Inkoo and Tahyna, the European California serogroup bunyaviruses: sequence and phylogeny of the S RNA segment

Olli Vapalahti, Alexander Plyusnin, Ying Cheng, Tytti Manni, Markus Brummer-Korvenkontio and Antti Vaheri

Haartman Institute, Department of Virology, POB 21, FIN-00014 University of Helsinki, Finland

Inkoo (INK) and Tahyna (TAH) viruses, European representatives of the California serogroup (CAL), genus *Bunyavirus*, family *Bunyaviridae*, are transmitted by mosquitoes and frequently infect man. The S segments of INK and TAH prototype strains were amplified, cloned and sequenced. INK S consists of 986 and TAH S of 977 nucleotides (nt) coding for a nucleocapsid protein of 235 amino acids (aa) and, in an overlapping reading frame, for a nonstructural protein of 92 or 97 aa, respectively. By S segment sequences and phylogenetic analysis INK was seen to be most closely related to Jamestown Canyon virus, isolated in the USA (92.4% nt and 96.6% aa identity), which is currently classified in a different subcomplex within the CAL viruses. TAH was genetically closest to Lumbo virus, isolated in Mozambique (89.0% nt and 94.1% aa identity). The data suggest that genetic variation within the CAL viruses is less related to geographical distance than to similarity in ecological cycles.

Members of the genus *Bunyavirus* in the family *Bunyaviridae*, are enveloped viruses with a three-segmented negative-strand RNA genome. The L segment codes for the RNA polymerase, the M segment for a polyprotein precursor of the two virion glycoproteins, G1 and G2, and a nonstructural protein NSm; the S segment codes for the two proteins, the nucleocapsid protein (N) and a nonstructural protein (NSs) with unknown function in overlapping reading frames (ORFs), which are translated from the same mRNA as a result of alternative initiation of translation (Elliott, 1990).

The genus *Bunyavirus* was divided originally according to antigenic relatedness into the Bunyamwera, California and Simbu serogroups. The first California serogroup (CAL) virus, California encephalitis (CE) virus, was isolated from mosquitoes in 1943 and CAL viruses were found to be widely distributed in North America. The most important human pathogens in North America are La Crosse (LAC) virus, which can cause encephalitis, especially in children, and Jamestown Canyon (JC) virus, affecting mainly young adults (for review, see Grimstad, 1988; Beaty & Calisher, 1991). JC is geographically most widely distributed in North America with relatively high human seroprevalence and reported cases of meningitis and encephalitis (Grimstad, 1988). The ecological cycles of CAL viruses vary; for JC deer especially are important vertebrate hosts, while for LAC, snowshoe hare (SSH) and for CE viruses small mammals are most important.

Inkoo virus (INK) was isolated in mouse brain (and later BHK-21 cells) from a pool of *Aedes communis* mosquitoes from Inkoo, Finland in 1964 (Brummer-Korvenkontio et al., 1973). The Finnish population has been shown to have a high seroprevalence against INK: in southern Finland about 20% and in Lapland 70–90% of people have neutralizing antibodies to INK (Brummer-Korvenkontio, 1973). INK has also been isolated in Sweden and former USSR (Francy et al., 1989; Butenko et al., 1991). The virus is transmitted by *Aedes communis* mosquitoes feeding on large mammals, such as cows, reindeer and moose, which also show high INK antibody prevalence (Brummer-Korvenkontio, 1973). Human infections occur during the summer, and are mainly mild or asymptomatic. However, cases of pneumonia, meningitis and encephalitis have been reported (Karabatsos, 1985). The true incidence of clinical disease caused by INK remains unknown.

Tahyna virus (TAH) was originally isolated in former Czechoslovakia in 1958 from a pool of *Aedes caspius* mosquitoes. High antibody prevalences are found in Central Europe, with 30% seroprevalence, e.g. in Moravia, Czech Republic (Swanepoel, 1994). The virus is transmitted variably in *Aedes vexans* and *Culiceta annulata* mosquitoes; small mammals such as hares, rabbits and hedgehogs serve as amplification hosts (Karabatsos, 1985; Peters & LeDuc, 1991). Human infection results in a febrile illness with respiratory and gastrointestinal symptoms, and occasionally also meningitis, but probably most infections are subclinical (Grimstad, 1988; Swanepoel, 1994). The virus has also been isolated from human...
Table 1. Comparison of INK and TAH S segment nucleotide sequences (entire coding region) and N and NSs amino acid sequences with those of other bunyaviruses

Values are percentage identity. CE* and JC*, prototype strains; CE/E, CE virus E6071 strain; JC/D, JC virus DAV28 strain; KEY, Keystone virus; BUN, Bunyamwera virus. Accession numbers of sequences are given in the legend to Fig. 2.

<table>
<thead>
<tr>
<th></th>
<th>CE*</th>
<th>CE/E</th>
<th>JC*</th>
<th>JC/D</th>
<th>JS</th>
<th>INK</th>
<th>MEL</th>
<th>KEY</th>
<th>TVT</th>
<th>LAC</th>
<th>SSH</th>
<th>LUM</th>
<th>TAH</th>
<th>BUN</th>
<th>AINO</th>
</tr>
</thead>
<tbody>
<tr>
<td>S (nt identity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INK</td>
<td>83.8</td>
<td>79.3</td>
<td>79.9</td>
<td>79.0</td>
<td>80.0</td>
<td>79.7</td>
<td>81.0</td>
<td>79.6</td>
<td>72.4</td>
<td>83.7</td>
<td>83.1</td>
<td>89.0</td>
<td>100</td>
<td>55.0</td>
<td>56.3</td>
</tr>
<tr>
<td>TAH</td>
<td>79.3</td>
<td>78.9</td>
<td>92.3</td>
<td>92.4</td>
<td>92.4</td>
<td>100</td>
<td>83.0</td>
<td>82.1</td>
<td>73.0</td>
<td>80.0</td>
<td>77.6</td>
<td>78.9</td>
<td>79.7</td>
<td>63.7</td>
<td>56.1</td>
</tr>
<tr>
<td>N (aa identity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INK</td>
<td>84.3</td>
<td>84.3</td>
<td>96.0</td>
<td>96.0</td>
<td>97.0</td>
<td>100</td>
<td>86.9</td>
<td>85.9</td>
<td>75.0</td>
<td>82.2</td>
<td>80.9</td>
<td>84.3</td>
<td>82.2</td>
<td>43.2</td>
<td>42.4</td>
</tr>
<tr>
<td>TAH</td>
<td>69.0</td>
<td>70.0</td>
<td>71.0</td>
<td>71.0</td>
<td>71.0</td>
<td>72.0</td>
<td>73.0</td>
<td>76.0</td>
<td>59.0</td>
<td>73.0</td>
<td>73.0</td>
<td>88.0</td>
<td>100</td>
<td>32.7</td>
<td>42.4</td>
</tr>
<tr>
<td>NSs (aa identity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INK</td>
<td>68.0</td>
<td>67.0</td>
<td>90.5</td>
<td>89.5</td>
<td>90.5</td>
<td>72.0</td>
<td>83.0</td>
<td>82.1</td>
<td>65.0</td>
<td>71.6</td>
<td>69.5</td>
<td>73.0</td>
<td>72.0</td>
<td>31.7</td>
<td>31.6</td>
</tr>
</tbody>
</table>

The S RNA segments of INK and TAH prototype strains were amplified by RT–PCR, cloned and sequenced. The INK prototype strain KN 3641 was isolated in our laboratory in 1964 and had been passaged five times in mouse brain and once in Vero E6 cells before use in this study. The TAH prototype strain 92 was obtained by our laboratory in 1962 from J. S. Porterfield, London, UK, and since has been passaged five times in mouse brain and twice in BHK-C-13 cells. Total RNA was extracted from INK and TAH infected Vero E6 cells as described (Chomczynski & Sacchi, 1987). Both cDNA synthesis and PCR were primed according to Dunn et al. (1994) using a degenerate genus-specific primer, AGTAGTGT-A/GCTCCAC, complementary to the 3' end of both genomic or antigenomic S RNA. The amplified fragment was cloned into pGEM-T (Pharmacia) and DNA isolated from two independent colonies was sequenced.

The INK and TAH S segment N ORF nucleotide sequences were compared with those of available bunyaviral S segment sequences. Programs in the PHYLIP package (Felsenstein, 1993) were used to make 1000 bootstrap replicates of the sequence data (SEQBOOT); distance matrices were calculated using Kimura's 2-parameter model (DNADIST) with a transition to transversion ratio of 2, and analyzed using the Fitch–Margoliash tree fitting algorithm (FITCH) with global arrangements option set. The occurrence ratios of particular branchings were calculated from these trees with the program CONSENSE.

The S segment sequences of INK and TAH prototype strains have been deposited in GenBank (accession numbers Z68496 and Z68497). The INK S segment consists of 986 nt coding for an N protein of 235 aa (26.5 kDa) and an NSs protein of 92 aa (10.5 kDa). The N protein of INK showed 92% nt and 97% aa identity to the most closely related JC isolates, including Jerry Slough (JS) virus (Table 1), and showed somewhat more similarity with Keystone (KEY) and MEL viruses (82–83% nt identity) than with other CAL viruses (73–80% nt identity).

The TAH S segment has 977 nt and codes for an N protein of 235 aa (26.9 kDa) and an NSs protein of 97 aa (10.5 kDa). The TAH S segment has 89% nt identity with the LUM S segment, and the TAH N protein has 94% aa identity with the LUM N protein (Table 1). The TAH S segment has 72–84% nt identity with the S segments of other CAL viruses. The NSs protein encoded by the second ORF is less conserved.

The first 100 nt of the 3' noncoding region (nt 784–884, in the messenger sense) differed considerably between INK and JC (70% identity) and between TAH and LUM; the TAH S segment also lacks the duplicated region present in the LUM S RNA (Dunn et al., 1994; Fig. 1b). The shorter 5' noncoding
Fig. 1. Alignment of the noncoding regions of INK with JC strains 61V2235* (JC*), DAV28 (JC/D) and Jerry Slough* (JS), and TAH with LUM. (*) Reference strain. (a) 5'-noncoding region (viral complementary sense), initiation codons underlined; (b) 3'-noncoding regions (stop codons are in bold; the region with sequence duplication in LUM is underlined). (.) Identity. (-) Gap. Primer sequences are shown in lower-case letters.
region is somewhat less variable but shows also specific features for each virus (Fig. 1a).

Phylogenetic analysis (Fig. 2) of the bunyaviral S segment coding sequences showed that INK and JC are closely related and share a common ancestor; likewise, TAH and LUM are closely related to each other. KEY and MEL seemed also to have a common ancestor with the INK–JC clade; similarly, the TAH–LUM and LAC–SSH clades and CE lineage shared a common root. TVT was clearly separate from all other CAL viruses. In some cases, a reliably high probability of branching order was not obtained (i.e. 63% for the branching of CE lineage and 52% for a common root for KEY and MEL viruses). The viruses belonging to different serogroups of bunyaviruses were clearly separated with long branch lengths. A phylogenetic tree based on N protein amino acid sequences was essentially similar (data not shown).

Two representatives of the CAL complex, INK and TAH, have been reported to circulate in Europe. The S segment sequences of prototype strains of both of these viruses were seen to be closely related to other previously published CAL virus sequences. INK and JC were both isolated in the early 1960s from mosquitoes from Southern Finland and Colorado, USA, respectively (Karabatsos, 1985). Despite of the vast geographical distance between these two locations, the viruses differed by only 8% at the nucleotide and 3–3.4% at the amino acid level. According to the S segment sequence, JC had a greater identity with INK than with any CAL virus isolated in the USA (excluding variety Jerry Slough). Also, the ecological
cycles of these two viruses share common features. For both viruses, sentinel animals serve as important amplification hosts, especially deer for JC and reindeer and moose for INK. Both viruses are transmitted by Aedes communis mosquitoes (Brummer-Korvenkontio, 1973; Grimstad, 1988), which, however, is by no means unique to these two viruses. On the other hand, the 3'-noncoding regions varied considerably between the two viruses suggesting that the divergence into these two lineages is not recent and that stringent functional constraints exist in the N protein with few possibilities of altering even the primary structure of the protein.

In the earlier literature, INK and JC are referred to as separate viruses in the CAL group, belonging to different antigenic complexes: INK is classified into the CE complex and JC within the MEL complex. Indeed, reactivities of these two viruses differ against MAbs specific for the G1 protein of LAC and TAH (Kingsford, 1991). As the glycoproteins are encoded by the M segment, which was not studied here, sequencing of the M segments of these and related viruses, as well as a thorough antigenic cross-comparison, is needed to evaluate the true relationship between these two viruses: are the M segments merely more divergent than the S segments with enough changes in the encoded epitopes to explain the antigenic differences, or have the two segments evolved separately through reassortment? Reassortment, supported by S segment phylogeny by Dunn et al. (1994), has been suggested to account for the evolution of Guaroa (GRO) virus in the Bunyamwera group. Recombination events have been documented to have occurred in nature for other members of the genus Bunyavirus, namely viruses in group C, Gamboavirus and Patois groups (summarized by Peters & LeDuc, 1991).

LUM virus, on the other hand, was reported earlier to be antigenically very similar to TAH, and thus LUM has been considered taxonomically to be a variety of TAH (Grimstad, 1988). However the 3'-noncoding region is considerably different, suggesting that the LUM lineage evolved later and that stringent functional constraints exist in the LUM virus or differ considerably in their glycoprotein sequence and, e.g., in cross-neutralization, remains to be answered.

Ms Leena Kostamovaara is acknowledged for excellent technical assistance. This work was supported by the Medical Research Council of the Academy of Finland, Finnish Culture Foundation, Finnish Medical Society and Sigrid Jusélius Foundation.

References


Received 1 February 1996; Accepted 26 April 1996