Sequences of wild Puumala virus genes show a correlation of genetic variation with geographic origin of the strains

Alexander Plyusnin,1* Olli Vapalahti,1 Kristian Ulfves,1 Heikki Lehväslaiho,1 Natalia Apekina,2 Irene Gavrilovskaya,2 Vladimir Blinov3 and Antti Vaheri1

1 Department of Virology, University of Helsinki, P.O. Box 21 (Haartmaninkatu 3), FIN-00014 Helsinki, Finland, 2 Research Institute of Poliomyelitis and Viral Encephalitis, Moscow and 3 Research Institute of Molecular Biology, Koltsovo, Novostibirsk Region, Russia

An experimental scheme was developed for direct sequence analysis of Puumala virus-containing specimens from wild rodents (Clethrionomys glareolus). Total RNA isolated from rodent lung tissues was reverse-transcribed in the presence of a universal 11 nucleotide primer complementary to all three viral RNA segments followed by amplification in a PCR with gene-specific primers. A full-length PCR product of approximately 1800 bp from the S segment encoding the viral nucleocapsid protein and a product of approximately 900 bp from the M segment (encoding the C-terminal two-thirds of the G2 protein and including the 3' non-coding region) of Puumala virus (from C. glareolus trapped in Udmurtia) were prepared and sequenced. No pronounced differences to Vero cell-grown viruses were seen. The Udmurtia/894Cg/91 strain was more closely related to the Bashkiria/CG18-20/84 strain than to the Finnish prototype strain of Puumala virus, Sotkamo/V-2969/81. Thus there is a correlation with the geographic origin of the three strains. The results indicate the occurrence of genetic drift and different selection pressures leading to (i) clustering of mutations, (ii) a lower frequency of nucleotide substitutions in the coding than in the 3' non-coding regions and (iii) a higher frequency of amino acid substitutions in G2 than in the N protein.

Puumala virus, the causative agent of the mild type of haemorrhagic fever with renal syndrome (HFRS) known as nephropathia epidemica, belongs to the Hantavirus genus in the Bunyaviridae family. The first aetiological agent known to cause HFRS, the Hantaan virus, was identified in 1976 (Lee et al., 1990) and Puumala virus was discovered in 1980 from tissue sections of bank voles (Brummer-Korvenkontio et al., 1980).

Hantaviruses are enveloped, spherical, negative-stranded RNA viruses. The three genome segments of approximately 1.7 (S), 3.6 (M) and 6.5 kb (L) encode the nucleocapsid protein (N), two glycoproteins G1 and G2 (M) and the polymerase protein respectively, (Elliott et al., 1991). At present, sequences of five hantavirus S segments, 10 M segments and three L segments have been published (for review, see Antic et al., 1992). The overall deduced amino acid similarity between the different viruses is about 70% for the L, 64% for the S and 53% for the M segment, G1 showing more variation than G2. For the Puumala virus, the sequences of the M and S segments (Vapalahti et al., 1992) of the prototype strain Sotkamo/V-2969/81 derived from a bank vole trapped in 1981 in Northern Finland (Brummer-Korvenkontio et al., 1982), and the L, M and S segments of the strain CG18-20 (Giebel et al., 1989; Stohwasser et al., 1992), derived from a bank vole trapped in 1984 in Bashkiria, European part of Russian Federation (Tkatchenko et al., 1984), have been published; initially the latter was erroneously thought to be the Swedish Hällnäs strain (see Giebel et al., 1989). In the present article we have used for Puumala virus strains and isolates the nomenclature adopted for influenza viruses and have accordingly designated this strain Bashkiria/CG18-20/84 or Bashkiria for short. The similarity between these two isolates, Sotkamo and Bashkiria, is 83% at the nucleotide level and 94% and 96% at the amino acid level for the M and S segments, respectively (Vapalahti et al., 1992).

All the sequences published so far have been derived from hantavirus strains passaged in Vero E6 cells. The adaptation process is quite tedious, and the frequency of successful isolations per patient or wild rodent tissue specimen is particularly low for Puumala virus. Thus, it seemed desirable to obtain sequences directly from wild rodents. Owing to the difficulties in conventional cloning of hantaviruses, PCR has been used in cloning and

The sequence data have been deposited in the EMBL Data Library with the accession numbers Z21497 and Z21509 (for S and M segments, respectively).
Fig. 1. General scheme of experiments. Pr, primer.

- **Rodent lung suspension**
  - Extraction of RNA
  - **Total RNA**
  - RT with Pr 1
  - PCR with Pr 1
  - PCR with Pr 2
  - PCR with Pr 3/Pr 4
  - Cloning in pCR™ II
  - **cDNA**
  - Sequencing

- **Full-length S segment**
  - Cloning in pCR™ II
  - **3′-terminal part of M segment**
  - Cloning in pCR™ II
  - Cloned product, nt 1-1827
  - Cloned product, nt 2775-3683

- **Pr 1**, 5′ TAGTAGTAGAC 3′
- **Pr 2**, 5′ TTCTGCAGTAOTAOTAGACTCCTTGAAAAG 3′
- **Pr 3**, 5′ TTGAATTCTATGTCAGTTTGATGGGAATAC 3′
- **Pr 4**, 5′ TTAAGCTTAGTAGTAOACTCCGCAAG 3′
- **Pr 5**, 5′ AGATTTCGCACAATAGTATGTGG 3′
- **Pr 6**, 5′ TTAAGCTTCTCCTGTGCTGGTGTOCC 3′
- **Pr 7**, 5′ OOTATTTOCCTOTOCACCAO 3′
- **Pr 8**, 5′ CTGCCATTAGTAATTCCCTO 3′
- **Pr 9**, 5′ CTOCGTCGAACTCAATCAATGGGAATTCA 3′
- **Pr 10**, 5′ GATGCTGCTATGTGTTATGG 3′
- **Pr 11**, 5′ CTGCCATTAGTAATTCCCTG 3′
- **Pr 12**, 5′ CTGCCATTAGTAATTCCCTG 3′
- **Pr 13**, 5′ CTGCCATTAGTAATTCCCTG 3′
- **Pr 14**, 5′ CTGCCATTAGTAATTCCCTG 3′
- **M13f/r**, either M13 (−40) forward or M13 reverse primer.

- M13f/r, either M13 (−40) forward or M13 reverse primer. The bars and arrows are not drawn to scale. The yield of a correctly sized PCR product for the S segment (approx. 1800 bp) was sufficiently high for wild Puumala virus as well as for laboratory strains. A PCR product corresponding to the full-length M segment was prepared from the laboratory-grown Sotkamo strain of Puumala virus but could not be amplified from the wild Puumala virus strain, which was not isolated, is designated Udmurtia/894Cg/91.

- The S gene segment of wild Puumala virus recovered from rodent tissue (strain designated Udmurtia/894Cg/91) consisted of 1827 nucleotides (nt) and contained a single open reading frame (ORF) coding for a 433 amino acid protein. The segment is three nt shorter than that of the Sotkamo strain of Puumala virus and 43 nt longer than that of the Bashkiria/CG18-20/84 strain. The deduced N proteins are of the same length. The ORF terminated with a TGA codon (as in the Bashkiria but not in the Sotkamo strain). The sequenced part of the Udmurtia strain M segment starts with nt 2785, reaches the end of the segment and encodes amino acid residues 913 to 1148 (the C-terminal two-thirds of G2) and contains a 3′-terminal non-coding region (3′ NCR) which is one base longer than that in Sotkamo and Bashkiria viruses.

- When the data from direct sequencing of the S segment (about 300 bases) were compared with those
from cloned PCR products no differences were seen. A single nucleotide substitution was found in the sequence of one recombinant plasmid (from three clones sequenced in that region) most probably due to the low fidelity of Taq polymerase.

The sequence of the Udmurtia strain S and M segments compared to those of the Sotkamo and Bashkiria strains suggest that there are no pronounced differences (e.g. large insertions or deletions of nucleotides in coding regions) within the sequenced areas between the wild strain and strains passaged in Vero cells in the laboratory. The 5' non-coding region (5' NCR) of the S segment appears to be almost invariant: only three nt out of 42 are different between Sotkamo and the Bashkiria or Udmurtia strains (the latter two contain an identical 5' NCR). Comparison of the mutation frequencies between the coding regions and 3' NCRs of S and M segments (Table 1) suggests that there is a higher variability in the non-coding than the coding region of both genes. The 3' NCRs also contain non-overlapping insertions/deletions, more in the S than in the M segment.

The distribution of nucleotide and amino acid replacements in the S segment and N protein of three Puumala virus sequences (Fig. 2, solid line) is non-random and very similar to that for all six hantavirus sequences (dotted line).

In the coding region there are some areas of higher variability as well as more conservative ones. The longest variable region is located between nt 650 and 900 and is colocalized with the area of highest variability in the N protein. Another variable area is located in the 3' NCR. The most conservative regions are located at the termini of the S segment. The overall distributions of replacements in the sequenced part of the M segment (data not shown) also indicate a higher variability in the 3' NCR than in the coding region. However, compared to the S segment, the distribution in the coding region is more even.

Most of the amino acid substitutions (14 of 22 for the G2 and 13 of 17 for the N protein sequence) are conservative ones. As well as the only potential Asn-linked glycosylation site (amino acids 948 to 950) in the sequenced part of G2, the cysteine residues are conserved in all three strains. Some substitutions appear to be clustered (amino acids 991 to 998 in G2 and 233 to 238 and 260 to 272 in N). Comparison of the frequencies of amino acid substitutions (Table 1) shows that there is a twofold lower variability in the N protein than in the G2 protein.

The sequence data indicate that the Udmurtia strain of Puumala virus is more closely related to the Bashkiria strain than to the Finnish prototype Sotkamo strain (Table 2). The degree of identity in all cases is slightly higher in S/N than in M/G2, i.e. the nucleoprotein gene is less variable.

In our experience, the scheme developed is highly efficient in the analysis of Puumala viruses from a variety of natural rodent specimens. Thus the PCR approach seems to be useful for molecular epidemiological studies of hantaviruses, as shown for other viruses (Becker & Darai, 1992).

The 5' NCR of the S segment containing the viral promoter was found to be quite conserved. The frequency of nucleotide substitutions is higher for the coding regions and particularly for the 3' NCR of both S and M segments (see Table 1). The total frequency of silent substitutions in the coding regions of the two RNA segments was almost equal. On the other hand, the frequency of amino acid substitutions was lower in the N protein than in the G2 protein reflecting different selective pressures.

The overall distribution of nucleotide and amino acid replacements in the three Puumala virus sequences is very similar to that in all six hantavirus sequences. This is in agreement with the previous observations (Antic et al., 1992) on preservation of the main features of hantaviral proteins within the genus and, moreover, provides the opportunity to map different domains more precisely.

Our results suggest that there is no extensive host-range selection with respect to the N protein or to the sequenced part of the G2 protein. Thus viruses propagated in Vero E6 cells in the laboratory are probably quite representative of the viruses circulating in rodents and may
Table 1. Distribution of nucleotide and amino acid substitutions in the S/N and M/G2 segments of Udmurtia, Sotkamo and Bashkiria strains of Puumala virus

<table>
<thead>
<tr>
<th></th>
<th>S segment</th>
<th>M segment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nt</td>
<td>aa</td>
</tr>
<tr>
<td>Coding region</td>
<td>1302</td>
<td>433</td>
</tr>
<tr>
<td>Substitutions</td>
<td>230</td>
<td>17</td>
</tr>
<tr>
<td>Frequency (%)</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>3' NCR</td>
<td>501</td>
<td>197</td>
</tr>
<tr>
<td>Substitutions</td>
<td>107</td>
<td>67</td>
</tr>
<tr>
<td>Frequency (%)</td>
<td>21</td>
<td>34</td>
</tr>
<tr>
<td>Non-overlapping insertions and deletions</td>
<td>15</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2. Similarity of the S and M segment nucleotide and deduced amino acid sequences of the Udmurtia strain to those of the prototype Sotkamo strain and the Bashkiria strain

<table>
<thead>
<tr>
<th>Puumala virus strain</th>
<th>Udmurtia (%)</th>
<th>Bashkiria (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sotkamo S nucleotides</td>
<td>86</td>
<td>87</td>
</tr>
<tr>
<td>N amino acids</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>M nucleotides</td>
<td>85</td>
<td>84</td>
</tr>
<tr>
<td>G2 amino acids</td>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>Bashkiria S nucleotides</td>
<td>95</td>
<td>-</td>
</tr>
<tr>
<td>N amino acids</td>
<td>98</td>
<td>-</td>
</tr>
<tr>
<td>M nucleotides</td>
<td>87</td>
<td>-</td>
</tr>
<tr>
<td>G2 amino acids</td>
<td>97</td>
<td>-</td>
</tr>
</tbody>
</table>

* The figures are based on total S segment/N protein and partial M segment/G2 protein sequences.

be suitable for diagnostic purposes and vaccine development. For further conclusions the sequences of other regions including that (those) responsible for the cell surface receptor binding site should be studied.

The Udmurtia strain of Puumala virus is more closely related to the Bashkiria strain than to the prototype Sotkamo strain. Thus there is a positive correlation between the sequences and the geographic distribution of the three strains. This conclusion is in agreement with our unpublished data of sequencing of an S gene fragment of Puumala virus strains circulating in some natural HFRS foci in Finland and in Udmurtia. The closer the geographical regions, the higher is the degree of similarity observed at the nucleotide level between the viral strains.

A phylogenetic tree constructed from S segment nucleotide sequences of hantaviruses (Fig. 3) shows that Puumala virus strains form a branch separated from the rest of the hantaviruses. A phylogenetic tree based on the sequenced part of the M segment has the same branching order and almost identical branch lengths (data not shown), suggesting a similar evolutionary history for both viral genes.

The nearest neighbour to Puumala virus in the phylogenetic trees is Prospect Hill virus. As pointed out earlier (Vapalahti et al., 1992) the similarity of Puumala viruses to Prospect Hill virus may be explained by the host ranges: these viruses are carried by species of the family Cricetidae (voles) whereas Hantaan and Seoul viruses are carried by species of Muridae (including mice and rats). In similar phylogenetic trees based on N protein sequences (Antic et al., 1992; data not shown) the distance of Prospect Hill virus from Puumala viruses is about half of that in the nucleotide-based trees. The nucleotide comparison reflects more faithfully the genotypic changes, whereas the changes in the amino acid sequence are restricted by functional constraints. According to the amino acid sequences, Puumala viruses are functionally closer to Prospect Hill virus than the nucleotide sequences suggest.

We thank Dr Markus Brummer-Korvenkontio, Dr Tapani Hovi, Dr Reijo Pyhälä and Dr Tuija Pöyry for valuable comments and Ms Raija Leveelahti and Leena Kostamovaara for expert technical assistance.

This work was supported by grants from the Medical Research Council of the Academy of Finland. The Sigrid Juselius Foundation, Helsinki and Finnish Foundation for Virus Research.

References


(Received 13 May 1993; Accepted 8 September 1993)