Molecular analysis of the genome of Chuzan virus, a member of the Palyam serogroup viruses, and its phylogenetic relationships to other orbiviruses

Makoto Yamakawa,1 Masanori Kubo2 and Susumu Furuuchi3

1Department of Exotic Diseases, National Institute of Animal Health, 6-20-1, Josuihoncho, Kodaira, Tokyo 187-0022, Japan
2National Institute of Animal Health, Kannondai, Tsukuba, Ibaraki 305-0856, Japan
3National Institute of Animal Health, Kaset Klang, Bangkhen, Bangkok 10900, Thailand

The nucleotide sequence of the entire genome of Chuzan virus, which belongs to the Palyam serogroup orbiviruses and causes congenital abnormalities of cattle, has been completed by analysis of the genes encoding minor core proteins (VP1, VP4 and VP6) and non-structural proteins (NS1, NS2 and NS3). The genome of Chuzan virus is 18,915 bp in length and the coding capacity of its open reading frames is 6071 aa. Comparative sequence analysis with other serogroups of the genus Orbivirus indicated that the outer capsid protein VP2, which is the neutralizing antigen, appears to be the most variable and the major core protein VP3 is the most conserved. Overall, the structural proteins, with the exception of VP2, are more conserved than the non-structural proteins among orbiviruses. Chuzan virus is phylogenetically most related to African horsesickness virus.

During ecological studies on important veterinary arthropod-borne viruses in Japan, Chuzan virus, a member of the Palyam serogroup orbiviruses, was isolated from the biting midge Culicoides oxystoma and from sentinel calves in 1985 (Miura et al., 1988). The virus was subsequently implicated in an epizootic of congenital abnormalities with hydranencephaly-cerebellar hypoplasia syndrome of calves observed in Kyushu district from Autumn 1985 through Spring 1986 (Goto et al., 1988a, b; Miura et al., 1990).

The Palyam serogroup viruses are members of the genus Orbivirus in the family Reoviridae. They have been associated with a variety of haematophagous arthropod vectors and large mammals, principally cattle, in many parts of the world including Asia, Australia and Africa (Knudson et al., 1984; Miura et al., 1988; Whistler & Swanepoel, 1988). Although the Palyam serogroup viruses seem to be involved in producing abortion and congenital malformations in cattle, the pathogenic importance of most of them remains unknown. As found in other members of the genus Orbivirus, bluetongue virus (BTV), epizootic haemorrhagic disease virus (EHDV) and African horsesickness virus (AHSV) (Roy, 1996), the genomes of the Palyam serogroup viruses consist of 10 dsRNA segments encoding seven structural (VP1—VP7) and four non-structural (NS1, NS2, NS3 and NS3A) proteins (Eaton & Gould, 1987; van Staden & Huismans, 1991). However, very little is known about the genetics and molecular properties, including sequence information, of the Palyam serogroup.

In order to understand the molecular biology of the Palyam serogroup viruses and to contribute to the knowledge of orbivirus evolution, we have cloned the Chuzan virus genome and analysed RNA segments 2, 3, 6 and 7 encoding the major capsid proteins VP2, VP3, VP5 and VP7, respectively (Yamakawa et al., 1999). In this paper, we report the nucleotide sequences of the other Chuzan virus genome segments 1, 4, 5, 8, 9 and 10 encoding VP1, VP4, NS1, NS2, VP6 and NS3, respectively. Sequence analysis of the Chuzan virus genome is now complete.

The cDNA clones representing genome segments 1, 4, 5, 8, 9 and 10 were selected from the previously constructed library (Yamakawa et al., 1999). Full-length cDNA clones of RNA segments 4, 5, 8, 9 and 10 were obtained, but both ends (bases 1–409 and 3654–3930) and a central region (bases 1828–1871) of segment 1 were absent from the library. Since both ends of the gene have been amplified by the 5'–RACE system (Gibco BRL) and cloned into pCR-Script SK(+) (Stratagene) and analysed (Yamakawa et al., 1999), a region covering bases 1828–1871 was amplified by RT–PCR from genomic dsRNA according to the protocol reported previously (Yamakawa et al., 1999). Appropriate restriction fragments or deletion derivatives of each clone were subcloned and sequenced using an ALFexpress DNA sequencer (Pharmacia Biotech). The data
Features of Chuzan virus genes and their encoded proteins

Table 1. Features of Chuzan virus genes and their encoded proteins

<table>
<thead>
<tr>
<th>Genome segment</th>
<th>Length (bp)</th>
<th>GC content (%)</th>
<th>5' non-coding region (bp)</th>
<th>Predicted ORF (aa)</th>
<th>Encoded protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3530</td>
<td>38.12</td>
<td>27/23</td>
<td>1295</td>
<td>VP1</td>
</tr>
<tr>
<td>2</td>
<td>3055</td>
<td>38.49</td>
<td>11/35</td>
<td>1002</td>
<td>VP2</td>
</tr>
<tr>
<td>3</td>
<td>2774</td>
<td>39.79</td>
<td>9/48</td>
<td>946</td>
<td>VP4</td>
</tr>
<tr>
<td>4</td>
<td>842</td>
<td>44.09</td>
<td>9/36</td>
<td>86</td>
<td>NS1</td>
</tr>
<tr>
<td>5</td>
<td>942</td>
<td>45.02</td>
<td>28/27</td>
<td>1495</td>
<td>VP3</td>
</tr>
<tr>
<td>6</td>
<td>1034</td>
<td>41.99</td>
<td>29/14</td>
<td>156</td>
<td>NS2</td>
</tr>
<tr>
<td>7</td>
<td>906</td>
<td>44.69</td>
<td>32/17</td>
<td>176</td>
<td>NS3</td>
</tr>
<tr>
<td>8</td>
<td>906</td>
<td>44.69</td>
<td>32/17</td>
<td>176</td>
<td>NS3</td>
</tr>
<tr>
<td>9</td>
<td>1967</td>
<td>47.69</td>
<td>38/25</td>
<td>1957</td>
<td>VP5</td>
</tr>
<tr>
<td>10</td>
<td>19528</td>
<td>33.33</td>
<td>38/25</td>
<td>6071</td>
<td>NS1</td>
</tr>
</tbody>
</table>

Note: An ORF beginning with the first AUG codon at bases 13–15 encodes only 21 aa. Another ORF extending from the second AUG codon at bases 23–25 encodes VP1.

The lengths of Chuzan virus genome segments and their coding capacities are summarized in Table 1. The total genome is 18,915 bp in length, ranging from 3930 bp (segment 1) to 728 bp (segment 10), and thus is shorter than the genetic contents of BTV-10 (19,218 bp) (Roy et al., 1990) and AHSV (19,528 bp) (Vreede & Huismans, 1998). The base ratio of the genome is 32.92% A, 27.82% U, 23.73% G and 15.53% C. The base ratios of individual segments are similar and the GC contents vary from 37.26% (segment 4) to 44.48% (segment 7). The 5’ non-coding regions vary in length from 9 nt (segment 4) to 35 nt (segment 5). The 3’ non-coding regions, ranging from 23 nt (segment 1) to 94 nt (segment 5), are longer than the 5’ non-coding regions. Overall, only 3.7% (702/18,915) of the genome is non-coding. A large open reading frame (ORF) which initiates protein synthesis from the first AUG is found in the mRNA strands of individual segments with the exception of segment 1, which has an ORF beginning with the second AUG codon at bases 23–25. An additional ORF beginning with the first AUG codon (bases 13–15) in segment 1 encodes only 21 aa. The sequence contexts of the potential initiation codons of segments 2–8 are in agreement with the Kozak consensus sequence, (A/G)XXAUGG, for strong translational initiation by eukaryotic ribosomes (Kozak, 1987). Flanking sequences of the initiation codons of segments 1, 9 and 10 deviate from the consensus sequence. The coding capacity of the ORFs is 6071 aa (see Table 1). Although an ORF capable of encoding a polypeptide of 131 aa (a predicted molecular mass of 14.3 kDa) was found in the minus strand of segment 9, such an ORF has not been identified in the cognate genes of BTV and AHSV (Roy et al., 1990; Turnbull et al., 1996). It seems unlikely that this ORF is used for protein synthesis. Apart from segment 9, no additional ORF of significant length was detected in either the plus or minus strand of each segment.

Comparisons of the predicted amino acid sequences of Chuzan virus VP1, VP4, VP6, NS1, NS2 and NS3 with the cognate proteins of BTV-10 (Roy et al., 1990), EHDV-1 (Wilson, 1994a, b; Jensen & Wilson, 1995), AHSV of different serotypes (van Staden et al., 1991; Mizukoshi et al., 1992, 1993; Laviada et al., 1995; Turnbull et al., 1996; Vreede & Huismans, 1998) and Broadhaven virus (BRDV) belonging to the tick-borne Kemerovo serogroup (Moss et al., 1992; Moss & Nuttall, 1995) are shown in Table 2. Chuzan virus VP1 appears to be highly conserved second only to the major core protein VP3 among orbiviruses (Yamakawa et al., 1999). The distribution of the homologous sequences is evenly spread through the molecules. Like BTV and AHSV, the consensus sequence elements conserved in RNA-dependent RNA polymerases (Poch et al., 1989; Koonin, 1992; Nakashima et al., 1996) are observed between residues 511 and 801 of Chuzan virus VP1 (data not shown), suggesting that Chuzan virus RNA segment 1 encodes the viral RNA polymerase.
Table 2. Comparisons of the minor capsid and the non-structural protein sequences of Chuzan virus and other orbivirus serogroups

Corresponding comparisons of the major capsid proteins VP2, VP3, VP5 and VP7 have been made in Yamakawa et al. (1999). —, Data not available.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Amino acid sequence identity (%)</th>
<th>Minor core protein</th>
<th>Non-structural protein</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>VP1</td>
<td>VP4</td>
</tr>
<tr>
<td>BTV-10</td>
<td></td>
<td>55.0</td>
<td>49.8</td>
</tr>
<tr>
<td>EHDV-1</td>
<td></td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>AHSV</td>
<td></td>
<td>63.3*</td>
<td>51.2†</td>
</tr>
<tr>
<td>BRDV</td>
<td></td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

* Data for AHSV-9.
† Data for AHSV-4.
‡ Data for AHSV-3.

Chuzan virus VP4 also exhibits high sequence identity with the VP4 proteins of BTV and AHSV and, therefore, is considered to be the candidate guanylyl transferase of the virus (Roy, 1996). A putative leucine zipper motif observed in five US isolates of BTV (Huang et al., 1993) is not identified in either Chuzan virus or AHSV. In contrast to VP1 and VP4, Chuzan virus VP6 is less conserved among orbiviruses. However, the physical properties such as overall amino acid composition and hydrophobic profile of the protein are very similar to those of other serogroups (data not shown). A sequence (GDGGSGES at residues 92–99) similar to the guanine NTP-binding motif (A/G)XXXGGKS(S/T) (Gorbalenya et al., 1988) is also present in Chuzan virus VP6. These findings imply that Chuzan virus VP6 has ssRNA- and dsRNA-binding properties and is the putative RNA helicase of the virus (Roy, 1996). The C-terminal region (the last 58 aa) of Chuzan virus NS2 is also less conserved among different serogroups, may be structurally and/or functionally important.

The NS1 proteins of BTV, EHDV, AHSV and BRDV form virus-specific tubular structures (tubules) in the cytoplasm of the infected cells (Moss & Nuttall, 1995; Roy, 1996). Although Chuzan virus NS1 appears to be variable among orbiviruses, the protein is clearly related to those of other serogroups (data not shown). The positions of five of 15 cysteine residues in Chuzan virus NS1 are conserved among different serogroups. It has been demonstrated that two of these conserved cysteine residues at positions 337 and 340 in BTV-10 NS1 are essential for tubule formation (Monastyrskaya et al., 1994). Nevertheless, distinct structures, like tubules, have not been observed in Chuzan virus-infected cells at any stages of infection (our unpublished data). The C terminus (the last 57 aa) of Chuzan virus NS1 exhibits high sequence variability compared to those of other NS1 proteins, indicating that this region probably determines the tertiary structure of the protein. In particular, cysteine residue(s) may directly influence(s) the morphology of the tubular complex by intramolecular disulfide bonds, since only Chuzan virus NS1 possesses no cysteine in this region. It is necessary to confirm whether the protein forms a virus-specific structure by gene expression experiments using recombinant baculoviruses (Roy, 1996); localization and functional significance must be clarified by further analyses.

Chuzan virus NS2 is also less conserved among different orbiviruses. The middle region at residues 153–242 is highly hydrophilic and variable among the NS2 proteins. It seems that Chuzan virus NS2 is associated with viral mRNA species since the ssRNA-binding motif (I/L)XXM(I/L)(S/T)XXG, identified in the non-structural proteins of the members of the family Reoviridae (van Staden et al., 1991), is also found at residues 73–81 of Chuzan virus NS2 with the substitution of isoleucine for methionine (ISL[ISEEG]). Theron et al. (1996a, b) reported for EHDV that this motif plays an important role in the interaction of NS2 with ssRNA by affecting the structural integrity of the protein and is involved in the formation of viral inclusion bodies as an essential structural determinant. It will be of interest to see whether the amino acid substitution in this motif reflects structural and functional differences of NS2 between Chuzan virus and other orbiviruses.

van Staden & Huismans (1991) indicated that two non-structural proteins are synthesized from different in-frame initiation codons of the smallest gene of Palyam virus as well as BTV, EHDV and AHSV. Therefore, two related proteins, NS3 and NS3A, are likely to be translated from the AUG codons at bases 19–21 and 52–54, respectively, of Chuzan virus RNA segment 10. Like other orbiviruses, Chuzan virus NS3 has two characteristic hydrophobic domains (residues 114–137 and 146–167). The conserved domain among the NS3 proteins is located near the N terminus (residues 46–87). These regions of Chuzan virus NS3 are considered to be
functionally important. As speculated by van Staden et al. (1995) and Jensen & Wilson (1995), the protein is possibly retained on the cell membrane by these two hydrophobic domains and the N-terminal domain, predicted to be exposed in the cytoplasm, may mediate the transport and egress of the virion from infected cells in the final stages of virus morphogenesis.

These data, together with our previous report (Yamakawa et al., 1999), indicate that the outer capsid protein VP2, virus-neutralizing antigen, is the most variable and the major core protein VP3, which is considered to form the basic scaffold of the core particle (Roy, 1996), is the most conserved among orbiviruses. Overall, the non-structural proteins are less conserved than the structural proteins except for VP2. Phylogenetic profiles of individual proteins indicate that Chuzan virus and AHSV are closely related and form a distinct branch of the phylogenetic tree, while BTV and EHDV form a distinct cluster of the same tree (Fig. 1). However, only VP6 is more closely related to BTV VP6 than to AHSV VP6. Tick-borne BRDV is distantly related to these Culicoides-borne orbiviruses.

With the exception of Chuzan virus, only partial sequences of segment 3 (two internal fragments of 1095 and 245 bp of D’Aguilar virus) and segment 10 (the first 590 bp of Palyam virus) of the Palyam serogroup virus have been reported to date (Gould & Pritchard, 1991; van Staden & Huismans, 1991). Two fragments of D’Aguilar virus segment 3 exhibit identities of 88.0 and 84.9% with the corresponding regions at nt 650–1744 and 1811–2055, respectively, of Chuzan virus segment 3. At the amino acid level, these regions of Chuzan and D’Aguilar viruses show 94.5 and 97.5% identities, respectively. On the other hand, a relatively low identity is observed between segment 10 genes of Chuzan and Palyam viruses (77.6 and 86.4% identities in nucleotide and amino acid sequences, respectively). Further sequence data from other members of this serogroup are required to determine precise intra-serogroup phylogenetic relationships. The data obtained from the complete sequence of Chuzan virus dsRNA genome should facilitate the establishment of molecular diagnostic tools to study the molecular epidemiology of the Palyam serogroup viruses.

References


Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M.


Received 15 October 1998; Accepted 14 December 1998