Journal of General Virology (2010), **91**, 1229–1238 DOI 10.1099/vir.0.018747-0

Recombination dynamics of human parechoviruses: investigation of type-specific differences in frequency and epidemiological correlates

J. Calvert,[1] T. Chieochansin,[1,2] K. S. Benschop,[3] E. C. McWilliam Leitch,[1] J. F. Drexler,[4,5] K. Grywna,[4] H. da Costa Ribeiro, Jr,[6] C. Drosten,[4] H. Harvala,[7] Y. Poovorawan,[2] K. C. Wolthers[3] and P. Simmonds[1]

Correspondence
P. Simmonds
Peter.Simmonds@ed.ac.uk

[1]Centre for Infectious Diseases, University of Edinburgh, Summerhall, Edinburgh, UK

[2]Center of Excellence in Clinical Virology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

[3]Laboratory of Clinical Virology, Academic Medical Center, Amsterdam, The Netherlands

[4]Institute of Virology, Bonn, Germany

[5]Infectious Diseases Research Laboratory, University Hospital Professor Edgard Santos, Federal University of Bahia, Salvador, Brazil

[6]Department of Paediatrics, Hospital Professor Edgar Santos, Federal University of Bahia, Salvador, Bahia, Brazil

[7]Specialist Virology Laboratory, Royal Infirmary of Edinburgh, Edinburgh, UK

Received 26 November 2009
Accepted 13 January 2010

Human parechoviruses (HPeVs) are highly prevalent RNA viruses classified in the family *Picornaviridae*. Several antigenically distinct types circulate in human populations worldwide, whilst recombination additionally contributes to the genetic heterogeneity of the virus. To investigate factors influencing the likelihood of recombination and to compare its dynamics among types, 154 variants collected from four widely geographically separated referral centres (UK, The Netherlands, Thailand and Brazil) were typed by VP3/VP1 amplification/sequencing with recombination groups assigned by analysis of 3Dpol sequences. HPeV1B and HPeV3 were the most frequently detected types in each referral region, but with marked geographical differences in the frequencies of different recombinant forms (RFs) of types 1B, 5 and 6. HPeV1B showed more frequent recombination than HPeV3, in terms both of evolutionary divergence and of temporal/geographical indicators of population separation. HPeV1 variants showing between 10 and 20 % divergence in VP3/VP1 almost invariably fell into different recombination groups, compared with only one-third of similarly divergent HPeV3 variants. Substitution rates calculated by BEAST in the VP3/VP1 region of HPeV1 and HPeV3 allowed half-lives of the RFs of 4 and 20 years, respectively, to be calculated, estimates fitting closely with their observed lifespans based on population sampling. The variability in recombination dynamics between HPeV1B and HPeV3 offers an intriguing link with their markedly different seasonal patterns of transmission, age distributions of infection and clinical outcomes. Future investigation of the epidemiological and biological opportunities and constraints on intertypic recombination will provide more information about its influence on the longer term evolution and pathogenicity of parechoviruses.

INTRODUCTION

Human parechovirus (HPeV) is a member of the large and expanding family of positive-stranded RNA viruses, the *Picornaviridae*. Picornaviruses infect humans and a range of mammalian species, typically targeting the gastrointestinal or respiratory tracts with a broad range of disease manifestations and severities. This variability in host interaction is manifested very clearly among human parechoviruses. These

The GenBank/EMBL/DDBJ accession numbers for the sequences obtained in this study are GU946681–GU946974 (see Supplementary Table S1, available in JGV Online).

A full list of the locations, sample dates and sources for the sequences used in this study is available with the online version of this paper.

viruses were originally described as echovirus 22 and 23 and classified with human enteroviruses to which they show similarities in clinical presentations and an enterovirus-like cytopathology on virus isolation. However, sequencing of the genome of HPeV has revealed distinct structures of replication and translation elements, substantial sequence divergence of encoded proteins and frequent differences in biological properties (Hyypia *et al.*, 1992; Stanway *et al.*, 1994). Consequently, they have now been renamed and reclassified into their own genus, *Parechovirus* (Stanway *et al.*, 2005).

HPeV is a highly prevalent, enterically transmitted virus, usually causing clinically unapparent infections primarily among young children. However, HPeV has been associated with gastrointestinal and respiratory tract symptoms, as well as with occasional cases of encephalitis and flaccid paralysis (Figueroa *et al.*, 1989; Koskiniemi *et al.*, 1989; Stanway *et al.*, 2000; Joki-Korpela & Hyypia, 2001). Although differences in host response, and in particular the greater susceptibility of neonates to disseminated sepsis-like illness, account for some of the variability in the outcome of infections, biological differences between variants of HPeV may also contribute significantly to its wide spectrum of clinical presentations.

Human parechoviruses show genetic and antigenic heterogeneity, and a number of distinct HPeV types are known to circulate widely in human populations throughout the world. There are currently a total of eight classified HPeV types and an additional six for which partial genome sequences are available. HPeV types differ from each other by 30–40 % in the primary nucleotide sequence and are antigenically distinct. Indeed, the first two types discovered were classified as separate serotypes (echovirus 22 and 23) on the basis of cross-neutralization assays (Wigand & Sabin, 1961). More recently, several new types in association with a variety of disease presentations (pyrexia, enteritis, Reye's syndrome) have been isolated and genetically characterized; these include type 3 (HPeV3) (Ito *et al.*, 2004; Boivin *et al.*, 2005), HPeV4 (Benschop *et al.*, 2006a), HPeV5 (Oberste *et al.*, 1998; Al Sunaidi *et al.*, 2007) and HPeV6 (Watanabe *et al.*, 2007), whilst molecular methods have been used to detect HPeV7 (Li *et al.*, 2009) and HPeV8 (Drexler *et al.*, 2009). There are unpublished reports of five additional HPeV types from Bangladesh and one from The Netherlands (Benschop *et al.*, 2008a) based on VP1 sequence comparisons, and these have now been provisionally assigned as types 9–14 (http://www.picornastudygroup.com/types/parechovirus/hpev.htm). It has been problematic to firmly attribute specific disease outcomes from infections with different HPeV types. However, recent investigations have convincingly implicated HPeV3 as a cause of a sepsis-like illness in young children (Abed & Boivin, 2005; Boivin *et al.*, 2005; Benschop *et al.*, 2006b; van der Sanden *et al.*, 2008; Harvala *et al.*, 2009).

The evolutionary history and timescale for the emergence of HPeV different types remain uncertain. In common with other picornaviruses and frequently observed in other mammalian RNA viruses, parechoviruses show extremely rapid sequence change over time, and dates as recent as 400 years ago for the origin of the different types have been proposed (Faria *et al.*, 2009). At least two evolutionary processes occur in the capsid-encoding genes; population immunity is probably a potent factor underlying the differentiation of HPeV into serologically distinct types, whilst shorter-term sequence change, as observed in other picornaviruses, reflects a process of neutral sequence drift with substitutions occurring predominantly at synonymous sites (Simmonds, 2006; Bailly *et al.*, 2009). The pattern of sequence diversity in the non-structural (NS) region of parechoviruses is different, and there is little genetic linkage with the structural gene region; the grouping of parechoviruses into types evident from analyses of capsid genes breaks down almost completely in the NS region and the 5′-untranslated region (5′UTR) (Benschop *et al.*, 2008b, 2009; Williams *et al.*, 2009; Zoll *et al.*, 2009), an evolutionary pattern shared among almost all genera of picornaviruses (Minor *et al.*, 1986; Santti *et al.*, 1999; Lukashev, 2005; Heath *et al.*, 2006; Simmonds, 2006). A recent large-scale analysis of complete genome sequences of HPeV identified a number of recombination breakpoints in the genomes of HPeV1, -3, -4, -5 and -6 clustering around the P1/P2 junction and between P1 and the 5′UTR (Benschop *et al.*, 2009). Frequent recombination events in these regions and elsewhere in non-structural genes underlie the differences in phylogenetic groupings in different parts of the HPeV genome, although how and why this occurs remains uncertain.

In the current study, we have created a large dataset of 154 new paired sequences from the structural (VP3/VP1) and non-structural (3Dpol) regions of HPeV variants circulating over the last 5 years in four widely geographically spaced referral centres in Europe, Southeast Asia and South America. Combined with previously published sequences from HPeV variants circulating up to 15 years ago in these and other geographical regions, we have been able to conduct a large-scale investigation of the dynamics of recombination in different HPeV types, and the evolutionary and epidemiological factors underlying its occurrence. The findings of substantial differences in substitution rate and recombination frequency of HPeV1 and HPeV3 provide tantalizing new insights into their differing molecular epidemiologies and clinical associations of infection.

RESULTS

HPeV type and 3Dpol clade assignments

Phylogenetic analysis of the VP3/VP1 region of the 170 study sequences identified 105 type 1, 41 type 3, six type 4, eight type 5, seven type 6, two type 10 and one type 14 sequences from the four study centres (Fig. 1a). The type 1 sequences could be further assigned into three type 1A and

102 type 1B sequences as proposed previously (Benschop *et al.*, 2009), with the type 1B sequences used for the recombination analysis described in the rest of the study. Type 1B was the most common type detected in all four referral centres, followed by type 3 and more variable detection of types 4, 5 and 6. Three variants from Thailand additionally were assigned as types 10 and 14 (see Supplementary Table S1, available in JGV Online). For each sample, sequences from the 3Dpol region were obtained and phylogenetically analysed to identify recombination groups (Fig. 1b). 3Dpol groupings were assigned by identifying bootstrap-supported clades. Groups identified in this manner showed a minimum of 5.5 % nucleotide sequence divergence from each other, a threshold corresponding to a naturally occurring minimum value in the distribution of pairwise distances between 3Dpol sequences (Fig. 2). This definition of clades corresponds to that used previously for assigning recombination groups (Benschop *et al.*, 2008b). However, inspection of the distribution and of the phylogenetic tree revealed a second level of variability, with several clades being much more divergent from each other with a threshold divergence of approximately 0.155 dividing them. However, the distribution (and tree) was dominated by the large number of type 1 sequences, and the rather few available pairwise distance values within genotypes 4-, 5- and 6-associated clades (filled bars, labelled 'other' in Fig. 2) indicated frequent distances at the low point of the type 1 distribution (different clades). For the analysis in the rest of the paper, sequences were considered to be in the same or different clades based on the lower threshold value.

The 170 study sequences could be assigned to a total of 35 different clades for the main 1B, 3 and 4–6 types (Figs 1 and 3) and a further four for the rarer types (AV, D, T in type 1A variants and U in type 14, whilst the type 10 variants were BA; Fig. 1). Most 3Dpol clades were associated with specific genotypes (grey-filled boxes in Fig. 3); for example, 25 were exclusively or most commonly associated with type 1B variants analysed in the study and three with genotype 3. However, some 3Dpol groups were associated with multiple types, including the finding of BD in genotypes 1B and 6, J with genotypes 1B and 4, A with genotypes 1B and 3, and BA with genotypes 1B, 5 and 10.

There were marked geographical differences in the distribution of different recombinant forms (RFs). Although type 1B sequences were the most commonly detected type in all four centres, variants from Thailand and Brazil showed almost entirely distinct distributions of RFs from those of the UK and The Netherlands. Even the latter two countries showed substantial differences in 3Dpol clade frequencies (despite sampling over a similar study period) with only clades D and AD from the 21 genotype 1-associated clades in the two countries overlapping in geographical distribution. On the other hand, type 3 sequences were almost invariably in clade A in both countries. The number of clades per sequence provides an approximate measure of the diversity of each HPeV type: values of 0.27 and 0.53, respectively, for type 1 and types 4–6 combined were substantially higher than the 0.071 for type 3, where 41 sequences fell into only three clades.

Differences in recombination frequency among HPeV types

The evidence for less frequent recombination in HPeV3 compared with other HPeV types was formally investigated by separating out epidemiological and evolutionary factors that may also influence recombination frequency (Fig. 4). An indicator of the period of evolutionary separation of variants was the degree of accumulated sequence divergence in the VP3/VP1 region of the genome. Comparison of very recently diverged type 1 variants (pairwise distances in VP3/VP1 of <0.005 from each other) revealed a low proportion of pairwise comparisons between sequences with 3Dpol sequences assigned to different groups (16 %). In contrast, the more divergent type 1 sequences invariably contained different clades from each other, providing evidence for inevitable recombination in this extended period of divergent evolution. Dividing evolutionary distances into six bins revealed a consistent relationship between divergence and the occurrence of recombination (Fig. 4a).

There was also a relationship between sequence divergence and recombination frequency in type 3 sequences, but recombination occurred after a substantially longer period of divergent evolution; for sequences showing between 5 and 10 % divergence, approximately 30 % were recombinant, compared with almost invariable recombination between type 1 variants in this divergence range. Insufficient numbers of sequences of types 4–6 were available to include them in the analysis, but the high proportion of RFs relative to the number of variants sequenced (Fig. 3) suggested frequent recombination comparable to that of type 1.

To provide a time calibration for this analysis, substitution rates for this region of VP3/VP1 were determined by a Bayesian maximum-likelihood procedure for HPeV types 1B and 3 (Table 1). Variability in this region occurred overwhelmingly at synonymous sites (d_N/d_S ratios of 0.0063 and 0.0037), fitting the assumption of neutral sequence divergence required by the analysis method. The substitution rate for HPeV1B was 7.2×10^{-3} substitutions per site per year, over 2.5 times faster than HPeV3 (2.8×10^{-3} substitutions per site per year). Despite the greater diversity within the type 1B clade, the most recent common ancestor (MRCA) of this type was approximately contemporary to that of HPeV3 (24.0 and 22.7 years ago, respectively). Both viruses thus have very recent predicted times of emergence and spread. Applying these rates to compare recombination dynamics between these types, we estimated the divergence of VP3/VP1 sequences at 50 % recombination points for HPeV1 and HPeV3 (where half of pairwise comparisons between variants within each type shared the same 3Dpol clade; Fig. 4a). For HPeV1, this fell

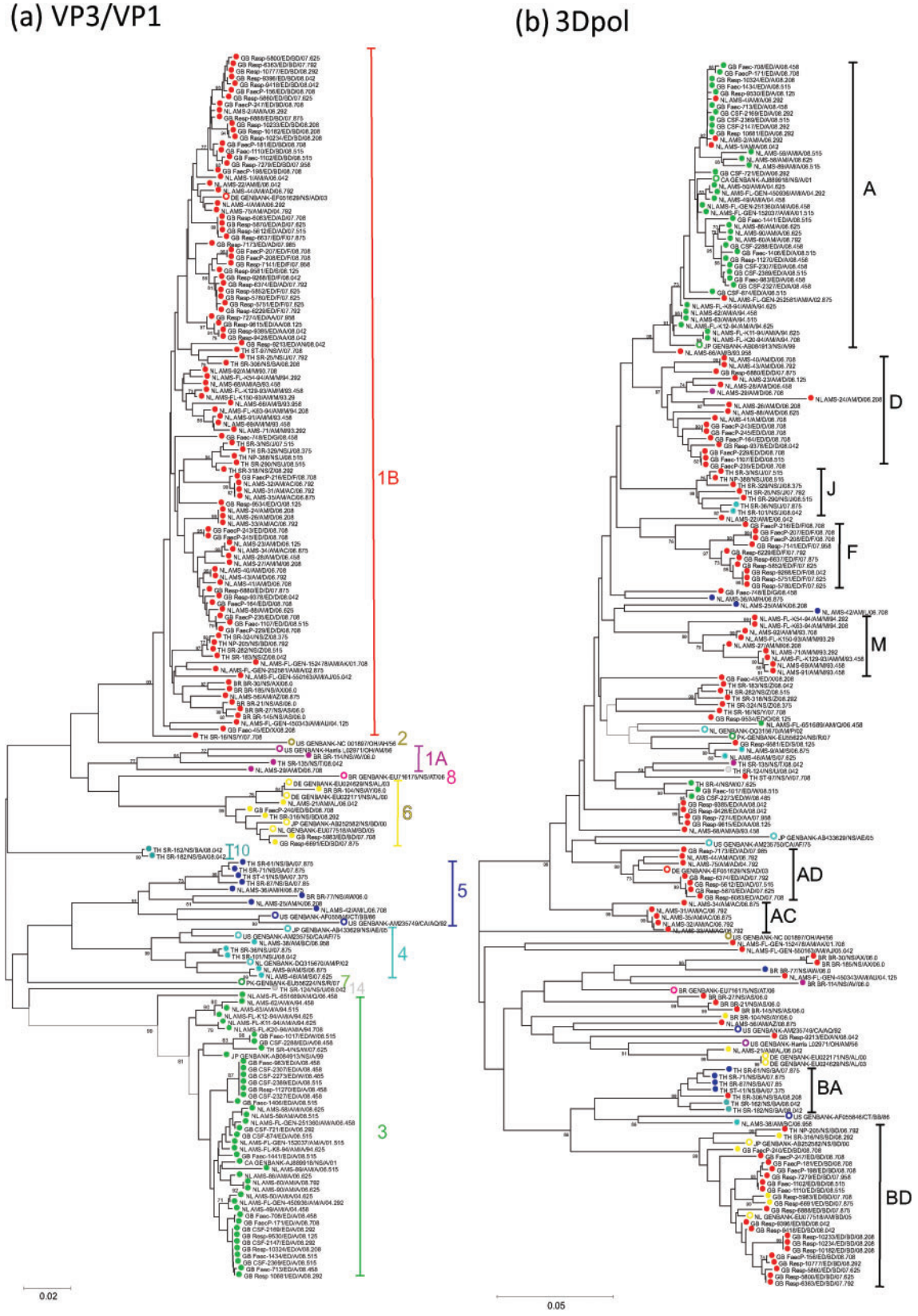
(a) VP3/VP1
(b) 3Dpol
1B
1A
2
8
6
10
5
4
14
3
A
D
J
F
M
AD
AC
BA
BD
0.02
0.05

Fig. 1. Phylogenetic comparisons of the VP3/VP1 region (a) and 3Dpol sequences (b) from the study samples (filled circles, coloured by type in both trees), and all available non-identical published sequences with sequence data in both genome regions (labelled by GenBank accession number; unfilled circles). Each sequence is labelled by country of origin, sample type, city abbreviation, 3Dpol clade assignment and decimalized year of collection. Trees were constructed by neighbour-joining of MCL (nucleotide) pairwise distances, with bootstrap resampling to demonstrate robustness of groupings; values $\geqslant$70 % are shown. For reasons of clarity, only the larger clades in 3Dpol (containing five or more members) have been labelled. Evolutionary distances are scaled according to the bars in the lower left-hand corner of each tree (number of nucleotide substitutions per site).

in the divergence range of 0.05–0.1 distance range (estimated at 0.06 by interpolation). Using the substitution rate for HPeV1 (Table 1), this predicted a 'half-life' of an RF of 4.1 years [3.1–6.3 years using the highest posterior density (HPD) interval of the substitution rate]. The greater 50 % divergence point (estimated at 0.125; Fig. 4a) and the slower substitution rate for HPeV3 (Table 1) combined to predict a much longer half-life than type 1 (22 years; range 15–37 years).

Epidemiological markers of separation (sample year difference and geographical distance) showed similar differences in recombination frequency between type 1 and 3 variants (Fig. 4b, c). For type 1, there was evidence of frequent co-circulation of different RFs within the same city and among samples collected in the same year (72 and 79 %, respectively, for recombination frequencies on pairwise comparison). These frequencies approached 100 % for temporally and geographically more widely spaced samples. There was, in marked contrast to HPeV1, evidence for the prolonged circulation of RFs of HPeV3, with only one-third of comparisons of variants sampled over 10 years apart showing different recombination groups (Fig. 4b). This observation is consistent with the long (22-year) half-life for RFs predicted for HPeV3 using divergence and substitution rates.

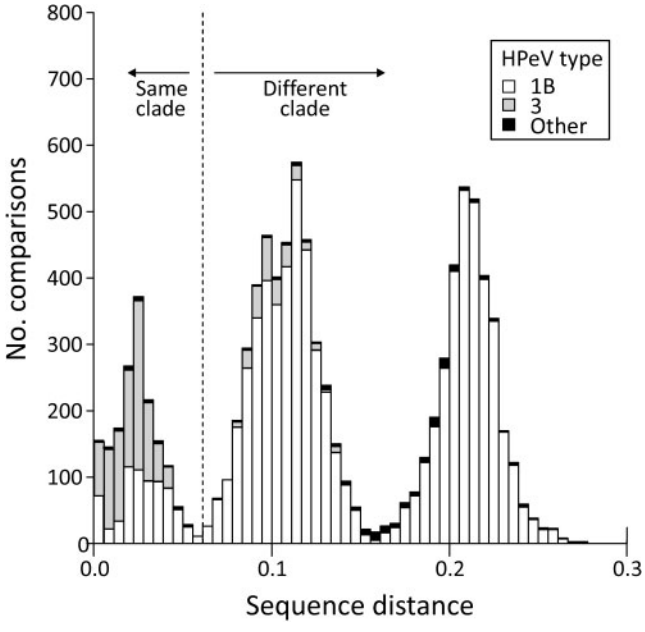

Fig. 2. Range of pairwise distances (uncorrected p distance) between VP3/VP1 sequences of genotypes 1B and 3 and of other genotypes, shown as a stacked histogram. The threshold value of 0.055 that divides comparisons between members of the same (phylogenetically defined) clades and between different clades is shown as a dashed line.

DISCUSSION

Recombination in parechoviruses

In agreement with previous studies (Benschop *et al.*, 2008b, 2009; Williams *et al.*, 2009; Zoll *et al.*, 2009), the uncoupling of phylogeny relationships between the structural and non-structural gene regions (Fig. 1) provides convincing evidence for the occurrence of recombination in parechoviruses. Although we have not sought to identify the position between VP3/VP1 and 3Dpol where the recombination events occurred, the marked differences in phylogenetic groupings between genome regions, and in particular the dispersal of monophyletic groups in VP3/VP1 (corresponding to HPeV types) into a plethora of intermingled clades in the 3Dpol region, cannot plausibly be explained by any other mechanism. The discordance in phylogenetic relationships across the HPeV genome resembles that of many other picornaviruses. For example, our previous phylogenetic analysis of 3Dpol sequences from a worldwide collection of 318 isolates of the human enterovirus, echovirus 30, revealed the existence of 38 phylogenetically distinct clades, interspersed with sequences from other species B enterovirus serotypes (McWilliam Leitch *et al.*, 2009).

The findings are consistent with previous descriptions of the mosaic nature of HPeV complete genome sequences and the mapping of recombination breakpoints most commonly to the 2A region (Benschop *et al.*, 2009; Williams *et al.*, 2009; Zoll *et al.*, 2009), and observation of inconsistent phylogenies between VP1 and 3Dpol regions of HPeV1 but not HPeV3 variants (Benschop *et al.*, 2008b). With the much larger dataset in the current study, evidence has been obtained for recombination in all five of the commonly circulating HPeV types in the four study countries and among published sequences.

Identification of RFs

Designation of HPeV RFs was dependent on the identification of distinct 3Dpol clades. In the current study, we have formalized the approach adopted in a previous investigation of incompatibility between VP1 and 3Dpol

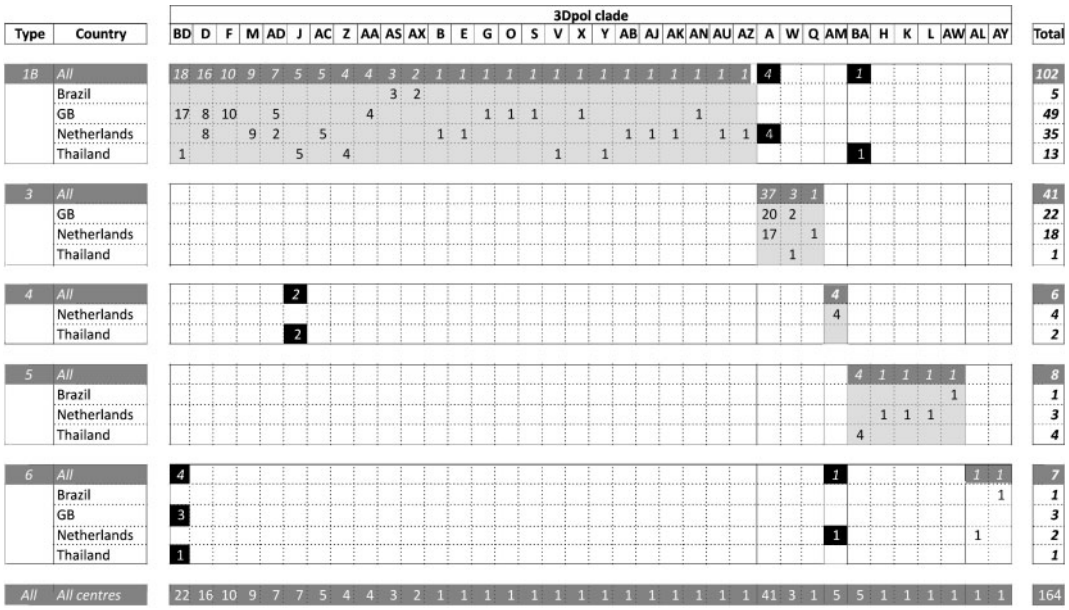

Fig. 3. Assignment of samples from the four study centres into types 1 and 3–6 (left hand columns) and recombination groups (BD–AY) based on phylogenetic analysis of the 3Dpol region (totals shown). Groups have been ordered by frequency of occurrence (from left to right); shaded grey boxes indicate 3Dpol groups associated with each type. Variants with 3Dpol groups more commonly found in another type are indicated by filled boxes.

regions (Benschop *et al.*, 2008b) through construction of bootstrap replicated trees and in analyses of the distribution of pairwise distances between sequences in this genomic region (Figs 1b and 3). Phylogenetic analysis revealed a strongly hierarchical structure with a series of approximately equally divergent, bootstrap-supported clades used for classification of HPeV variants into RFs. Pairwise distances between groups were typically in the range of 0.08–0.13, distinct from intra-group distances of 0.02–0.04. At this stage, it is difficult to hypothesize what

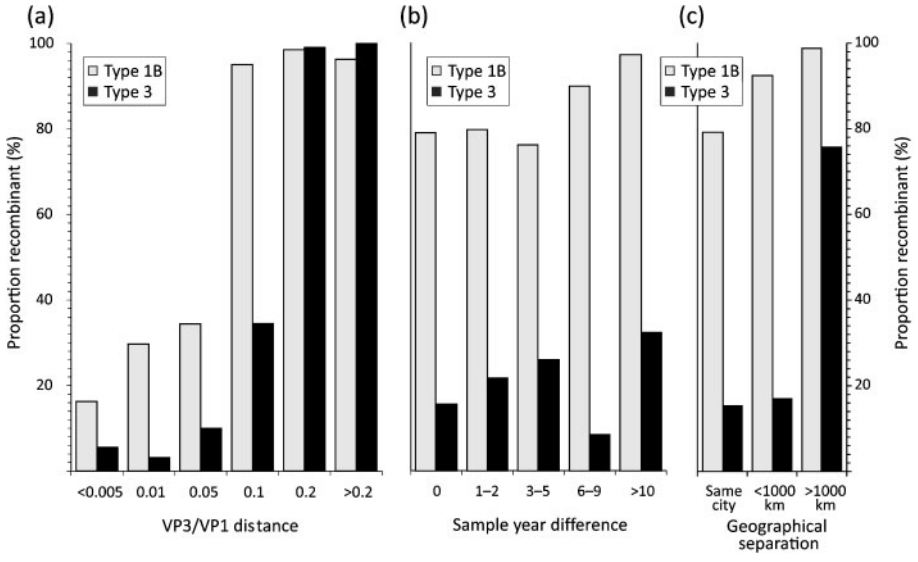

Fig. 4. Evolutionary and epidemiological factors influencing the frequency of recombination in HPeV1B and HPeV3. Pairwise comparison of the study sequences and other available HPeV sequences recorded the mean frequency of recombination in the different categories of sequence divergence (a), time between sampling (b) and geographical separation of the samples (c).

Table 1. Estimates of rates of sequence change of the VP3/VP1 region in HpeV1B and HPeV3

Type	Divergence*	d_N/d_S*	Substitution rate ($\times 10^{-3}$)†	MRCA†
HPeV1B	0.097	0.0063	7.24 (4.75–9.61)	1985 (1980–1990)
HPeV3	0.033	0.0037	2.83 (1.69–4.17)	1987 (1980–1992)

*d_N/d_S, Ratio of non-synonymous to synonymous substitutions per site, calculated using Jukes–Cantor corrected distances.
†Mean value with HPD interval in parentheses.

evolutionary events lie behind the tree structures of the 3Dpol regions of HPeV or human enteroviruses, their timescale for divergence or the underlying basis for their disjoint relationship with structural gene sequences. Nevertheless, the identification of RFs made possible by clade assignments provides the means to identify recombination events and differentiate clearly between different circulating virus populations.

3Dpol groups associated with one genotype may be observed in others, particularly in those occupying the same geographical region. Thus, type 1 and type 6 variants in the Edinburgh samples frequently shared 3Dpol groups and with these types in Thailand. A similar occurrence was observed for HPeV3, where the majority of samples had a group A 3Dpol sequence, one that spilled over into Dutch type 1 variants. These and other observations of intertypic recombination suggest that gene exchange between any pair of HPeV types can potentially create viable RFs. However, the extent to which hybrid viruses after a recombination event require further adaptive changes to gain full wild-type fitness and transmissibility remains unclear. Direct observational data on the recombination process and information on the effective population sizes and selection pressures operating on parechoviruses in nature are currently largely or entirely lacking.

Factors influencing the likelihood of recombination

The main correlate determining the likelihood of recombination was sequence divergence in VP3/VP1 (Fig. 4a). Divergence in this structural gene region provides an indication of the length of time any pair of viruses has been evolving apart from their last common ancestor. Unexpectedly, however, there was a marked difference in substitution rates calculated by BEAST between types 1 and 3 that complicated the calculations of the lifespans of their RFs. Of the two parechovirus types, the substitution rate for HPeV1 (7.2×10^{-3} substitutions per site per year) was more comparable to those determined previously in the structural gene regions for other picornaviruses, specifically human enteroviruses E30 (8×10^{-3}), EV71 (13.5×10^{-3}), poliovirus (9×10^{-3}) and EV70 (5×10^{-3}) (Takeda *et al.*, 1994; Kew *et al.*, 1995; Brown *et al.*, 1999; McWilliam Leitch *et al.*, 2009) and foot-and-mouth disease viruses (4×10^{-3}–14×10^{-3}; summarized in Cottam *et al.*, 2006). These contrast with the substantially lower substitution rates for parechoviruses determined in the P1 capsid and VP1 regions (2.2×10^{-3} and 2.8×10^{-3} substitutions per site per year; Faria *et al.*, 2009), closer to that determined for HPeV3 in the current study. However, this accuracy of the published estimate is potentially compromised by being based on a combined dataset of all HPeV types, each of which may show different substitution rates (as demonstrated for HPeV1 and HPeV3). Indeed, this may account for the substantial rate heterogeneity recorded among tree branches by BEAST analysis, the high coefficient of variation of 0.29 and 0.41 for the two genome regions (P1 and VP1) and rejection of a strict molecular clock (Faria *et al.*, 2009).

The mean half-lives calculated for RFs of HPeV1 and HPeV3 can only be regarded as approximate and are in any case likely to be variable and influenced by specific epidemiological circumstances. Nevertheless, the greater than fivefold difference in their predicted lifespans [4 and 22 years, respectively, with non-overlapping confidence (HPD) intervals] fits well with the epidemiological correlates of their recombination frequencies (Fig. 4b, c). In particular, observations that the same group A RFs of HPeV3 were detected throughout the period from 1994 to 2008 and the much more transient existence of HPeV1 RFs is consistent with these half-life predictions. The only comparable published analysis of recombination frequencies in picornaviruses described a 3–5-year turnover of individual RFs of the human enterovirus E30, based on longitudinal sampling in Europe (McWilliam Leitch *et al.*, 2009), and, using the same calculation method on the published data, a predicted half-life of 3.4 years, observations much more similar to HPeV1 than HPeV3. Similar also to E30 was the evidence for widespread geographical distributions of individual HPeV1 recombinants; the observation of substantial overlap in RF distributions between Edinburgh and Amsterdam suggests a shared transmission network that did not, however, extend to Thailand or Brazil. In the case of E30 (McWilliam Leitch *et al.*, 2009), RFs showed highly overlapping distributions in northern and western Europe (and Australia), whilst contemporary variants circulating in Southeast Asia were entirely distinct.

Although we do not understand the mechanisms underlying the differences in the evolutionary process of HPeV1 and HPeV3, and we have insufficient recombination data from other HPeV types for comparison, the marked differences

in rates and recombination frequencies offers an intriguing link with the known epidemiological and biological differences between these HPeV types. Recent data indicate different age distributions, seasonality and infection outcomes between HPeV1 and HPeV3 that may underlie or result from their differences in evolutionary dynamics. One example, and potentially connected to their difference in substitution rates, is the observation of a biannual pattern of infection in HPeV3, with infections almost entirely restricted to the summer months in even-numbered years in The Netherlands, the UK and probably elsewhere in Europe (Benschop *et al.*, 2006b, 2008a; van der Sanden *et al.*, 2008; Harvala *et al.*, 2009). In contrast, HPeV1 (and other types) recurs annually with peaks of incidence in the autumn (Benschop *et al.*, 2006b; van der Sanden *et al.*, 2008), similar to the seasonal pattern of human enteroviruses. HPeV3, along with the other less prevalent parechovirus types, may therefore experience much more extreme population contractions and periods of inactive environmental persistence during its transmission chain that would reduce its long-term substitution rate.

HPeV3 differs also in its clinical presentations, being specifically associated with neonatal sepsis and a potentially greater capacity to infect systemically and target the central nervous system (Abed & Boivin, 2005; Boivin *et al.*, 2005; Abed & Boivin, 2006; Benschop *et al.*, 2006b; van der Sanden *et al.*, 2008; Harvala *et al.*, 2009). It has been suggested that this difference may arise either through an absence of protective maternal antibody (although there is no data to show whether sepsis cases occur specifically in children of seronegative mothers) or through biological differences arising in the use of a different cellular receptor for entry. HPeV3 differs from HPeV1 and most other types by lacking the RGD motif in VP1 required for integrin binding (Ito *et al.*, 2004; Benschop *et al.*, 2008a), an entry mechanism shared with a number of other picornaviruses. It has indeed been suggested that the different cellular tropism conferred on type 3 through the use of alternative cellular receptors may be a significant factor limiting recombination with other HPeV types (Benschop *et al.*, 2009). Notwithstanding this, however, we have found evidence for recombination between HPeV3 and HPeV1 (several type 1 variants from Amsterdam had group A 3Dpol sequences that are found most commonly in HPeV3; Fig. 3). Direct *in vitro* investigation of the compatibility and constraints accompanying intertype recombination in parechoviruses would be of considerable value in resolving these unresolved biological issues.

In the longer term, investigations focusing on the relationship between the biology/pathogenesis of HPeV with its molecular epidemiology, recombination dynamics and evolution promise to provide important new insights into the nature of parechovirus infections and their prevention. This multidisciplinary approach will be of considerable value in the exploration of variability in circulation and outcome of other viral infectious diseases.

METHODS

Samples. HPeV-positive isolates were collected from diagnostic specimens referred to the Specialist Virology Laboratory, Royal Infirmary of Edinburgh, UK, and made available through the specimen archive. Samples comprised respiratory (n=37), faecal (n=26) and cerebrospinal fluid (CSF; n=10) specimen types. Further faecal and CSF samples were made available for the study from the Academic Medical Centre, Amsterdam, The Netherlands (n=48), and faecal and respiratory samples from Bangkok, Thailand (n=25). Sequences obtained from faecal samples from Brazilian subjects (n=8) and all available published nucleotide sequences from GenBank downloaded in August 2009 (31 in total) supplemented these data. A full list of the locations, sample dates and sources for the sequences used in the study is provided in Supplementary Table S1. Positive controls comprised serial dilutions of cell culture supernatant from two Dutch isolates of HPeV-1 (AMS-152478) and HPeV-3 (AMS-252277). RNA was extracted from controls and clinical samples using a Qiagen QIAamp Viral RNA kit, as described by the manufacturer.

Isolates were named using the following convention: two-letter country code/isolate number/two-letter city abbreviation/3Dpol clade/last two digits of the year of collection, followed by a decimalized fraction expressing the month of collection, if known. (e.g. GB/Faec-1406/ED/A/08.515 for faecal sample 1406 from Edinburgh, Great Britain (GB), isolated in July 2008 and belonging to the 3Dpol clade A). 3Dpol clade designations were based on the results of phylogenetic analysis of this region (see below). Cities of origin were used to calculate the geographical separation of samples.

Amplification of the VP3/VP1 and 3Dpol regions. The structural gene region was amplified as described previously (Harvala *et al.*, 2008). For the 3Dpol region, RT-nested-PCRs were performed using HPeV-specific primers designed to amplify a 700 nt region of the 3Dpol gene corresponding in position to that used in previous analyses of recombination (Benschop *et al.*, 2008b). Combined reverse transcription and first-round PCR used the Superscript III One-step RT-PCR system (Invitrogen) with the following first-round primers: 5′-GTNTAYARGATGATHATGATGGARA-3′ (outer sense, nt 6422–6446, numbered using the HPeV-1 Harris strain as a reference sequence; GenBank accession no. L02971) and 5′-YTTARTCAACACCATGGGCAYYA-3′ (outer antisense, nt 7253–7275). Thermal cycling for the first RT-PCR comprised heating at 43 °C for 1 h; 20 cycles of 53 °C for 1 min and 55 °C for 1 min; heating at 70 °C for 15 min and 94 °C for 2 min; 40 cycles of 94 °C for 30 s, 50 °C for 30 s and 68 °C for 1 min 45 s; and final extension at 68 °C for 5 min.

One microlitre of product was amplified in a second-round PCR using the following primers: 5′-GAYTGGCACTTYATGATYAAYGC-3′ (inner sense, nt 6512–6534) and 5′-ATNACMACWTCATA-ATCATCCAC-3′ (inner antisense, nt 7221–7243) with GoTaq DNA polymerase (Promega) in a 20 µl reaction volume with the following cycling conditions: 30 cycles of 94 °C for 18 s, 50 °C for 21 s and 72 °C for 90 s, with a final extension at 72 °C for 300 s and then at 25 °C for 600 s.

Amplicons were sequenced directly using a BigDye Terminator Cycle Sequencing kit (ABI) and the inner sense primers. Nucleotide sequences were annotated and aligned using the Simmonics sequencing package, version 1.8 (http://www.virus-evolution.org).

Evolutionary analysis. Phylogenetic trees were constructed by the neighbour-joining method from 1000 bootstrap samplings of maximum-composite-likelihood (MCL) distances, using the MEGA 4.0 software package with pairwise deletion for missing data. A Bayesian coalescent algorithm implemented in the Bayesian evolutionary

analysis by sampling trees (BEAST) software package version 1.5.2 (Drummond *et al.*, 2006) was used to estimate the rate of sequence change (Drummond & Rambaut, 2007). Dated sequence sets were run using a relaxed lognormal clock with a chain length of 10 million under the HKY/SRD06 model of substitution. The latter partitions codons into $(1+2)+3$ positions and parameters estimated separately for each (Shapiro *et al.*, 2006). All other parameters were optimized during the burn-in period. Virtually identical substitution rates and root heights were obtained using different coalescent priors, such as exponential growth, different substitution models (GTR) and different clocks (strict and relaxed – uncorrelated exponential; data not shown). Output from BEAST was analysed using the program TRACER (http://beast.bio.ed.ac.uk/Tracer).

ACKNOWLEDGEMENTS

The authors are grateful to Gillian Fewster and staff at the Microbiology Laboratory, Western General Hospital, Edinburgh, UK, for faecal surveillance samples and to Elly Gaunt, Kate Templeton and Carol Thomson for providing samples, data and other virus testing results from the respiratory sample archive. We would also like to thank Andrew Rambaut, Institute of Evolutionary Biology, University of Edinburgh, UK, for useful discussion. T. C.'s laboratory visit to Edinburgh was funded by the Royal Golden Jubilee PhD Program, the Thailand Research Fund, the Centre of Excellence in Clinical Virology, and the Graduate School, Chulalongkorn University, and the Commission on Higher Education, Ministry of Education, Thailand.

REFERENCES

Abed, Y. & Boivin, G. (2005). Molecular characterization of a Canadian human parechovirus (HPeV)-3 isolate and its relationship to other HPeVs. *J Med Virol* **77**, 566–570.

Abed, Y. & Boivin, G. (2006). Human parechovirus types 1, 2 and 3 infections in Canada. *Emerg Infect Dis* **12**, 969–975.

Al Sunaidi, M., Williams, C. H., Hughes, P. J., Schnurr, D. P. & Stanway, G. (2007). Analysis of a new human parechovirus allows the definition of parechovirus types and the identification of RNA structural domains. *J Virol* **81**, 1013–1021.

Bailly, J. L., Mirand, A., Henquell, C., Archimbaud, C., Chambon, M., Charbonne, F., Traore, O. & Peigue-Lafeuille, H. (2009). Phylogeography of circulating populations of human echovirus 30 over 50 years: nucleotide polymorphism and signature of purifying selection in the VP1 capsid protein gene. *Infect Genet Evol* **9**, 699–708.

Benschop, K. S., Schinkel, J., Luken, M. E., van den Broek, P. J., Beersma, M. F., Menelik, N., van Eijk, H. W., Zaaijer, H. L., VandenBroucke-Grauls, C. M. & other authors (2006a). Fourth human parechovirus serotype. *Emerg Infect Dis* **12**, 1572–1575.

Benschop, K. S., Schinkel, J., Minnaar, R. P., Pajkrt, D., Spanjerberg, L., Kraakman, H. C., Berkhout, B., Zaaijer, H. L., Beld, M. G. & Wolthers, K. C. (2006b). Human parechovirus infections in Dutch children and the association between serotype and disease severity. *Clin Infect Dis* **42**, 204–210.

Benschop, K. S., Thomas, X., Serpenti, C., Molenkamp, R. & Wolthers, K. (2008a). High prevalence of human parechovirus genotypes in the Amsterdam region and the identification of specific HPeV variants by direct genotyping of stool samples. *J Clin Microbiol* **46**, 3965–3970.

Benschop, K. S., Williams, C. H., Wolthers, K. C., Stanway, G. & Simmonds, P. (2008b). Widespread recombination within human parechoviruses: analysis of temporal dynamics and constraints. *J Gen Virol* **89**, 1030–1035.

Benschop, K. S., de Vries, M., Minaar, R., Stanway, G., van der Hoek, L., Wolthers, K. & Simmonds, P. (2009). Comprehensive full length sequence analyses of human parechoviruses: diversity and recombination. *J Gen Virol* **91**, 145–154.

Boivin, G., Abed, Y. & Boucher, F. D. (2005). Human parechovirus 3 and neonatal infections. *Emerg Infect Dis* **11**, 103–105.

Brown, B. A., Oberste, M. S., Alexander, J. P., Jr, Kennett, M. L. & Pallansch, M. A. (1999). Molecular epidemiology and evolution of enterovirus 71 strains isolated from 1970 to 1998. *J Virol* **73**, 9969–9975.

Cottam, E. M., Haydon, D. T., Paton, D. J., Gloster, J., Wilesmith, J. W., Ferris, N. P., Hutchings, G. H. & King, D. P. (2006). Molecular epidemiology of the foot-and-mouth disease virus outbreak in the United Kingdom in 2001. *J Virol* **80**, 11274–11282.

Drexler, J. F., Grywna, K., Stocker, A., Almeida, P. S., Medrado-Ribeiro, T. C., Eschbach-Bludau, M., Petersen, N., da Costa-Ribeiro, Jr, H. & Drosten, C. (2009). Novel human parechovirus from Brazil. *Emerg Infect Dis* **15**, 310–313.

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.

Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88.

Faria, N. R., de Vries, M., van Hemert, F. J., Benschop, K. & van der Hoek, H. L. (2009). Rooting human parechovirus evolution in time. *BMC Evol Biol* **9**, 164.

Figueroa, J. P., Ashley, D., King, D. & Hull, B. (1989). An outbreak of acute flaccid paralysis in Jamaica associated with echovirus type 22. *J Med Virol* **29**, 315–319.

Harvala, H., Robertson, I., McWilliam Leitch, E. C., Benschop, K., Wolthers, K. C., Templeton, K. & Simmonds, P. (2008). Epidemiology and clinical associations of human parechovirus respiratory infections. *J Clin Microbiol* **46**, 3446–3453.

Harvala, H., Robertson, I., McWilliam Leitch, C., Chieochansin, T., Templeton, K. & Simmonds, P. (2009). Aetiological role of human parechovirus type 3 in neonatal sepsis identified by direct typing assay on cerebrospinal fluid. *J Infect Dis* **199**, 1753–1760.

Heath, L., van der Walt, E., Varsani, A. & Martin, D. P. (2006). Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J Virol* **80**, 11827–11832.

Hyypia, T., Horsnell, C., Maaronen, M., Khan, M., Kalkkinen, N., Auvinen, P., Kinnunen, L. & Stanway, G. (1992). A distinct picornavirus group identified by sequence analysis. *Proc Natl Acad Sci U S A* **89**, 8847–8851.

Ito, M., Yamashita, T., Tsuzuki, H., Takeda, N. & Sakae, K. (2004). Isolation and identification of a novel human parechovirus. *J Gen Virol* **85**, 391–398.

Joki-Korpela, P. & Hyypia, T. (2001). Parechoviruses, a novel group of human picornaviruses. *Ann Med* **33**, 466–471.

Kew, O., Mulders, M. N., Lipskaya, G. Y., da Silva, E. E. & Pallansch, M. A. (1995). Molecular epidemiology of polioviruses. *Semin Virol* **6**, 401–414.

Koskiniemi, M., Paetau, R. & Linnavuori, K. (1989). Severe encephalitis associated with disseminated echovirus 22 infection. *Scand J Infect Dis* **21**, 463–466.

Li, L., Victoria, J., Kapoor, A., Naeem, A., Shaukat, S., Sharif, S., Alam, M. M., Angez, M., Zaidi, S. Z. & Delwart, E. (2009). Genomic characterization of novel human parechovirus type. *Emerg Infect Dis* **15**, 288–291.

Lukashev, A. N. (2005). Role of recombination in evolution of enteroviruses. *Rev Med Virol* **15**, 157–167.

McWilliam Leitch, E. C., Bendig, J. W., Cabrerizo, M., Cardosa, J., Hyypia, T., Ivanova, O. E., Kelly, A., Kroes, A. C. M., Lukashev, A. N. & other authors (2009). Transmission networks and population turnover of echovirus 30. *J Virol* **83**, 2109–2118.

Minor, P. D., John, A., Ferguson, M. & Icenogle, J. P. (1986). Antigenic and molecular evolution of the vaccine strain of type 3 poliovirus during the period of excretion by a primary vaccinee. *J Gen Virol* **67**, 693–706.

Oberste, M. S., Maher, K. & Pallansch, M. A. (1998). Complete sequence of echovirus 23 and its relationship to echovirus 22 and other human enteroviruses. *Virus Res* **56**, 217–223.

Santti, J., Hyypia, T., Kinnunen, L. & Salminen, M. (1999). Evidence of recombination among enteroviruses. *J Virol* **73**, 8741–8749.

Shapiro, B., Rambaut, A. & Drummond, A. J. (2006). Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol* **23**, 7–9.

Simmonds, P. (2006). Recombination and selection in the evolution of picornaviruses and other mammalian positive-stranded RNA viruses. *J Virol* **80**, 11124–11140.

Stanway, G., Kalkkinen, N., Roivainen, M., Ghazi, F., Khan, M., Smyth, M., Meurman, O. & Hyypia, T. (1994). Molecular and biological characteristics of echovirus 22, a representative of a new picornavirus group. *J Virol* **68**, 8232–8238.

Stanway, G., Joki-Korpela, P. & Hyypia, T. (2000). Human parechoviruses – biology and clinical significance. *Rev Med Virol* **10**, 57–69.

Stanway, G., Brown, F., Christian, P., Hovi, T., Hyypia, T., King, A. M. Q., Knowles, N. J., Lemon, S. M., Minor, P. D. & other authors (2005). Family *Picornaviridae*. In *Virus Taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses*, pp. 757–778. Edited by C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. San Diego: Elsevier/Academic Press.

Takeda, N., Tanimura, M. & Miyamura, K. (1994). Molecular evolution of the major capsid protein VP1 of enterovirus 70. *J Virol* **68**, 854–862.

van der Sanden, S., de Bruin, E., Vennema, H., Swanink, C., Koopmans, M. & van der Avoort, H. (2008). Prevalence of human parechovirus in The Netherlands in 2000 to 2007. *J Clin Microbiol* **46**, 2884–2889.

Watanabe, K., Oie, M., Higuchi, M., Nishikawa, M. & Fujii, M. (2007). Isolation and characterization of novel human parechovirus from clinical samples. *Emerg Infect Dis* **13**, 889–895.

Wigand, R. & Sabin, A. B. (1961). Properties of ECHO types 22, 23 and 24 viruses. *Arch Gesamte Virusforsch* **11**, 224–247.

Williams, C. H., Panayiotou, M., Girling, G. D., Peard, C. I., Oikarinen, S., Hyoty, H. & Stanway, G. (2009). Evolution and conservation in human parechovirus genomes. *J Gen Virol* **90**, 1702–1712.

Zoll, J., Galama, J. M. & van Kuppeveld, F. J. (2009). Identification of potential recombination breakpoints in human parechoviruses. *J Virol* **83**, 3379–3383.