Variability of the NS₅ protein among Rift Valley fever virus isolates

A. A. Sall,¹ P. M. de A. Zanotto,² H. G. Zeller,³ J. P. Digoutte,¹ Y. Thiongane⁴ and M. Bouloy⁵

¹ Institut Pasteur, 36 Avenue Pasteur, BP 220 Dakar, Senegal
² DIPA-UNIFESP, São Paulo, Brazil
³ Institut Pasteur de Madagascar, BP 1274, Tananarive, Madagascar
⁴ Institut Sénégalais de Recherche Agronomique, BP 2057 Dakar, Senegal
⁵ Laboratoire des Bunyavirides, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France

Eighteen strains of Rift Valley fever (RVF) virus collected over a period of 38 years and isolated from diverse localities in Africa and from various hosts (human, animal and arthropod) were investigated by RT–PCR followed by sequencing of the NS₅ protein coding region. This region was chosen to analyse variability because, in contrast to the N protein, the NS₅ protein differs in various phleboviruses and there exists an RVF virus (clone 13) in which 70% of the NS₅ ORF is deleted, suggesting that this sequence is under a weak selective pressure. Sequence data indicated that percentage divergence among isolates ranged from 0 to 9.6% at the nucleotide level and from 0 to 9.5% at the amino acid level. Phylogenetic analysis based on the NS₅ gene revealed two major lineages: Egyptian and sub-Saharan. This led to the establishment of the relatedness between strains and insights into the NS₅ protein, the function of which is still undetermined. Alignment of the deduced amino acid sequences indicated that the cysteine residues are conserved, as are several motifs representing potential phosphorylation sites.

Introduction

Rift Valley fever (RVF) is an arthropod-borne viral disease in Africa, primarily affecting ruminants, which results in high mortality in young and adult animals as well as abortions. Since the epidemic in Egypt in 1977, RVF virus has been shown to infect humans extensively, leading to several disease manifestations including acute febrile illness, hepatitis, encephalitis, haemorrhagic fever and ocular sequelae (Laughlin et al., 1979). Recent RVF epidemics in Mauritania (Digoutte & Peters, 1989), Madagascar (Morvan et al., 1992) and Egypt (Arthur et al., 1993), as well as the enzootic maintenance of the virus in Senegal (Zeller et al., 1997), have emphasized the emergence of this virus as a major public health threat in Africa. With the exception of Egypt, RVF virus is confined to sub-Saharan Africa, where it is widespread and maintained in endemic/ enzootic and epizootic forms (Meegan & Bailey, 1989). The survival of RVF virus during inter-epizootics is believed to depend on transovarial transmission of virus in floodwater Aedes mosquitoes (Linthicum et al., 1985).

RVF virus belongs to the genus Phlebovirus in the family Bunyaviridae, a genus which contains two serogroups: the sandfly fever group and the tick-transmitted uukuviruses (Murphy et al., 1995). According to serological data based on complement fixation, neutralization and haemagglutination inhibition tests, the sandfly fever group is composed of at least eight antigenic complexes: Bajaru, Candiru, Chilibre, Frijoles, Punta Toro, Rift Valley fever, Salehabad and an unassigned complex containing 16 viruses.

The genome of phleboviruses consists of three negative-sense single-stranded RNA segments designated L (large), M (medium) and S (small). The L segment codes for the L protein, the viral RNA polymerase. The M segment codes for a precursor to the envelope glycoproteins G1 and G2 and two non-structural proteins of 14 and 78 kDa. The S segment codes for nucleocapsid protein N and the non-structural protein NS₅ by an ambisense strategy (for reviews see Bouloy, 1991; Elliott et al., 1991; Giorgi, 1996; Schmaljohn, 1996).

Molecular and biological variability among phleboviruses has been studied for RVF virus (Battles & Dalrymple, 1988; Anderson & Peters, 1988; Saluzzo et al., 1989; Besselar et al., 1989).
1991) and to a lesser extent for Toscana virus (Schwarz et al., 1995). These reports indicated that among natural isolates of RVF virus, the antigenic properties of the glycoproteins and the nucleoprotein appeared to be stable, and sequences of the G2 epitopes in the M segment are relatively conserved. On the other hand, the Egyptian strains differed from the sub-Saharan strains by their pathogenicity in Wistar-Furth rats. Since the NS$_5$ protein is the most variable protein among phleboviruses (Giorgi et al., 1991), its use as a marker of variability was investigated. The extent of conservation of the NS$_5$ gene among RVF virus isolates is not known. This question was raised by the existence of the naturally occurring avirulent isolate ‘clone 13’ (Muller et al., 1995), which harbours a prominent deletion in the NS$_5$ region. This suggests that this domain is under weakly constraining evolutionary forces, resulting in the potential for accumulation of substitutions in this region. A panel of 18 strains of RVF virus, collected over a period of 38 years and isolated from various hosts in different countries, was selected and the NS$_5$ gene was amplified by RT–PCR and sequenced. This analysis was carried out to establish the relatedness between strains as well as to gain insights into the conserved motifs of this protein, the function of which is still undetermined.

**Methods**

- **Virus propagation and RNA extraction.** The origins of RVF virus isolates used in this study are shown in Table 1. Viruses were propagated in Vero cells cultivated in Leibovitz 15 growth medium supplemented with 5% foetal calf serum (Gibco BRL). At 48 to 72 h post-infection, infected cultures showed cytopathic effects and the presence of viral antigens could be detected by an indirect immunofluorescence assay. Cytoplastic extracts were obtained by the NP40 lysis method and total RNA was extracted with phenol–chloroform (1:1) and precipitated with ethanol as previously described (Bouloy et al., 1984).

- **RT–PCR and sequencing procedures.** Synthesis of first strand complementary DNA was primed with oligodeoxynucleotide NS3a (Muller et al., 1995) after denaturation of the RNA with 0.01 M methyl mercuric hydroxide for 10 min at room temperature. Methyl mercuric hydroxide was then complexed with 2 μl of 0.7 M β-mercaptoethanol, and reverse transcription was carried out with 100 ng of primer in a final volume of 20 μl containing 25 mM Tris–HCl pH 8.3, 25 mM KCl, 5 mM MgCl$_2$, 0.2 mM of each dNTP (Pharmacia) and 4 units of AMV reverse transcriptase (Promega).

  A 5 μl aliquot of the reverse transcription reaction mix was used for enzymatic amplification. PCR was carried out with 2.5 units of Taq DNA polymerase (Amersham) in a 100 μl final volume containing 10 mM Tris–HCl pH 8.3, 50 mM KCL, 1.5 mM MgCl$_2$, 0.2 mM of each dNTP and 0.5 μg of each primer (NS3a and NS2g) (Muller et al., 1995). Amplification was performed after a denaturation step at 95 °C for 3 min followed by 35 amplification cycles as follows: denaturation at 95 °C for 1 min, annealing at 53 °C for 1 min and extension at 72 °C for 1 min.

  Amplified DNA fragments were purified on an agarose gel with the Geneclean kit (Bio 101) according to manufacturer’s instructions and directly sequenced using the thermal cycle sequencing method with oligodeoxynucleotides labelled at the 5′ end with [γ-3P]ATP and polynucleotide kinase (Adams & Blakesley, 1991). Sequencing reactions were done with the primers used for PCR amplification. Most of the sequences were read on the two strands. The sequences were deposited in the EMBL database (accession numbers Y12739–Y12756).

- **Sequence alignments and phylogenetic analysis.** The amino acid sequences of the NS$_5$ protein were aligned using Clustal V (Higgins et al., 1992) in order to determine the positions of gaps in the alignment of the corresponding nucleotide sequence data. Phylogenetic trees for the NS$_5$ nucleotide sequences were estimated using the FastDNAmix program (Olsen et al., 1994). The transition/transversion substitution rate parameter (ts:tv) was calculated by searching for the ts:tv value which gives the highest likelihood. In order to establish the robustness of the topologies, bootstrap resampling was done (Swofford, 1993). In addition, given an optimal ts:tv parameter, all the trees with log likelihoods (lnL) not significantly different from the best maximum likelihood tree were collected. The level of resolution of the 50% majority rule consensus of this tree pool is then indicative of the topology stability at the tree space around the lnL maxima. These results were also compared to recon-

**Table 1. Characteristics of the RVF virus isolates used in this study**

<table>
<thead>
<tr>
<th>Code</th>
<th>Strain</th>
<th>Year</th>
<th>Source</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNS*</td>
<td>Smithburn</td>
<td>1944</td>
<td>Strain</td>
<td>Uganda</td>
</tr>
<tr>
<td>Ar UG 55</td>
<td>Lunyos</td>
<td>1955</td>
<td>Mosquito</td>
<td>Uganda</td>
</tr>
<tr>
<td>Ar CAR 69</td>
<td>Ar B 1976</td>
<td>1969</td>
<td>Mosquito</td>
<td>CAR</td>
</tr>
<tr>
<td>H EGY 77</td>
<td>ZH 548</td>
<td>1977</td>
<td>Human</td>
<td>Egypt</td>
</tr>
<tr>
<td>MP12*</td>
<td>MP12</td>
<td>1977</td>
<td>Strain ZH 548</td>
<td>Egypt</td>
</tr>
<tr>
<td>B EGY 93</td>
<td>B EGY 93</td>
<td>1993</td>
<td>Buffalo</td>
<td>Egypt</td>
</tr>
<tr>
<td>H EGY 93</td>
<td>H EGY 93</td>
<td>1993</td>
<td>Human</td>
<td>Egypt</td>
</tr>
<tr>
<td>Ar MAD 79</td>
<td>Ar Mg 811</td>
<td>1979</td>
<td>Mosquito</td>
<td>Madagascar</td>
</tr>
<tr>
<td>An MAD 91</td>
<td>An Mg 900</td>
<td>1991</td>
<td>Bovine</td>
<td>Madagascar</td>
</tr>
<tr>
<td>H1 MAU 87</td>
<td>H D 47502</td>
<td>1987</td>
<td>Human</td>
<td>Rosso</td>
</tr>
<tr>
<td>H2 MAU 87</td>
<td>H D 47311</td>
<td>1987</td>
<td>Human</td>
<td>(Mauritania)</td>
</tr>
<tr>
<td>H3 MAU 87</td>
<td>H D 47408</td>
<td>1987</td>
<td>Human</td>
<td>Yeur</td>
</tr>
<tr>
<td>H4 MAU 87</td>
<td>H D 48255</td>
<td>1987</td>
<td>Human</td>
<td>Tegmante</td>
</tr>
<tr>
<td>An GUI 84</td>
<td>An K 6087</td>
<td>1984</td>
<td>Bat</td>
<td>Kindia</td>
</tr>
<tr>
<td>Ar BUR 84</td>
<td>Ar D 38457</td>
<td>1984</td>
<td>Mosquito</td>
<td>Burkina</td>
</tr>
<tr>
<td>Ar SEN 84</td>
<td>Ar D 38661</td>
<td>1984</td>
<td>Mosquito</td>
<td>Faso</td>
</tr>
<tr>
<td>Ar SEN 93</td>
<td>Ar D 104769</td>
<td>1993</td>
<td>Mosquito</td>
<td>Kedougou</td>
</tr>
<tr>
<td>An SEN 93</td>
<td>An D 106417</td>
<td>1993</td>
<td>Bovine</td>
<td>Kolda</td>
</tr>
</tbody>
</table>

* Laboratory attenuated strain derived from a wild strain. SNS, Smithburn neurotropic strain.
The S genomic sequences of five phleboviruses, Punta Toro, RVF (MP12 strain), SFS, Toscana and Uukuniemi viruses, have already been published (Ihara et al., 1984; Marriott et al., 1989; Simons et al., 1990; Giorgi et al., 1991) and were shown to utilize an ambisense strategy to code for the NS₃ protein in the genomic sense and for the N protein in the antigenomic sense. The N and NS₃ ORFs represent two potential distinct targets for analysis of the S segment. The deduced N protein sequences exhibit homologies ranging from 30 to 54% (Giorgi et al., 1991); however, the NS₃ gene is much less conserved and cannot be aligned to analyse phylogenetic relationships among different phleboviruses. Noteworthy is that the function of the protein in the virus life cycle has not yet been determined and that its localization in infected cells, as well as its presence in viral particles, varies depending on the virus. With the exception of the NS₃ protein from RVF virus, which has been reported to form filaments in the nuclei of infected cells (Struthers & Swanepoel, 1982; Struthers et al., 1984), the NS₃ protein of the phleboviruses studied was detected in the cytoplasm. In the case of Uukuniemi virus, Simons et al. (1992) demonstrated that the protein was associated with the 40S ribosomal subunit and not detected in mature particles. In contrast, for Punta Toro virus, the term ‘non-structural’ does not seem appropriate since the protein was found associated with the ribonucleoproteins in mature particles (Overton et al., 1987).

The diversity of the NS₃ gene among phleboviruses raised the question of the extent of conservation of the NS₃ gene among strains of a particular virus. This was underlined by the existence of clone 13, a natural, attenuated isolate of RVF virus which harbours a large deletion in the NS₃ region and replicates in various hosts including mosquitoes, suggesting that the deleted region represents a potentially variable domain (Muller et al., 1995). Therefore, we decided to analyse the NS₃ gene of various strains of RVF virus and selected a panel of 18 RVF virus strains collected over 38 years in eight countries under epidemic and endemic conditions and isolated from a variety of host species (arthropods, cattle and humans). The characteristics of the isolates are described in Table 1. The attenuated strains MP12 (Caplen et al., 1985) and Smithburn neurotropic strain (SNS; Smithburn, 1949), which were obtained by cell passages of the parental virulent strains ZH 548 and Entebbe, respectively, were included in this study.

Two oligodeoxynucleotides, NS2g and NS3ag (Muller et al., 1995), were designed based on the published sequence of the MP12 strain and utilized as primers for RT–PCR amplification. They allowed amplification of a 601-nucleotide-long DNA fragment which was obtained for all the strains analysed and subsequently sequenced. Using these primers, clone 13 yielded a 121-nucleotide-long DNA fragment (Muller et al., 1995), which was too small to be included in this analysis. Nucleotide substitutions were scattered throughout the entire region and, with the exception of clone 13, no base insertion or deletion was observed. Pairwise comparison among these isolates showed percentage divergence ranging from 0 to 9.6% at the nucleotide level which corresponded to 0 to 9.5% at the amino acid level.

The phylogenetic tree of the nucleotide sequences was generated using the optimal ts:tv value of 2, which was determined empirically. The SFS virus NS₃ sequence was used as an outgroup to determine the coalescence node for RVF virus sequences and therefore indicate a root for the tree. Using ultrametric methods and excluding the SFS virus branch, a congruent root was obtained for Ar MAD 79 (Fig. 1). Similar trees were obtained with the maximum likelihood distance methods and maximum parsimony (not shown). Apart from Ar MAD 79, the isolates are distributed in at least two major lineages, Egyptian and sub-Saharan, with the latter lineage divided into two clusters, Ia and Ib. All the branchings of the tree are highly supported by bootstrap values. Group Ib strains were isolated only in West Africa (Senegal, Mauritania, Guinea and Burkina Faso), whereas group Ia appears more heterogeneous, including strains from Uganda, Madagascar, Central African Republic as well as strains from Senegal and Mauritania.

The existence of the Egyptian lineage is in good agreement with a set of data indicating that the strains from the 1977 outbreak could be distinguished from the sub-Saharan strains by several criteria: (i) the N protein of the Egyptian strains possesses a specific epitope recognized by the R1P2E7 monoclonal antibody (Saluzzo et al., 1989); (ii) the Egyptian strains were also highly pathogenic for Wistar-Furth rats; (iii) these strains were resistant to rat interferon; and (iv) they did not form plaques in primary rat hepatocytes (Anderson & Peters, 1988). As expected, the MP12 strain is closely related to its parent, H EGY 77, but surprisingly, is placed in a more ancestral position. The use of a mutagen to obtain MP12 may have altered the mutation regime in a very biased way for the NS₃ sequence. Moreover, this analysis shows that the strains isolated in 1977 and 1993 in Egypt are very close to each other, suggesting that the same virus remained endemic for a
Fig. 1. Phylogenetic tree for the NS₅ gene of RVF virus isolates. Values above branches indicate the level (%) of bootstrap support using maximum parsimony after 500 iterations. Values below branches indicate the number of times a given node was observed on a majority rule consensus of 50 trees with equivalent likelihood (lnL). Branch lengths shown are proportional to the number of substitutions per 100 residues. The rooting shown here was determined by the inclusion of the SFS virus NS₅ sequence as the outgroup.

long period in a large area or that the viruses circulating in 1977 and 1993 were introduced from the same source. Grouping based on geographical origin applies also to the two strains from Uganda, Ar UG 55 and SNS, a laboratory-attenuated strain derived from the Entebbe strain isolated in Uganda in 1944. Nevertheless, some of the groupings, especially those of some Mauritanian and Senegalese strains, are unexpected. The phylogenetic tree indicates that two phylogenetically different viruses corresponding to clusters Ia and Ib were circulating during the Mauritanian outbreak in 1987. However, the strains analysed originated from three places less than 50 km apart and from cases occurring over less than 1 month (see Table 1). It should be noted that strain H1 MAU 87 was isolated from a febrile case, and the others were isolated from fatal cases. Moreover, H1 MAU 87 is closely related to the Senegalese strain Ar SEN 93 which was isolated from mosquitoes trapped in 1993 in the Sahelian Ferlo, a northern region close to the Mauritanian border where the 1987 outbreak occurred (Zeller et al., 1997). On the other hand, the position of strain An SEN 93 within a different cluster could be explained by its origin; it was isolated from a healthy zebu in southern Senegal (Kolda, Casamance). Interestingly, this strain clustered with those from the neighbouring countries of Guinea and Burkina Faso. This suggests that Senegal is divided in two areas as far as RVF virus is concerned: the Sahelian zone where southern Mauritanian and northern Senegalese strains are circulating, probably through livestock and mosquitoes, and the Sudano–Guinean zone where southern Senegalese strains are in contact with those from bordering countries.

Regarding its geographical origin (i.e. from mosquitoes collected in Kedougou, a place in southern Senegal close to Guinea and Mali), Ar SEN 84 would be expected to cluster with An GUI 84 which was isolated from a bat in Kindia (Guinea), a bioclimatic area similar to Kedougou and in the vicinity of the Senegalese border. Its presence within the Egyptian lineage, although puzzling, may be due to a reassortment event. Indeed, the circulation of phylogenetically distinct strains in the same region and at the same period provides ideal conditions for a host to be co-infected and generate reassortants as occurs naturally for other members of the bunyavirus family like hantaviruses (Li et al., 1995; Henderson et al., 1995) and in other families of viruses with segmented genomes (Murphy & Webster, 1990). Further work will be necessary to test whether some of these viruses result from reassortment.

Finally, the strains isolated in Madagascar in 1991 and 1979 belong to different lineages; strain Ar MAD 79 is closely related to the Egyptian group. This relatedness was noted by Morvan et al. (1992) who analysed the antigenic properties of the N protein. The topology of the tree reported in Fig. 1 indicates that the virus isolated in the island in 1979 occupies the most ancestral position and shares a common ancestor with the Egyptian strains. For these reasons, Ar MAD 79 was included in the Egyptian lineage. On the other hand, strain An MAD 91 appears more closely related to the eastern or central African strains. As speculated by Morvan et al. (1992), it is possible that during this latter outbreak, RVF virus had been introduced to the island from the southern or eastern coast of Africa through livestock commerce.

Insights concerning the conserved amino acids in the NS₅ protein

Six potential phosphorylation sites for casein kinase 1 and 2, corresponding to the motifs S/T-X-R/K and S/T-X-X-D/E, were identified in the NS₅ protein of the MP12 strain. All of them are conserved in the strains SNS, H EGY 77, B and H EGY 93, and Ar SEN 84, the complete NS₅ ORF of which was sequenced. Among the other strains studied and within the region analysed, there are three potential sites at positions 98, 128 and 149 which are all conserved. This suggests that the protein could be phosphorylated at multiple sites.

The amino acid sequence alignment also indicates that the five cysteine residues at positions 39, 40, 150, 179 and 195 are
conserved. It is possible that these residues form either intra- or intermolecular disulfide bonds which are important for conservation of the three-dimensional structure of the protein and the formation of filamentous structures observed in cells infected with all the strains analysed so far, except for strain C13. Also noteworthy is the presence of the RGD motif at position 24, conserved in all isolates except for An MAD 91 in which arginine is replaced by lysine. The existence of such a motif, commonly present in extracellular adhesive proteins, in this intracellular protein is surprising.

Altogether these data indicate that the structure of the RVF virus NS₅ gene is conserved and that clone 13 is a very distinct isolate. The existence of this virus raises the question of whether the NS₅ protein is essential. In an attempt to investigate selective regimes on the NS₅ gene, ds and dv were calculated. The ds:dv ratio was found to be greater than one (8.36; \( P < 0.001 \)) indicating that during evolution more synonymous than non-synonymous changes accumulated. This suggests the existence of constraints against amino acid changes in the protein.

Finally, the oligonucleotides used for RT–PCR in this analysis appear to be good primers for amplifying the NS₅ specific region of the 18 isolates presented herein as well as other isolates tested in our laboratory (data not shown). Therefore, this work defines the basis for a rapid diagnostic test for RVF virus based on RT–PCR.

The authors thank P. Vialat for excellent technical expertise. We are grateful to Dr H. Bourhy for providing computer facilities and critically reading the manuscript. We thank Dr C. Prehaud, A. Billecocq and B. Le Guenno for fruitful discussions.

References


Felsenstein, J. (1993). PHYLIP (phylogeny inference package) version 3.5c. Distributed by the Author, Department of Genetics, University of Washington, Seattle, USA.


Murphy, F. A., Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis,


Received 29 April 1997; Accepted 19 June 1997