Nucleotide sequence of the NS5 gene of Banzi virus: comparison with other flaviviruses

L. D. Fulop,¹* A. D. T. Barrett² and R. W. Titball¹

¹Chemical and Biological Defence Establishment, Porton Down, Salisbury, SP4 0JQ, UK and ²Department of Pathology and Center for Tropical Diseases, University of Texas Medical Branch, Galveston, Texas, 77555-0605, USA

Banzi is a mosquito borne flavivirus which belongs to the Uganda S serocomplex. No nucleotide sequence data have previously been reported from any virus of this serocomplex. We have determined the nucleotide sequence of the NS5 gene from Banzi virus and the predicted amino acid sequence was elucidated. Previously identified conserved RNA polymerase, methyltransferase and flavivirus NS5 amino acid motifs were present in the Banzi virus NS5 protein. These data add to the evidence for the functional importance of the regions. The encoded amino acid sequence was compared with the predicted amino acid sequence of other flavivirus NS5 proteins. Analysis of these sequences suggested that Banzi virus is most closely related to the mosquito-borne flaviviruses and, in particular, yellow fever virus. This pattern of similarity is in accordance with the previously suggested serological classification of flaviviruses.

The flavivirus Banzi was originally isolated from a 9-year-old boy with a febrile illness in Tongaland (Smithburn et al., 1959). The virus was shown to be serologically related to Uganda S and yellow fever viruses and was classified as belonging to the Uganda S serocomplex (Calisher et al., 1989). The virus has also been associated with a case of febrile illness in Tanzania (Williams & Woodall, 1964). Neutralizing antibodies to Banzi virus have been found in human sera from South Africa, Mozambique, Angola, Namibia and Botswana (Smithburn et al., 1959; Kokernot et al., 1965). The virus has also been isolated from mosquitoes, cattle, rodents and hamsters in South Africa, Mozambique, Zimbabwe and Kenya (McIntosh et al., 1976a, b; Metslaar et al., 1974). The natural host of Banzi virus is probably rodents (McIntosh, 1961).

Little is known about the molecular biology of Banzi or related viruses and no nucleotide sequence data have been reported for genes from any virus belonging to the Uganda S serocomplex.

The flavivirus NS5 gene is thought to encode a protein with RNA-dependent RNA polymerase and methyltransferase activity (Bartholomeusz & Wright, 1993; Edward & Takegami, 1993; Koonin, 1993). Antibodies against the NS5 protein inhibit RNA synthesis in vitro (Edward & Takegami, 1993). The NS5 protein is therefore thought to be important in virus replication. It may form a complex with the NS3 protein, the latter possibly acting as a helicase. The nucleotide sequence of the NS5 gene has been determined for a number of flaviviruses. Previously we developed a reverse transcription/PCR which amplified a 1 kb fragment from the centre of the NS5 gene of a wide range of flaviviruses (Fulop et al., 1993). The purpose of this study was to determine the nucleotide sequence of the entire NS5 gene of Banzi virus. This has permitted comparison with other flavivirus NS5 gene sequences, in particular other mosquito-borne flaviviruses.

Banzi virus strain SAH 336 was propagated in Vero cells. Viral RNA was extracted, reverse transcribed and amplified as previously described (Fulop et al., 1993). Initially, a fragment of approximately 1 kb from the centre of the NS5 gene was generated as described by Fulop et al. (1993). The PCR fragment was purified and cloned into pUC18 or pT7-Blue. The nucleotide sequence of this fragment was determined. Nucleotide sequencing was performed by cycle sequencing in a Catalyst Molecular Biology LabStation followed by electrophoresis in a 373A DNA Sequencing System (Applied Biosystems). Duplicate clones of each plasmid were sequenced and analysis was carried out using the Lasergene software which contains a modified version of the CLUSTAL program and PAUP software.

*Author for correspondence. Fax +44 1980 613284. e-mail 100432.3200 @ compuserve.com

The sequence data reported in this article have been deposited in the EMBL Data Library under the accession number L40951.
Fig. 1. The nucleotide and predicted amino acid sequence of Banzi virus. The methyltransferase and GDD motifs are marked. Amino acids marked 1 to 10 are conserved either in tick-borne or mosquito-borne flaviviruses.
The flanking 5' and 3' regions of the gene were then amplified. The 5' fragment of the gene was amplified using a Banzi virus-specific antisense primer (5' TTCTGGTATGGATGATCAG 3') designed from the nucleotide sequence of the first fragment and a consensus sense primer (5' TGAAGAATCCAGTGTTGGA(T/C)GG 3') designed from published NS4B gene nucleotide sequences of mosquito-borne flaviviruses. The 3' fragment of the gene was amplified using a consensus antisense primer designed from a conserved region in the 3' non-coding region of mosquito-borne flaviviruses (5' GGGTCCTCTCAACCTCTAGTCC 3') and a Banzi virus-specific primer (5' CTGACCCCAGATTCTGGAGGCC- TAGTT 3') designed from the nucleotide sequence of the first fragment. These fragments were amplified, cloned and sequenced as described above.

The nucleotide sequence and deduced amino acid sequence of the NS5 gene are shown in Fig. 1. The start of the NS5 gene was identified by examining the deduced amino acid sequence and identification of a putative dibasic cleavage site conserved amongst all flaviviruses (Lin et al., 1993). Cleavage at this site would form the N terminus of the NS5 protein. In the case of Banzi virus the two basic amino acids were both arginine (sequence not shown). A TAA stop codon in the gene was identified 2716 nucleotides downstream of the NS5 gene start codon. The gene is 2715 nucleotides in length and the predicted protein contains 906 amino acids, with a predicted molecular mass of 103 kDa.

The Banzi virus NS5 predicted amino acid sequence contains most of the previously identified conserved motifs of the NS5 gene, including the GDD motif (position 666-668) found in all RNA-dependent RNA polymerase proteins (Kamer & Argos, 1984; Ishihama & Barbier, 1994) (Fig. 1). The GDD motif is preceded by a serine residue and followed by a cysteine residue, both these residues are strictly conserved in all of the predicted amino acid sequences of flavivirus NS5 proteins reported to date. Other motifs identified by Poch et al. (1989) and Koonin (1991, 1993) are also conserved. However, the putative methyltransferase motif 1 shows a functionally non-conservative substitution (Fig. 1), the arginine having been replaced with tryptophan (position 84). This substitution is found in other putative methyltransferases in yeast and bacteria but has not previously been reported in flavivirus NS5 proteins (Koonin, 1993).

Flaviviruses have been classified on the basis of their antigenic relationships as determined by cross-neutralization with polyclonal antisera (Calisher et al., 1989). Comparison of the nucleotide sequence and predicted amino acid sequence of the E gene of flaviviruses results in a similar pattern of classification (Heinz et al., 1990; Venugopal et al., 1994; Marin et al., 1994). Pierre et al. (1994) compared the nucleotide sequence of an RT/PCR...
product spanning 300 bases at the 3' end of the NS5 gene and a portion of the 3' non-coding region of selected mosquito-borne flaviviruses. They also were able to show a relationship between mosquito-borne flaviviruses similar to that obtained by serology.

We have compared the predicted amino acid sequences of all the NS5 proteins of flaviviruses reported to date. This analysis shows the percentage identity at the amino acid level between NS5 proteins within serocomplexes is 72.3% to 80.4% for the dengue serocomplex, for the Japanese encephalitis serocomplex, 81.0% to 94.6% and for the tick-borne encephalitis serocomplex, 82.9% to 86.9% (Fig. 2). The predicted amino acid sequence of Banzi NS5 protein was compared to reported NS5 sequences of other flaviviruses (Fig. 2). The deduced amino acid sequence of the Banzi virus NS5 protein shows greatest similarity with yellow fever virus NS5 protein (66.3% identity). However identity with the other flaviviruses is only slightly lower (55.6% to 59.7%).

A dendrogram of the phylogenetic relationship of all flavivirus NS5 genes reported was prepared (Fig. 3) in which the length of the horizontal lines corresponds to the relatedness of the viruses. We conclude that serological analysis and comparison of NS5 gene nucleotide sequence show a similar pattern of relationships between flaviviruses. Mosquito-borne and tick-borne viruses are separated into two groups. Within the mosquito-borne virus group all four dengue viruses are grouped together and Japanese encephalitis, Kunjin and West Nile viruses form another distinct group. The tick-borne viruses Langat, western tick-borne encephalitis (Neudoerfl), far eastern tick-borne encephalitis (Sofjin), and Powassan form a third separate group. Yellow fever virus and Banzi virus are members of two separate groups.

The current classification on the basis of serology leaves 17 viruses unassigned to a serocomplex. The classification of viruses based on nucleotide sequence may yield more precise data on the relationships between flaviviruses. Marin et al. (1994) showed that phylogenetic analysis of E and NS5 gene sequences resulted in a similar logical classification of flaviviruses. However, classification on the basis of E gene nucleotide sequence may result in similar problems to that based on serology.

Banzi virus is transmitted by mosquitoes (McIntosh et al., 1976b) and this is reflected in the predicted amino acid sequence of the Banzi NS5 protein. A comparison of flavivirus NS5 amino acid sequences revealed 10 residues at which an amino acid is strictly conserved in all the mosquito-borne viruses analysed to date, but a different amino acid is conserved in tick-borne viruses (Fig. 1). These changes are functionally non-conservative substitutions. Banzi virus shows identity with nine mosquito-borne conserved amino acids and only one tick-borne conserved amino acid. Also, in tick-borne viruses a conserved phenylalanine residue is found in amino acid position 183, this residue is not present in Banzi virus nor any mosquito-borne flavivirus analysed to date. These amino acids that are differently conserved in the NS5 protein of mosquito-borne and tick-borne flaviviruses could represent functional differences in virus replication in the arthropod vector.

We have reported here the first nucleotide sequence from a virus of the Uganda S serocomplex. Banzi virus is the most divergent mosquito-borne flavivirus for which nucleotide sequence data have been reported to date. The results presented here corroborate previous work on regions of the NS5 protein which are predicted to have functional importance. Our analysis of the NS5 gene and comparison with other flavivirus NS5 genes indicates that a phylogenetic comparison of flaviviruses based on NS5 gene sequence accurately reflects the previously suggested serological relationships. The Banzi NS5 amino acid sequence displays features characteristic of mosquito-borne flaviviruses. These findings, along with our previously reported NS5 pan-flavivirus RT/
PCR system (Fulop et al., 1993), provide a powerful tool for the identification of flaviviruses and analysis of uncharacterized newly emergent flaviviruses.

The authors wish to thank Dr D. Leslie (CBDE) for his useful advice and discussion during the preparation of the manuscript.

References


(Received 13 March 1995; Accepted 18 May 1995)