

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].

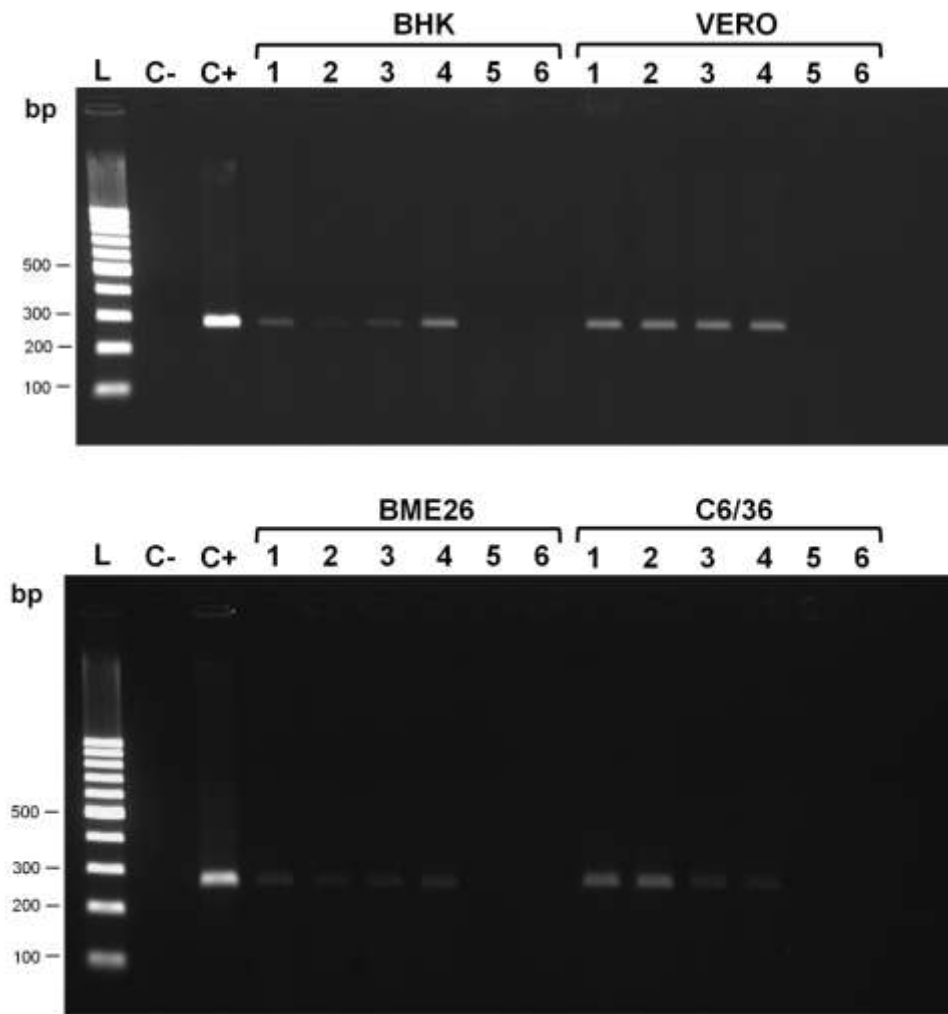


Fig. S2: detection of Mogiana tick virus in mammalian and arthropod cell cultures. Two percent agarose gel electrophoresis demonstrating the products generated by nonstructural 5 amplification (primer set 401-5-65 and 401-3-349) in baby hamster kidney (BHK), Vero (monkey), *Boophilus microplus* cattle 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.

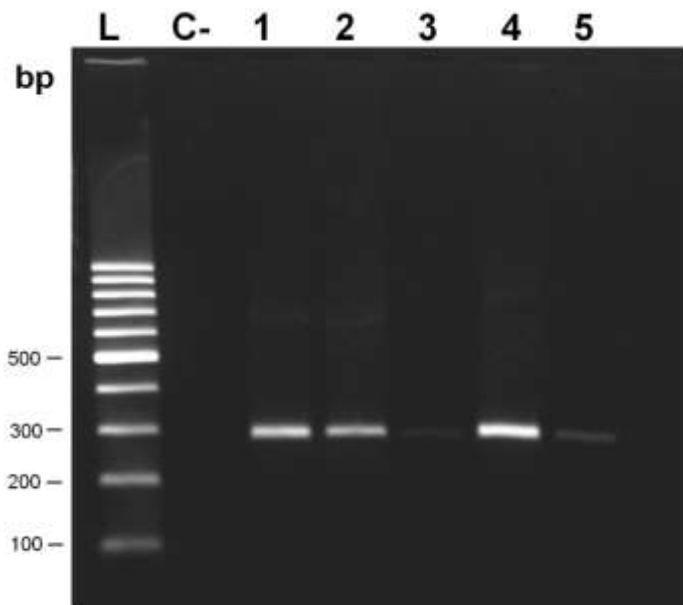


Fig. S3: molecular detection of purified Mogiana tick virus after ultracentrifugation of culture supernatant from second passage in Vero cells. Two percent agarose gel electrophoresis demonstrating the amplicons generated by nonstructural 5 amplification (primer set 401-5-65 and 401-3-349) in virus purified 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.

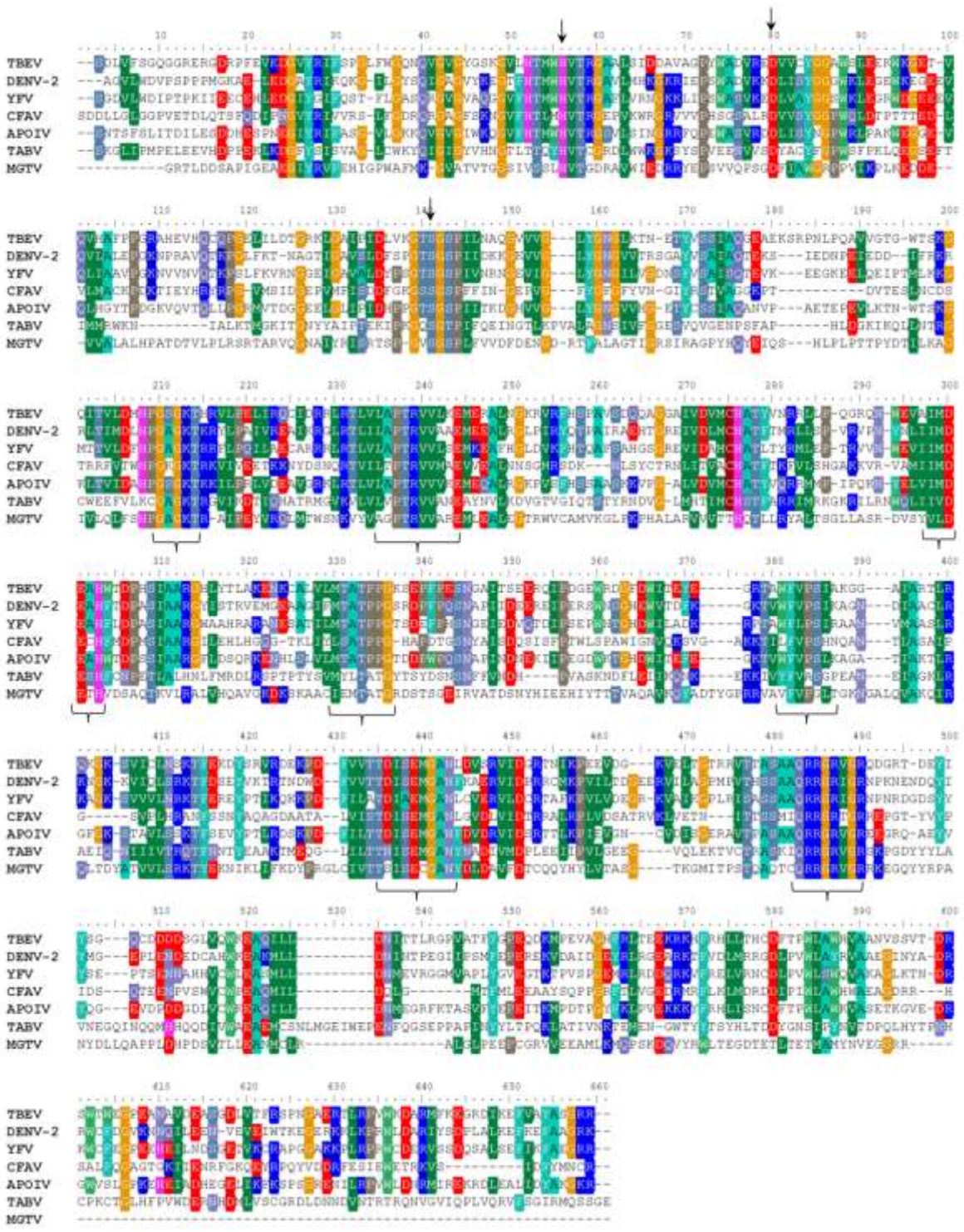


Fig. S4: multiple sequence alignment of nonstructural 3 from Mogiana tick virus and representative of tick-borne encephalitis virus (TBEV), mosquito-borne [dengue virus type 2 (DENV-2) and yellow fever virus (YFV)], insect-only [cell fusing agent virus (CFAV)] and not known vector [Apoi virus (APOIV)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

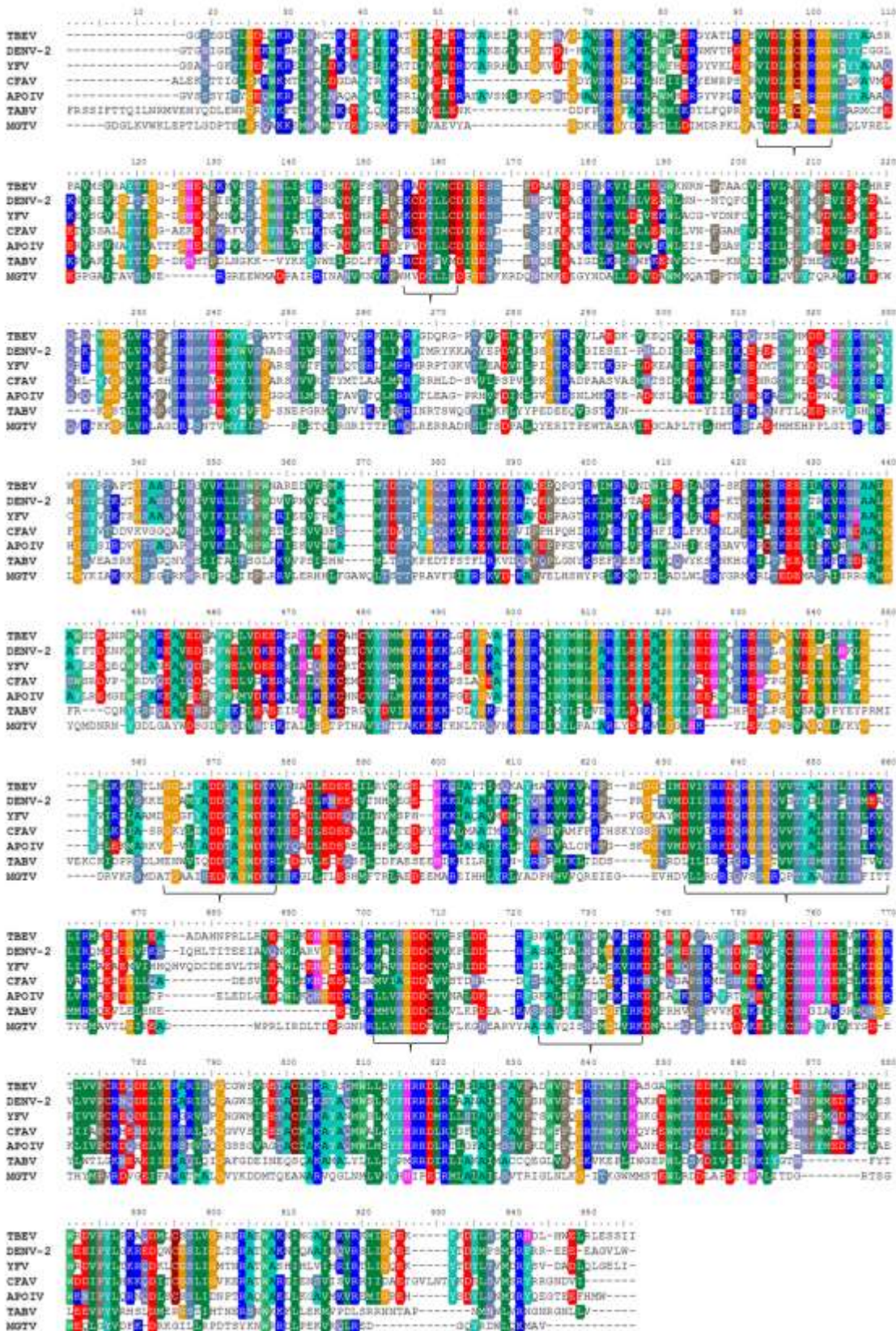


Fig. S5: multiple sequence alignment of nonstructural 5 from Mogiana tick virus and representative of tick-borne encephalitis virus (TBEV), mosquito-borne [dengue virus type 2 (DENV-2) and yellow fever virus (YFV)], insect-only [cell fusing agent virus (CFAV)] and not known vector [Apoi virus (APOIV) 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
<i>Aedes flavivirus</i>	-	I	NC_012932.1	YP_003084129.1	YP_003084132.1
Kamiti River virus	KRV	I	NC_005064.1	NP_937777.1	NP_937780.1
<i>Culex flavivirus</i>	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_003678.1	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**LAHSVVAARPLLPTRTLQGSLAHGVVPVRHLLA**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 ESLFPIFLTGLMAMQVSQNGDGWLLSP 119

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

>Contig3667_sequence_frame+2

AYSGCENESREVPLFGPTWPRALKRGGCRSWWGKPNQADQSRVCGVGVHDRLVA

EVCFPLPDLRSRSSLQVSRNVWGWYLSPTPRF AVELXXXXXXXXXXXXXXXXXPTLYST

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 ECEETMTYPCITLAATEEPVDLDCFCRDVKNVMVITYPTCKRNRTRRRDVTIQDHPPSVTL
99

Query: 206 ATSSL 220

SL

Sbjct: 100 TKPSL 104

>Contig547_sequence_frame+2_region_Flavi_propep

//QRFRIK**KCQKSKTLFVLSLMHKKYSICVNCRHRAVQQISRTSPSCITETAHP.TTSDSPHSQLLATSS**

LFSDS

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 PRRSRRSVSLPPHTEKKLETRHESWLETRS YLAHLEKTER 241

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

>Contig4248_sequence_frame+1_region_Flavi_M

DRQLSQWDSTVGIMAFWQKKTKKVTWATS YFLPVERHPHTPRCQRRPGNEPTQHLVSSXXXXXX
XXSTPARDTHTPHLTPNT.GSRGMAGSPRRTGHPVRSVQGYA**PQRGPRAGALSPKDEPVLTTPRSR**

WPPNRAPTSHIQRNEQVPAAHGAQAPRAQDEPTVQAWGRHQPREASEALPSNGHXEVTTQAHQE
FPHYAGFRSRDRGSGIVAQTSTDPMEMRAWLSLHPTSLHGYSILLPSHTSPHSHVPSFPPLGSSNQTS

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 VASRTGWMQSEVLSALSMFADVFLYKLFLETFCNCHAHNSTSYNRELTGVFGFTWMNV
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
QEASVPLRWPAFLHVA.AEGGRCWGTCSCSCFLIGNSRFLXXXXXXXXXXXXXXXXXXCVFFSAFR
IYTTWHLA VASRTGWMQSEVLSALSMFADVFLYKLFLETFCNCHAHNSTSYNRELTGVFGFTWMNV

>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

AYSGCENESREVPLFGPTWPRALKRGGCRSWWGKPNQADQSRVCGVGVHDRLVA EVCFPLPDL
RSRSSLQVSRNVWGWYLSPTPXXXXXSSHPLLSTDEPSSPTLYSTRIDLTVWLIPRVNNGRGQSTSC
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-----SGDEGWCQGPGRRGKGS--CLKAAME 58

AR GP+ + +G+ D GW G GKGS+ C+KAA E

Sbjct: 72 ARCPTMGPAVLTEERQIGTVCKRDQSDRGWGNHCGLFGKGSIVACVKAACE 122

>Contig553_sequence_frame-2_ Flavi_glycoprot

ARAERTGPSGAMRRQSVGSSGDEGWCQGPGRRGKGSMLKAAMELAVPAVKSGLQRTWTATRQ

>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

YTADMSGKWWLVKRDWYHDIALPWTAPSADFWHMDRLVEFSTPHATKQSVYTLGDQEGA 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

SGLPGFSVEGGHAQPVPPRSRKTTPREGWIPDWMTIAGDYFKRLQGKANVMGEGFDIILDEARD
LFPSTKEFALPGLIRQYLERRDVPLDSAFFLDVFLPLVLMALAITNRWTRMALAVGVYLTGYYYYV
AIMLAATSVVSLAFKAFAPREHKRGDVVIENGRLSVLGLTVVAVAATVGIHYYQSPSLIFAVASLA
GFAAILMALPTIQYHGATDVAKMMIAVLLFAGVIYIASSFDVDKDQLYLFVKTGTPTYHIPRGEREN
AVKMDQVYNLARYYAR**YPADLRARFVTEEAIWEWDPGATWTVAEMDYMATILHAVIFSLAVYT
MLTDSGETDRVLGD**IKMRILEKIKTTELFGDVTMAKITLAVSRVEWLVIAGNLLHTYFALGLPGV
AMEILLAAGLMVPTYHLWSRTRVMSYLRGTNGYRQPAEGLLPAPTIRSTQTQYTAGLATIVGIAA
VFNLYYYLTGSNPYVVAHSVLFVAGTFVMVQDNDQLNGYPLLMLMAYTFNSPSLALIGSIYKKC
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

WIP ++G +L W DK A G + LA V

Sbjct: 456 GMSWISPGLLGALLL--WMGVNARDKSIALAFLATGGVLLFLATNV 499

>Contig4362_sequence_frame-3_region_flavi_E_stem

APFLWRLPLPPALSSEAA GSAVGACRKGFECLVRTSSGTEACPLGATAWNHSLPGETRGSEESRDH
ALLGRQFHQYVLWKEWILPVMGVVLTPSWWPPPVPDKLTQRGSSAANGTVATLANRV TSVTWPP
GAS//

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-GHYRVQGGKF-----LGSRTWNKNPFPDPVEYESAKGRFP RRTEH-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
//VFGHRIWKRKRWFWAGWLPARICPLSTQNLAHTKQRLSPGRPAKHKCSYKAEKLEGVGLCSSGT
SVSQEDPGHYRVQGGKGLGSLRTWNKNPFPDPVEYESAKGRFP RRTEHCPTTPWYGYVITYVGMGM
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--WPITHTLGNRMVLDLILPKEMGGPASILNMVEGYSEQNKCPW 275

>Contig1327_sequence_frame-3
//KKAYNCMMPCSFQLSCGSFFVLLKCFPRRERCFSCIDEEPGFQMVNRNSAHTLESLVSWNIFYSSID
HMASKGWLILNTVIGDPASYLGWVAGTPQPSWVLWGGVAELWQPLPAEQGWGKLIFFPLVLTR//

➤ 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

//HHNEVAVSGT**SVSSRVMISLWREHSQLTFVAVLKIL**VFCSCLLFIMA

>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 LALVTIIAFLYMDQANVTMELEFLSTGDVPDGI ALEEDEGGNFRDLRGTY SDEGITIGQD 68

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

>Contig2431_sequence_frame-2

TRNYLLTA**IVVLTLASLLYSPSAGVLGAVVILTVSFLPRGDLRGRALDDSAPIGEA**EGIYRVFEHIGP
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---LKLDQSGSNRLKLDQNITDRFSPAQTGLIRVLTGSNQARPAQTGFDWLKPD
296

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

Sbjct: 1438 TGSSQRYDVEIDCDGNMCLMNDQGVPSIWAALRTGLILASAYNPYILPVTLGAYWM--
TT 1495

Query: 297 HRPR 308

H P+

Sbjct: 1496 HSPK 1499

*NS2B from Bagaza virus, region 1373..1500

>Contig3173_sequence_frame+3

DQLKPFLTGSNRTQSRPDRLKPVLNGSNRTQTGIDRLKQEQTGSNRLKLDQSGSNRLKLDQNITDRF
SPAQTGLIRVLTGSNQARPAQTGFDWLKPDShRPRP

➤ 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 FLMSGRGIGKMTLGMCCITASILLW 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

TAVTSRLHSGIHHQSPEATSCPHGKGGRDRRRHWAYRSPGWQPKEMRRARPRPWREMGTGLRY

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	Ticks collected per life stage			
		(n)			
		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
Itaçu/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, notwithstanding 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 ≥ 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 Mogjiana 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	-	-	-
	MGTV	51.8	61.7	55.4	49.1	66.9	53.8	-	-	-
Flavivirus	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.