Resolving ambiguities in genetic typing of human enterovirus species C clinical isolates and identification of enterovirus 96, 99 and 102

Betty A. Brown,1 Kaija Maher,1 Mary R. Flemister,1 Pejman Naraghi-Arani,1† Moyez Uddin,2 M. Steven Oberste1 and Mark A. Pallansch1

 Correspondence
Betty A. Brown
bzb2@cdc.gov

1Polio and Picornavirus Laboratory Branch, Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Mailstop G17, Atlanta, GA 30333, USA
2Institute of Public Health, Dhaka, Bangladesh

Molecular methods, based on sequencing the region encoding the VP1 major capsid protein, have recently become the gold standard for enterovirus typing. In the most commonly used scheme, sequences more than 75 % identical (>85 % amino acid identity) in complete or partial VP1 sequence are considered to represent the same type. However, as sequence data have accumulated, it has become clear that the ‘75 %/85 % rule’ may not be universally applicable. To address this issue, we have determined nucleotide sequences for the complete P1 capsid region of a collection of 53 isolates from the species Human enterovirus C (HEV-C), comparing them with each other and with those of 20 reference strains. Pairwise identities, similarity plots and phylogenetic reconstructions identified three potential new enterovirus types, EV96, EV99 and EV102. When pairwise sequence comparisons were considered in aggregate, there was overlap in percentage identity between comparisons of homotypic strains and heterotypic strains. In particular, the differences between coxsackievirus (CV) A13 and CVA17, CVA24 and EV99, and CVA20 and EV102 were difficult to discern, largely because of intratypic sequence diversity. Closer inspection revealed the minimum intratypic values and maximum intratypic values varied by type, suggesting that the rules were at least consistent within a type. By plotting VP1 amino acid identity vs nucleotide identity for each sequence pair and considering each type separately, members of each type were fully resolved from those of other types. This study suggests that a more stringent value of 88 % VP1 amino acid identity is more appropriate for routine typing and that other criteria may need to be applied, on a case by case basis, where lower values are seen.

INTRODUCTION

Based on sequence similarity and biological properties, human enteroviruses are divided into four species, Human enterovirus (HEV) A–D, with HEV-C comprising poliovirus (PV) types 1, 2 and 3, as well as coxsackievirus (CV) A1, CVA11, CVA13, CVA15, CVA17–22 and CVA24 (Knowles, 2006; Stanway et al., 2005). Although most infections are asymptomatic, enteroviruses can cause a variety of clinical diseases ranging from mild syndromes, such as common cold and rash, to more severe illnesses such as acute haemorrhagic conjunctivitis (AHC), aseptic meningitis, encephalitis, and acute flaccid paralysis (AFP). There are at least 30 000 hospitalizations per year in the USA due to enterovirus aseptic meningitis (Khettsuriani et al., 2003; Pallansch & Roos, 2001). Enteroviruses are among the leading causes of acute febrile illness in young children and similar symptoms are manifested by infections from almost all the HEV-C viruses.

Historically, enteroviruses were classified into serotypes by neutralization tests with type-specific antisera (Dalldorf & Sickles, 1956; Melnick & Hampil, 1965). Despite its historical and pragmatic importance, this virus typing system has numerous difficulties related to antigenic heterogeneity within virus types, which have made application of this classical approach increasingly problematic over time. In some cases, cross-neutralization may be non-reciprocal, resulting in so-called ‘prime strains’ (a prime strain is one which is neutralized poorly by antisera
made to the reference virus but antisera made to the prime strain effectively neutralizes both the prime strain and the reference virus (Melnick, 1996; Schmidt et al., 1966). There are several examples of antigenic variants within an HEV-C type. For example, there are three named variants of CVA20, termed CVA20a, CVA20b and CVA20c, in addition to those strains with ‘prototype’ neutralization properties (Behbehani et al., 1964). Echovirus 34, initially classified as a unique type (World Health Organization, 1968), was later shown to be a prime strain of CVA24, so the echovirus 34 designation is no longer used (Rosen et al., 1970). Similarly, another CVA24 isolate, EH24/70, which was associated with pandemic AHF, was poorly neutralized with antisera to the CVA24 prototype strain, Joseph (Mirkovic et al., 1974), so it is generally referred to as ‘CVA24 variant’ (CVA24v).

Recently, molecular sequencing has supplanted neutralization testing as the standard for enterovirus typing (Oberste & Pallansch, 2005). In the most commonly used molecular typing scheme, based on comparison of sequences encoding the VP1 capsid protein, homotypic viruses generally share at least 75% nucleotide identity and at least 85% amino acid sequence identity in VP1. Sequence data have helped to clarify the classification of a number of strains. For example, sequence comparisons have revealed that CVA15 is a strain of CVA11 and CVA18 is a strain of CVA13 (Brown et al., 2003; Oberste et al., 1999b) and numerous new types have been identified by sequence analysis (Arita et al., 2005; Junttila et al., 2007; Norder et al., 2003; Oberste et al., 2001, 2004a, 2005, 2007a; Smura et al., 2007a, b).

With the recent increase in the number of available enterovirus sequences, it has become apparent that the distinctions between types, in terms of percentage identities between heterotypic viruses (and identities among homotypic viruses), are not as discrete as originally thought. For example, in some cases the range of VP1 identities within a serotype may fall slightly below the 75% /85% limit, suggesting that the current quantitative typing standards may not be adequate to resolve all strains of all serotypes (Oberste et al., 2000). To begin to address this issue within a single enterovirus species, we have sequenced and analysed the P1 capsid-coding region of 53 recent HEV-C clinical isolates and the CVA24 prototype strain, Joseph. The sequences were compared to one another and to those of 20 previously sequenced strains. These comparisons have helped to refine the molecular definition of type and identified three new types within HEV-C: EV96, EV99 and EV102.

METHODS

Viruses. Strains sequenced for this study were isolated between 1975 and 2004 in 11 countries (see Supplementary Table S1, available in JGV Online). Twenty-seven isolates were obtained from the enterovirus collection at Centers for Disease Control and Prevention, Atlanta, GA, USA (CDC), while the remaining 26 strains were isolated from stool specimens obtained from AFP cases in Bangladesh between 1999 and 2003, collected as part of surveillance activities of the global Polio Eradication Initiative (Oberste et al., 2006). Viruses were propagated in human rhadomyosarcoma (RD) and/or human lung fibroblast (HLF) cells by standard methods (Melnick et al., 1979; World Health Organization, 2001). Historical strains from the CDC enterovirus collection (pre-1998) were originally typed using serotype-specific antisera, whereas strains isolated after 1998 were identified by sequencing a ~320 bp fragment from the S’ region of VP1 (Oberste et al., 2003). RNA was extracted from 140 µl infected cell culture supernatant using a QIAamp Viral RNA kit (Qiagen) and cDNA was synthesized using Superscript II reverse transcriptase (Invitrogen) with random hexamers or virus-specific primers.

Nucleotide sequencing. To facilitate complete capsid sequencing, RT-PCR primers were designed to anneal to sites encoding amino acid motifs that are highly conserved among enteroviruses (Brown et al., 2003; Oberste et al., 2004b, c, d, e, 2005, 2006) (Supplementary Table S2, available in JGV Online). Specific primers were designed from preliminary sequences to close gaps between the original PCR products. For all sequence determinations, the PCR products were purified for sequencing by using a High-Pure PCR product purification kit (Roche Molecular Biochemicals), and both strands were sequenced by automated methods, using fluorescent dideoxy chain terminators (Applied Biosystems). The sequences of 20 additional strains were obtained from GenBank (Supplementary Table S1).

Sequence analysis. Multiple sequence alignments were generated using the PILEUP program (Wisconsin Sequence Analysis Package, version 10.3; Accelrys) and, when required, the nucleotide sequence alignment was adjusted manually to conform to the optimized alignment of deduced amino acid sequences. Pairwise identities among the nucleotide and deduced amino acid sequences were calculated by using the programs Gap and Distances (Wisconsin Package). For each virus pair, amino acid identity was plotted vs nucleotide identity for the complete P1 region and for individual functional subregions, VP1–4. Local conservation between individual amino acid sequence pairs was visualized by calculating pairwise sequence identity in a sliding window of 200 residues and progressively advancing the window in 10-residue steps across the length of the alignment using SimPlot (Lole et al., 1999). Phylogenetic relationships were reconstructed using the neighbour-joining method implemented in MEGA4 (Kumar et al., 2001), with the Kimura two-parameter substitution model for nucleotide sequences (Kimura, 1980) and the PAM001 model for amino acid sequences (Dayhoff et al., 1979).

Neutralization assay. Twelve CVA24 and ‘CVA24-like’ isolates (see text) were characterized using standard antigenic methods (Grandien et al., 1995). Tests employed antisera specific for three different well-characterized CVA24 reference strains, CVA24-Joseph (CVA24 prototype strain; NIH horse antiserum, V027-501-560), CVA24-EH24/70 (CVA24v; CDC rabbit antiserum D4261), and CVA24-DN19 (echovirus 34; CDC rabbit antiserum 79-0095). Tenfold serial dilutions of antisera were mixed with an equal volume of constant virus concentration [100 CCID 50 (50% cell culture infectious dose) per well] in a 96-well plate and incubated at 37°C for 2 h before addition of 5 × 10 3 RD or HLF cells per well. The plates were incubated at 37°C in a humidified 5% CO 2 atmosphere and observed for up to 7 days for the appearance of viral cytopathic effect. Controls included the homologous strains to the antisera.

Nucleotide sequence accession numbers. The sequences described here have been deposited in the GenBank sequence database, accession numbers DQ995633–DQ995648, EF015008–
RESULTS

Overall sequence diversity

A total of 10 types were identified among the 53 isolates, based on the criterion of homotypic viruses sharing at least 85 % amino acid identity in VP1. The most abundant types were CVA13 (n=12), CVA21 (n=12), CVA24 (n=11) and CVA20 (n=10), while CVA11, CVA17 and CVA22 accounted for three, two and two strains, respectively, and no isolates of CVA1 and CVA19 were detected. Eight viruses were identified as putative members of three new types, EV96 (n=1), EV99 (n=6), and EV102 (n=1).

The P1 amino acid lengths ranged from 875 to 890 residues, with most insertions or deletions occurring in regions that were also variable in sequence between viruses of different serotypes. VP1 lengths were the most variable (296–309 aa), while VP2 and VP3 lengths varied by no more than three residues (238–241 aa and 269–272 aa, respectively), and all of the VP4 proteins were 69 residues. Insertions or deletions in three variable regions, the amino terminus, the B-C loop and the D-E loop resulted in VP1 length variation with the types CVA20 (P1 residues 601 and 676) and CVA13 (P1 residue 687).

Phylogenetic relationships

Phylogenetic trees were constructed from aligned P1 amino acid sequences (Fig. 1). Strains generally segregated into distinct clusters according to type, as expected. Several serotypes, such as each of the three poliovirus types, CVA11 and CVA17, formed compact, well defined clusters. Other serotypes, such as CVA21 and CVA13, formed apparent subclusters, but were nonetheless monophyletic as a whole. CVA24 strains were divided into two clusters, a tight CVA24v-like subgroup (strains from Dominican Republic, China, Japan, Jamaica, Brazil and the USA) and a loose Joseph-like subgroup (strains from South Africa, Puerto Rico and the USA, including the DN-19 strain). Eight of the CVA20 strains formed a tight cluster, with two additional Bangladesh strains somewhat separate from the main cluster. Strain 10424-BAN99 (proposed EV102; see below) was related to the CVA20 cluster, but it was a clear outlier. Strain 10488-BAN00 was phylogenetically distinct from all of the serotype clusters, suggesting that it may represent a new serotype (proposed EV96). Similarly, a group of six isolates, from Bangladesh, Oman and the USA, formed a distinct cluster which comprises another probable new serotype (proposed EV99; see below).

Phylogenetic trees were also constructed for each of the mature capsid proteins, VP1 (Fig. 2), VP2, VP3 and VP4 (Supplementary Fig. S1, available in JGV Online). Isolates did not cluster according to type in VP4 with the exception of the polioviruses, CVA13 and CVA21. The VP1 tree (Fig. 2) appears to have the best correspondence to the complete P1 tree (Fig. 1), with a similar topology and the same outlier sequences that suggest new serotypes. The VP2 and VP3 phylogenetic groupings also resembled the P1 tree, but with some slight variations. For example, CVA13 is more distant from CVA17 in VP2 than in P1, VP1 or VP3. In VP2, strain 10424-BAN99 (EV102) is more closely related to the CVA20 strains than in the P1 and VP1 trees, and it is included in the CVA20 cluster in VP3. EV96 and EV99 remain distinct from all other strains in all capsid subregions except VP4 (Figs 1, 2 and Supplementary Fig. S1).

Pairwise sequence identities

Nucleotide and amino acid sequence identities were calculated for all pairs of viruses (n=2628 comparisons), for the complete P1 region and for each of the mature capsid subregions, VP1–4 (Supplementary Tables S3–S12, available in JGV Online). Because the large number of synonymous nucleotide substitutions in enterovirus sequences tends to introduce ‘noise’ into nucleotide identity comparisons among more distantly related sequences (Jorba et al., 2008), we first focused on amino acid identities to try to discriminate between intratypic and intertypic comparisons (Supplementary Tables S4, S6, S8, S10 and S12). The P1 peptide sequences of the individual viruses in this collection differ from one another by up to 36 % (Supplementary Table S4), while the component capsid proteins VP1–4 differed from one another by up to 45, 33, 35 and 25 %, respectively (Supplementary Tables S6, S8, S10 and S12, respectively). In complete P1 comparisons, viruses of the same type are at least 88.5 % identical (Supplementary Table S4). Based on previous studies, we expected the lower limit of intraserotypic VP1 amino acid identity to be approximately 85–88 %, and comparisons between viruses of different types to show an identity lower than 85 % (Oberste et al., 1999a, b). However, the lower bound for intratypic VP1 comparisons was 86.6 %, whereas the upper bound for intertypic comparisons was 87.4 % (Supplementary Table S6). The lower bounds for intratypic values for VP2, VP3 and VP4 were 88.2, 91.5 and 91.3 %, respectively, whereas the upper bound for intertypic comparisons was 92.9, 94.1 and 100 %, respectively (Supplementary Tables S8, S10 and S12, respectively). The intertypic values for VP4 (Supplementary Table S12) reinforce the observation that strains are not monophyletic in that gene region (Supplementary Fig. S1c).

To resolve better the intratypic and intertypic comparison values, we plotted amino acid identity values vs nucleotide identity values for each pair of sequences, reasoning that the ‘noise’ inherent in the nucleotide comparisons would be greater for distantly related sequence pairs (i.e. for viruses of different serotypes) than for more closely related pairs (viruses of the same serotype), thus increasing the resolution between intratypic and intertypic comparisons.

EF015040, EF026081, EF015886, EF555644, and EF555645 (Supplementary Table S1).

Resolved ambiguities in HEV-C molecular typing
These plots were generated for the complete P1 comparisons as well as for comparisons within each of the capsid subregions VP1–3 (Fig. 3 and Supplementary Fig. S2). In the P1 analysis, 56 comparisons (~2% of the total) defined an ‘overlap area’ in which some intertypic values exceed the minimum value for intratypic comparisons (Fig. 3a). The overlap region (88.5–93.4% identity) included intratypic values for CVA22 and CVA13, and intertypic comparisons between CVA24 and EV99 and between CVA20 strains and EV102. In contrast to that of the complete P1 region, the VP1 overlap region of intratypic and intertypic comparisons was relatively small, with a range of 86.6–87.4% (Fig. 3b and Supplementary Table S6). Twenty-three VP1 comparisons (~1% of the total comparisons) were in this overlap zone. Seventeen of these were CVA13 intratypic comparisons and six were comparisons of CVA24 strains with EV99 strains or CVA20 strains with EV102. There were 33 intertypic comparisons whose values exceeded

---

**Fig. 1.** Phylogenetic relationships of HEV-C clinical isolates and other members of HEV-C, based on the P1 region. Distances were calculated using the PAM-001 matrix (Dayhoff et al., 1979) and relationships were reconstructed using the neighbour-joining algorithm implemented in MEGA 4.0 (Kimura, 1980), with 1000 bootstrap pseudoreplicates. Bar, amino acid distance as substitutions per site.
85\% (range, 85–87.4\%) and 15 intratypic comparisons that were less than 88\% identical (range, 86.6–88\%) (Fig. 3b and Supplementary Table S6). In the amino acid vs nucleotide sequence identity plots for VP2 and VP3 (Supplementary Fig. S2), the overlap regions were larger than those in the VP1 plot (Fig. 3b). There were 94 comparisons (~3.6\% of the total) in the overlap region (87.8–92.9\% identity) for VP2, and 97 comparisons (~3.7\% of the total) in the overlap region (88.2–94.1\% identity) for VP3. As shown for VP1, intratypic values for CVA22 and CVA13, and intertypic comparisons between CVA24 and EV99 and comparisons of CVA20 strains to EV102 were largely responsible for these overlap regions in VP2 and VP3 (Supplementary Fig. S2).

Because only a few comparisons appeared to be in the overlap zone for VP1, we analysed these relationships further by plotting the VP1 comparisons separately for each individual type for which multiple sequences were available (Fig. 4). That is, these plots included only intratypic comparisons for a single type (enclosed in solid ellipses), plus comparisons of strains of that type to all other strains. Only single VP1 sequences are available for CVA1, CVA19 and EV102, so there were no intratypic values to plot for those types. For EV96, additional sequences were obtained from GenBank to expand the number of available comparisons. Unlike the plot based on all 2628 comparisons, these twelve plots resolved all intratypic comparisons from the intertypic comparisons, with varying degrees of resolution (Fig. 4). For nine types (CVA11, 17, 20, 21, 22, EV96 and PV1, 2 and 3), the intratypic comparison values were well resolved from the intertypic values (minimum of 5\% difference in amino acid identity). For three types (CVA13, CVA24 and EV99), the resolution was only 1–3\%. As in the phylogenetic analysis (Fig. 2) CVA24 and EV99 were
Fig. 4. Analysis of intratypic and intertypic VP1 sequence relationships by plotting amino acid sequence identity vs nucleotide sequence identity, comparing members of a given type to one another and to members of other types. Solid ellipses indicate points representing intratypic comparisons. Dashed circles in the CVA24 and EV99 panels indicate CVA24-EV99 comparisons.
The capsid regions of all 15 types were analysed with Simplot using each prototype strain P1 peptide sequence as the query sequence, with a sliding window of 200 residues.
and a step of 10 residues (Fig. 5). For most types, overall conservation of amino acid sequence was observed (>90 % similarity) compared with their respective prototype strain, with the exception of certain strains of CVA13 and CVA22. CVA13 strains that were members of the prototype-like cluster (Fig. 2) were >90 % identical to the prototype strain throughout the capsid, whereas strains in the other cluster were <90 % identical to the prototype in most of VP2, but >90 % identical in other regions. Recent CVA22 isolates were less than 90 % identical to the prototype strain in the entire capsid region, except for sequences near the carboxyl terminus of VP3. In three cases, intertypic comparisons overlapped intratypic comparisons. CVA17 and CVA20 strains were about 90 % identical to CVA13 strains near the carboxyl terminus of VP1, while CVA24 and EV99 strains were about 90 % identical to one another in the middle of VP1. The EV102 isolate was at least 90 % identical to CVA20 isolates in the VP2 and VP3 regions, but diverged from CVA20 in VP1.

Antigenic relationships of CVA24 and EV99 isolates

Because EV99 was closely related to CVA24 and because of the historic observation of antigenic variation within CVA24, we investigated the antigenic relationships of EV99 to CVA24 using neutralization assays with well-characterized CVA24 antisera raised against CVA24-Joseph, CVA24v and CVA24-DN19 (Table 1). All nine of the CVA24 viruses tested were neutralized by at least one of the three antisera. By contrast, four of the six EV99 isolates (EV99-10627-OK85, EV99-10636-GA84, EV99-10696-OMA99 and EV99-10697-BAN04) could not be neutralized with any of the CVA24 antisera, but two other EV99 strains that clustered near these four (EV99-10461-BAN00 and EV99-10582-BAN01) were neutralized with CVA24-Joseph antisera; these strains shared as little as 81.5–82.1 % VP1 amino acid identity with the prototype CVA24-Joseph.

Table 1. Results of neutralization tests with CVA24-Joseph, CVA24v and CVA24-DN19 (E34) antisera

<table>
<thead>
<tr>
<th>Isolate</th>
<th>CVA24-Joseph</th>
<th>CVA24v (EH24)</th>
<th>CVA24-DN19</th>
</tr>
</thead>
<tbody>
<tr>
<td>EV99-10636-USA-GA84</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>EV99-10582-BAN01</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>EV99-10461-BAN00</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>EV99-10697-BAN04</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>EV99-10696-OMA99</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>EV99-10627-USA-OK85</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>CVA24-10625-USA-TX79</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CVA24-10626-PUR82</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CVA24-10628-JAM87</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CVA24-10629-BRA87</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CVA24-10631-USA-FL98</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>CVA24-10630-DOR93</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>CVA24-Joseph</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CVA24-DN19(E34)</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>CVA24V(EH24)</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
</tbody>
</table>

* +, Virus neutralized by antiserum; –, virus not neutralized by antiserum.

DISCUSSION

Analysis of complete VP1 sequences of enterovirus prototype strains and a limited number of additional strains suggested molecular sequence analysis could substitute for antigenic typing (Oberste et al., 1999a). In that system, strains sharing at least 75 % VP1 nucleotide identity and 88 % amino acid identity were presumed to be of the same serotype. Viruses of different serotypes were usually less than 70 % identical in VP1 nucleotide sequence (<85 % amino acid identity). Subsequently, we and others applied this approach to enterovirus typing using partial VP1 sequences (Caro et al., 2001; Casas et al., 2001; Norder et al., 2001; Oberste et al., 1999b, 2000). As additional data accumulated, the quantitative criteria were modified slightly, such that 85 % was considered the minimum value for intratypic VP1 amino acid identity (Oberste et al., 2002, 2004c). Additional studies, based on complete or partial VP1 sequences, have generally supported these empirical criteria and have provided some information on the range of intertypic VP1 amino acid identity. For example in a study with partial VP1 amino acid sequences, one echovirus 13 isolate was 96.3 % identical to the echovirus 13 reference strain, but it was also 87.1 % identical to EV69 (Rakoto-Andrianarivelo et al., 2005).

The recent identification of more than 20 new enterovirus types has relied almost exclusively on analysis of capsid region sequences using quantitative or phylogenetic molecular typing criteria, rather than on traditional antigenic comparisons (Arita et al., 2005; Junttila et al., 2007; Norder et al., 2003; Oberste et al., 2001, 2004c, 2005, 2007a; Smura et al., 2007a, b). In the present study, eight isolates were classified as members of three new types, EV96, EV99 and EV102. Even by the original molecular typing criteria (<70 % nucleotide identity and <85 % amino acid identity to known serotypes), EV96 is clearly distinct from all known enterovirus types and clearly constitutes a new type (Supplementary Tables S4 and S5; Figs 2 and 3). Our EV96 strain from Bangladesh resembles other recently identified EV96 strains from Asia and Europe, sharing 91–94 % VP1 amino acid identity (Arita et al., 2005; Smura et al., 2007a). All reported EV96 strains share at least 89 % VP1 amino acid identity with one another (minimum 75.6 % nucleotide identity) and are less than 78 % identical to strains of other types.

The six EV99 strains were 88–97 % identical to one another in VP1 amino acid sequence. EV99 isolates shared 56–87 % VP1 amino acid identity with other types, with the closest pair being CVA24-10625-TX79 and EV99-10697-BAN04.
(82–87% identity for all comparisons of EV99 and CVA24 strains). Extending this analysis to the entire capsid, the six EV99 isolates shared P1 identities with other types that ranged from 68 to 89%. P1 identity values have not been directly applied to molecular typing but, generally, enteroviruses of the same type share at least 90% P1 amino acid identity (Fig. 3a and Supplementary Table S4). Using this 90% P1 identity value as a guide, the EV99 strains can be defined as a new type despite their greater than 85% VP1 amino acid identity to CVA24 strains, since they are 90% identical to one another and <90% identical to CVA24 strains in complete P1. The distinction between EV99 and CVA24 strains is further supported by the phylogenetic and antigenic data. EV99 strains clustered together and were clearly distinct phylogenetically from CVA24 strains (Fig. 2). Because neutralizing epitopes may also be present on VP2 and VP3, the apparent immunological cross-reactivity between CVA24 and certain EV99 strains may be due to conserved epitopes on these proteins – EV99 and CVA24 strains were 87–91% identical in VP2 and 82–88% in VP3, compared with amino acid identities within CVA24 of 94–99% and 93–99% for VP2 and VP3, respectively (Supplementary Tables S8 and S10).

Comparison of our EV99 sequences with HEV-C sequences in GenBank identified two studies that describe untypable sequences. Some of the cluster 3 Madagascar isolates were also 82–86% identical to CVA24 strains. In the second study, enterovirus cases in France, one untypable HEV-C strain (GenBank accession no. AY208106) shared 94.4% VP1 identity to EV99-10636-GA84 in a region spanning the carboxyl-terminal one-third of VP1 (nt 592–903; aa 198–301). These isolates were also 82–86% identical to CVA24 strains. In the second study, of enterovirus cases in France, one untypable HEV-C strain (GenBank accession no. AY208106) shared 94.4% VP1 identity to EV99-10636-GA84 in a region spanning the carboxyl-terminal two-thirds of VP1 (Norder et al., 2003). These two reports are consistent with our Madagascar strains in that strains from Madagascar and France also share a high similarity to CVA24 and, like two of our EV99 strains, the French EV99 isolate was reportedly neutralized by anti-CVA24 antiserum. Similarly, CVA20-Tulane was 97% identical to the CVA20 prototype strain, but also shared 87% identity to EV102. EV102 is phylogenetically distinct in VP1, but possesses VP2 and VP3 proteins which are not recognizable different from those of CVA20 strains. Whether this pattern of identity between EV102 and CVA20 is characteristic of all EV102 strains can only be determined once more examples of EV102 have been isolated and characterized.

Phylogenetic analysis of CVA13 based on the VP1 region revealed two subclusters, one consisting of Bangladesh isolates and one US isolate, which diverged from the second CVA13 prototype-like cluster, with VP1 peptide identities ranging from 87 to 89%. We also compared these sequences to 20 published partial CVA13 sequences available in GenBank (data not shown). Sixteen CVA13 partial VP1 sequences (321 nt/107 aa at the 3’ terminus of VP1) were isolated during investigations of acute flaccid paralysis cases in Madagascar between 1994 and 2002 (Rakoto-Andrianarivelo et al., 2005, 2007). Of these, one clustered with the CVA13 prototype strain, while the remaining 15 formed yet a third CVA13 subcluster (data not shown). The remaining four CVA13 sequences from GenBank, from Bangladesh (Oberste et al., 2005), Cambodia (Arita et al., 2005), China (Bingjun et al., 2008), and Democratic Republic of Congo (Junttila et al., 2007), grouped into the same two clusters as our CVA13 sequences. Some of the cluster 3 Madagascar isolates were only 86.9% identical to a CVA13 isolate from Bangladesh and as much as 86.8% identical to two CVA17 isolates from Bangladesh, confirming the data from our Simplot analyses that suggested the 3’ one-third of VP1 may not fully resolve CVA13 and CVA17 (Fig. 5 and data not shown). In the same study as the China CVA13, three strains that were closely related to one another (97.4–100% amino acid identity) were considered untypable based on comparisons of a small partial VP1 sequence (aa 48–166), using 88% amino acid identity to the prototype as the intratypic boundary (Bingjun et al., 2008). One of the strains, however, had additional sequence available and the longer sequence (amino acids 48–293) was 87% identical to the CVA13 reference strain, identifying them as probable CVA13 strains. These data illustrate the importance of obtaining more complete VP1 sequence for molecular typing and suggest that analyses of short partial sequences should be interpreted cautiously, especially when they are used to infer the existence of a new type.

Plots of amino acid identity vs nucleotide identity have been used to better resolve differences between closely related sequences (Bailly et al., 2004; Oberste et al., 2007b). However, when a large number of sequences are compared, it may become difficult to resolve intratypic from intertypic comparisons (Fig. 3). Simplified comparisons, considering relationships one serotype at a time, suggest the absolute limits of intratypic comparisons – and their resolution from intertypic comparisons – may vary by serotype (Fig. 4). This approach helped to resolve CVA13 strains from those of CVA17 and CVA20, and EV99 strains from those of CVA24. One might expect the criteria to vary by species, as well. Future research may suggest additional computational approaches to permit definitive resolution of closely related serotypes by defined quantitative criteria.

The molecular typing criteria (75% VP1 nucleotide identity, and 85 or 88% VP1 amino acid identity) were originally established on the basis of empirical observations using comparisons with prototype strains (Oberste et al., 1999a, b, 2000, 2003; Oberste & Pallansch, 2005). Comparison with only prototype strains is probably sufficient for most applications, where sequence identities are clearly resolved, but the use of additional reference sequences may help to reveal relationships when the query
strain is only distantly related to the homologous prototype. The detailed analysis of additional clinical isolates, presented in this study, suggests that the change from 88 to 85% as a cutoff value may have been too aggressive, and that the more stringent value of 88% VP1 amino acid identity may be more appropriate for routine typing. Given that use of a higher threshold criterion will make it more difficult to unambiguously assign a type to a divergent strain within a known type, one must concede the existence of a ‘grey zone’ (~85–88% amino acid identity, for example), for which additional information (more reference sequences, complete P1 sequences, neutralization data, etc.) must be obtained before the strain can be definitively identified. Conversely, the ‘grey zone’ concept must also apply to the identification of potentially new types. Some strains, EV96, for example, are clearly resolved from known types by comparison with the VP1 sequences of prototype strains, whereas others (e.g. EV99 and EV102) appear to require more detailed analysis. Additional complete P1 sequences, coupled with antigenic, receptor and structural data, are needed from a wide range of enterovirus serotypes to better refine the empirical quantitative molecular typing criteria and to correlate specific sequences with biological function.

ACKNOWLEDGEMENTS

We are grateful to the Bangladesh staff of the Expanded Programme on Immunization, WHO, and the Bangladesh national polio eradication program for investigating cases of acute flaccid paralysis and obtaining stool specimens. We are indebted to Dr Nalini Withana, WHO/SEARO, for helping to facilitate our collaboration. We appreciate the technical assistance of Naomi Dybdahl-Sissoko and help from Ray Campagnoli and David Kilpatrick in designing the study group for discussions of enterovirus serotype definition based on molecular sequence data.

REFERENCES


