Nucleotide sequence and coding strategy of the Uukuniemi virus L RNA segment

Richard M. Elliott, Ewan Dunn, Jan F. Simons and Ralf F. Pettersson

Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, U.K. and Ludwig Institute for Cancer Research, Stockholm Branch, Box 60202, S-10401 Stockholm, Sweden

The complete nucleotide sequence of the L RNA segment of Uukuniemi virus has been determined from cloned cDNA. The L RNA is 6423 nucleotides in length, and is of negative polarity. The viral-complementary RNA contains a single large open reading frame of 2104 codons which corresponds to the L protein (Mr 241039). Comparison with the L protein sequences of other members of the Bunyaviridae showed homology with the Rift Valley fever phlebovirus L protein (38% amino acid identity), but no detectable similarity with bunyavirus, hantavirus or tospovirus L proteins. These data lend further support for the recent reclassification of uukuviruses and phleboviruses into the same genus, Phlebovirus, in the family Bunyaviridae. The L RNA sequence completes the determination of the Uukuniemi virus genome: since the M RNA segment is 3229 and the S RNA segment 1720 nucleotides, the whole genome comprises 11372 nucleotides.

Introduction

The Bunyaviridae family contains more than 300 viruses which share the common properties of a tripartite ssRNA genome and intracellular site of virus maturation in the Golgi complex; most of the viruses are transmitted by arthropods (reviewed by Elliott, 1990). However, in view of the large number of individuals in the family it is perhaps not surprising that considerable diversity exists in terms of genome structures and expression strategies, which have been reflected in the fluidity in the subclassification into genera. The present classification (Calisher, 1991) subdivides the Bunyaviridae into five genera: Bunyavirus, Hantavirus, Nairovirus, Phlebovirus and Tospovirus. Bunyaviruses and hantaviruses have a negative-sense coding strategy for all three genome segments (L, M and S), whereas the S RNA segments of phleboviruses and tospoviruses have an ambisense strategy (Bishop, 1986; de Haan et al., 1990; Elliott, 1990).

Uukuniemi (UUK) virus is the prototype of the uukuvirus group, formerly the Uukuvirus genus, and was isolated from an Ixodes ricinus tick in Finland (Oker-Blom et al., 1964). Biochemical and molecular biological studies revealed similar coding strategies and common 5' and 3' terminal sequences of the genomes of UUK virus and phleboviruses, such as Punta Toro (PT) virus and Rift Valley fever (RVF) virus; in addition, significant identity was detected between the amino acid sequences of the nucleocapsid proteins and to a lesser extent between the glycoproteins (Rönnholm & Pettersson, 1987; Simons et al., 1990). Thus, the uukuviruses were recently subsumed into the Phlebovirus genus (Calisher, 1991). However, biologically, the two groups are distinct: phleboviruses are generally transmitted by phlebotomine flies whereas uukuviruses are transmitted by ticks. Several phleboviruses are important human pathogens, but none of the uukuviruses have been associated with disease in man although antibodies to UUK virus have been demonstrated in humans (Beaty & Calisher, 1991).

We have previously determined the nucleotide sequence of the M (Rönnholm & Pettersson, 1987) and S (Simons et al., 1990) RNA segments of UUK virus. The M segment, which encodes the p110 precursor for the membrane glycoproteins G1 and G2, is 3229 nucleotides (resequencing of the 5' end of the virion RNA has revealed that it is two residues shorter than the previously reported sequence; R. Rönnholm, personal communication). The S RNA, which is transcribed into two subsegmental mRNA species encoding the nucleocapsid (N) and a non-structural (NSs) protein in an ambisense fashion, is 1720 nucleotides in length (Simons et al., 1990). By exclusion, and in analogy to other bunyaviruses, it is clear that the L RNA segment must encode the RNA-dependent RNA polymerase (L).
In this paper, we present the nucleotide sequence of the UUK virus L RNA segment. This completes the sequence determination of the genome. Comparison of the UUK virus L RNA sequence and the amino acid sequence of the predicted L protein with those of the recently published RVF phlebovirus L segment (Muller et al., 1991) reveals substantial identity, supporting the classification of phleboviruses and uukuviruses into the same genus.

**Methods**

**Preparation of viral RNA**. The prototype strain, S23, of UUK virus was grown in BHK21 clone 13 cells as described previously (Pettersson & Kääriäinen, 1973). The L RNA segment was isolated from purified virus as detailed before (Pettersson, 1987). Briefly, viral nucleocapsids were isolated from Triton X-100-solubilized virus by banding in a CsCl gradient and then disrupted by treatment with 1% SDS. The released RNA segments were separated on a 15 to 30% sucrose gradient, the three RNA peaks were pooled, and then concentrated by ethanol precipitation. The L RNA from four preparations was pooled; about 25 µg was obtained from a total of about 12 mg of purified virus.

**cDNA synthesis, cloning and sequencing**. Two cDNA syntheses were performed essentially following the protocol of Gubler & Hoffman (1983). In the first, reverse transcription of about 5 µg of purified L RNA was primed by 5 µg of random hexamers and 2 µg of a synthetic oligonucleotide complementary to the 3'-terminal 11 bases of the UUK virus M RNA segment (Rönholm & Pettersson, 1987). Double-stranded cDNA was digested with EcoRI and ligated to EcoRI digested M13mpl8 DNA. Nucleotide sequences were determined by the dideoxynucleotide chain termination method (Sanger et al., 1977; Rönnholm & Pettersson, 1987).

**Preparation of viral RNA.** The prototype strain, S23, of UUK virus was grown in BHK21 clone 13 cells as described previously (Pettersson & Kääriäinen, 1973). The L RNA segment was isolated from purified virus as detailed before (Pettersson et al., 1977; Rönnholm & Pettersson, 1987). Briefly, viral nucleocapsids were isolated from Triton X-100-solubilized virus by banding in a CsCl gradient and then disrupted by treatment with 1% SDS. The released RNA segments were separated on a 15 to 30% sucrose gradient, the three RNA peaks were collected and pooled individually, and then concentrated by ethanol precipitation. The L RNA from four preparations was pooled; about 25 µg was obtained from a total of about 12 mg of purified virus.

**Methods**

**Preparation of viral RNA.** The prototype strain, S23, of UUK virus was grown in BHK21 clone 13 cells as described previously (Pettersson & Kääriäinen, 1973). The L RNA segment was isolated from purified virus as detailed before (Pettersson et al., 1977; Rönnholm & Pettersson, 1987). Briefly, viral nucleocapsids were isolated from Triton X-100-solubilized virus by banding in a CsCl gradient and then disrupted by treatment with 1% SDS. The released RNA segments were separated on a 15 to 30% sucrose gradient, the three RNA peaks were collected and pooled individually, and then concentrated by ethanol precipitation. The L RNA from four preparations was pooled; about 25 µg was obtained from a total of about 12 mg of purified virus.

**Methods**

**Preparation of viral RNA.** The prototype strain, S23, of UUK virus was grown in BHK21 clone 13 cells as described previously (Pettersson & Kääriäinen, 1973). The L RNA segment was isolated from purified virus as detailed before (Pettersson et al., 1977; Rönnholm & Pettersson, 1987). Briefly, viral nucleocapsids were isolated from Triton X-100-solubilized virus by banding in a CsCl gradient and then disrupted by treatment with 1% SDS. The released RNA segments were separated on a 15 to 30% sucrose gradient, the three RNA peaks were collected and pooled individually, and then concentrated by ethanol precipitation. The L RNA from four preparations was pooled; about 25 µg was obtained from a total of about 12 mg of purified virus.
The nucleotide sequence data were aligned, overlapped and assembled using the database programs of Staden (1982, 1984). The 5'- (sense) terminal 17 nucleotides were obtained by primer extension sequencing using a synthetic oligonucleotide (5'GGTTGGATGCTGGCATT3'), corresponding to bases 6342 to 6359, and purified virion L RNA as template, following the method of Geliebter et al. (1986).

Analyses of the completed nucleotide sequence used programs in the University of Wisconsin Genetics Computer Group (UWGCG) package (Devereux et al., 1984).

Results

Nucleotide sequence of the UUK virus L RNA segment

Digestion with EcoRI of randomly primed cDNA to purified UUK virus L RNA yielded a number of pBR322-based recombinants with inserts in the range of 400 to 1300 bases. Six recombinants were selected for further analyses, and the L segment specificity of their inserts was confirmed by Northern blot hybridization of total infected cell RNA (data not shown). The inserts from these plasmids were partially sequenced, and also used as hybridization probes to detect other L segment-specific clones in the cDNA library prepared using the EcoRI/NotI adaptor strategy. The clones with the longest cDNA inserts were selected for sequencing. The majority of the sequence (nucleotides 5 to 6407) was obtained from at least two independent plasmid clones, except for the region between nucleotides 1900 and 3300, which was obtained from a single plasmid.

The 5'-terminal 17 nucleotides of the virion RNA (3' end of the viral-complementary strand) were determined by primer extension sequencing of a synthetic oligonucleotide, corresponding to bases 6342 to 6359 (Fig. 1), using purified genomic L RNA segment as the template (see Methods). The 3'-terminal four nucleotides (3' UUGUG . . . ) of the virion RNA were not actually determined, but were assumed considering the complementarity of the downstream nucleotides with the 5' end of the RNA, and the fact that these nucleotides are conserved in the UUK virus M and S RNA segments (Rönnholm & Pettersson, 1987; Simons et al., 1990; Simons & Pettersson, 1991). The complete sequence of the UUK virus L RNA segment expressed as mRNA-sense is shown in Fig. 1. The L RNA is 6423 nucleotides long and has a base composition of 27.8% A, 22.4% C, 24.6% G and 25.2% U. The terminal eight nucleotides of the 3' and 5' ends of the RNA are exactly complementary, and a base-paired stem of some 20 nucleotides can be predicted (Fig. 2). This is similar to the stems postulated for the M and S segments (Fig. 2; Rönnholm & Pettersson, 1987; Simons et al., 1990; Simons & Pettersson, 1991) though position 9 at the 5' end of the L RNA is a mismatched U residue, compared to a base-paired A residue in the M and S segments. Note that in the revised S RNA sequence (Simons & Pettersson, 1991), there is a G residue at position 11 from the 3' end of virion RNA (instead of a C residue as originally reported), which is base-paired with a C at the 5' end.

Examination of the L RNA sequence revealed a single long open reading frame (ORF) in the viral-complementary RNA, extending from the 5'-proximal AUG codon at positions 17 to 19 to the UGA stop codon at positions 6326 to 6328 (Fig. 1). The non-coding sequences are therefore 16 nucleotides at the 5' end and 95 nucleotides at the 3' end, and the ORF, which represents the L protein, comprises 2104 codons. Thus the UUK virus L protein has a predicted Mr of 241039 and an overall charge of +34 at neutral pH. The next largest ORF in either the genomic or viral-complementary RNA is only 121 codons.

Homologies of the UUK virus L RNA and L protein

Complete sequences of L RNA segments of representatives of four Bunyaviridae genera are now available (Elliott 1989b; de Haan et al., 1991; Schmaljohn, 1990; Muller et al., 1991). The nucleotide and predicted L protein sequences of these were compared to the UUK virus sequences using the COMPARE program in the UWGCG package to generate dot matrices (Fig. 3). No homology was detected at the nucleotide level with the bunyavirus, hantavirus or tospovirus sequences (not shown), but an interrupted diagonal was observed in the plot with RVF phlebovirus (Fig. 3a) which indicates a degree of genetic relatedness. At the amino acid level the similarity between the UUK and RVF virus L proteins is striking (Fig. 3b). Little homology was detected with the
Uukuniemi virus L RNA sequence

Fig. 3. Dot matrix comparisons of the UUK virus L RNA (a) and L protein (b to f) with other Bunyaviridae L sequences using the COMPARE program (Devereux et al., 1984). The parameters used were (a) window 21, stringency 14, and (b to f) window 30, stringency 16. The comparisons show (sequence on x-axis first) (a) UUK and RVF, (b) UUK and RVF, (c) UUK and TSW, (d) BUN and HTN, (e) BUN and TSW, (f) HTN and TSW viruses. The patterns generated using RVF virus L protein were similar to those generated using the UUK virus L sequence; the pattern shown in (e) is typical of comparisons showing 'little homology' (see text). The sequences were reported as follows: RVF virus, Muller et al. (1991); BUN virus, Elliott (1989b); TSW virus, de Haan et al. (1991); HTN virus, Schmaljohn (1990).

tospovirus (tomato spotted wilt, TSW) L protein (Fig. 3c) nor with the bunyavirus (Bunyamwera, BUN) or hantavirus (Hantaan, HTN) L proteins (data not shown); similarly the RVF virus L protein showed little homology with the L proteins of the other genera (data not shown). BUN, HTN and TSW virus L proteins showed a region of similarity in the central part of the molecule (Fig. 3d, e, f) which corresponds to the putative polymerase domain previously predicted in the BUN virus L protein (Elliott, 1989b).

UUK, RVF and TSW viruses share the property of having an ambisense coding strategy for the S RNA segment, in contrast to BUN or HTN viruses where the S segment shows a conventional negative-sense strategy. Hence it was rather surprising that the TSW virus L protein showed more similarity with the BUN virus L protein (Fig. 3e) than with the L proteins of UUK (Fig. 3c) and RVF viruses.

An optimal alignment of the UUK virus L protein with the RVF virus L protein is given in Fig. 4. The two sequences can be readily aligned with a few minor gaps, and only one major gap, towards the carboxyl terminus, inserted to optimize the alignment. The two L proteins show an overall 58% similarity (including conserved amino acid changes), with 38% of the residues being identical. The similarity is greatest in the central third of the molecule (between residues 800 and 1600 approximately) and blocks of up to 17 identical residues are found. There is less identity towards the amino and carboxy termini; this is also evident from the dot matrix plot (Fig. 3b). The middle third of the L protein contains the polymerase motifs identified by Poch et al. (1989) in
Fig. 4. Alignment of the L protein sequences of UUK virus (upper line) and RVF virus (lower line) using the BESTFIT program (Devereux et al., 1984). Identical residues are indicated by [ ], and conserved amino acid substitutions (similarities) by :. Gaps inserted in the sequences to maximize homology are indicated by dots. The four polymerase motifs identified by Poch et al. (1989) in RNA-dependent RNA polymerases are underlined.

Comparison of the other analogous proteins of UUK virus and RVF virus reveals 40% identity between the N proteins, 16% between the NSs proteins, 18% between UUK virus G1 and RVF virus proteins, and 40% between the polymerase proteins of UUK virus L and RVF virus P (unpublished results).
Table 1. Coding capacity of the UUK virus genome

<table>
<thead>
<tr>
<th>RNA segment</th>
<th>No. of nucleotides</th>
<th>Recognized gene products</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total 5'NC* 3'NC</td>
<td>Protein</td>
</tr>
<tr>
<td>L</td>
<td>6423 16 95</td>
<td>L 2103</td>
</tr>
<tr>
<td>M</td>
<td>3229 17 185</td>
<td>G1, G2 1008t</td>
</tr>
<tr>
<td>S</td>
<td>1720 V‡, 25 IR§, 74</td>
<td>NSs 273</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>N 254</td>
</tr>
<tr>
<td>Total</td>
<td>11372 92 354</td>
<td>3638 415500</td>
</tr>
</tbody>
</table>

* NC, Non-coding.
† Corresponds to the ORF of the p110 precursor.
‡ V, virion sense RNA.
§ IR, Intergenic region.
|| VC, Virus-complementary RNA.

Discussion

The nucleotide sequences of the M and S RNA segments of the UUK virus genome have been reported previously (Rönnholm & Pettersson, 1987; Simons et al., 1990; Giorgi et al., 1991). The determination of the sequence of the L RNA segment thus completes the genome sequence of the prototype uukuvirus. The coding capacity of the genome is summarized in Table 1. The complete genome comprises 11372 nucleotides of which 10926 (96%) code for the five recognized UUK virus proteins. In common with other Bunyaviridae genome segments, the 5' non-coding regions of the positive-sense RNA, at least for the L and M segments, are considerably shorter than the 3' non-coding regions. Also in common with other Bunyaviridae genome segments the terminal sequences at the 3' and 5' ends of the L RNA are complementary, permitting the RNA to fold into a panhandle. This is in accord with the electron microscopic observations of circular L nucleocapsids and RNA molecules reported previously (Pettersson & von Bonsdorff, 1975; Hewlett et al., 1977). The terminal eight nucleotides are identical in all three UUK virus RNA segments, and are also identical to the terminal eight nucleotides of the RVF virus genome segments (Muller et al., 1991). In addition, partial complementarity beyond this perfectly matched region continues for some 10 to 20 residues in all RNA segments.

The completion of the sequencing of the UUK virus genome and that of the RVF virus genome (Muller et al., 1991) provides strong evidence to support the classification of these viruses into the same genus, Phlebovirus, of the family Bunyaviridae (Calisher, 1991). The two internal viral proteins, L and N, show about 40% similarity between the two viruses; this is at the level seen, for example, when comparing the N proteins of viruses from different serogroups within the Bunyavirus genus (Elliott, 1989a). Essentially no similarity is detected when analogous proteins of viruses from different genera are compared (Elliott, 1990). The glycoproteins of UUK and RVF viruses show a lower similarity, but the positions of cysteine residues are highly conserved (Rönnholm & Pettersson, 1987). In addition, the RVF virus M segment encodes a non-structural protein for which UUK virus does not have an equivalent. The NSs proteins, the function of which is not known, are also dissimilar between UUK, PT and RVF viruses (Simons et al., 1990; Giorgi et al., 1991).

Hence there is evidence at the genome level that UUK and RVF viruses are related, but these viruses are clearly different with regard to their biology. It is tempting to speculate that these differences can be ascribed to the glycoproteins, and the non-structural proteins NSm and NSs.

R.M.E. is a Medical Research Council Senior Fellow and the work in Glasgow was supported by an MRC Project Grant.

References


(Received 14 February 1992; Accepted 20 March 1992)