Short Communication

Analysis of the medium (M) segment sequence of *Guaroa virus* and its comparison to other orthobunyaviruses

Thomas Briese,[1] Andrew Rambaut[2] and W. Ian Lipkin[1]

[1]Jerome L. and Dawn Greene Infectious Disease Laboratory, Mailman School of Public Health, Columbia University, New York, NY 10032, USA

[2]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Correspondence
Thomas Briese
thomas.briese@columbia.edu

Received 20 March 2004
Accepted 21 June 2004

Guaroa virus (GROV), a segmented virus in the genus *Orthobunyavirus*, has been linked to the Bunyamwera serogroup (BUN) through cross-reactivity in complement fixation assays of S segment-encoded nucleocapsid protein determinants, and also to the California serogroup (CAL) through cross-reactivity in neutralization assays of M segment-encoded glycoprotein determinants. Phylogenetic analysis of the S-segment sequence supported a closer relationship to the BUN serogroup for this segment and it was hypothesized that the serological reaction may indicate genome-segment reassortment. Here, cloning and sequencing of the GROV M segment are reported. Sequence analysis indicates an organization similar to that of other orthobunyaviruses, with genes in the order GN–NSm–GC, and mature proteins generated by protease cleavage at one, and by signalase at possibly three, sites. A potential role of motifs that are more similar to CAL than to BUN virus sequences with respect to the serological reaction is discussed. No discernable evidence for reassortment was identified.

Like other orthobunyaviruses, *Guaroa virus* (GROV) has a segmented, negative-strand RNA genome that is comprised of three segments, named small (S), medium (M) and large (L) (Bouloy *et al.*, 1973; Gentsch & Bishop, 1976; Clewley *et al.*, 1977; El Said *et al.*, 1979). The S segment of orthobunyaviruses encodes the nucleocapsid protein (N) and a non-structural protein (NSs) that may modulate viral polymerase activity, and acts as an alpha/beta interferon antagonist (Gentsch & Bishop, 1978; Bouloy *et al.*, 1984; Elliott, 1985; Bridgen *et al.*, 2001; Weber *et al.*, 2001). The genome-complementary strand of the M segment includes one open reading frame (ORF) for a polyprotein that yields the two surface glycoproteins GN and GC (G2 and G1, respectively; Lappin *et al.*, 1994) and a non-structural protein (NSm) of unknown function (Gentsch & Bishop, 1979; Fuller & Bishop, 1982; Elliott, 1985; Fazakerley *et al.*, 1988; Nakitare & Elliott, 1993). The L segment directs expression of a large, virion-associated protein with RNA-dependent RNA polymerase activity (Bouloy & Hannoun, 1976; Obijeski *et al.*, 1976; Elliott, 1989; Endres *et al.*, 1989; Jin & Elliott, 1991).

The International Committee on Taxonomy of Viruses considers *Guaroa virus* to be a species distinct from the species *California encephalitis virus* (CEV) and *Bunyamwera virus* (BUNV) within the genus *Orthobunyavirus* of the family *Bunyaviridae* (http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/index.htm); some investigators have suggested that GROV should not be included in either the California serogroup (CAL) or the Bunyamwera serogroup (BUN) (Whitman & Shope, 1962; Calisher & Maness, 1970; Wellings *et al.*, 1971; Hunt & Calisher, 1979; Klimas *et al.*, 1981). Serological assays have shown some link of GROV to both serogroups. In complement fixation (CF) assays, serological cross-reactivity was observed with BUN, but not CAL, members. In contrast, in haemagglutination–inhibition (HI) and neutralization (NT) assays, cross-reactivity was evident with CAL, but not BUN, members (Groot *et al.*, 1959; Casals & Whitman, 1960; Whitman & Shope, 1962; Tauraso, 1969). Results similar to those of CF assays were obtained in immunodiffusion, showing no cross-reactivity between GROV and CAL viruses, but weak cross-reaction with BUNV and Tensaw virus (Calisher & Maness, 1970; Wellings *et al.*, 1970). Immunoelectrophoresis, however, indicated common determinants between GROV and CAL viruses (Wellings *et al.*, 1971). Reaction in CF assays is determined by N, whereas reaction in NT/HI assays is determined by the glycoproteins (Lindsey *et al.*, 1977; Gentsch *et al.*, 1980; González-Scarano *et al.*, 1982; Kingsford & Hill, 1983; Ludwig *et al.*, 1991). Discordant serological reaction may therefore indicate different phylogenetic relationships for GROV N (S segment) and the glycoproteins (M segment). S-segment sequencing suggested a closer relationship to BUN than to CAL viruses;

The GenBank/EMBL/DDBJ accession number for the sequence reported in this paper is AY380581.

it has thus been hypothesized that GROV may be a reassortant virus (Dunn *et al.*, 1994). Here, we report the GROV M-segment sequence and its analysis in comparison to other M-segment sequences.

GROV RNA was reverse-transcribed by using Superscript II (Invitrogen) and amplified by PCR (Saiki *et al.*, 1985) using primers (1·6 μM; Table 1), dNTPs (200 μM), $MgCl_2$ (Table 1) and BIO-X-ACT polymerase (Bioline) in a PTC-200 thermocycler (MJ Research) for 45 cycles of 1 min at 92 °C, 1 min at 45–53 °C and 1–2·5 min at 68 °C (Table 1). Products were cloned and sequenced (Sanger *et al.*, 1977); analysis using the Wisconsin GCG package (Accelrys) indicated one ORF of 4254 nt (1418 aa) for the assembled sequence (GenBank accession no. AY380581).

Downstream of an untranslated region, the antigenomic strand encodes a protein that is related to GN of other BUN and CAL viruses (nt 20–943; 35 kDa). The N-terminal sequence is consistent with a functional signal peptide for membrane translocation (Blobel & Dobberstein, 1975; Lingappa *et al.*, 1978; von Heijne, 1988), similar to other viruses of the genus (Fazakerley *et al.*, 1988). In contrast to other orthobunyavirus M-segment sequences, GROV contains three potential AUG codons, with the first one being in the best context according to Kozak's rules (−3 = A, +4 = G; Kozak, 1986, 1991). This potentially results in a N-terminally extended product (Fig. 1). Cleavage of the signal peptide at T_{21} with respect to the first methionine is compatible with conservation of terminal tripeptides, as suggested by Lees *et al.* (1986), although SPV would change to TPV and RCF to KCF. Prediction of signalase cleavage at T_{21} or P_{23} by SignalP-NN/–HMM (http://www.cbs.dtu.dk/services/; Nielsen *et al.*, 1997) supports this view (data not shown).

The predicted amino acid sequence for GN contains the sequence KSLRV/AAR, allowing protease cleavage to separate mature GN from the downstream NSm analogue (xxxxl; Fig. 1) (Fazakerley *et al.*, 1988). The NSm-like sequence is characterized by a conserved, N-terminal, hydrophobic sequence followed by a short deletion, when compared to other M-segment sequences, and a motif that is conserved amongst BUN and CAL viruses (G_{416}DFc/t/sNKCg/rf/qC_{425}). Little conservation was observed around the NSm/GC junction, so a potential site for cleavage, possibly executed by signalase (Fazakerley *et al.*, 1988), is not apparent. Cleavage after a conserved alanine residue (A_{475}), analogous to the termination of NSm in CAL viruses (Campbell & Huang, 1999), is possible. This would result in nt 944–1444 encoding NSm (19 kDa). Analysis of the junction by SignalP predicts three cleavage sites with similar likelihoods: A_{472}, A_{475} and A_{479} (data not shown). Cleavage after A_{472} would result in positions −3 = V and −1 = A, one of the most frequent combinations in signalase sites. Cleavage after A_{479} would imply a long c-region, but would result in an N-terminal GC-tripeptide E_{480}EP, similar to BUNV and Cache Valley virus (CVV) [Fig. 1; Germiston virus (GERV) N-terminus deduced from alignments (Lees *et al.*, 1986; Gerbaud *et al.*, 1992; Lappin *et al.*, 1994); SignalP prediction, ATM-LV or VVA-GE]. Cleavage after any of the three alanines in GROV occurs close to a potential glycosylation site (N_{492}), but even cleavage at A_{479} would be at the 'minimum glycosylation distance' of 13 aa that has been determined for cleaved internal signals (Nilsson *et al.*, 1994).

The N-terminal portion of the next protein (nt 1445–4273; 108 kDa) is surprisingly divergent from other GC proteins. The C-terminal moiety, beginning about 150 aa after a

Table 1. Amplification primers

Primer	Sequence	3′ position	Annealing temperature (°C)	Extension time (min)
BUN-S5-F	5′-*GCCGC*AGTAGTGTACTACCGATAyA*	0	48	1
M940B-R	5′-GCwGCTCTwAGrCTTTTrTAmCCAG	915		
M590-F	5′-CATGCnTGyTTyAAdCAvCAyATG	580	50	1
M1300-R	5′-rbCyrCAyTTrTTdGwGAArTCACC	1265		
M940-F	5′-TGCCCTGGkTAyAAAAGyCTwAGAGC	936	51	2·5
GRO-D253R	5′-CCACAGCTATCAACTGCGCAT	2706		
M970-F	5′-ATGTGyAArTChArrGGnCCwGC	2450†	52	2
M3270-R	5′-CAnCCrAAyTCyTCrCAnCCC	3144		
GRO-D149F	5′-CTGGCTGCAAAGATGGACG	2640	53	2
GRO-E464R	5′-CCTTGCTGTCAATTGTTTTGCA	3908		
M3560-F	5′-TCnAArGGhTGyGGnAATGT	3465	47	1
M4170-R	5′-CCyTCATCyCknACyCkryACATCCA	4061		
GRO-U1092	5′-GGATGTTGTCGTCCGTAAATGTTATAATAA	3570	45	1‡
BUN-S3-R	5′-*CGCGCC*AGTAGTGTGCTACC	4453		

*Non-authentic bases added to some primers are indicated by italics.
†Position of misannealing is given.
‡Performed at 3·5 mM $MgCl_2$ (all other reactions were performed at 3 mM $MgCl_2$).

conserved potential trypsin cleavage site (tttt, Fig. 1; Fazakerley *et al.*, 1988), is again conserved when compared to other M-segment sequences. The 3′ non-coding region shows little conservation and is longer than those of the other M segments.

Five potential glycosylation sites were identified (+++; Fig. 1), including an N-terminal site in GC that is conserved among GC sequences of the CAL viruses and is found in approximately the same position in GROV (N_{492}); an N-terminal glycosylation site that is conserved in GC sequences of sequenced BUN viruses was not found in the majority of GROV clones analysed. Among 10 clones, six carried AATG-ACAtA for N_{616}DI (---; Fig. 1), whereas four carried AATG-ACAcA, encoding the potential glycosylation site N_{616}DT.

Analysis by TopPred2 (http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html; Claros & von Heijne, 1994) predicts six major transmembrane regions. Two are predicted for GN: the signal peptide (aa 6–26) and a second region that is compatible with a stop-transfer signal/membrane anchor (aa 210–230), which would result in a cytoplasmic location for a strictly conserved downstream KTY motif, a stretch of mostly hydrophobic residues that includes two proline residues, and the protease cleavage motif. Three transmembrane regions are predicted for NSm (aa 315–335, 368–388 and 455–475); the first, located 10 aa downstream of the GN C-terminus, is compatible with a third internal signal sequence, with cleavage predicted after G_{337} by SignalP. Signalase cleavage close to a cytoplasmic protease site may resemble the proposed situation at the C/prM cleavage site of flaviviruses (Stocks & Lobigs, 1998; Amberg & Rice, 1999). One transmembrane region (aa 1372–1392), a potential membrane anchor (Fazakerley *et al.*, 1988; Pekosz *et al.*, 1995), is predicted for GC. The overall topology appears to be well-conserved, as indicated by conservation of the same cysteine residues as in the polyproteins of all other BUN and CAL viruses (Fig. 1) (Lees *et al.*, 1986; Grady *et al.*, 1987; Pardigon *et al.*, 1988). Conservation of sequence motifs with respect to CAL but not BUN virus sequences is noted for K_{149}, Q_{163} and P_{299} in GN, P_{347} and N_{455}/F_{456} in NSm, and H_{591}QH, G_{601}EKCNSA$_{607}$, E_{958}, K_{1036}, G_{1243} and K_{1411}/K_{1412} in GC.

Pairwise, sliding-window distance analysis (SimPlot; http://sray.med.som.jhmi.edu/RaySoft/SimPlot/; Lole *et al.*, 1999) between GROV and BUN and CAL viruses indicated an almost equidistant position of GROV, with lowest distance scores in the GN region (approx. position 1–300; Fig. 2a) and highest scores obtained in the NSm sequence (approx. position 300–500) and in the N-terminal portion of GC (approx. position 500–1400). Serogroup-specific differences appear to be most pronounced in three regions (approx. positions 100–200, 550–650 and 1200–1275), where a separation between sequences of BUNV, CVV and GERV and CEV, Melao virus (MELV) and Trivittatus virus (TVTV) is observed. GROV appears to be less distant from BUNV/GERV/CVV than from CEV/MELV/TVTV in all regions except for the second part of the second region (at the N-terminus of GC toward N_{616}), where CEV/MELV/TVTV are less distant from GROV than BUNV/GERV; however, CVV remains the most closely related sequence throughout. In a reconstructed phylogenetic tree, the GROV M-segment sequence is placed in a closer relationship to sequences of BUN than of CAL viruses (62 % bootstrap support; Fig. 2b). Phylogenetic relationships of each individual ORF are similar to that of the entire sequence (data not shown).

A genetic distance of GROV from both serogroups is compatible with biological features. Genetic interference has been observed between CAL viruses in experiments that mimic interrupted feeding behaviour of mosquitoes (Beaty *et al.*, 1985; Sundin & Beaty, 1988). Interference was not observed between CAL viruses and GROV (Beaty *et al.*, 1983). Cell-culture experiments, however, indicated genetic distance not only from CAL, but also from BUN, viruses. Whilst genome-segment reassortment between GROV mutants was observed, heterologous reassortment was not observed between GROV and La Crosse virus (LACV), snowshoe hare virus (SSHV), TVTV or Tahyna virus (TAHV) (Gentsch *et al.*, 1980), but also not between GROV and BUNV, Maguari virus or Batai virus (Iroegbu & Pringle, 1981).

The structural determinants of GROV's unique serological reaction pattern are obscure. GN is not a major target of neutralizing antibodies that interfere with infection of mammalian cells (Ludwig *et al.*, 1989; Cheng *et al.*, 2000) and the few amino acids that are conserved in GN between GROV and CAL viruses do not correlate with identified epitopes of GN (/; Fig. 1; Cheng *et al.*, 2000). Therefore, GN is unlikely to form major determinants of the reaction of GROV in NT/HI assays. Likewise, NSm is unlikely to be involved. Epitopes detected in NT/HI assays have been mapped to the N-terminal portion of GC, mainly in relation to the trypsin site of LACV/SSHV (González-Scarano *et al.*, 1982; Kingsford *et al.*, 1983; Najjar *et al.*, 1985; Kingsford & Boucquey, 1990). However, their relation to primary amino acid sequence is not defined and only in one case has a particular amino acid that is involved in neutralization been identified (residue 29 of LACV GC; #, Fig. 1; Bupp & González-Scarano, 1998). The divergence of this region in comparison to available BUN virus sequences may explain the lack of cross-reaction between GROV and these viruses in NT/HI assays. Although also divergent when compared to sequences of CAL viruses, this region does contain motifs that are conserved with respect to CAL, but not BUN, viruses (H_{591}QH and G_{601}EKCNSA$_{607}$) and two glycosylation sites. N_{492}, which is conserved amongst CAL but not BUN virus sequences, is present, whereas N_{616}, which is conserved amongst BUN but not CAL virus sequences, is only present in a minority of GROV clones. N_{492} flanks the first putative antigenic domain that was proposed by Brockus & Grimstad (2001) and both conserved amino acid motifs and N_{616} are located in their second putative antigenic domain. It is conceivable that these positions contribute to the serological reaction of GROV. The mutation at

```
     1                                                                                                                                                                                                   160
 JCV .......MEV LNLIWLLALC SASPV.FQRC FQDGAIVKQN PSKEAVTEVC LKDDVSMIKT EAKYVKNTTG TFANNIAMRK WLVADWHDCR PQKSSGGHIN VIEVGDDLSL HTESYICSAD CVITVDKETA QIRLQTDKTN HFEVAGTTVK SGWFKSTTYI
MELV .........M IALIFLIAIA AASPV.YQRC FQDGAIVKQN PSKDAVTEVC LKDDVSMIKT EAKYVKNDTG TFSNNVAMRK WLVADWHDCR PQKSSGGHIN VIEVGDDLSL HTEAYICSAD CMISVDKETA QIRLQTDSTT HFEIAGTTVK SGWFKSTTYI
KEYV ........MMI WILLTVVTVT IANPV.YQRC FQDGAIVKQN PSKEAVTEVC LKDDVSMIKT EAKYVKNLTG TFSNNVAMRK WLVADWHDCR PQKRAGGHIN VIEVGDDLSL HTESYICSAD CVISVDKETA QIRLQTDTTN HFEVAGTTVK SGWFKSTTYI
LUMV .........M ILVLFFVALS AASPV.YQRC FQDGAIVKQS PSKQAVTEVC LKDDVSMIKT EARYFKNTTG IYANNVAIRK WLVADWHDCR PKKVSGGHIN VIEVGDDLTL HTEAYVCNAD CIIGVDKETA QVRLQTDTTN HFEIAGTTVK SGWFKSTTYI
LACV .........M ICILVLITVA AASPV.YQRC FQDGAIVKQN PSKEAVTEVC LKDDVSMIKT EARYVRNATG VFSNNVAIRK WLVSDWHDCR PKKIVGGHIN VIEVGDDLSL HTESYVCSAD CTIGVDKETA QVRLQTDTTN HFEIAGTTVK SGWFKSTTYI
 CEV .........M IFILILFAVA AASPV.YQRC FQDGAIVKQN PSKEAVTEVC LKDDVSMIKT EARYTKNSTG VFSNNVAIRK WLVSDWHDCR PRKVVGGHIN VIEVGDDLTL HTESYVCSAD CTIGVDKETA QVRLQTDTTN HFEIAGTTVK SGWFKSTTYI
TVTV .......MRN MIVFALLTIV AASPV.YKRC FQDGAIVRQE YSKNAVTEVC LKDDVSMIKT EAEYFKNSTG VFSKNIAYRK WLVSDWNECR PQKSANGHIN VIEVQPDLSL QTNSYVCAAT CVISIDKETA QVRLQTDTTN HFEISGTTVK SGWFKSTTYL
                ________h  |  Gc    *        // ////      *          /  ///////-++                //*/  ////         ///     ////  *///  *                                 v
GROV .MDISKHMQM LVVLYLISLS GTTPVVFNKC FTGGSLVFSK KSDTGMSEIC LKDDISMIKT SIDFKKENNT IVATNKVYKQ WIVQDWKDCH PMKMVGGTIN IIEVTADLMF KSVSYTCTTD CTIGIDRENA LIIFQTNKLN HFEVSGTTIK TGWFKNKASV 159
GERV MAISTSLLIV ALLIKLCLVN TAPPI..SKC FQDGILIAEL KSSSGISEFC IKDDISILKS EITYSKNDTG IFMHSKVFRH WTVADWKQCN H.TSAGGSTN VLEVDKNLNL VAKNYMCTRP CVITIDKENA QLLFQTEQLN QFEVTGTTIS TGWFKSKTSV
 CVV .......MLF LLILAFLGFS TSNPIA.TRC FHGGQLIAES KSATSISEFC LKDDVSMLKS EVVYIKNDTG LFGHNKVFRS WTIKDWKLCN PIPTAGGTIN VIEVNKDLSL MTKTYICSRD CTITIDKEEA QIIFQTDKLN HFEVSGTTLS SGWFKSKASV
BUNV .....MRIL ILLLAVTQLA VSSPVI.TRC FHGGQLIAER KSQTSISEFC IKDDVSMLKS EIVYTKNDTG IFGHSKVFRH WTITDWKACN PVVTAGGSIN VIEVDKNLNL VTRNYVCTGD CTITVDRKNA QIIFQTDKLN HFEVTGTTIS TGWFKSKASV

     161                                                                                                                                                                                                 320
 JCV TLDQTCEHLK VSCGPKSIQF HACFNQHMSC VRFLHRTILP GSIANSICQN IEIIILVTLS LVIFIFLSIV SKTYVCYLLM PIFTPIAFIY GWVYNKSCKK CKLCGLVYHP FTECGTHCVC GARYENSDRM KLHRASGLCP GYKSLRAARV MCKSKGPASI
MELV TLDQTCEHLK VSCGPKSIQF HACFNQHMSC VRFLHRTILP GSIANSICQN IEIIILVTLT LLIFIFLSIL SKTYICYILM PMFIPVAYIY GWIYNKSCKK CKLCGLVYHP FTECGTHCVC GARYETSDRM KLHRSSGLCP GYKSLRAARV MCKSKGPASI
KEYV TLDQTCERLK VSCGPKSIQF HACFNQHMSC VRFLHRTILL GSIANSICQN IEIIILVTLT LLIFILLSIL SKTYICYILM PISIPIAYIY GWVYNKSCKK CKLCGLVYHP FTECGTHCVC GARYETSDRM KLHRASGLCP GYKSLRAARV MCKSKGPASI
LUMV TLGQTCEHLK VSCGPKSIQF HACFNQHMSC VRFLHRTILP GSIANSVCQN IEIIILTVLT LLIFILLSIL SKTYICYILM PIFIPLAYIY GLIYNKSCKK CKLCGLVYHP FTECGTHCVC GARYDTSDRM KLHRASGLCP GYKSLRAARV MCKSKGPASV
LACV TLDQTCEHLK VSCGPKSVQF HACFNQHMSC VRFLHRTILP GSIANSICQN IEIIILVTLT LLIFILLSIL SKTYICYLLM PIFIPIAYIY GIIYNKSCKK CKLCGLVYHP FTECGTHCVC GARYDTSDRM KLHRASGLCP GYKSLRAARV MCKSKGPASI
 CEV SLDQTCEHLK VSCGPKSIQF HACFNQHMSC VRFLHRTILP GSIANSICQN VEIIILVTLA LLIFILLSII SKTYICYLLM PIFIPIAYIY GLIYNKSCKK CKLCGLVYHP ITECGTHCVC GARYDTSDRM KLHRASGLCP GYKSLRAARV MCKSKGPASI
TVTV TLDQTCEHLK ISCGSKSVQL HACFNQHMSC VRFLHRTILP GKHRKSICQN IEIIILTSLS LFIFLFLTII SKTYICYVLI PIFIPIAYVY GFIYNKSCKK CKLCGLVYHP FTECGSHCVC GAKYDTSDRM KMERSSGLCP GYKSLRAARS MCKSKGPAPI
         v *        *             *        *                        *        ________________________________       *                               +-+*     *    *                   * *                          *v      xxxxxx|     * NSm    _
GROV SLDQTCEHIK ITCGRKSMQL HACFKDHMSC IRFLHRTILP GYMATSICQN IEVILLTVFT LIIFGVLCIL IKTYICYLFL PLYMIPAYIY GCVYNKSCKI CINCGLAYHP FTNCGSHCVC GAKFETSDRM RIHRQSGFCP GYKSLRAARS MCKSKTCSLI 319
GERV SLDNTCEHIK VTCGKKSLQF HACFKQHMSC VRFLHRSVLP GYMASSICQN IELIIIIILT LAIFIFMCII TRTYICYLML PLFAPIAYLY GWLYNRSCKK CICCGLAYHP FTNCGSYCVC GSRFETSDRM RLERESGLCQ GFKSLRVARS LCKSKGSSLV
 CVV TLDRTCEHVK VTCGKKSLQF HACFKHHMSC VRFFHNTILP GNMITSMCQN IELIIILGLT LAIFILMVML TKTYICYLLM PIFMPIAYLY GWAYNKSCKK CNCCGLAYHP FTNCGSHCVC GLKFEASDRM RIHRESGLCQ GYKSLRIARL LCKSKGSSLV
BUNV TLDRTCEHIK VSCGKKTLQF HACFKQHMSC VRFLHRSILP GSMAISICQN IELIIITILA LCIFIIMIIL TKTYICYVLI PVFMPIAFAY GWAYNRSCKK CTCCGLAYHP FTNCGSYCVC GSKFETSDRM RMERESGLCQ GFKSLRVARR LCKSKGSSLI

     321                                                                                                                                                                                                 480
 JCV LSIMTAILIL TFVTPISAMV I.GESSELYT LDDLPNDMVQ MAERVNIYYQ ITIINYGVTW A.FFCLGLIL AYFFKKYQHK FLDFYAMYCE ECDMYHEKSG LTHWGDFTNK CRQCTCGQFE DVAGLMIHKK SYNCLVKYKA KWIRDILIIY LAMILIKDSL
MELV LSVLTAVLIL TFVTPMNALT L.GSSEENYK LENLPDDMLD MAERIDIYYL SIILNYSTTW G.LVIILLGIT FILFRRFQHR FLNCYAMYCS ECDMYHDKNG LVRNGDFTNK CRQCTCGQYE DAAGLIIHQK SYNCLIKYKA KWIRNILLTY IILILLTKDSV
KEYV LSILTAVLIL TFITPMNAMV L.GSSEEVYK LEDLPDDMLN MAEKIDIYYL SIVLNYSTTW G.LLILLLIV SCLFKKYHLV FLNYYALYCH ECDMYHEKKN LVYNGDFSNK CLQCTCGQYE DATGLIVHQK SYNCLINYKA KWFINILIIY TILVLLKDSV
LUMV LSIITAVLIL TFITPINAMI A.GETQETFK LEDLPDDMLS MAMKINMYYI MCMLNYVATW S.FIFCALII ALIFKKYQHK FLSYYAMYCQ ECDMYHDRSG LKYNGGFTNK CRQCTCGQYE DATGLMAHRK TYNCLIQYKS KWLVNFLITY IILMLIKDSV
LACV LSIITAVLVL TFVTPINSMV L.GESKETFE LEDLPDDMLE MASRINSYYL TCILNYAVSW G.LVIIGLLI GLLFKKYQHR FLNVYAMYCE ECDMYHDKSG LKRHGDFTNK CRQCTCGQYE DAAGLMAHRK TYNCLVQYKA KWMMNFLIIY IFLILIKDSA
 CEV LSIITAVLIL TFVTPINAMM V.GGSKETFE LEELPDDMLE MANNLNTYYM LCVTNYAVTW S.LLIIALAI GLLFKRYQHR FLDIYAMYCE ECDMYHDRSG IKWNGDFTNK CGQCTCGQYE DASGLLVHRK TYNCLIQYKS KWIMNFLITY VILVLLKDSV
TVTV LSLLSAVLIL TFVTPITAIS LXDEEKETYK LPELPDDKVN LETKIDLYFQ MCLINYATTW A.LIFCVLIT ACLFKKLQHK FLDYYAVYCH ECDMYHEKTS LKYNGDFSNK CRQCTCEQYE DATGLLTHRK SYNRLVQYKA KWTLNFLITY AILVIIKDSV
                    h                      v                                                                     *    *                       *   *  *                                 vv                 h
GROV FSIGGAILIL SFVTPIAG.. .....EKLYT LDELPEAYTT EINSIKIMQN YLELKLWITI TALFIIALIM MISNFIVVNI .CKLFYFHCY ECNMYHTKRN IRYNGDFCNK CGFCMCGVLE DVEGLKIHKT TESCKIKMHY RSTNFALILF GLLLILISVS 471
GERV ISILTAMLIL SFITPLEAMT TNYPDDKKFT LKEVNDIVLG RDMEQEL.KS SILILMSICG IGIILIFFGL TVLLEIVLEL IAKRSTIFCK ECNLIHDKKS MTYRGDFTNK CGFCPCGELE DPEGLVIHTT RKSCTYYIKI RNLKLIMLIF SIVILMQNAT
 CVV ISGLLSMLIL SFVTPIEGTI TSYPESRKYT LEEITDVVEG LVSEQNI.KE YIVFYTSLFG SLILLLAFIM SITLSKITSI LAKRNVIYCE ECSMYHSKRG IKYNGDFTNK CGFCTCNEQE DTEGLVIHKV SKRCTYKYQL KWAKVLMITL VLLLIAQNTI
BUNV ISILLSVLIL SFVTPIECTL TNYPTDQKYT LDEIADVLQA KTHEDST.KY YIILYTSLFC ACLTIIFAGV ALCLTIILEV LTKINVIFCN ECNMYHSKKS IKYVCDFTNK CGFCTCGLLE DPEGVVVHKA KKSCTYSYQI NWVRCIMIFV AFLFVIQNTI

     481                                                                                                                                                                                                 640
 JCV LFVGASGTDF AACTDEKVVT WNCTGPFLNL GRCDKSQKKA SYACIADQLK GLGVISVLDI PMIKNMPDEI SGSLKYIDSL ETYHEQLTAE YAFLTRYCDY YTQYTDKSGY SQTVWRTYLR SHDFGACVVY PNQHFC.RCV KYGDKCNSAK WDFANEMKTF
MELV LLTSAEGPDF SVCIEQKAIS WNCTGPFLNL GRCEKTQKKA SYATIATQLR SIDAISVLDI PMIAKIPEDI SGALKYIESL KTYHEQLTAE YAMLTRYCDY YTQFSDNSGY SQTSWRKFLR SHDFESCIIY PNQHFC.RCV KQDEKCSSAK WDFASEMKGF
KEYV LMAAAERPDF NTCIEQRSIS LNCTGPFINL GGCKKPQKRA SYQSIAEQLK AINAISILDM PMISKIPDDI SGALRHIENL KTYHEQLTAE FAMLTRYCDY YTQFSDNSGY SQTSWRTYLR SHDFEVCIAY PNQHFC.RCV KNGEKCTSAN WDFANEMKNY
LUMV LIVHAEGTDF EKCSSDINIV WNCTGPFLSL GNCEKKQKKE SYNDIAAQLK GLDAISVLDI PMISKIPEDI AGALRFIEEQ NSYHVQLTVE YAMLTRYCDY YTQFSDNSGY SQTTWRTYLR SHDFEACILY PNQHFC.RCV KHGDKCSRSN WDFANEMKNY
LACV IVVQAAGTDF TTCLETESIN WNCTGPFLNL GNCQKQQKKE PYTNIATQLK GLKAISVLDV PIITGIPDDI AGALRYIEEK EDFHVQLTIE YAMLSKYCDY YTQFSDNSGY SQTTWRVYLR SHNFEACILY PNQHFC.RCV KNGEKCSSSN WDFANEMKDY
 CEV VIVQAKGTDF SECIDADKIK WNCTGPFVNL GNCEKKQKKE PYSNIAAQLK GLSAISILDV PIIASIPEDI AGALRYIEDQ RSYHSQLTVE YAMLTRYCDY YTQFSDNSGY SQTTWRVYLR SHVFEACILY PNQHFC.RCV KHGDKCSSSN WDFASEMKNY
TVTV MIVGASNTDV KDCLDQEEIT WNCTGPFLNM GRCTRPQKKE NYGNIATQLK GLDVLSALDM PAVGKIPEDI AGALSYIESK KTYHEQLTLE FAMLTRYCDF YKQYADNSGY SQTTWRAYLR SHDFGVCIMY PNQHFC.RCV KMGEKCTSSN WDFANQMANY
        |  Gc          *              *       -  ++*#                                                                              *                                                 vvv  *   *      vvv*vvv        ---
GROV ASLAS..... ..CDA.EEPN INCIG.YKIN LTC.NKPLT. NEEA.VNYLV ENNIIEDVER EKALELTTDL ANMKKQLNFQ TNFKTCLVLE EVFLLNQCGY YKDLESNNEI KQIPWRTIAK TEAFDICSIR KHQHFC.RCM ATGEKCNSAN WDFAKEMNDI 618
GERV MLVVAG.... ENCWTNTEIK ADCVGPLIGP SACTNKGSK. TYKTVAQELV TASKITQLDA DKYVLLGDTI ESALDAITSQ KHYSAMHLLE IMFLMKHCDY YKVYEHNSGY SQTKWRLIAI ANSFDICTNT PTPNFC.KCL SDSS.CSTTT LNFATSMNAT
 CVV LIVAAE.... TDCWTKKSLD IDCIGPMLDI GSCTNKNTR. SYDAEAQKLV SQSKISQLDA DQASLLGNSI DSAIKATRAQ KTYPTMHLLE TIFLMKNCDY YTSFNHNSGY SQAKWRLTAK TGHFDICSRH STHHFC.QCI NDGTKCQNGD WDFAGEMNET
BUNV IMVAAE.... EDCWKNEELK EDCVGPLIAP KDCTDKDHK. TYLSEASLLA TAKKITQVDA ENVEILGKTM ESAIRVIERQ KTYHRMHLLE AVFLNKHCDY YKMFEHNSGY SQVKWRMMIK TQHFDICALQ ANSPFCAQCI ADNSCAQ.GS WEFDTHMNST

     640                                                                                                                                                                                                 800
 JCV YTGQKVKFNK DLNLALLVLH KAFRGTGTAY VAEAMAREDN GSMAHFALEI RKKFPTNALL AALIDYIVYL QSLGEIATFE LDEDWEDFKY KPEPTESTPQ VRSHRNDQYN FKNAESNSDT KVCKNVQKVL CISPRSKAKF ENIIACGEHA NPSVYMMPNV
MELV YLNKQAKFNK DLNLALSTFH KAFRGTPISY VTELISKNKN SSLLMYVSEI KRKFPGNALL VALLDFIMYL QSLSEISGFE LGDEWEDFKY DPEPTEPALM PRSHRSDQYN FKNAASVSKT KTCENIKRVL CLSPRSRATF DDVVACGEFA NPSVYMLPNI
KEYV YTGKQIRFDK DLNLALTTLH TAFRGTSTYY MSDILAETKA NSLRDYALAL KSKYPGNALL KAILDFIMYL QPLNEATKFK LSSDWEDFKY NSEPEARALM PRARSSGDYN FSNAKSTNKV KVCKNIQRVM CLSPRSKATV EDIIACGEHE EPSVYMVPNT
LUMV YSGKQSKYNK DLNLALMSFH HAFRGTSSFY TAKLLREKKT EELKNYIDKI RTEYPGNALL KAVIDYIKYM MGLTEVSNFK HDEMWDDFIF ENPSTRSSLS .RNSES.SYD FKTNINSDPT NTCKNIKAVT CLSPKSGVSF DSVIACGESN QPSIYQKPRK
LACV YSGKQTKFDK DLNLALTALH HAFRGTSSAY IATMLSKKSN DDLIAYTNKI KTKFPGNALL KAIIDYIAYM KSLPGMANFK YDEFWDELLY KPNPAKASNL ARGKE.SSYN FKLAISSKSI KICKNVKDVA CLSPRSGAIY ASIIACGEPN GPSVYRKPSG
 CEV YSGKQNKFNK DLNLALTSLH HAFRGTSSSY IAGLLSTKAN DELVQYTSKI REKFPGNALL KAIIDYIAYM KGLTEMSNFK KDEFWDELVY VPDPTKPPKL SRGSET.VYD FKSATSNLGE KNCKDVKGVV CLSPRSGVTY DSIIACGEGS SPNIYKIPKT
TVTV YNGKQNRFNK DLNLALYTMH HAFRGSTSYL ISKVLHEKKY DQLTNLLTQL KTKYPGNALL KALIGYQIYL KSLSEMSSFT PDELWDDLQY VPAPTFKPNL GRSGDN.EYK FDKSRIVDPQ TTCTSLKGVN CLSPRSRASV SDIIACGQAT QPALFYVPND
                                                                                                                                                                 *             *tttt                 *
GROV YNFKPTFYRR DVNLFLRIFS EAFPGSTYRY LYSILKNKDI EKIKELLTDI EKRFPYNQLL VGFIKFGHIL LIDN...... .......LAY SLPPDVSRMG AESRSKVE.. TSGYQVGQPI LSCNNPLIIR CMSPKLDVVI PALLKCHRNN KDSLYYIDH. 762
GERV YTSKVEFFNH DFTLFLDIFE AAFPGSATAF LFKKIKEKNP YQAFEMMGKI ANKYPNNKLL VVILKYGQYM VGLSHASTYQ LKQE....WV AKSLSLTRAQ R.TGLKMS.. MANAEPGPAT KECSDAKTIA CLTPKFQVEV NNLMSCGASP NFKIYVKTG.
 CVV YNSKNDFYSH DLNLFYTIFE NAFPGTTESL FYQLIQKKNT TGASKVLGKL ITKYGNNKMF VGIWKFGQYL LTLPYINNTE LSQNRRLTIL ATQELLSRSH ..SGRQES.. MSNAIQGSIT KECHNAKGVG CISPQFGIPI GNLTACGDSP NYKIYKTPE.
BUNV YSSKVDNFKH DFSLFLRIFE AAFPGTAYVE LLTNIKEKKP YQAVSMIEKI KKKFPNNKLL IGYLDFGKYL LGLSHASTYE L.QQRQLDKL YQPTELTRS. ..GGQQTS.. LANSVVGQAT KECKKYKDVS CLSPRFGIPL EDLISCCDQP NYNIYKKPK.

     801                                                                                                                                                                                                 960
 JCV SIYQSNSERS EYCIAGSHCL EKYEVVETEL LNALKKSKCW AKEMDDIILH KQSDGLRSCR IKDTGNCN.V QENDWTIVLC ENGNYYYSEV EKDYDKDQDV G.HFCLSPRC TTIRYPISKR HIKKCEWQIS HSTIDKISVQ ELADIEQYKK AISQKLQTSL
MELV TIYQSNIERS EYCIADSHCL EEFELVKPEL LTALKKSTCW AGEIDDIVLH KQSDGLRSCR IKDTGTCD.V LGNEWTIVLC ENNNFYYSEV ERDYDKDQDI G.HFCLSPRC NIIRYPINPR HIRTCKWQIS HSTIDKITVH ELADIEQYKK AITQKLQNSL
KEYV TIYQSNIERS EYCIADSHCL EEYELVQNEL LTALKKSTCW TGELENVIPH KQSDGIRSCR TKNTGRCN.V KGEDWTVILC ENDKYYYSEL NKDYDKDQDI G.HVCLSPRC NSIRYPISPR HITNCKWENP AGTISRIETH SLADIEQYKK AISQKLQNSL
LUMV QVYQSNSEQT EFCLLDTHCL EGYETVGEEV LSAIKKSKCW ETDFTEYIQF KQSDGVRSCR MKDSGECT.V TTNKWPIILC ENNRYYYSEL EKDYDKDQDI G.HFCLSPRC NTIRYPINQN HISNCTWQVR HNNIDKIEVH ELEDLEQYRK AITQKLQTSL
LACV GVFQSSTDRS IYCLLDSHCL EEFEAIGQEE LDAVKKSKCW EIEYPDVKLI QEGDGTKSCR MKDSGNCN.V ATNRWPVIQC ENDKFYYSEL QKDYDKAQDI G.HYCLSPGC TTVRYPINPK HISNCNWQVS RSSIAKIDVH NIEDIEQYKK AITQKLQTSL
 CEV GVYQSNSEPS NYCVSDSHCL EDFEVISQEE LDAIKKSRCW EVDYLYANPP KQSDGVRSCR MKDAGHCN.V TTNRWPIIQC DDNKYYYSEL QRDYDKEQDI G.HFCLSPRC ATVRFPIHPK HISNCDWQVS HSTVDKVNVH NLEDIEQYRK AITQKLQNSL
TVTV KIHQSNTDQN QYCIGDSHCL TXFTPITEEL LAAVKKSKCQ AFDLDNIVKT STSTGVRSCL VKKFGNCN.T TNQVWPITIC SDDQYYYTDT KQSYDSYQDI G.HFCLSPRC STIRYPVNPK HLSACTWTEE YTPVDKISVT NIEEIEQYKK ALLQKLQTGL
                    *         *                        *                        *          *                *                              *     *             -*+
GROV KVYTSFQHPQ SWCFADKHCL IHFRPVPGSE LNAVKKRNCF SSQ.VTEDQD QYNTKQAFCR AIKFGICTNI LNQQSRSIEC DNNLIYNTDS KLTNELTGDV .NEICLNTGC KDGRYPLNPI NSVNCTWEEN VLKSKHAQLI AVENIEELKR LTQDRLTHTL 920
GERV ELYKAHDRNS VWCLNDMHCL TPYTPANAEI ITTMKKMDCW QDN.PKQPTD EYAIPKRSCQ MKDRGLCN.S GADKWKIIKC DNHKLFYTDA LERRDPASIV GSNHCFSEKC QIERYPINPT SLTNCEWLYR AVRPEYIKKL SLQTIEEYKK AIADKLTHTL
 CVV KLYKSNNKGE VWCSNDVHCL NDFEPADEEV VNKIKKLTCF LTD.VSVEVD IFSVAASTCK MADKGVCT.V NGAKWNIIKC DSGLYYFTDH REGQDTGNDF G.EYCISHSC NTERFPINPD TLKDCTWEFH SRKSKYISTI SLESLEEFRR AMAEKLSHTL
BUNV KVYKAHDKEE TWCINDQHCL VDFVPAEADT VEKLKPMKCW LVD.PGKNDD VYSIAIKTCR VVDKGVCT.V NSQKWNIIKC DSGPLYYSDH IPGEDTGNDI G.HYCVSAGC KTDRYPINPD VVTDCVWEFT SRKSQYIGKI SMQSLEDYEK ALTDRLTHTL

     961                                                                                                                                                                                                1120
 JCV SIFKYVRT.K NLPHIKPAYK YITVEGVETI EGIENAFIES EIPALAGTAI GFKINSREGN YLMDVIGYVK SASYSAVYTK LYTTGPTIGI NTKHDEKCTG PCPAIIDHKD GWLTFSKERT STWGCEEFGC LAVSDGCVFG SCQDIIKDEL AVYRKTTEEL
MELV SIFKYKRT.K NLPHIKPIYK YIAIEGTETV EGIENAYIES EIPALAGTAI GFKINSKDGN HLMDVIGYVK SASYSSTYTK LYTTGPTIGI NTKHDEKCTG SCPTQVKHKT GWLTFAKERT STWGCEEFGC LAISDGCVFG SCQDIIKDEL TVYRKTTDEV
KEYV TIFKYRKT.K NLPHIKPIYK YITIEGTETV EGVENAFIEA EIPALAGTSI GFKVNSRDGD HLMDIIAYIK VASYSSTYTK LYTTGPTIGI NTKHDEKCTG NCPARLEHKV GWLTFAKERT STWGCEEFGC LAVSDGCVFG SCQDIIKEEL SVYRKDTEEV
LUMV TLFKYSRT.K NLPHIKPIYK YITIEGTETA EGIESAYIES EIPALAGTSI GFKITSKSGK HLLDVISYVK SASYSAVYTK LYTTGPTIGV NTKHDELCTG SCPVNVPHNT GWLTFARERT SSWGCEEFGC LAISDGCVFG SCQDIIKDEL TVYREETEET
LACV SLFKYAKT.K NLPHIKPIYK YITIEGTETA EGIESAYIES EVPALAGTSI GFKINSKEGK HLLDVIAYVK SASYSSVYTK LYSTGPTSGI NTKHDELCTG PCPANINHQV GWLTFARERT SSWGCEEFGC LAVSDGCVFG SCQDIIKEEL SVYRKETEEV
 CEV SIFKYART.K NLPHIKPIYK YITIEGTETA EGIESAYIES EIPALAGTSI GFKINTKEGK HLLDVIAYVK QASYSSLYNK MYVTGPTVGI NTKHDELCTG PCPVNVPHSD GWLTFARERT SSWGCEEFGC LAISDGCVFG SCQDIIKEEL TVYRKETEEA
TVTV NIFKYRKT.S NLPHIKSMYK YITVQGTGTA EGVENAYIES EMPALAGSSS GFKVLSKEGN HLFDVVAYIK SASYSSVYNK LYSTGPTLGI NVKHDEKCTG PCPANIPHKT GWLTFARERT STWGREEYGC LAISDGCVFG SCQDIIKPEM TVYRKQTDEI
                                            v                                                                                             *      *                v                    *          *        *
GROV DIYREKKTL. NEPKIIPSKK YISLQGIDSS DGIDSSYIEV DITARTNEAL GIQVRAPDNQ PIMDLIIHIT LASTKADYLR IYFTGPTVGI NVNEDEHCTG SCPEQIPHKQ NWLCFSKERT SQWGCEEEGC LAINEGCLFG SCQDLIKSEA RVEKKTTVEK 1079
GERV QLYHFAPLLE NLPHIKPTYK YITAQGTYTA DGIEGASITT SIPALSGTSV GFKINAKDGT DLLDIVVYIK ASVVKSIYNH IYDTGPTINI NSKHDELCTG QCPKKIPADP NWLTFSQERT SRWGCEEFGC LAINTGCVYG SCQDVIRTET KVYRKANEET
 CVV TIYNFKPT.A NLPHIRPTYK FITANGVENS DGIESAYILS SIPALGGISV GYNVQTKDNF PLLDIIVFIK SAVIRTTYNH IYDTGPTIGI NTQHDEHCTG PCPAVVPHKE NWMTFSQERT SRWGCEEFGC LAINTGCVFG SCQDIIHPET KVYRKAVEEE
BUNV ETYSFAP.LE NLPHIKPVYK YITAQGVENS DGIEGAFITA SIPAAGGTSI GYNVRSKDGF PLLDLIVFVK SAVIKSTYNH IYDTGPTISI NTKHDEHCTG QCPSNIEHEA NWLTFSQERT SRWGCEEFGC LAVNTGCVFG SCQDVIRPET KVYRKAVDEV

     1121                                                                                                                                                                                               1280
 JCV TTVELCLTFS DKTYCTNLNA ITPIITDKFE VQFKTVESYT LPRVIAIKNH EIKVGQINDL GVYSKGCGNV QKVNGTNYGN GSPRFDYLCH LASRKEVILR KCFDNDYQAC KFLQSPSSYR LEEHQGTVTV IDYKKILGTI KMKAILGDVK YKNYQNNIEI
MELV TELELCLTFA DKTYCTNLNA VTPIITDKFE VQFKTVESYS LPRIIAVKNH EIKVGQINDI GVYSKGCGNV QKVNGTSYGN GVPKFDYLCH LASRKEVTIR KCFDNDYQAC KFLQTPSSYR LEEHQGTVTI IDYKKILGSI KMKATLGDVK YKNYASNIDT
KEYV TNVELCLTFA DKTYCTSLNA ITPIITEKFE VQFKTVESYS LPRIVGIRNH EIRIGQINDL GSYSKGCGNV QRVNGTNYGS GQPKFDYLCH LASRKEVIVR KCFDNDYQAC KFLQTPSSYR LEEHEGTVTI IDYKKILGTI KMKAILGDVR YKNYASAIDI
LUMV TNVELCLTFA DKTYCTNLNA ITPIITDQFE VQFKTVEAYS LPRIIAVRNH EIMVGQINDI GVYSKGCGNV QKINNTNYGN GVPKFDYLCH LASRKEVIIR KCFDNDYQAC KFLQTPASYR LEEEGGSVTV IDYKKILGTI KMKAILGDVK YKTFADNVDI
LACV TDVELCLTFS DKTYCTNLNP VTPIITDLFE VQFKTVETYS LPRIVAVQNH EIKIGQINDL GVYSKGCGNV QKVNGTIYGN GVPRFDYLCH LASRKEVIVR KCFDNDYQAC KFLQSPASYR LEEDSGTVTI IDYKKILGTI KMKAILGDVK YKTFADSVDI
 CEV TNVELCLTFA DKTYCTELNP VTPIITNLFE VQFKTVETYS LPRVVAVQNH EIRIGQINDL GVYSKGCGNV QKVNNTIYGN GVPKFDYLCH LASRKEVIVR KCFDNDFQAC KFLQSPPSYR LEEDDGSITV IDYKKILGTI KMKAILGDVK YKTFAENIDI
TVTV VTVELCLTFV DKTYCTELNA VTPIISDLFE VQFKTVESYS LPKIVAIKNH QIFVGQINVI GTNSKGCGNV QVFNGTNYGS GTPRFDYLCH LASRREVIMR KCFDNDYQAC KFLQQPSSFR LEEEGETVTI IDYKKILGTI KMKAIFGDVK YRTFSEKVDI
               *                 *                                                              *           +++                           *                      *        *
GROV HIVRLCIIHA NEDKCTDLDP EEPLITESLE LQLDSITTKN LPELILIKDH KVYTGQINDL CAVSSGCCDV QVVNNTIFCC GTVKFDYLCY GAHRKDVVVR KCYNNHYESC KLLAEQKELT FLDESSAVSL EHNNYIIGSL KLKIVLGDIR YKLFRKDIQL 1239
GERV VMLTVCITYP GHTFCTDVNA HEPKITDELE LQFKTIDIKS LPNLVAVTNH KLYTGQINDL GTFGQMCGNV QKTNTSHTGA GTPKFDYTCY SASRKDIIIR RCYNNNYDSC RLLNQESDLL FDDNHETLVV YNNKRLNGEL ALKLLGDIQ YKLYTENMEL
 CVV VVLTVCINFP GNNYCTEINA IEPKITDEIE LQFKTVDTKT LPNLLAVQNH KLYSGQINDL GSFSQGCGNV QKTNHSILGM GTVKFDYVCH GASRKDVIVR RCYNNNFDSC KLLKEETSLI FADNHETLEV AHTKHILIGEL QFKIMLGDIR YKSFAESPEL
BUNV VILTVCITYP GHTFCTEINA IEPKITEEIE LQFKTVDTKT LPYIVAVNNH KLYSGQINDL GTFGQMCGNV QKTNSSILGT GTPKFDYTCH GASRKDIIVR RCYNNNFDSC KLLKEETQLI FNDDHDTITV YNTNHLIGEL AIKLILGDIQ YKLFTETLDL
```

Fig. 1. cont.

```
      1281                                                                                                                                          1440
 JCV  SAEGTCAGCI DCFENIHCEF TIHTTVEASC PITTECVSFH DRILITPDEH KYALKIVCHD KPKVSLKFKI CNTHVDAAMT IVDSKPILEL APVDQTTYIR EKDERCKTWM CRVRDEGFQV ILEPFKNLFG SYIGI.FYTG ILSIICLLIV VYIILPICFK
MELV  NAEGSCAGCI NCFENIHCEF TIHSTVETSC QIESDCTSFH DRVLITPNEH KYALKLICHD KPASSLKFKI CNTKVEAAMT LVDARPIIEL APVDQTAYIR EKDERCKTWM CRVRDEGIQV IFEPFKNLFG SYIGI.FYAI VICIIVLFLI VYIVLPICFK
KEYV  NAEGVCAGCI NCFENIHCEF TIHTTIEASC PIESDCVSFH DRMLVTPNEH KYALKLICED KPGSALKFKI CNTKIDAAMT LVDAKPILEL APVDQTAYIR EKDERCKTWM CRVRDEGIKV LLEPFKNLFG SYIGI.VYSI ILIIIAAIII IYIILPICFK
LUMV  TVEGACTGCI NCFDNIHCEL TIHSTVEASC PITSSCTVFH DRILITPNEH KYAIKVICNG KPGISLPFKI CNTKVDAAMT LTDAKQILEL APVDQTTYIK EKDERCKTWM CRVRDEGLQV IFEPFKNLFG SYIGI.FYTF VIFLVVIFIV IYIVLPICFK
LACV  TAEGSCTGCI NCFENIHCEL TLHTTIEASC PIKSSCTVFH DRILVTPNEH KYALKMVCTE KPGNTLTIKV CNTKVEASMA LVDAKPIIEL APVDQTAYIR EKDERCKTWM CRVRDEGLQV ILEPFKNLFG SYIGI.FYTF IISIVVLLVI IYVLLPICFK
 CEV  TAEGVCTGCI NCFENIHCEL TIHTTIEASC PVTSPCTLFH DRILITPDEH KYAIKVICTE KQTGTLPFKI CNSKIDATLT LVEAKPILEL APVDQTAYIR EKDERCKTWM YRVRDEGIQV ILEPFKNLFG SYIGI.FYTS ILVVIAMFII VYVILPICFK
TVTV  NAEGVCAGCI NCFERINCEF TIHSTVETSC PLVADCQLFH DRLLITPNEH KYAIKAICKD KPSNTLRFSI CNTKVEATVT IIDAKPILEL APVDQTAYIK EKDERCKTWL CRVKDEGVAV LLEPFKNLFG SYIGI.FYGI ILVVLAILVC IYIVLPVCFK
         v *  *   *     *                                           *              *                                     *
GROV  EVEGECIGCK NCLSGIDCNF KVTSPSETSC KIESPCSMFI ENLIIKAGHT DTNFKIHCKT IDSKVIEFKI CNNIYKVHIT ETKGNDKIEL NTGDQTSYVQ QHDNRCGTWL CRLKEEGIGL LLEPLKAIFG TYIT.WFWVI LGIIVFLFIG VYIFLPMIFK 1398
GERV  ELEAKCVGCV GCFESYQCNL QITSSLDETA LYLVPVSHFH DRIQIKTTKK DYAMKISCTR DPGDKASFRV CGKSYDFNFH TVPKNDKIEV NVGDETSYIK EKDNRCGRWL CRVRDEGLSV IFEPLNNFFG NYLNMFLYIL GGIIL.LFLA LYILMPMCAR
 CVV  EIDAKCVGCP SCFESYSCNF QIVSNIDTVC SVEGPCTLFH NRIIISANKQ SYGLKMSCQT KPNQNEEFII CNRKYSVLFT TIDKNDKIEV NTGDQTSYIY EKDSRCKTWL CRVRDEGISV LFEPLKAFFG SYFSIIFYVV VGIIV.LFLI IYIFLPMFFK
BUNV  QIDAKCVGCP DCFESYSCNF QIVSNIDTIC SLEGPCDTFH NRISIKAMQQ NYAVKLSCQK DPRPSGTFKI CNREYTVVFH TVAKDDKIEI NVGDQTSFIK EKDDRCKTWL CRVRDEGISV IFEPIKAFFG SYFSIFFYII VVVVV.GFLI IYIFMPMFMK

      1441       1461
 JCV  LRDTLRQHED AYKKEMKIR* .
MELV  LKDTLKQHED AYKREMKIR* .
KEYV  LRDTLKKHED AYRKEMKIR* *
LUMV  LRDTLRQHED AYKREMKIR* .
LACV  LRDTLRKHED AYKREMKIR* .
 CEV  LKDTLQRHED AYKREMKIR* .
TVTV  IKDTLKKHED AYKREMKIR* .
                   vv
GROV  VRDILKKNEE EYKKDIRYRE * 1418
GERV  LRDELKRNER LHQMEMKKR* .
 CVV  LKDVLKRNEY LYLQELKHK* .
BUNV  LKEVLKANEK LYLQEIKQK* .
```

Fig. 1. Alignment of M-segment sequences of selected CAL [Jamestown Canyon virus (JCV), MELV, Keystone virus (KEYV), Lumbo virus (LUMV), LACV, CEV and TVTV] and BUN (GERV, CVV and BUNV) viruses with that of GROV. ___, Potential transmembrane regions; h, potential h-region of predicted signal sequence; |, potential proteolytic cleavage; *, conserved cysteine; ///, epitopes identified by Cheng *et al.* (2000); + + +, potential glycosylation site; v, conservation between GROV and CAL serogroup virus sequences; xxx, conserved cleavage motif at C-terminus of GN; bold letters indicate predicted N-terminus of GC; ttt, potential trypsin cleavage site; #, amino acid position involved in neutralizing epitope of LACV (Bupp & González-Scarano, 1998). Amino acid positions for the GROV sequence are indicated at the end of each line.

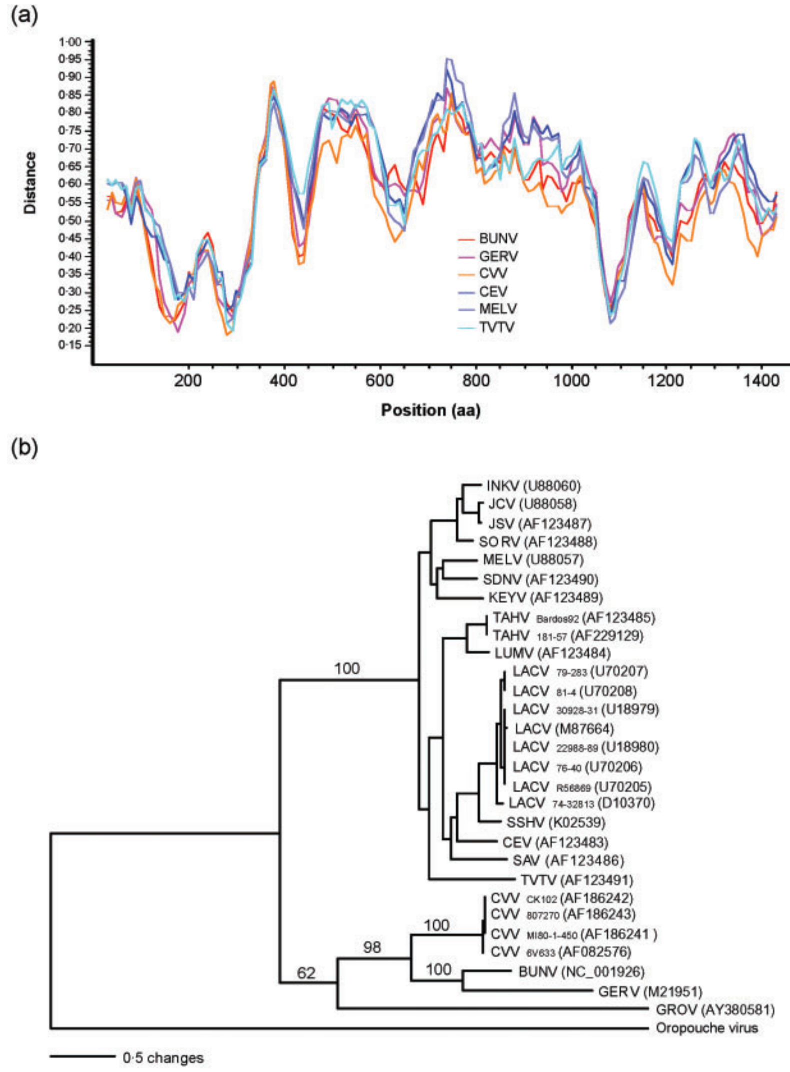

Fig. 2. Phylogenetic analysis of the GROV M-segment sequence. (a) Sliding-window distance analysis between GROV, BUN viruses BUNV, GERV and CVV and CAL viruses CEV, MELV and TVTV (amino acid sequence; window, 60 aa, step, 10 aa). (b) Reconstructed phylogenetic maximum-likelihood tree for nucleotide sequence. The tree was constructed by using the 'subtree prune regraft' (SPR) heuristic-search strategy under the general time-reversible model of nucleotide substitution with site-specific rate heterogeneity, modelled by using the discrete gamma distribution (Yang, 1994). Parameters were initially estimated on a neighbour-joining tree. Bootstrap support resulting from 500 SPR heuristic-search replicates is indicated for relevant branches. Oropouche virus (Simbu serogroup) served as an outgroup to root the tree. GenBank accession numbers are shown in parentheses. Abbreviations: INKV, Inkoo virus; JSV, Jerry Slough virus; SAV, San Angelo virus; SDNV, Serra do Navio virus; SORV, South River virus.

the glycosylation site N_{616}DI/T is intriguing, given that M-segment sequence has been associated with plaque size (Iroegbu & Pringle, 1981) and the observation that GROV can generate both large and small plaque morphologies, of which only the small variant elicited antibodies that were cross-reactive with CEV and TAHV (Tauraso, 1969).

In summary, our analysis of the GROV M-segment sequence indicates a relative phylogenetic relationship that is comparable to that reported for the GROV S-segment sequence (Fig. 2b; Dunn *et al.*, 1994), and does not provide evidence for genome-segment reassortment. Instead, in a sequence that is almost equidistant to published CAL and BUN virus sequences, isolated determinants in the N-terminal portion of GC were identified that potentially relate to the unique serological reaction pattern of GROV and are more compatible with GROV forming a bridge between both serogroups, as originally proposed by Whitman & Shope (1962).

Acknowledgements

We thank Bob Tesh for supplying the *Guaroa virus* stock, Wuxia Fu for excellent technical assistance and Charles Calisher for many helpful suggestions and discussions. This work was supported by awards from the Ellison Medical Foundation and NIH (AI 056118) to T. B. and W. I. L.

References

Amberg, S. M. & Rice, C. M. (1999). Mutagenesis of the NS2B-NS3-mediated cleavage site in the flavivirus capsid protein demonstrates a requirement for coordinated processing. *J Virol* **73**, 8083–8094.

Beaty, B. J., Bishop, D. H. L., Gay, M. & Fuller, F. (1983). Interference between bunyaviruses in *Aedes triseriatus* mosquitoes. *Virology* **127**, 83–90.

Beaty, B. J., Sundin, D. R., Chandler, L. J. & Bishop, D. H. L. (1985). Evolution of bunyaviruses by genome reassortment in dually infected mosquitoes (*Aedes triseriatus*). *Science* **230**, 548–550.

Blobel, G. & Dobberstein, B. (1975). Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J Cell Biol* **67**, 835–851.

Bouloy, M. & Hannoun, C. (1976). Studies on Lumbo virus replication. I. RNA-dependent RNA polymerase associated with virions. *Virology* **69**, 258–264.

Bouloy, M., Krams-Ozden, S., Horodniceanu, F. & Hannoun, C. (1973). Three-segment RNA genome of Lumbo virus (Bunyavirus). *Intervirology* **2**, 173–180.

Bouloy, M., Vialat, P., Girard, M. & Pardigon, N. (1984). A transcript from the S segment of the Germiston bunyavirus is uncapped and codes for the nucleoprotein and a nonstructural protein. *J Virol* **49**, 717–723.

Bridgen, A., Weber, F., Fazakerley, J. K. & Elliott, R. M. (2001). Bunyamwera bunyavirus nonstructural protein NSs is a nonessential gene product that contributes to viral pathogenesis. *Proc Natl Acad Sci U S A* **98**, 664–669.

Brockus, C. L. & Grimstad, P. R. (2001). Comparative analysis of G1 glycoprotein-coding sequences of Cache Valley virus (*Bunyaviridae*: *Bunyavirus*) isolates. *Virus Genes* **22**, 133–139.

Bupp, K. & González-Scarano, F. (1998). Pseudotype formation with La Crosse virus glycoproteins. *J Gen Virol* **79**, 667–671.

Calisher, C. H. & Maness, K. S. C. (1970). Arbovirus identification by an agar-gel diffusion technique. *Appl Microbiol* **19**, 557–564.

Campbell, W. P. & Huang, C. (1999). Sequence comparisons of medium RNA segment among 15 California serogroup viruses. *Virus Res* **61**, 137–144.

Casals, J. & Whitman, L. (1960). A new antigenic group of arthropod-borne viruses: the Bunyamwera group. *Am J Trop Med Hyg* **9**, 73–77.

Cheng, L. L., Schultz, K. T., Yuill, T. M. & Israel, B. A. (2000). Identification and localization of conserved antigenic epitopes on the G2 proteins of California serogroup bunyaviruses. *Viral Immunol* **13**, 201–213.

Claros, M. G. & von Heijne, G. (1994). TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci* **10**, 685–686.

Clewley, J., Gentsch, J. & Bishop, D. H. L. (1977). Three unique viral RNA species of snowshoe hare and La Crosse bunyaviruses. *J Virol* **22**, 459–468.

Dunn, E. F., Pritlove, D. C. & Elliott, R. M. (1994). The S RNA genome segments of Batai, Cache Valley, Guaroa, Kairi, Lumbo, Main Drain and Northway bunyaviruses: sequence determination and analysis. *J Gen Virol* **75**, 597–608.

Elliott, R. M. (1985). Identification of nonstructural proteins encoded by viruses of the Bunyamwera serogroup (family *Bunyaviridae*). *Virology* **143**, 119–126.

Elliott, R. M. (1989). Nucleotide sequence analysis of the small (S) RNA segment of Bunyamwera virus, the prototype of the family Bunyaviridae. *J Gen Virol* **70**, 1281–1285.

El Said, L. H., Vorndam, V., Gentsch, J. R. & 7 other authors (1979). A comparison of La Crosse virus isolates obtained from different ecological niches and an analysis of the structural components of California encephalitis serogroup viruses and other bunyaviruses. *Am J Trop Med Hyg* **28**, 364–386.

Endres, M. J., Jacoby, D. R., Janssen, R. S., González-Scarano, F. & Nathanson, N. (1989). The large viral RNA segment of California serogroup bunyaviruses encodes the large viral protein. *J Gen Virol* **70**, 223–228.

Fazakerley, J. K., González-Scarano, F., Strickler, J., Dietzschold, B., Karush, F. & Nathanson, N. (1988). Organization of the middle RNA segment of snowshoe hare bunyavirus. *Virology* **167**, 422–432.

Fuller, F. & Bishop, D. H. L. (1982). Identification of virus-coded nonstructural polypeptides in bunyavirus-infected cells. *J Virol* **41**, 643–648.

Gentsch, J. & Bishop, D. H. L. (1976). Recombination and complementation between temperature-sensitive mutants of a bunyavirus, snowshoe hare virus. *J Virol* **20**, 351–354.

Gentsch, J. R. & Bishop, D. H. L. (1978). Small viral RNA segment of bunyaviruses codes for viral nucleocapsid protein. *J Virol* **28**, 417–419.

Gentsch, J. R. & Bishop, D. H. L. (1979). M viral RNA segment of bunyaviruses codes for two glycoproteins, G1 and G2. *J Virol* **30**, 767–770.

Gentsch, J. R., Rozhon, E. J., Klimas, R. A., El Said, L. H., Shope, R. E. & Bishop, D. H. (1980). Evidence from recombinant bunyavirus studies that the M RNA gene products elicit neutralizing antibodies. *Virology* **102**, 190–204.

Gerbaud, S., Pardigon, N., Vialat, P. & Bouloy, M. (1992). Organization of Germiston bunyavirus M open reading frame and physicochemical properties of the envelope glycoproteins. *J Gen Virol* **73**, 2245–2254.

González-Scarano, F., Shope, R. E., Calisher, C. E. & Nathanson, N. (1982). Characterization of monoclonal antibodies against the G1 and N proteins of LaCrosse and Tahyna, two California serogroup bunyaviruses. *Virology* **120**, 42–53.

Grady, L. J., Sanders, M. L. & Campbell, W. P. (1987). The sequence of the M RNA of an isolate of La Crosse virus. *J Gen Virol* **68**, 3057–3071.

Groot, H., Oya, A., Bernal, C. & Barreto-Reyes, P. (1959). Guaroa virus, a new agent isolated in Colombia, South America. *Am J Trop Med Hyg* **8**, 604–609.

Hunt, A. R. & Calisher, C. H. (1979). Relationships of bunyamwera group viruses by neutralization. *Am J Trop Med Hyg* **28**, 740–749.

Iroegbu, C. U. & Pringle, C. R. (1981). Genetic interactions among viruses of the Bunyamwera complex. *J Virol* **37**, 383–394.

Jin, H. & Elliott, R. M. (1991). Expression of functional Bunyamwera virus L protein by recombinant vaccinia viruses. *J Virol* **65**, 4182–4189.

Kingsford, L. & Hill, D. W. (1983). The effect of proteolytic cleavage of La Crosse virus G1 glycoprotein on antibody neutralization. *J Gen Virol* **64**, 2147–2156.

Kingsford, L. & Boucquey, K. H. (1990). Monoclonal antibodies specific for the G1 glycoprotein of La Crosse virus that react with other California serogroup viruses. *J Gen Virol* **71**, 523–530.

Kingsford, L., Ishizawa, L. D. & Hill, D. W. (1983). Biological activities of monoclonal antibodies reactive with antigenic sites mapped on the G1 glycoprotein of La Crosse virus. *Virology* **129**, 443–455.

Klimas, R. A., Ushijima, U., Clerx-van Haaster, C. M. & Bishop, D. H. L. (1981). Radioimmune assays and molecular studies that place Anopheles B and Turlock serogroup viruses in the Bunyavirus genus (Bunyaviridae). *Am J Trop Med Hyg* **30**, 876–887.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.

Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem* **266**, 19867–19870.

Lappin, D. F., Nakitare, G. W., Palfreyman, J. W. & Elliott, R. M. (1994). Localization of Bunyamwera bunyavirus G1 glycoprotein to the Golgi requires association with G2 but not with NSm. *J Gen Virol* **75**, 3441–3451.

Lees, J. F., Pringle, C. R. & Elliott, R. M. (1986). Nucleotide sequence of the Bunyamwera virus M RNA segment: conservation of structural features in the Bunyavirus glycoprotein gene product. *Virology* **148**, 1–14.

Lindsey, H. S., Klimas, R. A. & Obijeski, J. F. (1977). La Crosse virus soluble cell culture antigen. *J Clin Microbiol* **6**, 618–626.

Lingappa, V. R., Katz, F. N., Lodish, H. F. & Blobel, G. (1978). A signal sequence for the insertion of a transmembrane glycoprotein. Similarities to the signals of secretory proteins in primary structure and function. *J Biol Chem* **253**, 8667–8670.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152–160.

Ludwig, G. V., Christensen, B. M., Yuill, T. M. & Schultz, K. T. (1989). Enzyme processing of La Crosse virus glycoprotein G1: a bunyavirus-vector infection model. *Virology* **171**, 108–113.

Ludwig, G. V., Israel, B. A., Christensen, B. M., Yuill, T. M. & Schultz, K. T. (1991). Monoclonal antibodies directed against the envelope glycoproteins of La Crosse virus. *Microb Pathog* **11**, 411–421.

Najjar, J. A., Gentsch, J. R., Nathanson, N. & González-Scarano, F. (1985). Epitopes of the G1 glycoprotein of La Crosse virus form overlapping clusters within a single antigenic site. *Virology* **144**, 426–432.

Nakitare, G. W. & Elliott, R. M. (1993). Expression of the Bunyamwera virus M genome segment and intracellular localization of NSm. *Virology* **195**, 511–520.

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**, 1–6.

Nilsson, I., Whitley, P. & von Heijne, G. (1994). The COOH-terminal ends of internal signal and signal–anchor sequences are positioned differently in the ER translocase. *J Cell Biol* **126**, 1127–1132.

Obijeski, J. F., Bishop, D. H. L., Murphy, F. A. & Palmer, E. L. (1976). Structural proteins of La Crosse virus. *J Virol* **19**, 985–997.

Pardigon, N., Vialat, P., Gerbaud, S., Girard, M. & Bouloy, M. (1988). Nucleotide sequence of the M segment of Germiston virus: comparison of the M gene product of several bunyaviruses. *Virus Res* **11**, 73–85.

Pekosz, A., Griot, C., Stillmock, K., Nathanson, N. & González-Scarano, F. (1995). Protection from La Crosse virus encephalitis with recombinant glycoproteins: role of neutralizing anti-G1 antibodies. *J Virol* **69**, 3475–3481.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–1354.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463–5467.

Stocks, C. E. & Lobigs, M. (1998). Signal peptidase cleavage at the flavivirus C-prM junction: dependence on the viral NS2B-3 protease for efficient processing requires determinants in C, the signal peptide, and prM. *J Virol* **72**, 2141–2149.

Sundin, D. R. & Beaty, B. J. (1988). Interference to oral superinfection of *Aedes triseriatus* infected with La Crosse virus. *Am J Trop Med Hyg* **38**, 428–432.

Tauraso, N. M. (1969). Identification of two plaque variants of Guaroa virus. *Arch Gesamte Virusforsch* **28**, 212–218.

von Heijne, G. (1988). Transcending the impenetrable: how proteins come to terms with membranes. *Biochim Biophys Acta* **947**, 307–333.

Weber, F., Dunn, E. F., Bridgen, A. & Elliott, R. M. (2001). The Bunyamwera virus nonstructural protein NSs inhibits viral RNA synthesis in a minireplicon system. *Virology* **281**, 67–74.

Wellings, F. M., Sather, G. E. & Hammon, W. M. (1970). A type-specific immunodiffusion technique for the California encephalitis virus group. *J Immunol* **105**, 1194–1200.

Wellings, F. M., Sather, G. E. & Hammon, W. M. (1971). Immunoelectrophoretic studies of the California encephalitis virus group. *J Immunol* **107**, 252–259.

Whitman, L. & Shope, R. (1962). The California complex of arthropod-borne viruses and its relationship to the Bunyamwera group through Guaroa virus. *Am J Trop Med Hyg* **11**, 691–696.

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* **39**, 306–314.