig2370_sequence_frame+2
LTYQWDDFLLQFIGNGQHSIGKCVCLSVSSILWVPPRAQVLGRLRSIQQAM.KRRGRFKR

>Contig2651-blastx
==> gi|158516888|ref|YP_001527877.1| polyprotein* [West Nile virus]
    Length = 3433

 Score = 28.9 bits (63), Expect = 0.014
 Identities = 11/20 (55%), Positives = 14/20 (70%)
 Frame = -2

Query: 94 HWAYRSPGWQPKEMRRARPR 35

  H+AY PGWQ + MR A+ R

Sbjct: 2394 HYAYMVPGWQAEAMRSAQRR 2413

*NS4B from West Nile virus, region 2274..2528

>Contig2651_sequence_frame-2

TAVTSRLHSGIHHQSPEATSCPHGKGGRDRRR<mark>HWAYRSPGWQPKEMRRARPR</mark>PWREMGTGLRY

Fig. S6: BLASTX results for contigs of deep sequencing data (Supplementary data, Table SII) which presented similarities to structural (anchc, prem and envelope) and nonstructural (NS) proteins (N1, NS2A-NS2B, NS4A-NS4B) from flaviviruses. Some of contig sequences are displayed as partial sequence (// symbol). The aligned regions in each contig sequence is highlighted in cyan.

TABLE SIII

Description of the location and number of collected tick samples

City/state/Region	Farm	Unfed larvae (g)[a]	Male	Female < 4 mm	Engorged female
Presidente Prudente/SP/SE	A	1.0[b]	17	17	10
Rezende/RJ/SE	B	-	17	16	6
Montes Claros/MG/SE	C	1.2[b]	22	10	11
Quirinópolis/Góias/CW	D	0.5	19	15	10
PR or RS/S	G	0.5	20	15	8
PR or RS/S	H	0.5	19	15	10
Uberlândia/MG/SE	I	1.0[b]	18	15	10
Araguari/MG/SE	J	0.5	20	12	10
Catalão/Góias/CW	L	0.5	20	5	8
Piracanjuba/Góias/CW	M	0.6	16	16	10
Itauçu/Góias/CW	N	0.8	17	15	10
Pará de Minas/MG/SE	P	0.7	19	16	9
Santa Vitória/Góias/CW	Q	1.0[b]	18	14	10
Cassilândia/MS/CW	R	1.0[b]	20	20	10
Paranaíba/MS/CW	S	1.0[b]	18	15	9
Água Clara/MS/CW	T	0.7	15	15	10
Ribas do Rio Pardo/MS/CW	U	-	20	16	10
Birigui/SP/SE	V	-	18	15	8
Ribeirão Preto/SP/SE	X	1.0	-	80	60

a: each 0.5 g of tick egg mass corresponds to approximately 10,000 unfed larvae; *b*: samples with double representation for those farms. Farms are alphabetically labelled for easier designation. Pools of ticks were done according life stage. CW: Central-West Region; GO: Góias; MG: Minas Gerais; MS: Mato Grosso do Sul; PR: Paraná; RJ: Rio de Janeiro; RS: Rio Grande do Sul; S: South Region; SE: Southeast Region; SP: São Paulo.

TABLE SIV

Primer sequences

Target	Identification	Forward primer 5'-3'	Reverse primer 5'-3'
TriplEx2 vector	PT2F1/PT2R1	AAGTACTCTAGC AATTGTGAGC	CTCTTCGCTATTACGCCAGCTG
TriplEx2 vector	PT2F3	TCTCGGGAAGCGCGCCATTGT	-
Tick actin	ActinF/ActinR	CACGGTATCGTCACCAACT	TGATCTGCGTCATCTTCTG
Contig 317[a]	317-5-126/317-3-383	GTTACG GCTTCAGGAACCAA	GGAGGGTTGCATTTTTAGCA
Contig 401[b]	401-5-65/401-3-349	GGGTTGGAGCCAGTTAGTCA	ATC ACGAAGTTTGTGGGAGG
Contig 2743[c]	2743-5-126/2743-3-378	TCCACCACCTTTACCGACTC	AGGAGACGTCTGTTTCCCCT

a: similarity with nonstructural (NS)3 protein-NTPase; b: similarity with NS5-methyltransferase; c: similarity with NS5-RNA-dependent RNA polymerase.

TABLE SV

Protein statistics for nonstructural (NS)3 and NS5 from
Mogiana tick virus

	Amino acids (aa) n (%)	
	NS3	NS5
Size of full-length sequence	554	866
Strongly basic(+) aa (K.R)	64 (11.5)	131 (15.1)
Strongly acidic(-) aa (D.E)	59 (10.6)	115 (13.3)
Hydrophobic aa (A.I.L.F.W.V)	187 (33.7)	284 (32.8)
Polar aa (N.C.Q.S.T.Y)	145 (26.2)	185 (21.4)

Fig. S7: supplemental results for codon usage session.

We explored the sequences by performing analysis of base composition in nonstructural (NS)3 and NS5 nucleotide (nt) sequences from Mogiana tick virus (MGTV) and from other flaviviruses using four different approaches: (i) a statistic measure for the effective number of codons in a gene (Nc) (Wright 1990). The values range from 20 (implying that only one codon is used for each amino acid) to 60 (all codons are used equally). (ii) Overall GC content and the GC3 content (G + C content of 3rd-base codon position). High content of GC3 has been related with synonymous codon usage bias (Carbone et al. 2003, Wan et al. 2004); (iii) the codon usage adaptation index (CAI) reflects the codon bias related to a reference set (Sharp & Li 1987).

The overall GC content presented lowest variation when compared to Nc and GC3 values. The latter values were different for NS3 and NS5 genes, nothwithstanding the variation among the Flaviviridae members analysed (Supplementary data, Table SVI). The value for Nc is inversely related to codon bias, thus low Nc values mean that codon usage is highly biased. The Nc-$_{NS3}$ for MGTV was lower than those observed for mosquito-borne, tick-borne and no known vector (NKV) flavivirus groups, whilst Nc-$_{NS5}$ for MGTV was the lowest value observed for all members analysed.

Also, the GC3 content for MGTV is the highest among flaviviruses for both proteins [NS3 (61.7%) or NS5 (66.9%)]. We then compared the Nc and GC3 values obtained for NS3 and NS5 genes (only within flavivirus groups) with those reported by Schubert and Putonti (2010) that evaluated the same measures for full-length polyprotein sequences. The majority of measures for entire polyproteins are more similar to values for NS5 than NS3, except for Nc for mosquito-borne group and GC3 content for NKV group which is incomparable, either NS3 or NS5 values. Perhaps the NS5 measures in this context might be accounted for as being for the whole polyprotein and though the preliminary results Nc-$_{MGTV}$ and GC3$_{MGTV}$ reveal a marked codon bias among flavivirus groups.

The CAI indices range from 0-1; CAI = 1 if a gene always uses the most frequently used synonymous codons in the reference set. Besides the CAI, an expected CAI (eCAI) was obtained, based on NS3 and NS5 nt sequences using the Flavivirus sp. codon usage (TABLE). The eCAI is a threshold value for discerning if CAI values are statistically significant (Puigbò et al. 2008b). Then, CAI values are normalised with eCAI (ratio CAI:eCAI). If the normalised value $\geq$ 1.0, it means that the observed CAI is equal or higher than an eCAI and, consequently, it could be interpreted as a codon usage adaptation towards the Flavivirus genus.

TABLE SVI

Codon bias in Mogiana tick virus (MGTV) and other Flaviviridae viruses evaluated by effective number of codons (Nc) values and GC content

Flaviviridae genus	Group	NS3			NS5			Flavivirus polyprotein[a] (Schubert & Putonti 2010)		
		Nc	GC3	GC	Nc	GC3	GC	Nc	GC3	GC
Hepacivirus[b]	HCV	49.9	69.7	59.1	53.2	66.2	57.8	-	-	-
Flavivirus	MGTV	51.8	61.7	55.4	49.1	66.9	53.8	-	-	-
	Insect-only[c]	50.18 (± 0.97)	54.15 (± 1.45)	51.12 (± 0.80)	57.38 (± 1.21)	57.55 (± 2.63)	50.97 (± 1.2)	56.44 (± 0.05)	57.5	51.62
	Tick-borne[d]	55.64 (± 0.58)	56.96 (± 0.72)	54.40 (± 0.39)	53.34 (± 1.06)	60.18 (± 0.67)	53.62 (± 0.19)	53.96 (± 0.016)	59.0	54.07
	Mosquito-borne[e]	52.87 (± 0.52)	48.97 (± 1.84)	49.22 (± 0.78)	50.83 (± 0.74)	51.81 (± 1.64)	48.67 (± 0.88)	52.40 (± 0.026)	52.5	49.47
	Non known vector[f]	58.68 (± 1.38)	40.56 (± 1.63)	44.56 (± 1.19)	50.86 (± 1.01)	43.22 (± 2.72)	43.90 (± 1.69)	50.24 (± 0.075)	24.7	43.84
Pestivirus[g]		51.40 (± 0.38)	49.46 (± 0.63)	46.10 (± 0.39)	51.16 (± 0.43)	49.22 (± 1.15)	45.12 (± 0.41)	-	-	-

a: thirty-seven Flavivirus genomes were used in this study; *b*: the only RefSeq information available (up to June 2012) for this genus belongs to hepatitis C virus (HCV) (genotype 1 was used); *c*: cell fusing agent virus, *Culex* flavivirus, Aedes flavivirus, Kamiti River virus; *d*: Alkhurma virus, tick-borne encephalitis virus, Powassan virus, Langat virus, Louping ill virus; *e*: dengue virus 1-4, yellow fever virus, Murray Valley encephalitis virus, Usutu virus, Japanese encephalitis virus, West Nile virus; *f*: Montana myotis leukoencephalitis virus, Rio Bravo virus, Modoc virus, Apoi virus, Tamana bat virus; *g*: Border disease virus X818, Classical swine fever virus, Bovine viral diarrhoea virus genotype 1-2, Pestivirus giraffe-1.