Phylogeny of the Simbu serogroup of the genus *Bunyavirus*

Mohammad F. Saeed,1,3 Li Li,2,3 Heiman Wang,2,3 Scott C. Weaver1,2,3 and Alan D. T. Barrett1,2,3

1 Department of Microbiology & Immunology, 2 Department of Pathology and 3 Center for Tropical Diseases, The University of Texas Medical Branch, Galveston, TX 77555, USA

The Simbu serogroup of the genus *Bunyavirus*, family *Bunyaviridae* contains 25 viruses. Previous serological studies provided important information regarding some but not all of the relationships among Simbu serogroup viruses. This report describes the nucleotide sequence determination of the nucleocapsid (N) gene of the small genomic segment of 14 Simbu serogroup viruses and partial nucleotide sequence determination of the G2 glycoprotein-coding region (encoded by the medium RNA segment) of 19 viruses. The overall phylogeny of the Simbu serogroup inferred from analyses of the N gene was similar to that inferred from analyses of the G2 protein-coding region. Both analyses revealed that the Simbu serogroup viruses have evolved into at least five major phylogenetic lineages. In general, these phylogenetic lineages were consistent with the previous serological data, but provided a more detailed understanding of the relatedness amongst many viruses. In comparison to previous phylogenetic studies on the California and Bunyamwera serogroups of the *Bunyavirus* genus, the Simbu serogroup displays much larger genetic variation in the N gene (up to 40% amino acid sequence divergence).

Introduction

The family *Bunyaviridae* contains over 300 viruses, which are classified into five genera, including the *Bunyavirus* genus, which comprises more than 170 viruses. The genome of bunyaviruses consists of three segments of single-stranded, negative-sense RNA, designated large (L), medium (M) and small (S). The L segment encodes a large polyepitope, the L protein, which has been shown to have replicate and transcriptase activities (Jin & Elliott, 1991, 1992). The M segment encodes a polyprotein which undergoes post-translational proteolytic cleavage to give rise to virion surface glycoproteins G1 and G2, and a non-structural protein called NSm (Gentsch & Bishop, 1979; Fuller & Bishop, 1982; Elliott, 1985; Fazakerley et al., 1988; Gerbaud et al., 1992). The S segment encodes two proteins, the nucleocapsid (N) protein and a non-structural protein (NSs). These proteins are encoded in overlapping reading frames from the same mRNA (Elliott, 1990). The virion surface glycoproteins have been implicated in many of the important biological properties of bunyaviruses, including virulence, attachment, cell fusion and haemagglutination (Schmaljohn, 1996). Studies have indicated that neutralizing antibodies are directed against epitopes on the G1 glycoprotein. The N protein encapsidates one copy of each RNA segment, forming nucleocapsids that each contain a few molecules of L protein. It is believed that the N protein induces the formation of complement-fixing antibodies upon infection of an appropriate mammalian host.

Based on the serological relationships, viruses in the genus *Bunyavirus* have been divided into 18 serogroups. One of the largest serogroups within the genus is the Simbu serogroup, named after the prototype virus. This group, first described by Casals (1957), currently contains 25 related viruses, which have been isolated from all continents, except Europe (Calisher, 1996). Most of the Simbu serogroup viruses have been isolated from arthropods such as mosquitoes and culicoid midges, as well as from vertebrate hosts. Two members of this serogroup, Oropouche and Akabane, are of particular importance. Oropouche virus has been responsible for several large outbreaks of a dengue-like illness in human populations in South America (LeDuc & Pinheiro, 1989; Tesh, 1994; Pinheiro et al., 1998), while Akabane virus causes epizootics of congenital defects in cattle in Australia, Japan and the Middle East, and has been a serious threat to the livestock industry for many decades (Gonzalez-Scarano et al., 1991; Gonzalez-Scarano, 1996).

Classification of Simbu serogroup viruses is based upon antigenic relationships determined by plaque reduction neutralization, haemagglutination inhibition, complement fixation and radial immunodiffusion tests. Although a number of
reports describe serological relationships among some of the Simbu serogroup viruses (Takahashi et al., 1968; Calisher et al., 1969; Reeves et al., 1970; David-West, 1972; Causey et al., 1972), the most comprehensive analysis has been performed by Kinney & Calisher (1981), who studied the serological relationships among all the recognized members of the serogroup at the time of publication. These authors showed that although most Simbu serogroup viruses were readily distinguished in neutralization assays, they exhibited complex relationships by complement fixation tests. On the basis of cross reactivity in complement fixation tests, these authors divided the serogroup into five serocomplexes, Simbu, Manzannila, Oropouche, Thimiri and Nola. However, certain questions remain to be answered. For example, what are the evolutionary relationships among these viruses and what is the degree of relatedness among viruses in the same serocomplex? Furthermore, two viruses, Para and Jatobal were added to the Simbu serogroup after the establishment of serological classification and therefore the overall relationship of these viruses to other Simbu serogroup viruses has not been determined.

Comparative analysis of nucleic acid and protein sequences has become the primary method for determining virus relationships and examining virus evolution. Application of these methods to study the Simbu serogroup bunyaviruses has been hindered by a lack of sequence data. To date, the S RNA nucleotide sequences of only Aino, Akabane, Tinaroo and Oropouche viruses and the N gene nucleotide sequence of Jatobal virus have been published (Akashi et al., 1984, 1997a, b; Saeed et al., 2000, 2001). In this study, the nucleotide sequences of the N gene of 14 additional Simbu serogroup viruses, as well as partial nucleotide sequences of the M RNA segment (corresponding to the N-terminal half of the G2 glycoprotein) of 19 Simbu serogroup viruses were determined. Both N and G2 nucleotide sequence data were analysed to investigate the phylogeny of the Simbu serogroup.

Methods

■ Viruses and their propagation. Simbu serogroup viruses used in this study are listed in Table 1 along with their geographical origins, host sources and years of isolation. These viruses represent low passage isolates obtained from the World Arbovirus Reference Center at the University of Texas Medical Branch (UTMB), Galveston, TX, USA. Before use, each virus was propagated once in monolayer cultures of Vero (green monkey kidney) cells cultured in Eagle’s minimum essential medium (Sigma) supplemented with 2% foetal bovine serum (Gibco-BRL) and antibiotics (penicillin–streptomycin, Sigma). When 70–80% of cells exhibited cytopathic effects, the virus-containing cell culture supernatants were clarified of cellular debris by centrifugation, and stored frozen at −70 °C in 0.5–1.0 ml aliquots.

■ Amplification and sequencing of genomic sequences. Viral RNA was extracted from cell culture supernatant, using the method described by Ni & Barrett (1995). The purified viral RNA was then subjected to RT–PCR to amplify the genomic sequences. The N gene cDNA was amplified using the primer pair ORO1A/ORO2S or ORON5/ORON3, as described previously (Saeed et al., 2000). For some viruses a combination of ORO1A and ORON3 or ORON5 and ORO2S was used. To obtain M segment sequences, a partial length region encoding the N-terminal half of the G2 glycoprotein was amplified using primers M14C (5’ CGGAATTCGATGGACTAC1ACC 3’) and M619R (5’ GACATAGT(C/T)TGATGGAACGATC 3’), described by Fulhorst et al. (1996). The cDNA products were analysed by electrophoresis through a 1–2% agarose gel and purified using Qiagen gel extraction kit (QIAGEN), according to the manufacturer’s instructions. In most cases the amount of cDNA was sufficient for direct sequencing, which was carried out by dye termination cycle sequencing technique (Applied Biosystems) with the same primers used for PCR product amplification and the nucleotide sequences were resolved in an ABI 377 DNA sequencer. Both strands were sequenced for each cDNA. In cases where the amount of cDNA was insufficient for direct sequencing, the cDNAs were first ligated into a bacterial vector (pGEMT-easy, Promega), which was then amplified in E. coli JM109 cells and recovered using a SNAP plasmid miniprep kit (Invitrogen). Subsequently, the nucleotide sequence of the cloned cDNA was determined using primers directed towards the T7 and SP6 promoter regions of the vector flanking the cDNA cloning site. Both cDNA strands were sequenced for at least three plasmid clones.

■ Analysis of the nucleotide sequence data. The majority of the nucleotide sequence analyses, such as pairwise comparison, multiple sequence analyses, translation etc., were performed using various programs implemented in the PCGENE (Inteligentic) and Vector NTI software packages. For phylogenetic analyses, nucleotide sequences were aligned using either the ‘PileUp’ program of the University of Wisconsin Genetics Computer Group (UWGGC) package (Devereux et al., 1984) or the Vector NTI software with default settings. Phylogenetic analyses were carried out using neighbour-joining (NJ), Saitou & Nei, 1987) and maximum parsimony (MP) methods implemented in PAUP (Phylogenetic Analyses Using Parsimony) 4.01b software (Swofford, 1999).

For NJ analysis, a distance matrix was calculated from the aligned sequences using the Kimura 2-parameter formula (Kimura, 1980) and an NJ tree was computed. The reliability of the inferred tree was tested by bootstrap analysis (Felsenstein, 1985), performed on 1000 pseudoreplicate data sets generated from the original sequence alignment, and a bootstrap consensus tree was generated. For MP analysis, nucleotide characters were either assigned equal weights or transversions were assigned six times the weight of transitions. A search for the most parsimonious tree was performed using the heuristic algorithm. If more than one equally parsimonious tree was obtained, a consensus tree was computed. The reliability of the inferred consensus tree was tested by a bootstrap test performed on 1000 pseudoreplicate data sets using a heuristic search option and a bootstrap consensus tree was computed.

Results

Nucleotide sequence of the S RNA segment and/or N gene of Simbu serogroup viruses

Twenty-four Simbu serogroup viruses used in these studies are listed in Table 1, along with their place, source and year of isolation. Due to the unavailability of a viable isolate, Utive virus was not included in these studies. Of the 24 viruses examined, the S genomic RNA sequence of Oropouche (ORO) virus and N gene sequence of Jatobal (JAT) virus were reported in earlier studies (Saeed et al., 2000, 2001), while the S RNA sequences of Aino (AINO), Akabane (AKA) and Tinaroo (TIN) viruses were published by Akashi et al. (1984, 1997a, b) and were obtained from GenBank for comparative analyses. The remaining 19 Simbu serogroup viruses used for sequence
Table 1. Geographical origins, years and sources of isolation of Simbu serogroup viruses used in these studies

<table>
<thead>
<tr>
<th>Virus</th>
<th>Strain</th>
<th>Abbreviation</th>
<th>Year</th>
<th>Geographical origin</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aino*</td>
<td>JaNAr 28</td>
<td>AINO</td>
<td>1964</td>
<td>Japan</td>
<td>Mosquitoes</td>
</tr>
<tr>
<td>Akabane*</td>
<td>OBE-1</td>
<td>AKA</td>
<td>1974</td>
<td>Japan</td>
<td>Cattle</td>
</tr>
<tr>
<td>Buttonwillow</td>
<td>A 7956</td>
<td>BUT</td>
<td>1962</td>
<td>USA</td>
<td>Rabbits</td>
</tr>
<tr>
<td>Douglas</td>
<td>CSIRO 150</td>
<td>DOU</td>
<td>1978</td>
<td>Australia</td>
<td>Cattle</td>
</tr>
<tr>
<td>Facey’s Paddock</td>
<td>Aus Ch 16129</td>
<td>FP</td>
<td>1974</td>
<td>Australia</td>
<td>Mosquitoes</td>
</tr>
<tr>
<td>Ingwavuma</td>
<td>An 4165</td>
<td>ING</td>
<td>1959</td>
<td>South Africa</td>
<td>Birds</td>
</tr>
<tr>
<td>Inini</td>
<td>CaAn-128d</td>
<td>INI</td>
<td>1974</td>
<td>Brazil</td>
<td>Rodent</td>
</tr>
<tr>
<td>Jatobal</td>
<td>BeAn 423380</td>
<td>JAT</td>
<td>1985</td>
<td>India</td>
<td>Mosquitoes</td>
</tr>
<tr>
<td>Kaikalur</td>
<td>VRC 713423-2</td>
<td>KAI</td>
<td>1971</td>
<td>India</td>
<td>Mosquitoes</td>
</tr>
<tr>
<td>Manzanilla</td>
<td>AV-782</td>
<td>MER</td>
<td>1964</td>
<td>USA</td>
<td>Monkey</td>
</tr>
<tr>
<td>Mermet</td>
<td>TRVL 9760</td>
<td>ORO</td>
<td>1955</td>
<td>Trinidad</td>
<td>Human</td>
</tr>
<tr>
<td>Oropouche</td>
<td>IOT 4083</td>
<td>ORO</td>
<td>1997</td>
<td>Peru</td>
<td>Human</td>
</tr>
<tr>
<td>Oropouche</td>
<td>GML 445252</td>
<td>ORO</td>
<td>1989</td>
<td>Panama</td>
<td>Human</td>
</tr>
<tr>
<td>Para</td>
<td>CSIRO 110</td>
<td>PEA</td>
<td>1976</td>
<td>Australia</td>
<td>Midges</td>
</tr>
<tr>
<td>Sabo</td>
<td>AN 9398</td>
<td>SABO</td>
<td>1966</td>
<td>Nigeria</td>
<td>Goat</td>
</tr>
<tr>
<td>Sango</td>
<td>An 5077</td>
<td>SAN</td>
<td>1965</td>
<td>Nigeria</td>
<td>Cattle</td>
</tr>
<tr>
<td>Sathuperi</td>
<td>I-11155</td>
<td>SAT</td>
<td>1957</td>
<td>India</td>
<td>Mosquitoes</td>
</tr>
<tr>
<td>Shamonda</td>
<td>An 5550</td>
<td>SHA</td>
<td>1965</td>
<td>Nigeria</td>
<td>Cattle</td>
</tr>
<tr>
<td>Shuni</td>
<td>An 10107</td>
<td>SHU</td>
<td>1966</td>
<td>Nigeria</td>
<td>Cattle</td>
</tr>
<tr>
<td>Simbu</td>
<td>SA Ar 53</td>
<td>SIM</td>
<td>1955</td>
<td>South Africa</td>
<td>Mosquitoes</td>
</tr>
<tr>
<td>Tinaroo*</td>
<td>SCIRO 153</td>
<td>TIN</td>
<td>1978</td>
<td>Australia</td>
<td>Midges</td>
</tr>
<tr>
<td>Thimiri</td>
<td></td>
<td>THI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yaba-7</td>
<td>Y-7</td>
<td>YABA</td>
<td>1963</td>
<td>Nigeria</td>
<td>Mosquitoes</td>
</tr>
<tr>
<td>Utinga</td>
<td></td>
<td>UTI</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Sequence data obtained from GenBank. Accession nos AINO, M22011; AKA, AB000851; TIN, AB000819.

Determination were obtained from the World Arbovirus Reference Center at UTMB, Galveston, TX, USA and represented low passage isolates.

To amplify the S RNA-encoded N open reading frame (ORF), genomic RNA was extracted from purified viruses and subjected to RT–PCR, using the S RNA-specific (ORO1A and ORO2S) or N gene-specific (ORON5 and ORON3) primers or a combination thereof, as described previously (Saeed et al., 2000). The RT–PCR reactions yielded a cDNA product for all viruses used, except Manzanilla (MAN), Inini (INI), Para (PARA), Utinga (UTI) and Thimiri (THI). Subsequently, the cDNAs were sequenced either directly or following cloning and amplification in a bacterial vector (pGEMTeasy or pCR2.1). Analysis of the sequence data confirmed that six of them [Ingwavuma (ING), Kaikalur (KAI), Sabo (SABO), Yaba-7 (YABA), Simbu (SIM) and Douglas (DOU)] that were amplified by the S RNA-specific primers represented the complete S genomic RNA segment. Buttonwillow (BUT), which was amplified using a combination of ORON5/ORO2S primers, represented the complete N gene sequence plus the 3′NCR, while seven [Mermet (MER), Sathuperi (SAT), Shamonda (SHA), Shuni (SHU), Peaton (PEA), Sango (SAN) and Facey’s Paddock (FP)] that were amplified by the N gene-specific primers represented the complete N ORF (GenBank accession nos for these sequences are AF362392–AF362405). Further analyses of the sequence data indicated that in the genome-complementary orientation each cDNA contained two overlapping ORFs. The larger ORF corresponded to the N ORF, while the smaller ORF, which was in +1 frame and entirely contained within the N ORF, corresponded to the NSs ORF. Thus, overall genetic organization of the S RNA segment of these viruses was essentially the same as has been reported for other bunyaviruses, including Simbu serogroup viruses sequenced to date (Dunn et al., 1994; Bowen et al., 1995; Huang et al., 1996; Akashi et al., 1984, 1997a, b; Saeed et al., 2000). In each virus, except ORO and JAT, the N ORF consisted of 699 nucleotides and was predicted to encode a protein of 233 amino acids. The N ORF of ORO and JAT viruses consisted of 693 nucleotides, encoding a predicted protein product of 231 amino acids. The NSs ORF ranged in size from 273 to 288 nucleotides, and the predicted size of the NSs protein ranged from 91 to 96 amino acids (a table summarizing the features of S RNA and N ORF of Simbu serogroup viruses determined in this study and those published previously is provided on JGV Online as supplementary data, see http://vir.sgmjournals.org).
Comparison of the N ORF sequences of Simbu serogroup viruses

The nucleotide and predicted amino acid sequences of the N protein-coding regions of Simbu serogroup viruses determined in this study and those previously published were compared. The nucleotide sequence identity among these viruses ranged from 65% (between MER and SABO viruses) to 96% (between AINO and KAI viruses). Similarly, the amino acid sequence identity in the predicted N proteins ranged from 59-6% (between MER and DOU or between BUT and AKA viruses) to 99-1% (between SAN and PEA viruses). Several viruses displayed a very high nucleotide and amino acid sequence identity in the N ORF (see JGV Online for supplementary data, http://vir.sgmjournals.org). Based on a high nucleotide sequence identity (> 85%) and/or high amino acid sequence identity (> 90%), 16 of the 19 Simbu serogroup viruses were divided into five groups. Group I consisted of ORO and JAT viruses, group II contained MER and ING viruses, group III contained DOU, SAT and SHA viruses, group IV consisted of KAI, AINO, SHU, SAN and PEA viruses, and group V comprised AKA, TIN, YABA and SABO viruses. Viruses within a group exhibited very little (8% or less) amino acid sequence divergence, while variation between members of different groups ranged from 20 to 40%. Three viruses, BUT, SIM and FP, which did not fall into any of the five groups, exhibited significant (20–40%) amino acid sequence divergence from other Simbu serogroup viruses.

Despite significant amino acid sequence variation observed among several Simbu serogroup viruses, alignment of the N protein amino acid sequence indicated that among all Simbu serogroup viruses examined, 92 amino acids (approximately 40%) were identical, while 87 amino acids (approximately 37%) were conservative substitutions. Six regions (residues 5–14, 43–51, 71–81, 88–103, 123–138, 140–187) were highly conserved amongst the N proteins of all Simbu serogroup viruses and contained identical or conserved amino acids (see JGV Online for supplementary data, http://vir.sgmjournals.org).

Phylogenetic analyses of the N ORF nucleotide sequences

To examine the phylogeny of the Simbu serogroup, all available N ORF nucleotide sequences of Simbu serogroup viruses and those of representative members of California (Bowen et al., 1995) and Bunyamwera (Elliott, 1989; Dunn et al., 1994) serogroups were aligned and phylogenetic analyses were carried out. Since previous phylogenetic analysis of 28 ORO virus strains revealed the existence of three genotypes in South America (Saeed et al., 2000), three strains of ORO virus, representing each of the genotypes, were included in these analyses. Sequences of California and Bunyamwera viruses were used as outgroup to root the tree.

Phylogenetic analyses were carried out by NJ and MP methods (for MP analyses, weighted and unweighted parsimony methods were used, both of which yielded trees with similar topology and bootstrap support). The results indicated that the overall phylogenetic relationships among Simbu serogroup viruses determined by the NJ and MP methods were the same (Fig. 1). However, a few minor differences were noted between the results of NJ and MP analyses: (i) NJ analysis included the ING, MER, FP and BUT viruses in a single clade, while MP analysis indicated that inclusion of FP and BUT viruses in this clade is not well supported by bootstrap values and that these two viruses may constitute independent lineages; (ii) NJ analysis indicated that JAT virus is more closely related to the Peruvian genotype of ORO virus than to the Brazilian or Panamanian genotypes, but this relationship between JAT and Peruvian ORO strain was not...
Simbu serogroup bunyaviruses

resolved by MP analysis. To resolve these differences between the results of NJ and MP analyses, phylogenetic analysis using the maximum likelihood (ML) method was carried out, which indicated that the overall topology of the ML tree (not shown) was the same as those obtained by NJ and MP analyses; however, like MP analysis, ML analysis placed FP and BUT viruses in distinct lineages and JAT virus did not appear to be closely related to the Peruvian genotype of ORO.

Thus, from all these analyses it was deduced that, with respect to members of the California and Bunyamwera serogroups, all Simbu serogroup viruses clustered into a monophyletic group, which was further divided into five distinct phylogenetic lineages, designated I, II, III, IV and V.

Lineage I contained 13 viruses: TIN, AKA, YABA, SABO, SAT, SHA, DOU, SIM, KAI, AINO, SHU, PEA and SAN. Lineage II was represented by the ORO virus genotypes and JAT virus, while lineage III consisted of ING and MER viruses. Lineages IV and V were represented by FP and BUT viruses, respectively. Lineage I could be further divided into four smaller groups or clades (designated Ia to Id in Fig. 1). While the evolutionary relationships amongst clades were not always clear, the order of descent of several viruses within their respective clades was clearly established. For example, in clade 1a SHU virus occupied a basal position to AINO and KAI viruses; in clade 1b, SABO virus was basal to YABA virus, which in turn was basal to TIN and AKA viruses. Similarly, in clade 1c, DOU virus had a basal relationship to SAT and SHA viruses.

**Determination of partial N ORF nucleotide sequence of Inini, Para and Manzanilla viruses and their phylogenetic relationships to other Simbu serogroup viruses**

Despite repeated attempts, no N cDNA product was obtained for INI, MAN, PARA, THI or UTI viruses, using either S RNA-specific (ORO1A/ORO2S) or N gene-specific (ORON5/ORON3) primers or a combination thereof. Therefore, to obtain at least some sequence information a set of two new primers (SIMFOR2 and SIMREV2) was designed based on the conserved nucleotide sequences near the 5′ and 3′ termini of the N ORF of Simbu serogroup viruses sequenced above. The primer SIMFOR2 (5′ ATTTTCAACGATGTC-CAACCGGA 3′) consisted of 25 nucleotides representing the conserved region close to the 5′ terminus of the N ORF, while the primer SIMREV2 (5′ GAAGGCTCTAGCCTGCTG-GTGAATCC 3′) consisted of 27 nucleotides representing the complement of the conserved region near the 3′ terminus of the N ORF. Use of these primers in RT–PCR reactions yielded a cDNA product which migrated as expected (650 bp) for each of the INI, MAN and PARA viruses.

Analysis of the sequence data indicated that the cDNAs of INI, PARA and MAN viruses were 650, 660 and 625 nucleotides in length, respectively. The INI virus cDNA sequence exhibited highest nucleotide sequence identity with ORO virus N ORF (85%); however, the INI sequence was 43 nucleotides shorter, and with respect to the ORO virus N ORF, it was missing 13 nucleotides at the 5′ terminus and 30 nucleotides at the 3′ terminus. The PARA and MAN virus cDNA sequences displayed a very high nucleotide sequence identity with each other (99–4%) and both viruses displayed approximately 85% nucleotide sequence identity with N ORFs of ING and MER viruses. With respect to ING virus N ORF, the PARA virus cDNA was missing a total of 36 nucleotides (12 at the 5′ terminus and 24 at the 3′ terminus), while MAN virus cDNA was much shorter, with 62 nucleotides missing (42 from the 5′ terminus and 20 from the 3′ terminus).

To determine the relationships of INI, MAN and PARA viruses to other Simbu serogroup viruses, the partial N ORF nucleotide sequences of INI, PARA and MAN viruses were aligned with the homologous partial N ORF sequences of other Simbu serogroup viruses, and NJ analysis was performed. The resulting tree (Fig. 2) illustrated that INI virus is more...
M. F. Saeed and others

Fig. 3. Phylogeny of the Simbu serogroup based on a partial nucleotide sequence of the G2 glycoprotein gene. Partial G2 gene nucleotide sequences of Simbu serogroup were aligned using the Clustal W program implemented in Vector NTI software. The phylogram was estimated by NJ analysis using PAUP (version 4.01b). The distance matrix was calculated using the Kimura 2-parameter formula. Values adjacent to each node represent percentage bootstrap support calculated from 1000 replicates. Horizontal branch lengths are proportional to the scale bar, which represents 10% nucleotide sequence divergence.

Phylogeny of the Simbu serogroup based on a partial nucleotide sequence of the G2 glycoprotein gene. Partial G2 gene nucleotide sequences of Simbu serogroup were aligned using the Clustal W program implemented in Vector NTI software. The phylogram was estimated by NJ analysis using PAUP (version 4.01b). The distance matrix was calculated using the Kimura 2-parameter formula. Values adjacent to each node represent percentage bootstrap support calculated from 1000 replicates. Horizontal branch lengths are proportional to the scale bar, which represents 10% nucleotide sequence divergence.

closely related to ORO and JAT viruses, while PARA and MAN viruses (whose sequences were nearly identical to each other) exhibited a closer relationship to MER and ING viruses than to other viruses in the Simbu serogroup.

Partial nucleotide sequences of the G2 protein-coding region of Simbu serogroup viruses

Although the analyses of the N ORF sequences provided some information regarding the phylogenetic relationships amongst Simbu serogroup viruses, to obtain a better understanding of the evolutionary relationships among these viruses, nucleotide sequences from other genomic segments should be compared. Thus, purified RNA from all Simbu serogroup viruses, except AKA and Ulive, were subjected to RT–PCR using primers M14C and M619R (see Methods) to amplify a 570 nucleotide fragment representing the nucleotide sequence of the N-terminal half of the G2 glycoprotein, which is encoded in the 3’ end of the M genomic RNA segment. Using this approach a cDNA fragment for all but four viruses (UTI, PEA, INI and TIN) was obtained and subsequently sequenced. In pairwise comparisons, the nucleotide sequence identities among these sequences ranged from 51–9% (between SABO and JAT viruses) to 99–1% (between MAN and PARA viruses). However, in general, most viruses shared approximately 55–60% nucleotide sequence identity with each other (see JGV Online for supplementary data, http://vir.sgmjournals.org). Analysis of the predicted amino acid sequences indicated that following the translation initiation codon (Met) there is a signal peptide with a potential proteolytic cleavage site that, in all the sequences, preceded a conserved ‘P’ residue. The length of the signal peptide varied among these viruses and ranged from 13 to 18 amino acids. Alignment of the amino acid sequences revealed that only 32 (approximately 17%) amino acids were identical among all sequences; however, overall sequence similarity among the G2 protein of these viruses appears to be high as there were several stretches or blocks of amino acids that were represented by identical and/or conserved residues (see JGV Online for supplementary data, http://vir.sgmjournals.org).

Phylogenetic analysis of the G2 protein-coding nucleotide sequences

NJ analysis was performed to deduce relationships amongst Simbu serogroup viruses based on the partial nucleotide sequence of the G2 glycoprotein gene. The resulting phylogenetic tree (Fig. 3) depicts that although most Simbu serogroup viruses exhibit substantial genetic distance from each other, overall phylogeny corresponds to that obtained by the analyses of the N gene nucleotide sequences (compare Figs 1 and 2 with 3). As with N gene-based trees, NJ analysis of G2 gene sequences also revealed that the Simbu serogroup has evolved into five lineages, which corresponded well with the lineages inferred from the analysis of N ORF nucleotide sequences. Although distantly related, THI virus (for which the N gene sequence could not be determined) occupied lineage IV and displayed a relatively closer relationship to FP virus. As with the N gene-based tree, lineage I of the G2-based tree could also be divided into four clades. Clade Ia consisted of AINO, KAI and SHU viruses, clade Ib consisted of DOU and SAT viruses, clade Ic consisted of SABO, SHA and YABA viruses, while clade Id contained SIM virus.

Discussion

This study reports the first phylogenetic analyses of the Simbu serogroup. Complete nucleotide sequences of the N ORF (encoded by the S genomic segment) of 19 viruses (including 14 determined in this study and five that were previously published) and partial nucleotide sequences of the G2 glycoprotein gene (encoded by the M genomic segment)
of 19 viruses were compared phylogenetically. The overall phylogeny of the Simbu serogroup inferred from the phylogenetic analyses of the N ORF and the G2 protein-coding region was similar, but not identical (compare Figs 1 and 2 with 3). Furthermore, many of the viruses within lineages also exhibited a similar relationship by both G2- and N-based analyses.

It is important to note that with a few exceptions, the overall phylogeny of the Simbu serogroup deduced from N- and G2-based analyses corresponded very well with the serological classification proposed by Kinney & Calisher (1981). For example, lineage I of both N- and G2-based trees consisted of the same viruses and this lineage corresponded to the Simbu serocomplex. Similarly, lineage III of the G2-based tree corresponded to lineage III of the N-based tree; this lineage was congruent with the Manzanilla serocomplex. Among the notable discrepancies between the results of our phylogenetic analyses and Kinney & Calisher’s serological analyses was the placement of FP and BUT viruses. According to Kinney & Calisher, FP virus is a member of the Oropouche serocomplex and BUT virus belongs to the Manzanilla serocomplex. In contrast, our phylogenetic analyses (both N- and G2-based) indicated that FP and BUT viruses were members of separate lineages (lineages IV and V, respectively) distinct from those that corresponded to their proposed serogroups. However, FP and BUT viruses were serologically distinct (fourfold or greater difference in cross-neutralization and cross-complement fixation tests) from all other Simbu serogroup viruses (except that BUT virus exhibited a one-way reactivity to INI virus in neutralization tests). The inclusion of FP virus in the Oropouche serocomplex was based only on its very weak cross reactivity with UTI and Utive viruses in complement fixation tests, while BUT virus was included in the Manzanilla serocomplex because of its one-way reactivity to INI virus in neutralization tests (Kinney & Calisher, 1981). Also, G2-based phylogenetic analysis revealed that THI virus is distantly related to FP virus. In contrast, neutralization tests (that are directed towards the virion surface glycoproteins G1 and G2) indicated no cross reactivity between these two viruses and therefore these two viruses were placed in different serocomplexes (Kinney & Calisher, 1981). In the absence of nucleotide sequence data for the complete M segment, it is difficult to address the basis for this discrepancy.

As noted above, PARA virus is a relatively new addition to the Simbu serogroup and was discovered after the serological studies of Kinney & Calisher (1981). Therefore, the precise relationship of PARA virus to other Simbu serogroup viruses is unknown. Based on the partial nucleotide sequence data for the N ORF and G2 protein-coding region determined in this study, PARA virus was very closely related to MAN virus, and in the phylogenetic analyses the two viruses exhibited a very close relationship (Figs 2 and 3). Accordingly, it is likely that, antigenically, the two viruses are also closely related, and it can be speculated that one of them may be a variety of the other virus, as has been suggested for AINO and KAI viruses (Kinney & Calisher, 1981).

Despite an overall congruence between the N- and G2-based trees, certain marked differences in the topologies were also evident. For example, SAN virus exhibited a close relationship with AINO, KAI and SHU viruses based on N-derived tree, but was distantly related to these viruses in the G2-based tree and displayed a relatively closer relationship to the clade consisting of SABO, SHA and YABA viruses. Similarly, SHA virus, which exhibited a very close relationship with SAT virus in the N-based tree, was only distantly related to SAT virus in the G2-based tree and showed a very close relationship to YABA virus. A possible reason for these incongruences between N- and G2-based trees may be genetic reassortment among the ancestral viruses leading to emergence of new viruses with the same S genomic segment, but different M segment or vice versa. However, until a complete nucleotide sequence of the M segment is determined, this possibility will remain speculative.

Comparative analysis of the N ORF sequences of 19 California serogroup viruses indicated that amino acid sequence variation among these viruses ranged from < 1 % to 20%, with the exception of Trivittatus virus, which exhibited maximum (24–27%) variation in this serogroup (Bowen et al., 1995; Huang et al., 1996). In contrast, maximum amino acid sequence variation observed among the N protein amino acid sequences of Simbu serogroup viruses was > 40 %, suggesting that the extent of divergence among some Simbu serogroup viruses was greater than that observed even among the most distantly related California serogroup viruses reported to date. Further evidence for a greater divergence of Simbu serogroup viruses comes from the observation that only 40% of the amino acid residues in the N protein were identical amongst all Simbu serogroup viruses analysed. In contrast, alignment of the N protein amino acid sequences of either California or Bunyamwera serogroup viruses sequenced to date revealed that nearly 60 % amino acid residues were identical amongst members of a serogroup, while nearly 40% residues were identical between members of the two serogroups (Dunn et al., 1994; Bowen et al., 1995; Huang et al., 1996). The reason for this greater divergence among Simbu serogroup viruses is unclear; however, a possible explanation may be that the Simbu serogroup was established earlier than the California and Bunyamwera serogroups in the evolutionary history of the Bunyaviridae genus. The greater divergence among Simbu serogroup viruses may also be a reflection of the extent of their geographical distribution. Among the three serogroups (Simbu, Bunyamwera and California), the Simbu serogroup has the widest geographical distribution, while most viruses (12 of 14) in the California serogroup have been isolated from a relatively narrow geographical range (North and South America). Similarly, vector association may also have contributed to a greater divergence of Simbu serogroup. Most Simbu serogroup viruses are associated with biting midges,
while viruses of California and Bunyamwera serogroups are mainly associated with mosquitoes (Calisher, 1996).

Comparison of the nucleotide sequence representing the N-terminal half of the G2 glycoprotein-coding region revealed that there was a remarkably high degree of genetic diversity in this region of the genome among Simbu serogroup viruses. With the exception of a few viruses, most viruses exhibited a 40–45% divergence in the nucleotide sequence (and 40–50% divergence in the deduced amino acid sequence) with each other (see JGV Online for supplementary data, http://vir.sgmjournals.org). This is in sharp contrast to the data obtained for the N ORF sequences, where numerous Simbu serogroup viruses displayed a very close genetic relationship with each other (see JGV Online for supplementary data, http://vir.sgmjournals.org). This apparent incongruence may be related to the location and functions of the two proteins in virus particles. G2 glycoprotein, encoded by the M genomic segment, is located on the surface of virions, while the N protein, encoded by the S genomic segment, is an internal protein. In addition, Simbu serogroup viruses have been isolated from a variety of hosts including mammals, birds and insects. Therefore, it can be speculated that during evolution of these viruses, the need to infect an alternate host may also have played some role in greater genetic divergence in the M segment. Another possible explanation for the incongruence in the M and S segment-derived sequence data may be genetic reassortment. It is possible that during evolution of Simbu serogroup viruses, some of the ancestral viruses underwent genetic reassortment involving M genomic segments, resulting in new viruses with similar S segments, but diverse M segments. The explanations presented above are not mutually exclusive and it is possible that all may have contributed to a higher genetic diversity in the M segment of different Simbu serogroup viruses.

In conclusion, our phylogenetic analyses demonstrated relationships amongst the Simbu serogroup viruses that were consistent with the results of serological tests (Kinney & Calisher, 1981). However, more importantly, phylogenetic analyses resolved relationships among several viruses that could not be distinguished by complement fixation tests.

We wish to thank Drs Robert E. Shope and Robert B. Tesh for providing the Simbu serogroup viruses used in these studies. We also thank Drs Stuart T. Nichol, Michael D. Bowen, Pierre Rollin and C. J. Peters for helpful discussions. This work was supported in part by NIH grant AI 43336. Mohammad F. Saeed was supported in part by the James W. McLaughlin Fellowship Fund.

References


Received 13 December 2000; Accepted 10 May 2001