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

Thomas Briese,1 Andrew Rambaut2 and W. Ian Lipkin1

1Jerome L. and Dawn Greene Infectious Disease Laboratory, Mailman School of Public Health, Columbia University, New York, NY 10032, USA
2Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

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 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 viruses with respect to the serological reaction is discussed. No discernable evidence for reassortment was identified.
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), MgCl₂ (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 (Kozak, 1986). The predicted amino acid sequence for GN contains the sequence KSLRV/AAR, allowing protease cleavage to separate mature GN from the downstream NSm analogue (xxxxx; 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 (Gf416DFf/t/sNKGC/g/r/qCG25). 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 (A475), 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: A472, A475 and A479 (data not shown). Cleavage after A472 would result in positions −3 = V and −1 = A, one of the most frequent combinations in signalase sites. Cleavage after A479 would imply a long c-region, but would result in an N-terminal Gc-tripeptide E480EP, 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., 1994); SignalP prediction, ATM-LV or VVA-GE]. Cleavage after any of the three alanines in GROV occurs close to a potential glycosylation site (N492), but even cleavage at A479 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

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>3’ position</th>
<th>Annealing temperature (°C)</th>
<th>Extension time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUN-S5-F</td>
<td>5′-GCCGCAGTAGTGACTACCCGGAyA*</td>
<td>0</td>
<td>48</td>
<td>1</td>
</tr>
<tr>
<td>M940B-R</td>
<td>5′-CGwGCTGTwGrCTTTTrAmCCAG</td>
<td>915</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M590-F</td>
<td>5′-CATGCGwTGyTyyAdCwCAywATG</td>
<td>580</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>M1300-R</td>
<td>5′-rbCyrGAyTTrTdTgwGArTCACC</td>
<td>1265</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M940-F</td>
<td>5′-TGCCCCTGGKTAywAAAAyGTwAGAGC</td>
<td>936</td>
<td>51</td>
<td>2·5</td>
</tr>
<tr>
<td>GRO-D253R</td>
<td>5′-CCACAGCTATCAACTGCGCAT</td>
<td>2706</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M970-F</td>
<td>5′-ATGTGyATwTChArwGwGC</td>
<td>2450†</td>
<td>52</td>
<td>2</td>
</tr>
<tr>
<td>M3270-R</td>
<td>5′-CAnCGrAywTCywGrCwAncCC</td>
<td>3144</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRO-D149F</td>
<td>5′-CTGCGCTGCAAAAGATGGGAC</td>
<td>2640</td>
<td>53</td>
<td>2</td>
</tr>
<tr>
<td>GRO-E464R</td>
<td>5′-CTTGCGCTGCAATTTGTGTTTTCGA</td>
<td>3908</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3560-F</td>
<td>5′-TCAwArGyGyGyGwGnyATGT</td>
<td>3465</td>
<td>47</td>
<td>1</td>
</tr>
<tr>
<td>M4170-R</td>
<td>5′-CCywTACTCyCwCwACywCwkyACwATCCA</td>
<td>4061</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRO-U1092</td>
<td>5′-GwATGTTGTGCTGCGGAATTTGTTAAATAA</td>
<td>3570</td>
<td>45</td>
<td>1†</td>
</tr>
<tr>
<td>BUN-S3-R</td>
<td>5′-CGGCCAGTAGTGTGCTGCGGAATTTGTTAAATAA</td>
<td>4453</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Non-authentic bases added to some primers are indicated by italics.
†Position of misannealing is given.
‡Performed at 3·5 mM MgCl₂ (all other reactions were performed at 3 mM MgCl₂).
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 (N492); an N-terminal glycosylation site that is conserved in GC sequences of sequence BUN viruses was not found in the majority of GROV clones analysed. Among 10 clones, six carried AATG-ACAtA for N616DI (---; Fig. 1), whereas four carried AATG-ACAcA, encoding the potential glycosylation site N616DT.

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-termian, is compatible with a third internal signal sequence, with cleavage predicted after G337 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 K149, Q163 and P299 in GN, P347 and N455/F456 in NSm, and H591,QH, G601,ECNSA607, E958, K1036, G1243 and K1411/K1412 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 BUN/GERV/TVTV than from CEV/MELV/TVTV in all regions except for the second part of the second region (at the N-terminus of GC toward N616), 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 (H591,QH and G601,ECNSA607) and two glycosylation sites. N492, which is conserved amongst CAL but not BUN virus sequences, is present, whereas N616, which is conserved amongst BUN but not CAL virus sequences, is only present in a minority of GROV clones. N492 flanks the first putative antigenic domain that was proposed by Brockus & Grimstad (2001) and both conserved amino acid motifs and N616 are located in their second putative antigenic domain. It is conceivable that these positions contribute to the serological reaction of GROV. The mutation at
Fig. 1. cont.
**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; **l**, 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 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 N616DI/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 relatively 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


