Evolution and conservation in human parechovirus genomes

Çiğdem H. Williams,1 Maria Panayiotou,1 Gareth D. Girling,1 Curtis I. Peard,1 Sami Oikarinen,2 Heikki Hyöty2,3 and Glyn Stanway1

1Department of Biological Sciences, University of Essex, Colchester CO4 3SQ, UK
2Department of Virology, University of Tampere, Medical School, Tampere, Finland
3Center for Laboratory Medicine, Tampere University Hospital, Tampere, Finland

Human parechoviruses (HPeVs) are frequent pathogens with a seroprevalence of over 90% in adults. Recent studies on these viruses have increased the number of HPeV types to eight. Here we analyse the complete genome of one clinical isolate, PicoBank/HPeV1/a, and VP1 and 3D protein sequences of PicoBank/HPeV6/a, isolated from the same individual 13 months later. PicoBank/HPeV1/a is closely related to other recent HPeV1 isolates but is distinct from the HPeV1 Harris prototype isolated 50 years ago. The availability of an increasing number of HPeV sequences has allowed a detailed analysis of these viruses. The results add weight to the observations that recombination plays a role in the generation of HPeV diversity. An important finding is the presence of unexpected conservation of codons utilized in part of the 3D-encoding region, some of which can be explained by the presence of a phylogenetically conserved predicted secondary structure domain. This suggests that in addition to the cis-acting replication element, RNA secondary structure domains in coding regions play a key role in picornavirus replication.

INTRODUCTION

Picornaviruses are non-enveloped, positive-sense RNA viruses with icosahedral capsids, composed of 60 copies of three or four virus-encoded proteins (VP1–4 or VP0, VP1 and VP3) (Stanway et al., 2002, 2005). They have a genome of around 7000–8000 nt encoding one polyprotein which is cleaved by virus proteases to give the structural and non-structural (2A–C and 3A–D) proteins. Picornaviruses consist of economically and socially very important human and animal viruses such as polioviruses, other enteroviruses, rhinoviruses, hepatitis A virus, foot-and-mouth disease virus and parechoviruses. Currently, parechoviruses consist of two species: human parechovirus (HPeV) and Ljungan virus (LV) (Johansson et al., 2002; Joki-Korpela & Hyöty, 2001; Stanway & Hyöty, 1999; Stanway et al., 2000). HPeVs are frequent infectious agents and although they usually cause mild gastroenteritis and respiratory disease in young children, more serious cases, such as flaccid paralysis, encephalitis and myocarditis, have also been reported, particularly associated with HPeV3 infection (Baumgarte et al., 2008; Benschop et al., 2006a, 2008b; Ehrnst & Eriksson, 1993; Figueroa et al., 1989; Harvala et al., 2008, 2009; Joki-Korpela & Hyöty, 2001). Recently, HPeV1 has also been linked to otitis media (Tauriainen et al., 2008). With the isolation of several HPeV types in recent studies (Abed & Boivin, 2005; Al-Sunaidi et al., 2007; Baumgarte et al., 2008; Benschop et al., 2006b; Drexler et al., 2009; Ito et al., 2004; Li et al., 2009), eight types (1–8) of HPeVs are currently known.

HPeVs have several distinctive features compared with other picornaviruses (Ghazi et al., 1998; Hyöty et al., 1992; Stanway et al., 1994, 2000). Post-translational cleavage of the polyprotein results in only three structural proteins (VP0, VP3 and VP1), as the cleavage of VP0 to VP4 and VP2 seen in other picornaviruses does not occur (Stanway et al., 1994). In addition, the HPeV VP3 protein possesses a unique N-terminal extension of approximately 30 aa, which are predominantly positively charged (Stanway et al., 1994, 2000) and the 2A protein has homologous residues (NC and H box functional motifs) to cellular proteins that are involved in cell proliferation (Hughes & Stanway, 2000). However, until recently, HPeVs have been relatively little studied and poorly understood. HPeV1 is the most commonly isolated type, but the sequence information on recent HPeV1 isolates is limited and the
Human parechovirus evolution and conservation

Methods

Patients, virus isolation and clinical data. Patient 069960A is a control child from the DIPP study, a prospective study that recruits children at high genetic risk of developing type 1 diabetes (T1D). These children are first screened at birth for human leukocyte antigen (HLA)-risk alleles for T1D and those with increased genetic risk (14.9 % of all newborns) are invited to a prospective follow-up starting from birth. Stool samples are collected during the follow-up at monthly intervals for virus analyses (Lönrot et al., 2000). This DIPP child was born on 17 February 2000 and, altogether, 13 stool samples were collected during her follow-up (last sample at the age of 23 months). All stool samples were screened for the presence of HPeV RNA using an RT-PCR assay as previously described (Tauriainen et al., 2008); two samples were found to be HPeV positive. PicoBank/HPeV1/a was isolated from the first sample, collected on 17 October 2000, and PicoBank/HPeV6/a was isolated from the second sample, collected on 19 November 2001. Symptoms and diseases were recorded every 3 months during the visits to the DIPP clinic, when parents filled in a questionnaire recording all infections, symptoms and medications. The parents of this child did not report any symptoms of infections at the time the second sample was collected and the child was not reported to have any allergies or to be on permanent medication. The protocol of the DIPP study has been accepted by ethical committees of the participating university hospitals and a written consent has been obtained from all study subjects.

Monolayers of SW480-α5β1 cells, a human colon carcinoma cell line transfected to express the human full-length β1-integrin subunit (Weinacker et al., 1994), were infected with the PicoBank/HPeV1/a suspension and the viral RNA was extracted. The cells were maintained in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal calf serum (heat inactivated), 2 mM glutamine, 100 U penicillin ml−1, 100 μg streptomycin ml−1 and 1 mg geneticin ml−1 (Invitrogen). PicoBank/HPeV6/a was propagated in the same way, although the sequence data presented was from viral RNA extracted directly from the stool sample of patient 069960A.

Extraction of viral RNA and RT-PCR. Viral RNAs were extracted from cell cultures by the use of a QIAamp Viral RNA mini kit (Qiagen), according to the manufacturer’s instructions. The final eluate was obtained by the addition of 60 μl buffer AVE (0.04 % sodium azide containing RNase free water) to the QIAamp column and subsequent incubation at room temperature for 1 min followed by centrifugation at 6000 g for 1 min.

Extracted RNA was then mixed with oligonucleotide pairs that flank the region of interest and RT-PCRs were carried out by using a SuperScriptIII One-Step RT-PCR kit (Invitrogen) according to the manufacturer’s instructions. Briefly, 5 μl RNA, 100 pmol of each primer, 1 μl RT-Taq mix and 25 μl 2× reaction mix were mixed together in a total volume of 50 μl (made up with sterile ultrapure water). RT-PCR was carried out in an Applied Biosystems 9700 thermal cycler. Amplicons were isolated by agarose gel electrophoresis and purified using GFX PCR DNA and gel band purification kit (Amersham Pharmacia) according to the manufacturer’s instructions. The purified fragments were then sequenced commercially (Genservice).

Oligonucleotides. The oligonucleotides used in this study are listed in Supplementary Table S1 (available in JGV Online). In short, degenerate oligonucleotides generated by determining the conserved regions of the alignments of all the known HPeV sequences were used in order to amplify cDNA. The resulting PCR products were then sequenced directly and in instances where the sequence data were of poor quality, oligonucleotides were designed based on the available sequences and the amplification process and direct sequencing were repeated. All of the derived sequences were obtained in both orientations.

Virus strains and nucleotide sequence data. The GenBank accession numbers for the sequence data obtained in this study are as follows: PicoBank/HPeV1/a, FM242866; PicoBank/HPeV6/a, VP1, FM242868, 3D, FM242867. Altogether, 17 complete HPeV genome sequences and 61 HPeV VP1 and 3D sequences were used (GenBank accession numbers are given in Supplementary Table S2, available in JGV Online).

Computer analysis. Alignments of 17 HPeV complete genomes and 61 HPeV VP1 and 3D protein sequences were carried out using CLUSTAL W (http://align.genome.jp/). The neighbour-joining method in MEGA4 software (Tamura et al., 2007) was used to construct the phylogenetic trees, employing the maximum composite likelihood method for nucleotide sequences. SimPlot version 3.2 (Lole et al., 1999), using the Kimura two-parameter technique, was used to investigate possible recombination events. Synonymous codon variability was determined, using the Simmonics sequence editor (Tuplin et al., 2002), by measuring mean pair-wise distances at each codon position of the 17 complete sequences and 61 3D sequences. The analysis was restricted to codons where the encoded amino acids were the same in all sequences. RNA secondary structure predictions were performed using the Mfold site (Zuker, 2003).

Results

PicoBank/HPeV1/a and other complete HPeV sequences

The virtually complete sequence of the Finnish HPeV isolate, PicoBank/HPeV1/a, was obtained in both orientations by RT-PCR and sequencing of overlapping RT-PCR amplicons. The first part of the 5′ untranslated region (UTR) was not determined due to technical difficulties encountered in the amplification of this region; therefore, the genome sequence obtained starts at position 64 relative
to the HPeV1 prototype Harris. This omission results in a 7274 bp sequence, where the partial 5\' UTR sequence (646 bp) precedes an ORF encoding a 2180 aa polyprotein and an 88 bp 3\' UTR, followed by a poly (A) tail. Isolate PicoBank/HPeV6/a was obtained from the same child 13 months later. VP1 sequence analysis shows that it is a type 6 strain (Fig. 1a).

The PicoBank/HPeV1/a genome [excluding the poly (A) region] contains 2362 As, 2036 Us, 1496 Gs and 1380 Cs. This is similar to all the completely sequenced isolates. The high A/U content of HPeVs is partially related to a marked preponderance of NNU/NNA codons (32.0–36.4 \% NNU and 28.6–31.1 \% NNA in the 17 completely sequenced HPeVs). There is minor variation in codon usage between the different HPeVs (see Supplementary Fig. S1, available with the online version of this paper) but this does not correlate with date of isolation or type (data not shown).

**VP1 and 3D sequence comparisons**

Two distal regions, encoding parts of VP1 and 3D, have been widely used to study the evolution of picornaviruses. Phylogenetic trees showing the relationship between these regions in the 61 viruses where both have been sequenced are shown in Fig. 1. The VP1 analysis, based on 606 nt, shows that the eight HPeV types clearly segregate. Strains of the types other than HPeV3 cluster loosely and this is particularly evident for HPeV1, which seems to divide into two main genotypes (for which we propose the terms Clusters 1A and 1B). PicoBank/HPeV1/a belongs to Cluster 1A, while the other completely sequenced modern HPeV1 isolates, BNI-788St and 7555312, belong to Clusters 1A and 1B, respectively (De Souza Luna et al., 2008; Zoll et al., 2009). Several HPeV1 isolates branch separately, including the prototype Harris. The HPeV2 prototype, Williamson, the only HPeV2 representative in the analysis, groups with the HPeV1 strains, but is distinct from them. Within HPeV4, the older strains form a tighter cluster, while the more recent prototype, K251176-02, is more diverged.

Phylogenetic analysis of the corresponding HPeV 3D sequences (Fig. 1b) shows that the type-specific segregation seen in the VP1 phylogeny is not fully carried over to this region, supporting the suggestion that intertypic recombination events have taken place in the evolution of these viruses. Members of HPeV1 VP1 Clusters 1A and 1B still largely segregate together in the 3D comparisons, supporting the idea that these are distinct lineages. There are no examples of recombination between members of these clusters, although four Cluster A viruses are diverged in the 3D comparison, suggesting recombination with other lineages. Three of the HPeV6 strains, including PicoBank/HPeV6/a, cluster together but are also closely related to two HPeV1 strains dating from 2001 and 2003. The other HPeV6 isolate is distinct and its closest relative is the HPeV1 prototype Harris. In general, the older HPeV strains (1956–1992) tend to form a series of distinct branches. A small number of recent isolates also form more distinct branches, suggesting the co-circulation of several distinct 3D genotypes.

**SimPlot analysis**

Incongruence between the VP1 and 3D trees is indicative of recombination during HPeV evolution. To look at this further, SimPlot was used to analyse the relationship between the other 16 complete HPeV genomes and PicoBank/HPeV1/a (Fig. 2a). This confirms the high similarity to HPeV1 BNI-788St across the whole genome and to HPeV1 7555312 in the 5\' UTR and P1 regions (Fig. 1). Although PicoBank/HPeV1/a shows similarity to HPeV1 Harris before position 3000, after this point, the similarity decreases and there is much more similarity to the HPeV3 strains and to the HPeV4 prototype K251176-02. Also, from around position 5000, an increase in similarity to T75-4077 can be seen. Sequence alignment suggests that in PicoBank/HPeV1/a, recombination with HPeV3-related sequences occurred at around genome position 3071 (HPeV1 Harris numbering) as the number of differences from HPeV3 A308/99 decreases sharply after this point (see Supplementary Fig. S2, available with the online version of this paper). Recombination with HPeV4 sequences seems to have occurred slightly closer to the 5\' terminus, around position 2982.

Using HPeV1 Harris as the basis of the SimPlot analysis reveals two features (Fig. 2b). Within the capsid region, the similarity of HPeV1 Harris to the other two HPeV1 strains drops at around position 1500 and approaches the level seen with other HPeV strains, particularly HPeV2 Williamson. This suggests that during the evolution of HPeV1 Harris, intertypic recombination has occurred within the capsid region. Secondly, as the 3D tree indicates (Fig. 1b), there is similarity to the HPeV6 isolate BNI-67/03 in the 3\' part of the genome, while the SimPlot analysis shows similarity to the other HPeV6 isolates between positions 4000 and 5500. This suggests that, unlike the other completely sequenced viruses, the HPeV6 isolates preserve remnants of a genome relatively similar to HPeV1 Harris, even though this virus was isolated more than 50 years previously.
Human parechovirus evolution and conservation

Cluster 1A

Cluster 1B

http://vir.sgmjournals.org 1705

http://vir.sgmjournals.org

Downloaded from www.microbiologyresearch.org by IP: 54.70.40.11 On: Fri, 30 Nov 2018 05:28:36

Downloaded from www.microbiologyresearch.org by IP: 54.70.40.11 On: Fri, 30 Nov 2018 05:28:36

http://vir.sgmjournals.org
Suppression of synonymous site variation

The availability of 17 genome sequences for HPeVs provides significant information that can be mined to elucidate aspects of the biology and evolution of these viruses. The complete genome sequences were firstly analysed for synonymous sequence variability by measuring mean pair-wise distances at each codon, using the

![SimPlot analysis of PicoBank/HPeV1/a compared with the other 16 complete HPeV sequences (a) and of complete HPeV sequences using HPeV1 Harris as reference (b). A window of 400 and a step of 10 nt were used.](image-url)
Simmonic sequence editor (Simmonds & Smith, 1999) (Fig. 3). There was little conservation of synonymous codons utilized across most of the genome, although the predicted HPeV cis-acting replication element (cre), located in VP0, was observed as a trough in codon variability in the analysis. However, the 3D protein-encoding sequence shows a marked suppression of variability. To investigate this further, 61 available partial 3D-encoding sequences were analysed in the same way. This larger dataset confirmed the constrained variability and highlighted two regions of low variation, suggesting that the RNA sequence contains information in addition to its coding capacity. There are also two strongly conserved occurrences of CG dinucleotides (due to CGX codons), although these are not in the most constrained areas (Fig. 3b). CGX codons are relatively rare elsewhere in HPeV genomes.

![Fig. 3. Synonymous codon variation across the whole HPeV coding region for 17 complete genomes (a) and across part of the 3D-encoding region (shown by the bracket in Fig. 3a) of 61 HPeV sequences (b). Nucleotide positions are shown on the scale underneath in each case. Regions of low variability are indicated by bars. The positions of CG dinucleotides in each of the 61 sequences are mapped below the scan of the 3D-encoding region and the position of the stems (black) and kissing interaction (grey) of the predicted RNA structure are also shown. Two highly conserved CGX codons are marked with triangles.](image)

![Fig. 4. Predicted RNA secondary structure and tertiary, kissing interaction (indicated by arrows) of the HPeV1 Harris strain, between positions 7008 and 7125 (part of the 3D-encoding region). The structure was identified using the default parameters of Mfold, except that a value of 20% suboptimality was employed. The secondary structure shown was the most energetically favourable of the predicted structures obtained (AG = -20.40 kcal mol⁻¹ (≈ -85.35 kJ mol⁻¹)) and the potential tertiary structure interaction between two of the loops was observed by eye. Numbers and the symbols > or < adjacent to nucleotides indicate the occurrence and frequency of variations in that position in the 61 sequences. Invariant nucleotides among the 61 available sequences are shaded grey and base-pairs supported by covariance and A-U/G-U/G-C variation are indicated by white type on a black background. Non-covariant nucleotides are indicated with an asterisk (*). The structure shown at the top of domain I is seen in only three strains and two slightly different structures, generated by relative slippage of the stem, as shown in the box, are seen in 43 and 15 of the strains, respectively. Only residues within the stem–loops are indicated as conserved or variant.](image)

A potential RNA structure in the 3D-encoding region

Mfold analysis of the 3D-encoding region predicted a secondary structure in the 3′-most end of the two
constrained regions (Fig. 4). This is made up of three stem–loops and visual analysis suggested that two of the loop regions have the capability of base-pairing to give a tertiary (‘kissing’) interaction. Analysis of all the available sequences is compatible with the predicted structure (Fig. 4). Although this region is, by definition, constrained in terms of sequence variation meaning that there is little covariance, the vast majority of any variation seen is consistent with the structure, as it either is fully covariant or involves A/G or C/U differences which maintain the structure. No phylogenetically conserved structure was predicted within the 5’-most constrained region.

**HPeV proteins**

A consensus of all the predicted cleavage sites in the completely sequenced HPeVs (Fig. 5) indicates that, as in several picornaviruses, there is a preference for glutamine (Q) at the P1 position (positions upstream of the cleavage site are termed P1, P2 etc. and those downstream are P1’, P2’ etc.). The P1 position is less constrained than often seen in picornaviruses. There is a strong preference for aspartic acid (D) or asparagines (N) at the P2 position. As in several picornaviruses, the P4 position is usually occupied by a small hydrophobic amino acid. An unexpected observation is the frequent occurrence of lysine (K) at the P7 position.

Alignment of the HPeV 2A sequences indicates that the reported H box, NC and tm (putative transmembrane) motifs are completely conserved (data not shown). These motifs are also highly conserved in the 2A proteins of LV and duck hepatitis virus (DHV) sequences, adding further significance to the idea that they are of key functional importance. These motifs are completely conserved (data not shown). These motifs are completely conserved (data not shown). These motifs are completely conserved (data not shown). These motifs are completely conserved (data not shown).

The other distinctive HPeV protein is VP3, which has an N-terminal extension in parechoviruses compared with all other picornaviruses. The previously reported strongly basic nature of this region is maintained in all the completely sequenced HPeVs (Fig. 6). The N-terminal 14–16 aa are diverse, but these are followed by a highly conserved motif of the form BXBXXBXXB [B= basic (K/R)]. In all the HPeVs this has the sequence KXXXRRXXK and this motif is also conserved in LVs. Another copy of the BXBXXBXXB motif (KXXXXKXXR) occurs upstream in VP3 in the HPeV5 isolates and interestingly a motif of the same form (KXXXXKXXR) is conserved in the HPeV 2A (data not shown).

**DISCUSSION**

Human parechoviruses are common pathogens with a seroprevalence exceeding 90% reported in some studies (Stanway et al., 2000). Several diagnostic laboratories now screen for HPeV infection and together with the retrospective analysis of archived material, a growing number of HPeV associations with disease have been reported recently (Baumgart et al., 2008; Benschop et al., 2006; Benschop et al., 2008b; Boivin et al., 2005; De Souza Luna et al., 2008; De Vries et al., 2007; Drexler et al., 2007; Harvala et al., 2008, 2009; Li et al., 2009; Tauriainen et al., 2008; Wakatsuki et al., 2007; Wolthers et al., 2008). A greater understanding of the biology and molecular biology of these viruses is therefore required. The sequence information on modern HPeV1 isolates is particularly sparse, given that this is the most commonly isolated type. As the number of sequences available increases, there are greater opportunities for a more detailed analysis, which should reveal facets of the evolution and molecular biology of these viruses. Here, we report the genome sequence of a recent HPeV isolate (PicoBank/HPeV1/a) obtained from a child recruited in the Finnish DIPP study and the VP1 and 3D sequences of another isolate (PicoBank/HPeV6/a) obtained from the same child 13 months later, together with an analysis of all the available complete sequences and partial VP1 and 3D-encoding sequences.

Analysis of the VP1 region of HPeV1 isolates shows that there are two predominant clusters, for which we propose the terms Cluster 1A and Cluster 1B (Fig. 1a). These both contain viruses from the period 2001–2006 and include isolates from The Netherlands and Germany, while Cluster 1A also contains isolates from Finland. The Finnish strain
PicoBank/HPeV1/a belongs to Cluster 1A, as does the other completely sequenced recent HPeV1 isolate BNI-788St, while the other completely sequenced HPeV1 isolate, 7555312, belongs to Cluster 1B. The results indicate that these two predominant genotypes were co-circulating, at least in Northern Europe, for several years. There are also several more diverse HPeV1 isolates from this period, indicating a complex pattern of HPeV1 genotypes. This may be related to the extremely high incidence of infection shown by seroprevalence studies and possibly reflects an endemic spread rather than geographically or temporally related epidemics. One of the diverse HPeV1 isolates is the prototype Harris, which is distinct from all the other HPeV1 strains in the analysis. Nucleotide sequences were chosen for these comparisons, as they are more informative than amino acid sequences when isolates are closely related. Amino acid trees give similar results, although branches are more compressed (data not shown). In both nucleotide and amino acid analyses, Harris clusters with the other HPeVs, albeit loosely and there currently seems to be no reason to consider reclassification of HPeV1 isolates into different types.

Recent reports have started to show that, in common with other picornaviruses, notably the common human pathogens enteroviruses, recombination is an important evolutionary mechanism in HPeVs (Benschop et al., 2008a; Lindberg et al., 2003; Lukashev et al., 2003; Oberste et al., 2004; Santti et al., 1999; Simmonds & Welch, 2006). By using a larger panel of VP1 and 3D sequences than previously possible, it is clear that the trees based on the relationships within these regions are not congruent and this is consistent with extensive recombination during divergence of the isolates (Fig. 1). Sharp transitions in nucleotide similarity observed by SimPlot analysis are also consistent with recombination (Fig. 2). In the case of PicoBank/HPeV1/a, loss of sequences closely related to the older HPeV1 Harris isolate and gain of sequences more related to HPeV4 and HPeV3 isolates can be mapped relatively precisely (Supplementary Fig. S2). The nucleotide positions identified correspond to 16 aa before the VP1 C-terminus (HPeV4 sequences) and position 14 in 2A (HPeV3 sequences) i.e. both close to the VP1/2A boundary. It is potentially significant that recombination can occur within the VP1 protein itself, particularly as the RGD motif, which plays a role in receptor binding (Boonyakiat et al., 2001), is located close to the VP1 C-terminus and downstream of the putative HPeV4 recombination point. Differences in this region generated through recombination could potentially alter the integrin recognized and thus HPeV biology. Moreover, imprecise recombination could have been involved in the generation of the HPeV3/HPeV7/HPeV8 lineages, which lack RGD motifs through apparent deletions in this region (Ito et al., 2004), as well as HPeV1 and HPeV5 isolates which lack RGD motifs (Benschop et al., 2008b).

An analysis of synonymous codon variation in PicoBank/HPeV1/a and the other complete HPeV genomes shows little constraint across most of the genome, although the putative cre correlates with a minor region of low diversity (Fig. 3a). In contrast, in parts of 3D, some codons are perfectly or almost perfectly conserved among all 61 sequences, suggesting functions in addition to protein-coding (Fig. 3b). Through Mfold analysis, we predict that part of this region folds into a complex structure, made up of three stem–loops, with a tertiary structure (‘kissing’) interaction between two of the loops. The high degree of conservation means that there is limited covariance support for this structure, but virtually all the variation seen is covariant, providing evidence that the structure is functionally significant (Fig. 4). The data presented here, together with the location of the cre in the coding region in most picornaviruses and with thermodynamic and phylogenetic evidence for RNA structures in other viruses, highlight the potential significance of RNA structures in coding regions (Mathews et al., 1999; Tuplin et al., 2002). Predicted structures in caliciviruses have been tested recently and their disruption affects virus replication (Simmonds et al., 2008). It would be interesting to perform analogous manipulations of the 3D structure to understand the reasons for its strong conservation.

Fig. 6. Alignment of the N terminus of VP3 in the completely sequenced HPeVs. Lysine (K) and arginine (R) residues are shaded in grey. The conserved BXBXXBXB motif seen in all HPeVs is indicated by a double-headed arrow, while a motif of the same form seen in HPeV5 isolates is indicated by a dotted double-headed arrow.
In addition to this putative structure, there is suppression of synonymous codon variation in an adjacent region where we have been unable to predict a significant secondary structure. There is also almost perfect conservation of two CGX (arginine) codons in the 3D region, while there is a very strong preference for AGR arginine codons (typically >90% of arginine codons are AGR in HPeVs) throughout the rest of the genome. Rare codons are often correlated with low tRNA abundance and have been proposed to affect folding of protein domains due to pausing of translation, and this could be the reason for their conservation at this location (Buchan & Stansfield, 2007). Conservation in the 3D region may also represent functions such as protein binding. Thus, information located in the 3D-encoding region may play a major regulatory role in HPeV replication.

The increasing number of complete HPeV genome sequences also allows a more complete analysis of the proteins encoded. The processing sites in the polyprotein seem to be defined by similar patterns to those used in other picornaviruses, for instance Q at the P1 position and a small hydrophobic amino acid at the P4 position (Birtley et al., 2005). There is a preference for D or N at P2, which may also help to define cleavage site specificity. A preference for K at P7 seems rather distal from the cleavage site, but could be an additional determinant (Fig. 5). HPeV proteins 2A and VP3 have distinct features compared with those of other picornaviruses. The Hbox and NC motifs identified in 2A are perfectly conserved in all HPeVs, providing evidence that the observed similarity to a family of human proteins involved in cell proliferation is functionally significant. The VP3 protein has an N-terminal extension relative to other picornaviruses (Stanway et al., 1994). The first part of this is relatively variable in the HPeVs (Fig. 6) but there is a well-conserved motif in both HPeVs and the animal parechovirus LV of the form BXBXBXB (where B = K or R). This does not correspond to a known motif, but it is reminiscent of nuclear/nucleolar localization signals, which often have clusters of basic residues. It could also interact with the HPeV RNA. The only other occurrence of this motif in HPeVs is in 2A, which has been identified as an RNA-binding protein that could use the motif for this function (Samuilova et al., 2004).

In summary, the limited molecular information available and the increasing number of reports on the clinical significance of HPeVs indicate that further work on these pathogens is required. The data we present here significantly increase our understanding of the HPeV genome and its evolution. This should underpin further analyses of the biology of HPeVs and the molecular features associated with their high prevalence and pathogenesis.

ACKNOWLEDGEMENTS

This study was supported by grants from the Wellcome Trust, EC (VIRDIAB project, Quality of Life Programme, KA2, Contract number QLK 2-CT-2001-01910), the Juvenile Diabetes Research Foundation, Academy of Finland, and competitive research funding of the Pirkkamaa Hospital District. We acknowledge the important contribution of the DIPP study, especially Professors Olli Simell, Mikael Knip, Jorma Ilonen and Riitta Veijola, as well as all the families taking part in the DIPP study.

REFERENCES


