*Journal of General Virology* (2000), **81**, 803–812.  Printed in Great Britain

# Molecular epidemiology of coxsackievirus B4 and disclosure of the correct VP1/2A$^{pro}$ cleavage site: evidence for high genomic diversity and long-term endemicity of distinct genotypes

Mick N. Mulders,[1] Mika Salminen,[2] Nisse Kalkkinen[3] and Tapani Hovi[1]

[1,2] Department of Virology, Enterovirus Laboratory[1] and Department of Infectious Disease Epidemiology, HIV Laboratory[2],
National Public Health Institute (KTL), Mannerheimintie 166, 00300 Helsinki, Finland

[3] Protein Chemistry Laboratory, Institute of Biotechnology, University of Helsinki, PO Box 56, 00014 Helsinki, Finland

**Genetic diversity among 107 coxsackievirus B4 field isolates has been studied. These isolates included clinical and environmental isolates originating from Finland, the Netherlands and France, and also from several other countries, including the USA. Three genomic regions were used for phylogenetic analyses: the VP1/2A junction, the entire VP1 and the VP4/VP2 region. Alignment of the deduced amino acid sequence in the VP1/2A junction revealed extensive sequence variation at the previously proposed cleavage site. MS analysis of proteolytic fragments from VP1 revealed that the exact cleavage site is situated between amino acid residues Thr-849 and Gly-850. At least seven distinct genetic lineages, or genotypes, had been circulating in Europe during the period 1959–1998. Two genotypes were endemic in the Netherlands during most of the investigated period. Genetically closely related strains could be found in different countries, and different genotypes co-circulated at the same time in a given country. Clustering patterns were identical in the three genomic intervals. In the VP4/VP2 region, the intraserotypic variation approached interserotype variation. Sequence comparisons of the entire VP1 gene gave a reliable genetic identification of enterovirus serotype. It is suggested that, for genotype classification of previously serotyped coxsackievirus B4 isolates, comparison of VP1/2A sequences is sufficient, but for more detailed investigation of genetic relationships, and for 'genetic serotyping', the entire VP1 gene should be used. The VP4/VP2 region is less reliable for genetic serotyping and genotyping, although the primers are able to amplify many different serotypes.**

## Introduction

Coxsackievirus B4 (CV-B4) belongs to the family *Picornaviridae*, genus *Enterovirus*. In total, this genus comprises 64 serotypes isolated from humans, and an additional 28 serotypes that infect other mammals (King *et al.*, 1999). Transmission occurs through the faecal-oral or respiratory routes. In general, enterovirus infections go unnoticed, or symptoms are just 'flu'-like, and therefore not recognized as such. Occasionally, infection leads to severe neurological disorders or other diseases, e.g. meningitis, encephalitis, pleurodynia, myocarditis, conjunctivitis or severe systemic infections in neonates (Melnick, 1996). Furthermore, in several studies, enteroviruses, especially coxsackieviruses B, have been associated with juvenile type 1 diabetes (reviewed in Ramsingh *et al.*, 1997), although the actual mechanism of pathogenicity is yet to be elucidated. It is also known that CV-B4 and CV-B5 infections are relatively widespread in humans and these viruses are readily detected in the environment (Hovi *et al.*, 1996). Thus, despite the low attack rate, they can be considered important human pathogens. It is therefore important to understand their epidemiology and pathogenicity.

The single-stranded, messenger-sense viral RNA genome is replicated by the virus-encoded replicase, an RNA-dependent RNA polymerase, which is error-prone as it lacks proofreading

**Author for correspondence:** Mick Mulders.
Fax +358 9 47448355. e-mail Mick.Mulders@ktl.fi

The GenBank accession numbers of the coxsackievirus B4 sequences presented in this paper are AF159980–AF160168.

0001-6552 © 2000 SGM

activity (Holland *et al*., 1982). Therefore, enteroviruses exist as mixtures of micro-variants, so-called quasispecies (Holland *et al*., 1992). During replication in the human host or transmission from one person to the other, mutations accumulate, and some mutations become enriched or fixed, depending on various selection processes, e.g. by a genetic 'bottleneck'-mechanism. Furthermore, homologous recombination occurs frequently within a given serotype (intratypic) and between different serotypes (intertypic). Several studies in the past have demonstrated this mainly for polioviruses (reviewed by Agol, 1997). As a result, the polioviruses, like many other RNA viruses, have diverged into numerous independently evolving genetic lineages (reviewed by Kew *et al*., 1995).

In the global polio eradication campaign, partial nucleotide sequence analysis of a 150 nt stretch encompassing the junction region between VP1 (90 nt, 3′-end), the major capsid protein which contains the most important antigenic determinants, and the 2A protease (60 nt, 5′-end) is being routinely used to identify the likely geographic origin of genetic lineages (Rico-Hesse *et al*., 1987). Genetic clusters with less than 15% sequence divergence in the VP1/2A region have been designated genotypes (Rico-Hesse *et al*., 1987).

Our goal was to generate a sequence database of CV-B4 field isolates from several, mainly European countries, and to study the genotype distribution based on sequence variation observed in the VP1/2A region. Furthermore, we compared clustering patterns obtained from sequences from the entire VP1 gene. This capsid gene reflects the serotypic character of a virus to a high extent, as major antigenic sites are located on this protein. Previous studies have shown the use of one primer pair for the amplification of different enterovirus serotypes (Olive *et al*., 1990; Pulli *et al*., 1995; Arola *et al*., 1996; Huttunen *et al*., 1996) and demonstrated the usefulness of the VP4/VP2 region for the routine genetic classification of enteroviruses. We have also included this region in our analyses.

## Methods

■ **Viruses.** A total of 109 CV-B4 strains were used in this study, which included the ATCC prototype strain J. V. B., isolated from a stool sample of a 10-year-old child in New York, USA, in 1951, and a plaque variant of Edward's isolate (E-2), which was obtained in the USA in 1958 from the myocardial tissue of an infant with a generalized infection, focal necrosis and inflammation of the pancreas (Hartig *et al*., 1983). The genomes of the latter two viruses have been sequenced completely [J. V. B. GenBank accession no. X05690 (Jenkins *et al*., 1987); E-2 GenBank accession no. S76772 (Kang *et al*., 1994)]. Forty isolates were obtained from patients from the Netherlands, spanning a period of almost four decades (1959–1998). The 30 French strains were isolated in 1996 from sewage in Paris and the 21 Finnish isolates were from both patients and sewage (1986–1998). Six patient isolates were from the USA (1974–1988) and there were three strains from Denmark (1982, 1993), one Canadian strain (1983), one Romanian strain (1986), one from Latvia (1998) and three from Pakistan (1991–1992).

■ **Propagation of virus strains.** Freezer stocks were re-cultured on African Green Monkey Kidney (GMK) cells [1:10 dilution in maintenance medium (Eagle's MEM with 2% foetal calf serum)], three times freeze–thawed and clarified by centrifugation. Serotype was confirmed using the monovalent neutralizing antiserum.

Virus isolates obtained from sewage water samples were plaque-purified to resolve possible mixtures of genotypically different viruses (Mulders *et al*., 1999). In brief, 5-day-old monolayers of GMK cells in 6-well plates were inoculated with serial 10-fold virus dilutions prepared in maintenance medium. After 1 h at 37 °C, the inoculum was discarded and 2 ml semi-solid carboxymethylcellulose overlay was added to each well. After overnight incubation, plaques of 10–30 infected cells (approx. 5 µl) were picked under a microscope with low-power magnification using a pipette tip. The picked plaques were mixed with 100 µl maintenance medium and samples were stored at −20 °C until inoculation on cell culture.

■ **Virus purification.** GMK cell monolayers in Roux bottles were infected with $10^{-2}$ diluted stocks of selected CV-B4 strains and cultured overnight at 36 °C. Cultures were harvested at full CPE and virus was purified over sucrose gradients as described previously (Abraham & Colonno, 1984).

■ **RNA isolation.** Coxsackievirus RNA was purified from 50 µl cell culture supernatant using the RNeasy kit (Qiagen), dissolved in 50 µl diethyl pyrocarbonate-treated $H_2O$ and stored in aliquots at −70 °C.

■ **Primers.** For cDNA synthesis and PCR amplification, several combinations of different primer pairs were used (Table 1).

■ **Reverse transcription (RT).** cDNA synthesis was carried out for 1 h in a reaction volume of 10 µl, which contained 1 µl heat-denatured RNA, 25 mM Tris–HCl (pH 8·3 at 25 °C), 5 mM $MgCl_2$, 50 mM KCl, 2 mM DTT, 1 mM dNTPs (Boehringer Mannheim), 2 U avian myeloblastosis virus reverse transcriptase (Finnzymes), 4 U RNase inhibitor (Promega) and primer oligo$(dT)_{27}$CCG. A mixture of mineral oil (Sigma) and DynaWax (Finnzymes) 7:1 was applied as overlay to prevent evaporation and premature priming and formation of primer–dimers during the subsequent amplification step. Upon completion of the RT reaction, the sample was heated at 95 °C for 5 min to denature the enzyme, and then chilled on ice.

■ **PCR.** For enzymatic amplification, the reaction volume was adjusted to 50 µl by adding 4 l 10 × PCR buffer (0·1 M Tris–HCl, pH 8·8 at 25 °C, 0·5 M KCl, 1% Triton X-100), 0·75 l 50 mM $MgCl_2$, 12·5 pmol of both upstream and downstream primers, 1 U recombinant *Thermus brockianus* DNA polymerase (DyNAzyme II; Finnzymes) and distilled water. For simplicity, uniform cycling conditions were chosen: 1 min at 94 °C, followed by 35 cycles of 1 min at 94 °C, 1 min at 42 °C and 2 min 30 s at 72 °C. Table 1 shows a list of primers which were used in this study to amplify the VP1/2A junction, the entire VP1 and the VP4/VP2 region of the different CV-B4 strains. PCR products were visualized on an ethidium bromide-stained 2% agarose gel.

■ **Sequence analysis.** Depending on the presence of a single or multiple bands in the gels, PCR products were either directly purified using the Qiaex PCR purification kit (Qiagen), or excised from the gel and purified using the Qiaex gel extraction kit (Qiagen). Amplicons were eluted using 30–50 µl 10 mM Tris–HCl (pH 8·5) and stored at −20 °C. Cycle sequencing reactions were performed using the (Big) Dye terminator kit (PE Applied Biosystems) with the same forward or reverse primer as used in the preceding RT–PCR step. Both forward and reverse reactions were carried out to resolve possible ambiguous nucleotides. Comparison of forward and reverse sequences was made using Sequence Navigator (version 1; PE Applied Biosystems). Alignments were made using PILEUP, part of the GCG program suite (version 9.1). Programs DNAdist (maximum likelihood model of nucleotide substitution) and

**Table 1.** Primers used in this study

| Primer | Position* | Orientation | Sequence (5′ → 3′) |
|---|---|---|---|
| 71693† | 0449–0469 | F | CCT CCG GCC CCT GAA TGC GGC |
| 580 | 0583–0602 | F | GGC TGC TTA TGG TGA CAA T |
| 81294‡ | 1198–1182 | R | GGC AAC TTC CAC CAC CA |
| 72438 | 1198–1179 | R | GGC AAC TTC CAC CAC CAC CC |
| 08839 | 2367–2386 | F | TGC TTT GTG TCA GCA TGC AA |
| 71268 | 2373–2393 | F | GTG TCA GCA TGC AAC GAT TT |
| 71267 | 2397–2416 | F | GTA CGC ATG TTG AGG GAC AC |
| 73123 | 2865–2884 | F | CAG ACA CAC CAA ATA ATG TA |
| 08840 | 2868–2890 | F | ACA CAC CAA ATA ATG TAC GTG CC |
| 72241 | 2874–2894 | F | CAA ATA ATG TAC GTG CCA CCT |
| 72751 | 2891–2911 | F | ACC TGG AGG CCC CGT ACC AAC |
| 72242 | 2959–2940 | R | AAG ATG CTA GGG TTG GTG GA |
| 08841 | 2977–2958 | R | GCA TTG CCC TCT GTC CAA AA |
| 72353 | 2987–2968 | R | CTT GGT GGT GCA TTG CCC TC |
| 72752 | 3144–3160 | F | ACC AGC ACC ATC CGC AT |
| 08842 | 3156–3176 | F | CGC ATC TAC TTC AAA CCC AA |
| 71851 | 3343–3321 | R | ACT ACC TTG TAA TTG CCC ACA TA |
| 73124 | 3391–3372 | R | TCC CAC ACG CAA TTT TGC CA |
| 08843 | 3400–3381 | R | TTA TAA TCC TCC CAC ACG CA |
| 71497 | 3412–3388 | R | AGA AGG TCT CTA TTA TAA TCC TCC C |
| 71853 | 3468–3450 | R | TTG TGC ATT GGC ATC TGG C |
| 71933 | 7397–7395 | R | TTT TTT TTT TTT TTT TTT TTT TTT TTT CCG |

* Primer positions are given according to the orientation of the primer, either forward (F) or reverse (R) sense; numbering is according to Jenkins *et al.* (1987).

† 71693 by Rotbart (1990)

‡ 81294 modified from Olive *et al.* (1990).

Neighbor (UPGMA method of clustering), part of the PHYLIP (phylogeny inference) package (version 3.572c; Felsenstein, 1993) were used to construct dendrograms. Branch lengths are not optimized, but this method does assume an evolutionary clock. Graphic representations of dendrograms were obtained with NJplot or TreeView (version 1.5.3). Quartet puzzling (1000 replicates) was performed using Puzzle (version 4.0; Strimmer & von Haeseler, 1996).

Three different regions were analysed: (i) a 150 nt stretch encompassing the VP1/2A junction region (90 nt at the 3′-end of VP1 and the adjacent 60 nt at the 5′-end of 2A; position 3201–3350, numbering according to Jenkins *et al.*, 1987); (ii) the entire VP1 gene of 843 nt (position 2448–3290); and (iii) the VP4/VP2 region, encompassing the entire VP4 gene and 213 nt at the 5′-end of VP2 (position 744–1163). The VP1 sequences were constructed from overlapping amplicons. All CV-B4 sequences used in this study have been submitted to GenBank (accession numbers AF159980–AF160169). From the isolates for which the entire VP1 gene sequence was established, the submitted sequence also included the junction region with the two neighbouring genes, VP3 and 2A (position 2417–3371). In other cases, only the 150 nt VP1/2A sequence has been submitted (position 3201–3350). The sequences submitted to GenBank encompassing the VP4/VP2 region (position 603–1163) also included the hypervariable domain in the 5′NCR.

■ **Analysis of C terminus of VP1.** The capsid proteins of the CV-B4 strains J. V. B., E-2 and 15248NET72 were separated by 12% SDS–PAGE (Laemmli, 1970), and visualized by staining with Coomassie brilliant blue. The bands containing a mixture of VP1 and VP2 were cut out and the proteins were digested 'in-gel' (Hellman *et al.*, 1995) with endoproteinase LysC (Wako). The generated peptides were extracted from the gel pieces and subjected to mass-map analysis using delayed extraction matrix-assisted laser-desorption ionization/time-of-flight (MALDI–TOF) MS (Bruker Biflex; Bruker-Daltonics). Angiotensin 2 was used as an internal standard.

## Results

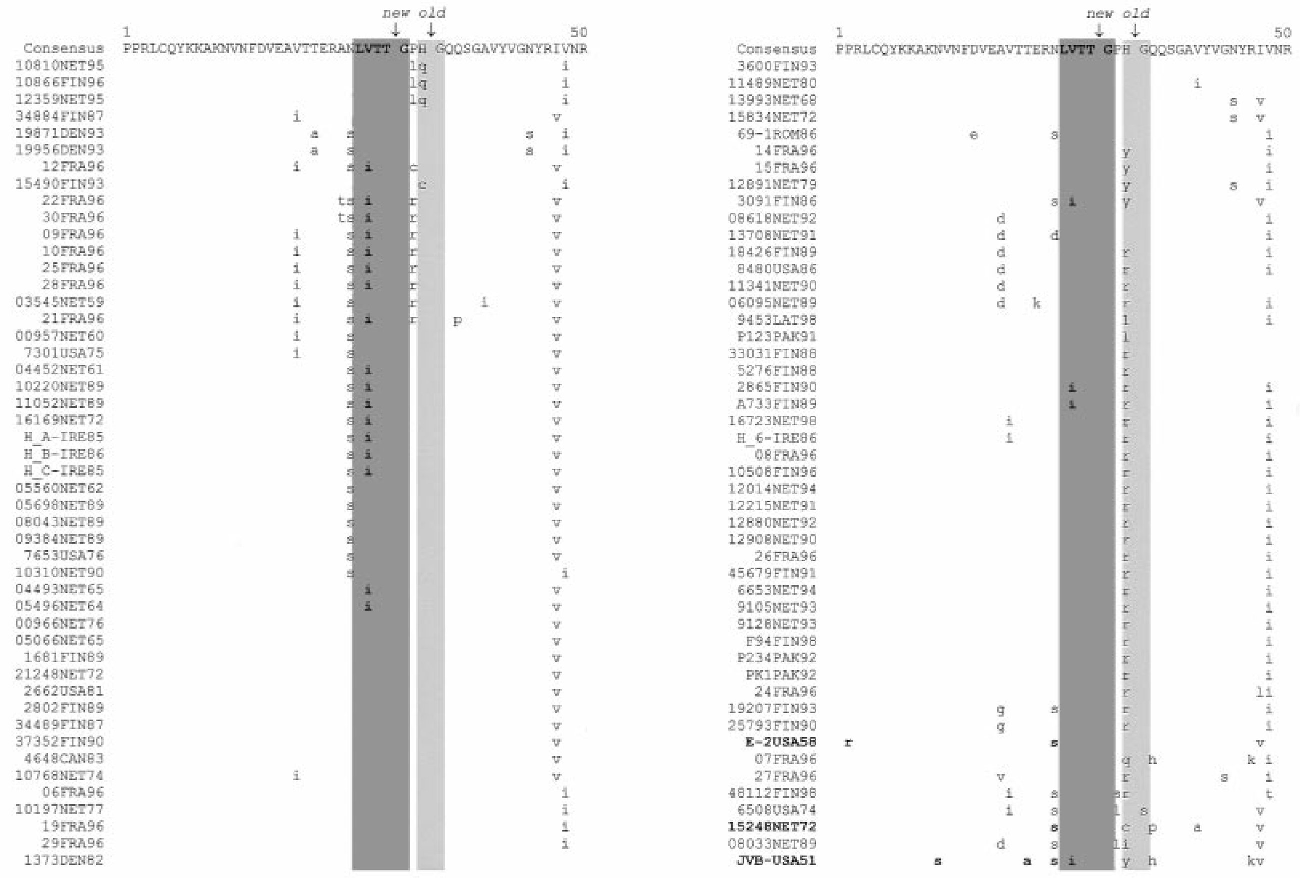
### General observations

In total, we obtained 107 CV-B4 field isolates from different regions of the world. All isolates were analysed in one or more of the genomic intervals used in this study. Of the total number, 72 were sequenced in all three of the aforementioned genomic intervals. From the remaining 35, all but five strains could be sequenced in the VP4/VP2 region; 15 strains could be sequenced in neither the VP1/2A region nor across the entire VP1 gene; and an additional 14 could only be sequenced in the VP1/2A region, and not across the VP1 gene. All nucleotide positions could be resolved if the sequencing in general was successful.

Based on the obtained distance matrices (not shown), the maximum genetic variation between CV-B4 isolates was as follows: 27·3% in the VP1/2A junction region; 19·3% in the entire VP1 gene; and 20·7 % in the VP4/VP2 region. However, the extent of maximum amino acid variation in the

**Table 2.** Amino acid sequence variation in the VP1/2A junction region

Nineteen amino acid residues at the C terminus of the VP1 protein from sequences in bold (JVB-USA51, E-2USA58 and 15248NET72) have been confirmed experimentally. Arrows indicate the location of the old (light grey box) and the newly (dark grey box) established proteolytic cleavage site between VP1 and 2A. Only codons that differ from the consensus sequence are shown (see text for further details).



```
                                         new old
            1                             ↓   ↓               50
  Consensus PPRLCQYKKAKNVNFDVEAVTTERANLVTT GPH GQQSGAVYVGNYRIVNR
 10810NET95                                 lq               i
 10866FIN96                                 lq               i
 12359NET95                                 lq               i
 34884FIN87                    i                            v
 19871DEN93                      a   s                   s   i
 19956DEN93                      a   s                   s   i
    12FRA96                    i     s i    c               v
 15490FIN93                                  c               i
    22FRA96                         ts i    r               v
    30FRA96                         ts i    r               v
    09FRA96                    i     s i    r               v
    10FRA96                    i     s i    r               v
    25FRA96                    i     s i    r               v
    28FRA96                    i     s i    r               v
 03545NET59                    i     s      r       i       v
    21FRA96                    i     s i    r    p          v
 00957NET60                    i     s                      v
  7301USA75                    i     s                      v
 04452NET61                          s i                    v
 10220NET89                          s i                    v
 11052NET89                          s i                    v
 16169NET72                          s i                    v
  H_A-IRE85                          s i                    v
  H_B-IRE86                          s i                    v
  H_C-IRE85                          s i                    v
 05560NET62                          s                      v
 05698NET89                          s                      v
 08043NET89                          s                      v
 09384NET89                          s                      v
  7653USA76                          s                      v
 10310NET90                          s                       i
 04493NET65                            i                    v
 05496NET64                            i                    v
 00966NET76                                                 v
 05066NET65                                                 v
  1681FIN89                                                 v
 21248NET72                                                 v
  2662USA81                                                 v
  2802FIN89                                                 v
 34489FIN87                                                 v
 37352FIN90                                                 v
  4648CAN83                                                 v
 10768NET74                    i                            v
    06FRA96                                                  i
 10197NET77                                                  i
    19FRA96                                                  i
    29FRA96                                                  i
  1373DEN82
```

```
                                         new old
            1                             ↓   ↓               50
  Consensus PPRLCQYKKAKNVNFDVEAVTTERANLVTT GPH GQQSGAVYVGNYRIVNR
  3600FIN93
 11489NET80                                          i
 13993NET68                                              s  v
 15834NET72                                              s  v
  69-1ROM86                e         s                       i
    14FRA96                                  y               i
    15FRA96                                  y               i
 12891NET79                                  y           s   i
  3091FIN86                          s i     y              v
 08618NET92                   d                              i
 13708NET91                   d      d                       i
 18426FIN89                   d              r               i
  8480USA86                   d              r               i
 11341NET90                   d              r
 06095NET89                   d   k          r               i
  9453LAT98                                  l               i
  P123PAK91                                  l
 33031FIN88                                  r
  5276FIN88                                  r
  2865FIN90                            i     r               i
  A733FIN89                            i     r               i
 16723NET98                    i             r               i
  H_6-IRE86                    i             r               i
    08FRA96                                  r               i
 10508FIN96                                  r               i
 12014NET94                                  r               i
 12215NET91                                  r               i
 12880NET92                                  r               i
 12908NET90                                  r               i
    26FRA96                                  r               i
 45679FIN91                                  r               i
  6653NET94                                  r               i
  9105NET93                                  r               i
  9128NET93                                  r               i
   F94FIN98                                  r               i
  P234PAK92                                  r               i
   PK1PAK92                                  r               i
    24FRA96                                  r              li
 19207FIN93                   g      s       r               i
 25793FIN90                   g              r               i
   E-2USA58  r                       s                      v
    07FRA96                                  q  h          k i
    27FRA96                   v              r          s    i
 48112FIN98                    i     s      sr               t
  6508USA74                    i     s      l  s            v
 15248NET72                          s       c  p    a      v
 08033NET89                   d      s      li              v
  JVB-USA51            s         a   s i     y  h          kv
```

VP1/2A region (50 aa) was surprisingly high at 22·0 %. In VP1 (length 281 aa), the maximum amino acid variation was only 4·3 % and in the VP4/VP2 region (140 aa), this value was 5·7 %.

### C terminus of the VP1 protein

Alignment of the deduced amino acid sequences at the VP1/2A junction revealed heterogeneity at the suggested VP1/2A cleavage site (Jenkins *et al*., 1987). Since the location of this site was based on sequence comparisons with other picornaviruses, rather than direct analysis, we wanted to identify the exact proteolytic cleavage site between VP1 and 2A. Thus, we separated the capsid proteins from strains J. V. B., E-2 and 15248NET72 by SDS–PAGE, excised the bands containing VP1 and VP2, which co-migrated with VP1, from the gel and subjected the proteins to ‘in-gel’ digestion followed by extraction of the generated peptides from the gel pieces. MALDI–TOF MS analysis of the digests revealed that the digest from J. V. B. contained a peptide with a monoisotopic mass of 2049·07. The only sequence to which this mass fits in the deduced J. V. B. sequence is the C-terminal LysC peptide from VP1 with a structure NVNFDVEAVTAERASLITT (calculated monoisotopic mass 2049·04). Similarly, the digest from strain E-2 contained a peptide with the monoisotopic mass of 2065·07 corresponding to the C-terminal peptide NVNFDVEAVTTERASLVTT (calculated monoisotopic mass 2065·07) of VP1. Furthermore, the digest from 15248NET72 contained a peptide with a monoisotopic mass of 2095·06 which corresponds to the C-terminal peptide NVNFDVEAVTTERTSLVTT (calculated monoisotopic mass 2095·04) of VP1. These results, together with the known nucleotide sequences across the VP1/2A junction region from the three analysed strains, unambiguously define the exact cleavage sites between VP1 and 2A in the analysed strains. The obtained sequences corresponded completely with the amino acid sequence deduced from the amplicon sequences. We found one amino acid difference with the published J. V. B. sequence (GenBank accession no. X05690; Jenkins *et al*., 1987), which was confirmed in the nucleotide sequence (position 3235: A

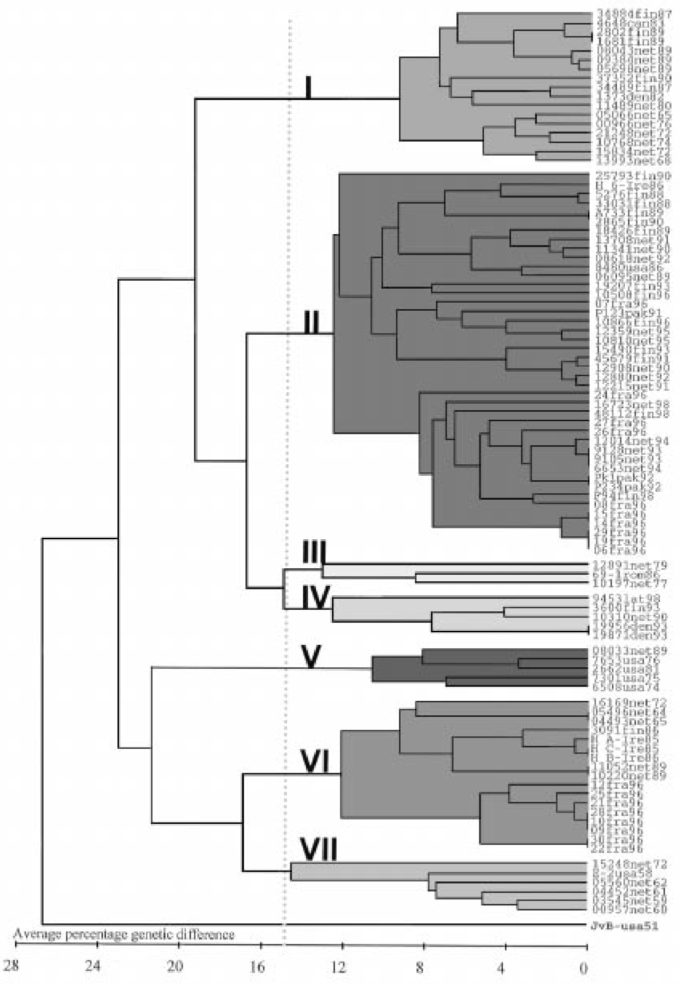


Fig. 1

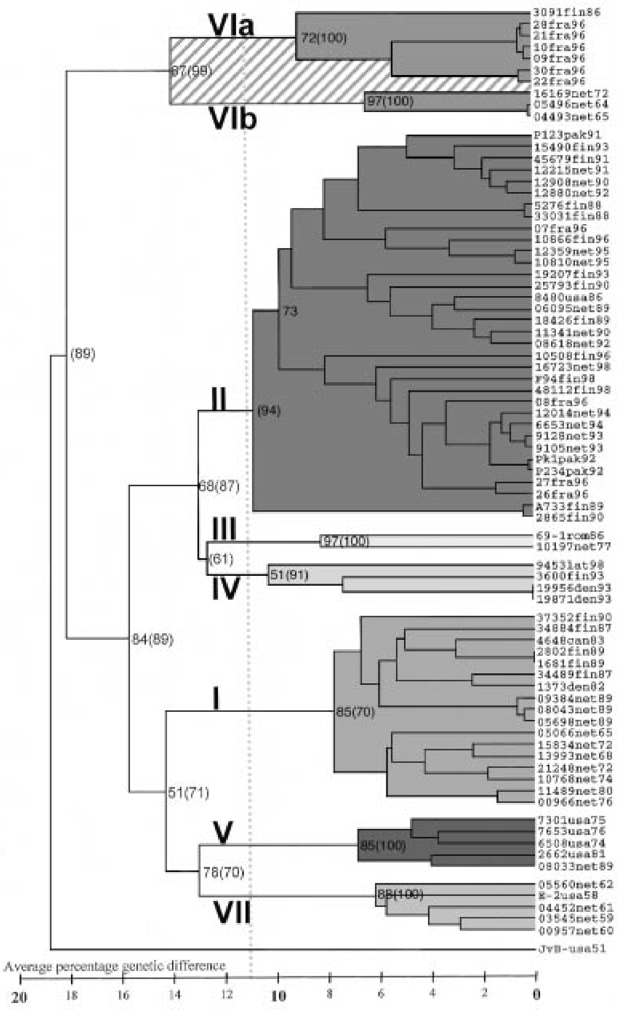


Fig. 2

**Fig. 1.** Genotype distribution of CV-B4 based on the 150 nt VP1/2A junction region; the dotted vertical line at 15% indicates the maximum variation within a designated genotype. Each shaded region indicates a genotype, numbered I–VII. A distance matrix was calculated using DNAdist (maximum likelihood) and the dendrogram was calculated with Neighbor (UPGMA). Four sequences had been published previously (HA-Ire85, HB-Ire86, HC-Ire85 and H6-Ire86) by Hughes *et al.* (1993); JvB-usa51 is the ATCC prototype strain J. V. B. and E-2usa58 is the diabetogenic strain. Country abbreviations: can, Canada; den, Denmark; fin, Finland; fra, France; ire, Ireland; lat, Latvia; net, the Netherlands; pak, Pakistan; rom, Romania; usa, United States of America.

**Fig. 2.** Dendrogram based on a sequence alignment of CV-B4 VP1 sequences. Numbers indicate quartet puzzling support values of each branch (1000 replicates). Numbers in parentheses are the corresponding bootstrap values as calculated using the PHYLIP package (1000 replicates). The dotted vertical line at 11% indicates the maximum variation within a designated genotype. Genotype VI (hatched) is subdivided into two separate genotypes, VIa and VIb (for abbreviations see legend to Fig. 1).

instead of G, resulting in Asn instead of Ser). The deduced amino acid sequence of five residues around the VP1/2A junction 'LxTT/G', where x is either V or I, and proteolytic cleavage occurs between T and G, is conserved among all sequenced CV-B4 isolates (Table 2).

## Sequence clustering in the VP1/2A region

Sequences were obtained from a total of 90 CV-B4 field isolates in the 150 nt VP1/2A junction region. Two entirely sequenced reference strains, J. V. B. and E-2, and four previously published Irish strains (Hughes *et al.*, 1993) were included in the analysis. Adopting the genotype demarcation of 15% as introduced by Rico-Hesse *et al.* (1987) for polioviruses, seven distinct genotypes, numbered I–VII, could be observed in the UPGMA dendrogram (Fig. 1).

Genotype I consisted of strains isolated in Canada, Finland, Denmark and the Netherlands (Fig. 1). Strains from the Netherlands belonging to this genotype had been isolated

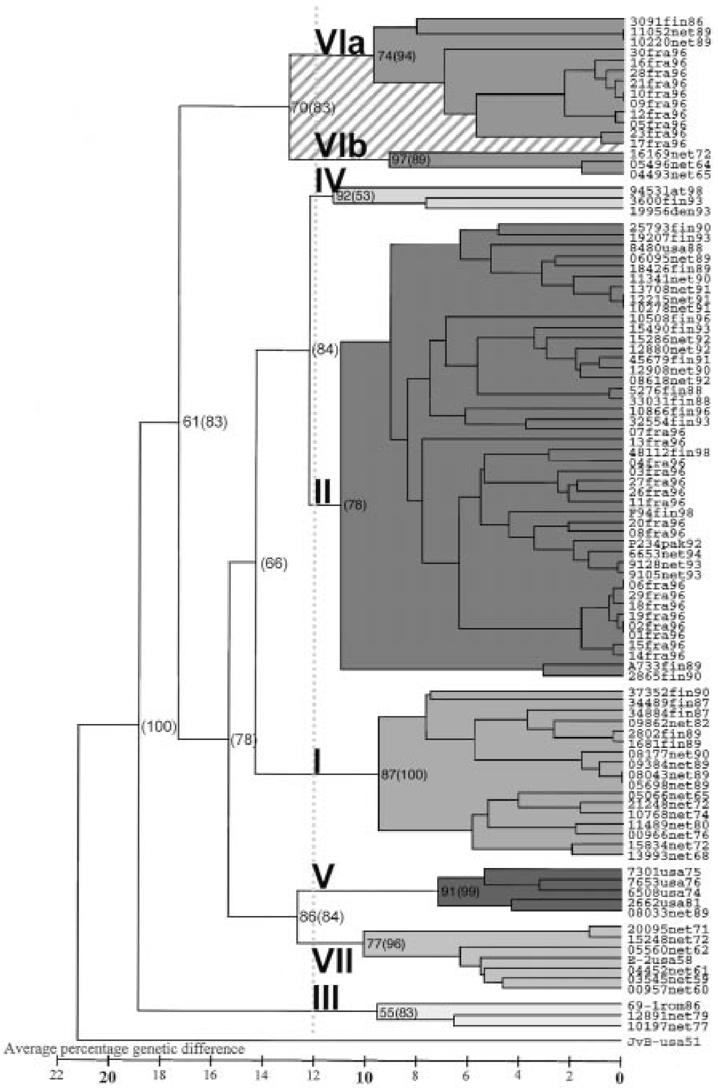


**Fig. 3.** Dendrogram based on the alignment of CV-B4 sequences of the VP4/VP2 region (420 nt). The dotted vertical line at 12% indicates the maximum variation within a designated genotype. Genotype VI (hatched) is subdivided into two separate genotypes, VIa and VIb (for abbreviations see legend to Fig. 1 and for additional information see Fig. 2).

throughout three decades, between 1965 and 1989. The second large genotype, II, contained strains from France, Finland, the Netherlands, Pakistan, the USA and published sequences from Ireland (Hughes *et al.*, 1993). Two other genotypes (III and IV) were closely related, with an average divergence of slightly more than 15%. Genotype III contained isolates from the Netherlands and Romania, and genotype IV contained isolates from Finland, Latvia, Denmark and the Netherlands. Genotype V comprised several older strains from the USA as well as one from the Netherlands, and genotype VI consisted of strains from the Netherlands, Finland, Ireland and France. Genotype VII contained the diabetogenic E-2 strain from the USA, 1958, as well as several isolates from the Netherlands of similar age (1959–1962). Genotypes I, II and VI were the most predominant, which is due to the fact that some countries were represented by a large number of strains (Finland, France and the Netherlands). No sequences clustered with the prototype J. V. B., and it therefore formed a genotype of its own. Within a 2 year period (1989–1990), five different genotypes, and within one genotype (II), three separately evolving sublineages, were co-circulating in the Netherlands. During the same period, two genotypes were found in Finland, where within one genotype (II), four different sublineages were found (Fig. 1). Identical clustering patterns were found using the Neighbor-joining method implemented in ClustalX (Thompson *et al.*, 1997), and with Puzzle (data not shown), and supported the above genotype classification based on the UPGMA method.

## Genetic variation in the VP1 gene

The entire 843 nt VP1 gene of 76 strains was sequenced (Fig. 2). As the observed maximum genetic variation in the

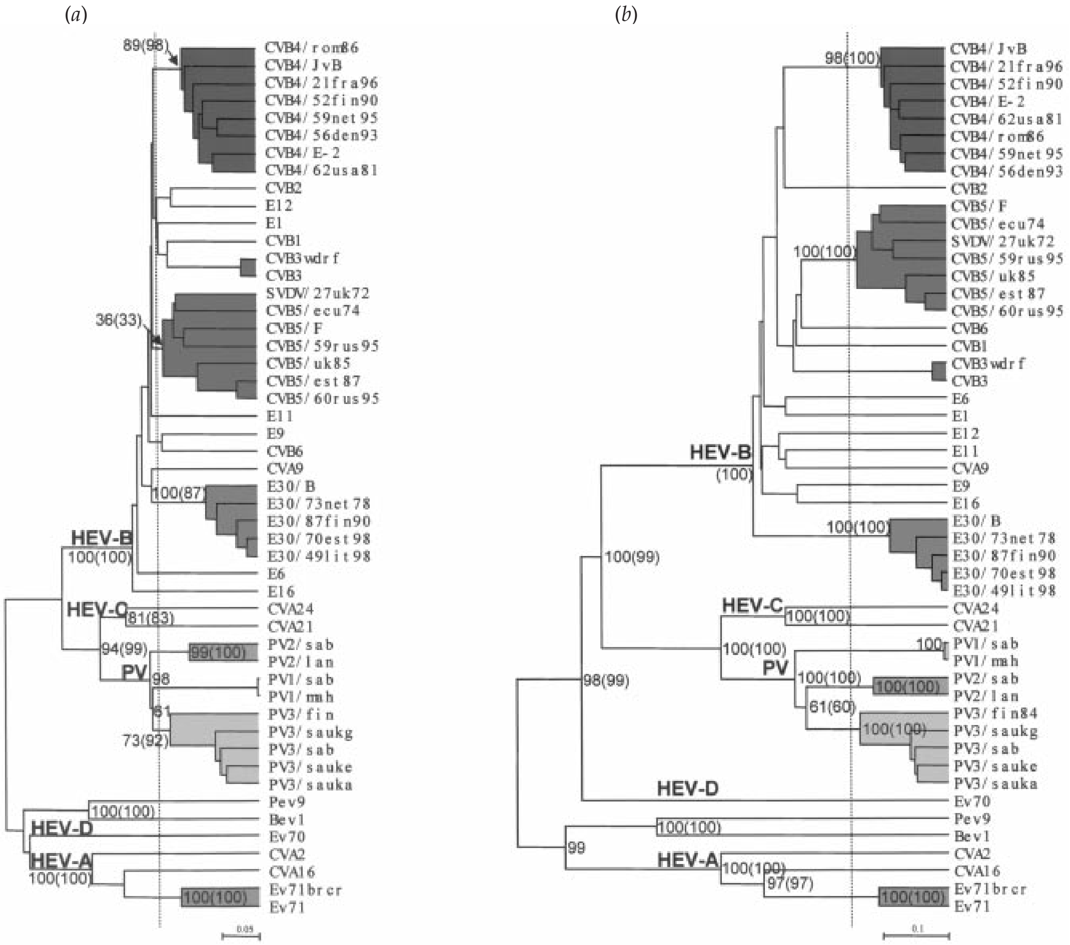


**Fig. 4.** Dendrograms of representative strains of each CV-B4 genotype compared with other enterovirus sequences available in GenBank, including porcine enterovirus (Pev9) and bovine enterovirus (Bev1), or obtained in our laboratory. (*a*) VP4/VP2 (420 nt); (*b*) entire VP1. The dotted line indicates serotype demarcation. Note that the scale of the VP4/VP2 dendrogram (*a*) is half the scale of the VP1 dendrogram (*b*). Numbers at the branchpoints (either left or right) indicate quartet puzzle values with bootstrap factors in parentheses. Both dendrograms were submitted to 1000 quartet puzzling and 1000 bootstrap replicates. Abbreviations: SVDV, swine vesicular disease virus; E, echovirus; Ev, enterovirus; PV, poliovirus; HEV, human enterovirus.

VP1 gene is lower than that in the VP1/2A junction, the extrapolated genotype demarcation had to be adjusted to 11%. Basically, clustering patterns were identical to those seen in the alignment of VP1/2A sequences. One exception was genotype VI, which divided up into two separate genotypes with a genetic difference approaching 14% between the two. Reliable quartet puzzling values were found for most branches (Fig. 2). Bootstrap analysis carried out using the PHYLIP package supported the Puzzle findings. Branching order of the genotype clusters was somewhat different: genotype VI grouped closely with genotype VII in the VP1/2A dendrogram, but formed a distinct, separate cluster in the VP1 tree. Genotypes V and VII now appeared more closely related, whereas genotype I branched with V and VII, and no longer with II, III and IV. Genotype distribution using ClustalX and Puzzle, again, gave identical results (not shown).

### VP4/VP2 region

The sequences of the VP4/VP2 region, encompassing the entire VP4 gene (207 nt) and 213 nt of the 5′-end of VP2, adjacent to VP4, were obtained from 104 strains, including J. V. B. and E-2. A dendrogram of the sequence alignment can be found in Fig. 3. Clustering patterns similar to those obtained with the VP1 sequence alignment were observed. Based on the distance matrix (not shown), the genotype demarcation was extrapolated to be 12%, which is less than the intragenotype variation in the VP1/2A region, but slightly larger than in the VP1 gene. Branching order was very similar with the VP1 tree. As in the VP1 tree, genotype VI again consisted of two separate clusters. The only exception was genotype III, which grouped closely with genotype IV in the VP1/2A and VP1 trees, but formed an entirely separate cluster in the VP4/VP2 tree. Identical genotype distribution patterns were observed using both ClustalX and Puzzle (not shown).

### Comparison with other human enterovirus serotypes

Fig. 4 shows a UPGMA comparison of nucleotide sequences of different enterovirus serotypes in both the VP4/VP2 region (*a*) and the VP1 (*b*). In both genomic intervals, CV-B4 isolates clustered together in a separate cluster as compared to other enterovirus serotypes, although statistical support (bootstrap and quartet puzzling) of the separate branches was lower in the VP4/VP2 region (Fig. 4). The maximum genetic distance between nucleotide sequences of different field isolates of a given serotype was 25% or less in the VP1 gene. In the VP4/VP2 region, this intraserotypic genetic distance approached the interserotype divergence. This can be seen most prominently in the human enterovirus B (HEV-B) species cluster, which contains all CV-B and echoviruses, as well as CV-A9 and human enterovirus 69. These observations were confirmed by the lower statistical support for the species and serotype clusters in the VP4/VP2 region.

## Discussion

We analysed the partial nucleotide sequence of 107 field isolates of CV-B4 from different European countries, from the Middle East and from the Americas spanning a period of four decades (1959–1998). Previously, Rico-Hesse *et al.* (1987) introduced the term ‘genotype’ for a cluster of epidemiologically related poliovirus sequences. Within a 150 nt stretch encompassing the VP1/2A junction (90 nt at the 3′-end of VP1 and 60 nt at the 5′-end of $2A^{pro}$), a genetic divergence of less than 15% indicated a likely epidemiological link. This region has been used with success in the global polio eradication initiative of the World Health Organization to establish transmission patterns of poliovirus genotypes and to locate endemic areas (Rico-Hesse *et al.*, 1987; Kew *et al.*, 1995; Mulders *et al.*, 1995). According to these principles, seven designated genotypes of CV-B4 were found to circulate during the investigated period (Fig. 1). We were not in a position to confirm epidemiological links of the CV-B4 isolates within any given 15% cluster, but the overall pattern of strain segregation in the dendrograms supported this criteria for a genotype (Fig. 1). The variability patterns were reminiscent of those of polioviruses in several other aspects as well. We analysed the entire VP1 (Fig. 2) and a region encompassing the entire VP4 and the 5′-end of VP2 (Fig. 3), in order to confirm clustering patterns obtained from the VP1/2A alignment. Clustering patterns were similar in all three of the analysed regions, except for genotype VI, which separated into two distinct clusters based on VP1 (Fig. 2) and VP4/VP2 sequences (Fig. 3), but not in the VP1/2A region (Fig. 1).

Another remarkable deviation was genotype III, which grouped closely to genotype IV in the VP1/2A and the VP1 tree, but formed a cluster of its own in the VP4/VP2 tree (Fig. 3). A possible explanation could be a recombination event between two CV-B4 viruses belonging to different genotypes. The cross-over point of recombination should have been located somewhere between VP2 and VP1. If this is true, it is a remarkable observation, as enterovirus recombination within the P1 capsid region has not been reported previously.

There was a striking difference in the range of sequence variability in the VP1/2A junction region as compared to the entire VP1 or the VP4/VP2 region. In the VP1/2A junction region, genotypes differed by less than 15% (Fig. 1), but only by 11% in the entire VP1 sequence (Fig. 2), and by 12% in the 420 nt VP4/VP2 region (Fig. 3).

Co-circulation of CV-B4 strains belonging to different genetic lineages was relatively common within one country (the Netherlands and Finland, 1989 and 1990, and France 1996; Fig. 1). On the other hand, on several other occasions genetically similar viruses were found to circulate in geographical regions that were far apart: Pakistan, France and the Netherlands (genotype II); in Finland, USA and the Netherlands (another variant of genotype II); Finland, the Netherlands and Canada (genotype I). This might reflect previous epidemic spread of the genotype or the existence of a globally persisting genotype.

CV-B4 can circulate uninterrupted in a given area for many decades: viruses belonging to the same genotype were found in the Netherlands between 1965 and 1990. This illustrates the endemic characteristics of the virus. This observation may be relevant to the hypothesis that populations with a high incidence of insulin-dependent diabetes mellitus (IDDM) might be endemically infected by a specific serotype, genotype or a set of variants of enteroviruses (Hovi, 1998). Finland has the highest incidence of IDDM in the world (Padaiga *et al.*, 1997). None of the CV-B4 genotypes found in this study were specifically enriched in Finland, but the sample of strains isolated from Finland was relatively limited. Viruses closely related to the E-2 diabetogenic variant (Hartig *et al.*, 1983; Kang *et al.*, 1994) had been isolated in the Netherlands during the same period (1959–1962) as the variant (1958), but not later (Fig. 1). This indicates that E-2 belonged to a commonly circulating genotype VII, now probably extinct. Hence, genotype VII cannot have any relevance to IDDM in general. Furthermore, since several different serotypes have been associated with IDDM recently (Roivainen *et al.*, 1998), there is no need to search for a ‘diabetogenic’ genotype of CV-B4 only.

The 150 nt VP1/2A region is sufficient to perform elementary molecular epidemiological studies, but due to its restricted length and its high nucleotide and amino acid variability, we recommend using the entire VP1 gene, as it appears to be a more reliable region for detailed molecular epidemiology (Fig. 2). The reliability of the clustering patterns was checked by performing quartet puzzling and bootstrap analyses, and by applying three different approaches for estimating the maximum likelihood distances: UPGMA, neighbour-joining and quartet puzzling. All methods gave identical clustering patterns, statistically supported by high quartet puzzling values for each genotype branch.

The high degree of nucleotide variation in the VP1/2A region as compared to the two other genomic intervals is remarkable (Fig. 1). One explanation could be the location of important antigenic sites at the C terminus of VP1. CV-A9, a closely related enterovirus serotype, contains an important antigenic site at the C terminus of VP1 (Pulli *et al*., 1998). The degree of amino acid sequence variation in an antigenic site is strikingly higher and might therefore account for the high sequence divergence in the VP1/2A junction region. One might have expected to see relatively more variation in the entire VP1 gene than in the VP4/VP2 region, because of the assumed abundance of antigenic sites located on the VP1 capsid protein. However, the reverse was true, although the differences remained small. The explanation may be that the VP1 gene contains large stretches of sequence conservation among isolates from the same serotype; both internalized structural regions as well as externalized regions, which define the virus serotype character. This view is supported by the optimal serotype-specific clustering of the strains using the VP1 sequences.

We sequenced regions of the genome that code for important motifs in the synthesis and processing of enterovirus proteins. The context of the initiator codon AUG in all of the sequenced field isolates was completely conserved, with A at position $-3$ and G at position $+4$ from the translation initiation sequence. In all but three cases, the sequence was **A**AAAUG**G** (once it was AGCAUGG and twice AAUAUGG). The cleavage site between VP4 and VP2 (N/S) was fully conserved among all field isolates, as was the VP3/VP1 cleavage site (Q/G). Alignment of the deduced amino acid sequences at the VP1/2A junction (Table 2) revealed unexpected heterogeneity at the putative VP1/2A cleavage site (Jenkins *et al*., 1987). Therefore, we decided to determine the true C-terminal sequence of VP1 on three strains. In all three cases, the VP1 was found to end with an LxTT motif, with x being V or I. The newly identified LxTT/G motif for the cleavage site was fully conserved among all sequenced CV-B4 strains.

Two positions in the P1 region have been previously associated with CV-B4 virulence: in VP4, the Ser residue at position 16 (Ramsingh & Collins, 1995), which is Arg in the non-virulent variant; and in VP1, Met-129, which becomes Thr in the non-virulent variant (Caggana *et al*., 1993). None of the sequenced CV-B4 field isolates had the avirulent sequence. In VP4, position 16 was mostly Ser (UCn), six isolates had Gly (GGn) and one had Asn (AAy). All isolates except one [05560NET62, which had Ile (AUA)] had Met (AUG) at position 129 in VP1.

The three genomic regions which were analysed were also assessed for their potential use in genetic typing of enteroviruses. Comparison of the VP1/2A junction region of different enterovirus serotypes is not very useful as the overall length of this fragment is very short and even within a serotype, there exists a high degree of variation. Using such highly variable and short genomic intervals would not give reliable phylogenetic branching. As the VP4/VP2 region is highly conserved among all enteroviruses, this region also appears to be less suited for genetic typing (Fig. 4 *a*). Therefore, of the three analysed regions, the VP1 region would be the region of choice (Fig. 4 *b*). This gene contains major antigenic sites as well as receptor recognition sequences and, therefore, resembles most optimally an enterovirus serotype. When comparing nucleotide sequences of different serotypes, a distinct serotype-specific clustering pattern can be observed for each of the enterovirus serotypes in which multiple isolates have been included (Fig. 4). The only drawback of this region is the necessity to use different primer combinations to obtain a PCR product, due to the high genetic variability of the VP1 gene.

As a starting point or for elementary epidemiological studies, the VP4/VP2 region is useful, as it can be amplified by one primer pair. However, more detailed information will be obtained from the VP1/2A region as a result of the higher sequence variability. Most reliable and detailed information can only be obtained from the VP1 region. VP1 sequence information can also be used to reliably define the serotype of a given strain.

In conclusion, CV-B4 behaves like many other RNA viruses with respect to high genetic variability. Many genetic lineages circulate worldwide and appear to have both endemic and epidemic forms of transmission. For genetic typing purposes, the use of the entire VP1, or at least a large proportion of it, is recommended, whereas for less-demanding studies, both the VP1/2A and the VP4/VP2 region can be used.

Mirja Stenvik (KTL) is thanked for everything; Päivi Hirttiö is acknowledged for her invaluable technical support to this project; Eija Penttilä for virus isolation and characterization. The authors acknowledge Albert Ras and Harrie van der Avoort (RIVM, Bilthoven) and Anton van Loon (AZU, Utrecht) for sending the isolates from the Netherlands; Laurent Andréoletti (University Hospital, Lille) and Michelle Aymard (Laboratory of Virology, Claude Bernard University, Lyon) for the French isolates; Mark Pallansch and Steven Oberste (CDC, Atlanta) for the American isolates; Radu Crainic and Jean Balanant (Pasteur Institute, Paris) for the Romanian isolate; Mubina Agboatwalla (Civil Hospital, Karachi) for the Pakistani isolates; Blenda Böttiger (State Serum Institute, Copenhagen) for the Danish strains; Veronika Bubovich (State Environmental Centre, Riga) for the Latvian isolate; Antero Airaksinen (KTL) and Roland Zell (Institute for Virology, Jena) for their fruitful discussions. This work was supported by the Commission of the European Communities (contract ERBCH BGCT93-0360, and partly by ERBIC 15CT96-0912).

## References

**Abraham, G. & Colonno, R. J. (1984).** Many rhinovirus serotypes share the same cellular receptor. *Journal of Virology* **51**, 340–345.

**Agol, V. I. (1997).** Recombination and other genomic rearrangements in picornaviruses. *Seminars in Virology* **8**, 77–84.

**Arola, A., Santti, J., Ruuskanen, O., Halonen, P. & Hyypiä, T. (1996).** Identification of enteroviruses in clinical specimens by competitive PCR followed by genetic typing using sequence analysis. *Journal of Clinical Microbiology* **34**, 313–318.

**Caggana, M., Chan, P. & Ramsingh, A. (1993).** Identification of a single amino acid residue in the capsid protein VP1 of coxsackievirus B4 that determines the virulent phenotype. *Journal of Virology* **67**, 4797–4803.

**Felsenstein, J. (1993).** PHYLIP (Phylogeny Inference Package) version 3.5c. Department of Genetics, University of Washington, Seattle, WA, USA.

**Hartig, P. C., Madge, G. E. & Webb, S. R. (1983).** Diversity within a human isolate of coxsackie B4: relationship to viral-induced diabetes. *Journal of Medical Virology* **11**, 23–30.

**Hellman, U., Wernstedt, C., Gonez, J. & Heldin, C. H. (1995).** Improvement of an 'In-Gel' digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Analytical Biochemistry* **224**, 451–455.

**Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & VandePol, S. (1982).** Rapid evolution of RNA genomes. *Science* **215**, 1577–1585.

**Holland, J. J., de la Torre, J. C. & Steinhauer, D. A. (1992).** RNA virus populations as quasispecies. *Current Topics in Microbiology & Immunology* **176**, 1–20.

**Hovi, T. (1998).** Molecular epidemiology of enteroviruses with special reference to their potential role in the etiology of insulin-dependent diabetes mellitus (IDDM) – a review. *Clinical & Diagnostic Virology* **9**, 89–98.

**Hovi, T., Stenvik, M. & Rosenlew, M. (1996).** Relative abundance of enterovirus serotypes in sewage differs from that in patients: clinical and epidemiological implications. *Epidemiology & Infection* **116**, 91–97.

**Hughes, M. S., Hoey, E. M. & Coyle, P. V. (1993).** A nucleotide sequence comparison of coxsackievirus B4 isolates from aquatic samples and clinical specimens. *Epidemiology & Infection* **110**, 389–398.

**Huttunen, P., Santti, J., Pulli, T. & Hyypiä, T. (1996).** The major echovirus group is genetically coherent and related to coxsackie B viruses. *Journal of General Virology* **77**, 715–725.

**Jenkins, O., Booth, J. D., Minor, P. D. & Almond, J. W. (1987).** The complete nucleotide sequence of coxsackievirus B4 and its comparison to other members of the *Picornaviridae*. *Journal of General Virology* **68**, 1835–1848.

**Kang, Y., Chatterjee, N. K., Nodwell, M. J. & Yoon, J. W. (1994).** Complete nucleotide sequence of a strain of coxsackie B4 virus of human origin that induces diabetes in mice and its comparison with nondiabetogenic coxsackie B4 JVB strain. *Journal of Medical Virology* **44**, 353–361.

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

**King, A. M. Q., Brown, F., Christian, P., Hovi, T., Hyypiä, T., Knowles, N. J., Lemon, S. M., Minor, P. D., Palmenberg, A. C., Skern, T. & Stanway, G. (1999).** *Picornaviridae*. In *Virus Taxonomy. Seventh Report of the International Committee for the Taxonomy of Viruses*, pp. 996. Edited by M. H. V. Van Regenmortel, C. M. Fauquet, D. H. L. Bishop, C. H. Calisher, E. B. Carsten, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. New York/San Diego: Academic Press.

**Laemmli, U. K. (1970).** Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

**Melnick, J. L. (1996).** Enteroviruses: polioviruses, coxsackieviruses, echoviruses, and newer enteroviruses. In *Fields Virology*, 3rd edn, pp. 655–712. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott–Raven.

**Mulders, M. N., Lipskaya, G. Y., van der Avoort, H. G. A. M., Koopmans, M. P. G., Kew, O. M. & van Loon, A. M. (1995).** Molecular epidemiology of wild poliovirus type 1 in Europe, the Middle East, and the Indian subcontinent. *Journal of Infectious Diseases* **171**, 1399–1405.

**Mulders, M. N., Reimerink, J. H. J., Stenvik, M., Alaeddinoglu, I., van der Avoort, H. G. A. M., Hovi, T. & Koopmans, M. P. G. (1999).** A Sabin vaccine-derived field isolate of poliovirus type 1 displaying aberrant phenotypic and genetic features, including a deletion in antigenic site 1. *Journal of General Virology* **80**, 907–916.

**Olive, D. M., Al-Mufti, S., Al-Mulla, W., Khan, M. A., Pasca, A., Stanway, G. & Al-Nakib, W. (1990).** Detection and differentiation of picornaviruses in clinical samples following genomic amplification. *Journal of General Virology* **71**, 2141–2147.

**Padaiga, Z., Tuomilehto, J., Karvonen, M., Podar, T., Brigis, G., Urbonaite, B., Kohtamaki, K., Lounamaa, R., Tuomilehto-Wolf, E. & Reunanen, A. (1997).** Incidence trends in childhood onset IDDM in four countries around the Baltic sea during 1983–1992. *Diabetologia* **40**, 187–192.

**Pulli, T., Koskimies, P. & Hyypiä, T. (1995).** Molecular comparison of coxsackie A virus serotypes. *Virology* **212**, 30–38.

**Pulli, T., Lankinen, H., Roivainen, M. & Hyypiä, T. (1998).** Antigenic sites of coxsackievirus A9. *Virology* **240**, 202–212.

**Ramsingh, A. I. & Collins, D. N. (1995).** A point mutation in the VP4 coding sequence of coxsackievirus B4 influences virulence. *Journal of Virology* **69**, 7278–7281.

**Ramsingh, A. I., Chapman, N. & Tracy, S. (1997).** Coxsackieviruses and diabetes. *Bioessays* **19**, 793–800.

**Rico-Hesse, R., Pallansch, M. A., Nottay, B. K. & Kew, O. M. (1987).** Geographic distribution of wild poliovirus type 1 genotypes. *Virology* **160**, 311–322.

**Roivainen, M., Knip, M., Hyöty, H., Kulmala, P., Hiltunen, M., Vähäsalo, P., Hovi, T., Åkerblom, H. K. & the Childhood Diabetes in Finland (DiMe) Study Group (1998).** Several different enterovirus serotypes can be associated with prediabetic autoimmune episodes and onset of overt IDDM. *Journal of Medical Virology* **56**, 74–78.

**Rotbart, H. A. (1990).** Enzymatic RNA amplification of the enteroviruses. *Journal of Clinical Microbiology* **28**, 438–442.

**Strimmer, K. & von Haeseler, A. (1996).** Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Molecular Biology & Evolution* **13**, 964–969.

**Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997).** The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882.

---

Received 5 July 1999; Accepted 10 November 1999