Comparison of capsid sequences from human and animal astroviruses

Christine M. Jonassen,¹ Tom Ø. Jonassen,¹ Yehia M. Saif,² David R. Snodgrass,³ Hiroshi Ushijima,⁴ Mitsugu Shimizu⁵ and Bjørn Grinde¹

¹ Department of Virology, National Institute of Public Health, PO Box 4404 Torshov, N-0403 Oslo, Norway
² Ohio Agricultural Research and Development Center, Ohio, USA
³ Moredun Research Institute, Edinburgh, UK
⁴ Faculty of Medicine, University of Tokyo, Japan
⁵ National Institute of Animal Health, Ibaraki, Japan

We have sequenced the genomic 3′-end, including the structural gene, of human astrovirus (HAstV) serotype 7 and morphologically related viruses infecting pig (PAstV), sheep (OAstV) and turkey (TAstV-1). These sequences were compared with corresponding astrovirus sequences available in the nucleic acid databases, including sequences of the seven other HAstV serotypes, two other avian astroviruses (TAstV-2 and avian nephritis virus) and astrovirus from cat (FAstV). A 35 nt stem–loop motif near the 3′-end of the genome, previously described as being highly conserved, was present in all of the astroviruses except TAstV-2. In the N-terminal half of the capsid precursor protein, there were several short conserved peptide motifs. Otherwise the capsid proteins of astroviruses infecting different hosts were highly divergent. Calculation of genetic distances revealed that the distance between FAstV and HAstV is comparable to the largest distances between different HAstV serotypes. Higher similarities between the HAstV, FAstV and PAstV capsid sequences suggest interspecies transmissions involving humans, cats and pigs relatively recently in the evolutionary history of astroviruses.

Introduction

Astroviruses infect the gastrointestinal tract of humans (Madeley & Cosgrove, 1975) and various animals, e.g. sheep (Snodgrass & Gray, 1977), cattle (Woode & Bridger, 1978), dog (Williams, 1980), domestic cat (Hoshino et al., 1981), red deer (Tzipori et al., 1981), duck (Gough et al., 1984), mouse (Kjeldsberg, 1985), turkey (McNulty et al., 1980) and pig (Shimizu et al., 1990). The viruses are assigned to the family Astroviridae (Jiang et al., 1993; Monroe et al., 1993; Cubitt, 1996) based on common and unique genomic traits and/or on a distinct morphology, i.e. typical particles have a characteristic five- or six-pointed star motif on their surface. Recently an avian nephritis virus (ANV) isolated from chicken was shown by genomic analyses to be a member of the family Astroviridae (Imada et al., 2000). Astroviruses appear to be species-specific, and to our knowledge there is no record of documented interspecies transmission.

Human astroviruses (HAstV) are a common cause of diarrhoea in children, the elderly and the immunosuppressed (reviewed by Glass et al., 1996). They are non-enveloped particles with a plus-strand RNA genome of approximately 7 kb (Carter & Willcocks, 1996). The genome contains three open reading frames (ORFs), as well as terminal non-coding regions (NCR). ORFs 1a and 1b are presumably linked by a translational frameshift, and encode non-structural proteins. ORF2 is transcribed into a subgenomic mRNA (Monroe et al., 1991, 1993; Cubitt, 1996) and encodes the capsid proteins. HAstV is divided into different serotypes (Kurtz & Lee, 1984) that are designated HAstV-1 to HAstV-8. The serotypes correspond well with phylogenetic reconstructions based on genome sequences (Belliot et al., 1997c; Monceyron et al., 1997).

Only limited sequence information has been available for animal astroviruses. A cat astrovirus (FAstV) ORF2 sequence is available in the nucleic acid databases, and recently full-length genomic sequences of an isolate of turkey astrovirus
The astrovirus capsid proteins are presumably synthesized as a single precursor (Lewis et al., 1994) that is subsequently processed, but the maturation process is not fully understood. Different experimental conditions, as well as strain-specific differences, may account for the differences in the capsid compositions reported. For HAstV-1 and -2 an initial capsid...
precursor of 86–90 kDa is observed (Monroe et al., 1991; Lewis et al., 1994; Bass & Qiu, 2000). This is processed to a 79 kDa protein, which, when cultured in the absence of trypsin, is the major HAstV-1 structural protein. However, the low infectivity of these particles suggests that further extracellular processing does occur in vivo, in the presence of trypsin, probably to three distinct proteins of 20–34 kDa (Bass & Qiu, 2000; Monroe et al., 1991; Sanchez-Fauquier et al., 1994). The N terminus of the 79 kDa HAstV-1 protein (Bass & Qiu, 2000), as well as two HAstV-2 overlapping capsid proteins, VP26 and VP29 (Sanchez-Fauquier et al., 1994), have been determined by amino acid sequencing. In other HAstV serotypes (Bellioli et al., 1997a; Kurtz & Lee, 1987) and in OAstV (Herring et al., 1981) and PAstV (Shimizu et al., 1990) different numbers and sizes of capsid proteins are reported.

A 35 nt stem–loop, previously referred to as s2m, near the astrovirus ORF2 stop codon, is a highly conserved RNA structure. Sequence differences in base-paired nucleotides of the 3′ ends of virus RNA were amplified and sequenced as described earlier (Jonassen et al., 1998). Longer amplicons were generated using the 5′-RACE kit (Gibco BRL) using a minus-strand primer based on the initially obtained 3′-sequence both for reverse transcription and in the subsequent 5′-RACE PCR. The PCR products obtained were sequenced in both directions using a primer walking strategy. Sequencing was performed using the ABI PRISM BigDye Terminator Cycle Sequencing kit (Applied Biosystems) and analysed on an ABI PRISM 310 Genetic Analyser (Applied Biosystems).

### Sequencing

Isolation of RNA was performed either with Trizol (Gibco BRL) or QIAamp Viral RNA Mini kit (QIAGEN). Initially the 3′-ends of virus RNA were amplified and sequenced as described earlier (Jonassen et al., 1998). Longer amplicons were generated using the 5′-RACE kit (Gibco BRL) using a minus-strand primer based on the initially obtained 3′-sequence both for reverse transcription and in the subsequent 5′-RACE PCR. The PCR products obtained were sequenced in both directions using a primer walking strategy. Sequencing was performed using the ABI PRISM BigDye Terminator Cycle Sequencing kit (Applied Biosystems) and analysed on an ABI PRISM 310 Genetic Analyser (Applied Biosystems).

### Biocomputing

Sequence analyses were performed using the Wisconsin sequence analysis package version 10.0 (Genetics Computer Group, Madison, Wisconsin, USA) and the PHYLIP package (Joseph Felsenstein, Department of Genetics, University of Washington, Seattle, Washington, USA; http://evolution.genetics.washington.edu/). These programs were accessed via the Norwegian EMNet node (http://www2.no.embnet.org/). To avoid bias caused by over-representation of HAstV sequences, the weight of these sequences was reduced during aligning. Alignments were manually adjusted using the BioEdit sequence alignment editor (Tom Hall, Department of Microbiology, North Carolina State University, North Carolina, USA; http://www.mbio.ncsu.edu). Similarity plot analysis was done with the SimPlot program (Stuart C. Ray, Department of Medicine, Johns Hopkins University, Baltimore, Maryland, USA; http://www.med.jhu.edu/deptmed/sray/download/).

### Accession numbers

The sequences used in the figures have the following accession numbers: HAstV-1, S68561; HAstV-2, L06802; HAstV-3, AF117209; HAstV-4, Z38863; HAstV-5, HA15136; HAstV-6, Z40658; HAstV-7, Y08632; HAstV-8, Z66541; TAstV-1, T03169; TAstV-2, AF206593; ANV, AB033998; OAstV, Y15397; PAstV, Y15398 and FastV, AF056197.

### Results

We sequenced more than 2.7 kb, covering the 3′ ORF and the 3′-NCR, of the OAstV, TAstV-1, PAstV and HAstV-7 genomes. The 3′ ORF of these viruses shows similarity to the

### Methods

#### Virus samples

Faecal samples were diagnosed as containing astroviruses by electron microscopy. The virus samples have been described previously: OAstV (Snodgrass & Gray, 1977), PAstV (Shimizu et al., 1990), TAstV-1 (Reynolds & Saif, 1986) and HAstV-7 (Lee & Kurtz, 1994).

#### Amino acid sequences

The sequences used in the figures have the following accession numbers: HAstV-1, S68561; HAstV-2, L06802; HAstV-3, AF117209; HAstV-4, Z38863; HAstV-5, HA15136; HAstV-6, Z40658; HAstV-7, Y08632; HAstV-8, Z66541; TAstV-1, T03169; TAstV-2, AF206593; ANV, AB033998; OAstV, Y15397; PAstV, Y15398 and FastV, AF056197.

#### Phylogenetic analysis

Amino acids that are identical or similar (determined by the GCG program PRETTYPLOT) in all available sequences are indicated by colons below the consensus sequence. It should be noted that the consensus sequence is biased by HAstV sequences being over-represented. Amino acids that are identical or similar (determined by the GCG program PRETTYPLOT) in all available sequences are indicated by colons below the consensus sequence. It should be noted that the consensus sequence is biased by HAstV sequences being over-represented.
ORF2 of the previously published astrovirus sequences. This indicates that the 3′ ORF of OAstV, PAstV and TAstV-1 encodes the structural proteins. In nucleotide database searches with astrovirus capsid as query sequences, other astroviruses were the only ones that gave significant scores (data not shown).

The astrovirus ORF2 encodes between 671 aa in TAstV-1 and 816 aa in FAstV, corresponding to deduced molecular masses (M_r) of 72.5 kDa and 89.8 kDa respectively. ORF2 was shortest in the three avian astroviruses. The amino acid sequences inferred from the capsid genes of FAstV, PAstV, OAstV, TAstV-1, TAstV-2, ANV and the eight HAstV serotypes were aligned. An alignment of the N-terminal half of the putative capsid precursor of representative astroviruses is shown in Fig. 1. In the C-terminal half, the difference between the astroviruses was so extensive that the alignment is probably suboptimal and is therefore not shown. This is illustrated in Fig. 2, in which the per cent identical residues along our best alignment are shown. As the C-terminal parts of the capsid precursor sequences were difficult to align, and the s2m is most likely homologous, the alignment was adjusted to keep the s2m aligned.

The initial methionine appeared to be homologous in the non-avian hosts. Judging from alignment of the 3′-end of the upstream ORF (data not shown), the avian astroviruses seemed to have the initial methionine in the same position as the other astroviruses. In all but the TAstV-1 sequence, the neighbouring downstream codon starts with a G residue, suggesting that this is a strong start codon (Kozak, 1997).

The 3′-NCR was about 80 nt in most of the astroviruses. In the avian astroviruses, however, it ranged from 140 nt in TAstV-1 to 305 nt in ANV (Imada et al., 2000). The ANV s2m was 117 nt farther from the poly(A) tail than the HAstV s2m.

**Table 1. Distances between capsid proteins of astroviruses infecting different hosts**

<table>
<thead>
<tr>
<th></th>
<th>Avian</th>
<th>Ovine</th>
<th>Human</th>
<th>Feline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porcine</td>
<td>2.28–2.41</td>
<td>1.75</td>
<td>0.74–1.00</td>
<td>0.77</td>
</tr>
<tr>
<td>Avian</td>
<td>0.91–1.49</td>
<td>2.31–2.54</td>
<td>2.00–2.32</td>
<td>2.28–2.37</td>
</tr>
<tr>
<td>Ovine</td>
<td>1.54–1.04</td>
<td>1.78</td>
<td>0.74–1.00</td>
<td>0.77</td>
</tr>
<tr>
<td>Human</td>
<td>&lt; 0.54</td>
<td>0.46–0.54</td>
<td>0.74–1.00</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Imada et al. (2000) calculated a 91% sequence identity between the ANV s2m and other astrovirus s2m, but all the differences were in base-paired nucleotides which covaried to maintain the RNA folding predicted earlier (Monroe et al., 1993; Jonassen et al., 1998). Except for the avian astroviruses, the ORF2 stop codon was located in the middle of the s2m. In TAstV-1 and ANV the stop codons were respectively 21 and 103 nt upstream of this position.

The basic residues in the N-terminal part of the capsid precursor, described as being similar to basic stretches found in the capsid proteins of two coronaviruses (Carter, 1994), were highly conserved between HAstV, FAstV, PAstV and TAstV-2 (position 19–34 in Fig. 1). A pattern of alternating small and basic amino acids, mainly serine (S) and arginine (R), was apparent in these sequences, most significantly in TAstV-2, where an SR dipeptide was repeated six times. In OAstV, ANV and TAstV-1 the homologous residues were also highly basic, but otherwise different from the astrovirus consensus sequence.

The N terminus of the 79 kDa capsid protein found in HAstV-1 cultivated without trypsin (Bass & Qiu, 2000) is at position 75 in Fig. 1. In this region the amino acid sequences of HAstV, FAstV and PAstV were highly conserved, while the avian astroviruses and OAstV diverged from the consensus. The N termini of the HAstV-2 structural proteins VP29 and VP26 reported by Sanchez-Fauquier et al. (1994) are in positions 379 and 416, respectively, in Fig. 1. These putative trypsin cleavage sites do not seem to be conserved in all astroviruses.

A similarity plot comparing the amino acid sequence of HAstV-2 with HAstV-7 and the animal astroviruses is shown in Fig. 2. ANV and TAstV-2 showed profiles similar to TAstV-1, but were not included in the figure. The N-terminal half of the capsid precursor was considerably more conserved than the C-terminal half. In the N-terminal half and the C-terminal end of the capsid precursor, FAstV was as close to HAstV-2 as HAstV-7 was. The same profile of similarity with HAstV-2 was seen for PAstV, while OAstV did not display the C-terminal similarity. As is also shown in Table 1, the overall
similarities to the HAstV sequences decreased in the following order: FAstV > PAstV > OAstV > avian astroviruses.

A phylogenetic tree based on an amino acid distance matrix verified the relationship between the different astroviruses suggested above (Fig. 3). The most striking feature of the tree was that FAstV and PAstV cluster with HAstV, while OAstV and the avian astroviruses are more distantly related. The three known avian astroviruses are widely separated evolutionarily, but still constitute a monophyletic group.

Discussion

Even though there is very limited sequence similarity left between the structural genes of the avian astroviruses, OAstV and the HAstV/FAstV/PAstV clade (Table 1), there are still some capsid peptide motifs that are common to all the investigated astroviruses (Fig. 1).

Viruses infecting different hosts have either coevolved with the hosts since the hosts diverged, or they are the result of inter-species transmission. The phylogenetic tree (Fig. 3) reveals that FAstV, PAstV and HAstV are more closely related to each other than to OAstV. Sheep, cats and pigs all parted with human ancestors some 100 million years ago. As the latest common ancestor of sheep and pigs presumably is more recent (Novacek, 1992), the most likely explanation for the discrepancy between the family trees of the viruses and their hosts is that astroviruses have been transmitted between cats, pigs and humans, possibly involving intermediary hosts. The observation that FAstV is as related to HAstV as the most distantly related HAstVs are to each other suggests that zooneses involving pigs, cats and humans have occurred relatively recently compared to the latest common ancestor of all astroviruses. It follows that although zooneses appear to be rare, one should be aware of the possibility and take proper precautions.

The absence of the s2m from TAstV-2 may reflect that the TAstV-2 lineage split from the other astroviruses before an ancestor of the other astroviruses acquired the s2m. Alternatively, the s2m was once present in TAstV-2 as well, but was later lost. The present data do not support one of these explanations over the other.

It has proven feasible to detect all HAstV serotypes with a single PCR (Jonassen et al., 1995; Sakon et al., 2000). Conserved nucleotide elements could facilitate the design of an Astroviridae-specific PCR to be used for the detection of viruses from animals for which no sequence information is available. Alignment of the nucleotide sequences from the 3’-end of the available astrovirus genomes revealed that the only viable option for a pan-astrovirus PCR rests with the s2m, in combination with a downstream primer that binds to the poly(A) tail. A PCR with an s2m sense primer and an NV(T)n anti-sense primer may function, even for Astroviridae other than the ones sequenced. One should keep in mind, however, that the s2m is present in certain other viruses (Jonassen et al., 1998), while it is absent in TAstV-2.

As most astroviruses share the same morphology, we expected to find conserved features in the capsid proteins. Conversely, the presence of motifs conserved at the amino acid level suggests an essential function, and thus possibly a role in the assembly and function of the virus particle.

It has been suggested that in the case of HAstV, unlike Norwalk virus, post-translational processing is required for the assembly of virus particles (Carter & Willcocks, 1996). In HAstV-2 cultured in the presence of trypsin, the three structural proteins are referred to as VP34, VP29 and VP26, where VP26 is an N-terminal truncated version of VP29 (Sanchez-Fauquier et al., 1994). Bass & Upadhyayula (1997) found two neutralizing monoclonal antibodies that reacted only with VP29, while a third also attached to VP26. Another neutralizing antibody developed by Sanchez-Fauquier et al. (1994) reacted with both proteins. The results suggest that VP26 and VP29 contain at least one common motif on the surface of the virus capsid that may be involved in cell binding, while the N-terminal part of VP29 has an additional epitope of similar characteristics. The lack of cross-reactivity with monoclonal antibodies against VP26/VP29 suggests that the third, larger structural protein (VP34) is independently coded. The HAstV peptide between the suggested N terminus of the 79 kDa protein (Bass & Qiu, 2000) and the N terminus of the 29 kDa protein (Sanchez-Fauquier et al., 1994) has a deduced Ms of 31 kDa, which may be compatible with VP34. The peptide from the N terminus of the 29 kDa protein to the end of ORF2 has a deduced Ms of 46–49 kDa, which probably does not give room for a 34 kDa protein between the C
terminus of VP29 and the end of ORF2. The observation that the neutralizing monoclonal antibodies all react to VP26/VP29, rather than to VP34, suggests that VP26/VP29 dominates the surface of the particle. This may explain the higher heterogeneity of VP26/VP29 compared to the N-terminal part of the capsid precursor.

Unlike other astroviruses, ANV can be cultured to high titres in the absence of trypsin. This suggests that the processing of the capsid precursor is different from that of other astroviruses, and may explain the lack of conservation of the trypsin cleavage sites that are suggested for HAstV. The lack of requirement for trypsin-like proteases may be related to the different tissue tropism of ANV compared to most astroviruses.

The basic region in the N-terminal part of the capsid precursor of the astroviruses contains, to a variable extent, an SR-repeat motif. In addition to certain coronaviruses that are related to transmissible gastroenteritis virus, SR repeats are also found in, for example, the E2 protein of some human papillomaviruses (Lai et al., 1999) and some baculoviruses (Oellig et al., 1987). In multicellular organisms long stretches of alternating S and R residues are found in the SR protein family of pre-mRNA splicing factors. The SR region of the papillomavirus E2 protein has been shown to be important for nuclear localization and regulation of gene expression. It is possible that this motif serves a regulatory function also in the astroviruses.

The region beyond the presumptive 3′-end of the VP26/VP29 gene is hypervariable, with large inserts or deletions. The C-terminal 6 aa are highly conserved, with the exception of the avian astroviruses. This motif, however, is within the s2m, and is thus presumably conserved by selection at the RNA level. This assumption is strengthened by the location of the s2m downstream of ORF2 in TAstV-1 and ANV.

We wish to thank Torunn Merete Sveen for excellent technical assistance.

References


Received 8 September 2000; Accepted 12 January 2001