## A distinct picornavirus group identified by sequence analysis

Timo Hyypiä\*<sup>†</sup>, Christine Horsnell<sup>‡</sup>, Marita Maaronen\*, Mahboob Khan<sup>‡</sup>, Nisse Kalkkinen<sup>§</sup>, Petri Auvinen\*, Leena Kinnunen<sup>¶</sup>, and Glyn Stanway<sup>‡</sup>

Department of \*Virology, University of Turku, Kiinamyllynkatu 13, SF-20520 Turku, Finland; <sup>‡</sup>Department of Biology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom; <sup>§</sup>Institute of Biotechnology, University of Helsinki, Karvaamonkuja 3, SF-00380 Helsinki, Finland; and <sup>¶</sup>Molecular Biology Unit, National Public Health Laboratory, Mannerheimintie 166, SF-00300 Helsinki, Finland

Communicated by Robert H. Purcell, June 5, 1992

ABSTRACT Although echovirus 22 is presently classified as a member of the enterovirus group in the family of picornaviruses, it has been reported to have exceptional biological properties when compared with other representatives of the group. We have determined the complete nucleotide sequence of the echovirus 22 (Harris strain) genome, which appears to be significantly different from all the other studied picornaviruses. However, the organization of the genome [7339 nucleotides, excluding the poly(A) tract] is similar to that of previously sequenced picornaviruses. This genome includes a 5' untranslated region, relatively well-conserved when compared with aphtho- and cardioviruses, followed by an open reading frame coding for a 2180-amino acid-long polyprotein. The amino termini of capsid polypeptides VP1 and VP3 were determined by direct sequencing, and the other proteolytic cleavage sites in the polyprotein were predicted by comparison with other picornavirus proteins. The amino acid identities of echovirus 22 polypeptides with the corresponding proteins of other picornaviruses are in the 14-35% range, similar to those percentages seen when representatives of the five picornavirus groups (entero-, rhino-, cardio-, aphtho-, and hepatoviruses) are compared. Our results suggest that echovirus 22 belongs to an independent group of picornaviruses.

Picornaviruses are small, naked, icosahedral particles containing a single-stranded RNA genome with mRNA polarity. This group includes a number of important human and animal pathogens and has been studied extensively. Currently picornaviruses are divided into five main groups, largely on the basis of physicochemical properties and pathogenesis in the host. These groups are rhinoviruses, aphthoviruses [footand-mouth disease viruses (FMDVs)], cardioviruses (e.g., encephalomyocarditis virus), hepatoviruses [hepatitis A virus (HAV)], and enteroviruses. The latter are subgrouped into polioviruses, coxsackie A and B viruses, echoviruses, and enteroviruses 68-71. Recently echoviruses have been shown to contain genetically distinct subgroups, the major one closely resembling coxsackie B viruses and the minor one, including echovirus 22, exhibiting specific molecular characteristics (1, 2). Echovirus 22 is also known to have other exceptional biological properties when compared with members of the enterovirus group (3-7).

Nucleotide-sequence analysis of picornaviruses has cast light on the classification of the family (8). In most cases, the nucleotide and amino acid identities among the group members correspond well with the previous divisions. For instance, polioviruses and coxsackie B viruses have been shown to be quite homogeneous subgroups and also relatively closely related. On the other hand, HAV, previously classified as enterovirus 72, shows only minimal sequence homology with enteroviruses and, indeed, any picornavirus studied so far (9). Classification of picornaviruses on the basis of genetic properties may soon become essential because their detection in the disease context can be expected to be done increasingly by the PCR, which uses conserved and specific regions of the genome (10). Molecular characteristics are also essential when detailed pathogenetic mechanisms of viral diseases are concerned. We show here that echovirus 22 has a distinctive genome when compared with the presently known picornaviruses, suggesting that it should be classified in an independent group among these viruses.

## MATERIALS AND METHODS

Echovirus 22 (EV22; Harris strain) was obtained from the American Type Culture Collection (ATCC). It was plaquepurified three times and shown to be neutralized by a EV22-specific antiserum (ATCC). The propagated virus was purified in successive sucrose and CsCl gradients, as described (2). Standard methods, including negative staining by 2% potassium phosphotungstate or 1% aqueous uranyl acetate, were used in electron microscopy. Virus RNA was purified and cloned as described (2), and two cDNA clones, together representing almost the entire genome, were sequenced by the dideoxynucleotide chain-termination method. Three strategies were used: double-stranded sequencing after subclone generation by the ExoIII deletion method (11, 12), direct sequencing with oligonucleotide primers (13), and sequencing of templates generated by cloning restriction fragments into M13mp18/19 vectors (Pharmacia). Approximately 98% of the sequence, including the entire capsid region, was determined on both strands, and all the nucleotides were sequenced at least twice. Sequencing of 75 nucleotides (nt) at the 5' terminus was done by primer extension with virus RNA as a template (14). For direct amino-terminal sequence analysis, the capsid polypeptides from purified virus were separated in a SDS/12.5% polyacrylamide gel, electroblotted on a poly(vinylidene difluoride) membrane (15), and subjected to sequence analysis in a gas/pulsed liquid sequencer equipped with an on-line phenylthiohydantoin-amino acid analyzer (16). Sequences were assembled by using Genetics Computer Group (17) and Staden-Plus (Amersham) software; the analysis was carried out by programs GAP, PRETTY, and WORDSEARCH from the Genetics Computer Group package. Picornavirus sequences used in the computer comparisons have been presented in a recent review article (8). Dendrograms were derived by using the KITSCH package of the PHYLIP suite of programs (18).

## RESULTS

Structure and Organization of the Genome. Electron microscopy of purified EV22 revealed spherical particles 28  $\pm$ 

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HAV, hepatitis A virus; UTR, untranslated region; nt, nucleotide(s); FMDV, foot-and-mouth disease virus.

<sup>&</sup>lt;sup>†</sup>To whom reprint requests should be addressed.

The sequence reported in this paper has been deposited in the GenBank data base (accession no. L00675).

1 nm (n = 58) in diameter with the appearance of a picornavirus (Fig. 1). Complete nucleotide-sequence analysis showed that the length of the EV22 genome (Fig. 2), excluding the poly(A)-tract, is 7339 nt, which is in the range found for rhinoviruses and enteroviruses. The base composition (adenine = 32%, cytosine = 19%, guanine = 20%, and uracil = 29%) is relatively similar to rhinoviruses (in HRV14, G+C = 41%) and HAV (G+C = 38%) rather than to enteroviruses (G+C =  $\approx 46\%$ ). As seen in most other picornaviruses, the occurrence of the dinucleotide CG is very low. The predicted 5' untranslated region (UTR) extends to nt 709 and is followed by an open reading frame coding for a 2180-amino acid-long polyprotein. Codon usage analysis shows a preponderance of XXU and XXA (65%), similar to that of rhinoviruses and HAV. The length of the 3' UTR is 90 nt.

The codon that initiates the open reading frame was identified by two main criteria: the occurrence of a 9-nucleotide stretch of pyrimidine residues located 19 nt upstream and the methionine codon being in an optimal Kozak context (AXXAUGG). A polypyrimidine motif is found in all studied picornaviruses, and that seen in EV22 is very similar in sequence, and also in spacing from the AUG, to that of cardioviruses, aphthoviruses, and HAV. Picornavirus VP4 polypeptides, studied so far, have an amino-terminal glycine, and they are myristoylated (8). The first glycine in the EV22 polyprotein is located 13 residues from the predicted amino terminus and appears within a myristoylation consensus motif. It is, therefore, possible a short leader is present in EV22. The putative leader sequence is more reminiscent of the predicted one in HAV, where apparently the true amino terminus of VP4 is seven amino acids downstream, than of the long leader polypeptides found in aphthoviruses and cardioviruses (19).

The genome organization of EV22 is analogous to that of other picornaviruses because amino acid identities can be observed throughout the polyprotein, and this makes it possible to suggest most of the cleavage sites in the polyprotein. For direct detection of some of the cleavage sites, the capsid proteins in the purified virus were separated by SDS/PAGE, and the proteins were electroblotted on a poly-(vinylidene difluoride) membrane. Protein bands corresponding to apparent molecular masses of 30 kDa, 30.5 kDa, and 38 kDa were subjected to amino-terminal sequence analysis. The 30-kDa and 30.5-kDa proteins gave the results APNGK-KKN and NSWGSQMDL, respectively, unambiquously localizing these structures on the deduced sequence (VP3 and



FIG. 1. Electron micrograph of purified echovirus 22. (×300,000.)

VP1). The largest polypeptide repeatedly gave no result in Edman degradation, and no signal representing the predicted amino terminus of VP2 was seen in any bands analyzed. Tryptic peptide analysis of the 38-kDa band generated sequences localized in the predicted VP4 region (VESVG-NEIGGNLLTK), in the VP2 region (NVVQATTTVNT-TNL), and spanning the predicted VP2/VP4 cleavage site (VADDASNILGPNXFATTA). Together with the inability to find any trace of VP2 sequences in the other capsid polypeptide bands, this result strongly suggests that no VP4/VP2 cleavage occurs, and VP0 is the predominant species in mature virions.

In the polypeptide region, homologous to the VP2 protein in other picornaviruses, there is an apparent deletion in the "puff" region (20) when compared with entero-, rhino-, and cardioviruses, making the EV22 polypeptide structurally more similar to those of aphthoviruses and HAV. VP3 is larger than in other picornaviruses, mainly from a hydrophilic extension of  $\approx 20$  amino acids at its amino terminus. An interesting finding in the VP1 polypeptide is an RGD sequence close to its carboxyl terminus. Such a motif is known to be involved in cell attachment in various systems (21).

Of the nonstructural proteins, some (e.g., 2C, 3D) can easily be aligned with the corresponding polypeptides of other picornaviruses, whereas others (2A, 3A) exhibit only minimal identity (Table 1). 2A is an important protein to consider because the strategy of its action divides picornaviruses into two categories: in enteroviruses and rhinoviruses it is responsible for an autocatalytic cleavage activity directed to its amino terminus, whereas in aphthoviruses and cardioviruses, 2A cleaves at its carboxyl terminus (8). However, deducing the mode of action of EV22 2A from the predicted sequence is not possible. The function of polypeptides 2B and 3A in virus replication is still poorly understood, and they have a relatively low degree of conservation. Although the function of polypeptide 2C has not yet been completely elucidated, it is one of the most conserved proteins among picornaviruses. As this protein aligned well with other picornaviruses, polypeptide 2C was used in detailed comparative analyses. The localization of polypeptide 3B (VPg) was predicted by the genomic position, presence of the conserved tyrosine (Y) as the third amino acid, and comparisons with other picornavirus VPgs. The 3C protease contains the conserved sequence GXCGG, proposed to form part of the active site. Previously observed, highly conserved motifs (KDELR, YGDD, FLKR) are found in the 3D polymerase.

**Relationship of Echovirus 22 to Other Picornaviruses.** The nucleotide identities of the 5' and 3' UTRs and the amino acid identity of the predicted polypeptides were compared with representatives of picornavirus groups (Table 1). In parts of the 5' UTR, EV22 RNA exhibits a high degree of sequence conservation relative to aphtho- and cardioviruses; for instance, there are several blocks of, at least, an identical 10 nt. These viruses, together with HAV, share a similar predicted 5' UTR secondary structure, which is distinct from that of entero- and rhinoviruses (8). The identity of the EV22 polypeptides with corresponding proteins of other picornaviruses varied between 14 and 35%; this value is of the same order as that seen when HAV is compared with other picornavirus groups (9).

Because the most pronounced identity was seen in polypeptide 2C, this protein was analyzed in more detail. When its amino acid sequence was compared with the National Biomedical Research Foundation data base, this EV22 polypeptide was found to be more similar to the 2C protein of 22 other sequenced picornaviruses than to any other proteins in the data base (data not shown). In a dendrogram, illustrating the relationships between picornaviruses, EV22 and HAV join to the other major groups at an amino acid-difference level exceeding 70% when polypeptide 2C is used for com-

170 50 SCCTAGGGCCTTATAACCCGACTT 210 110 130 10 CCTATGCCTGGTCCCCACTATTCGAAGGCAACTT 330 350 CCTAAGACAATTTGATCAA TTTGAAAG 210 ATGCTTAAAGCATAGTGTAAATGA 390 TGCTTAATAGCATCCTAGTGCCAG 270 TGCCAT 450 290 310 230 GCAATAAGAAGA CTTATG CCTCTAGT TGTCTGGTGAG TGTG 470 490 510 530 410 4 30 470 ATTAGG 650 NAACGI ATTCAGT TAGTGGTGC GGAACTTATTGGAA 670 690 710 570 590 610 6 30 ATTATGGAGACAAT TAGTTGTAAGGCCCACGAAGGATGCCCAGAAGGTACCCGCAGGTAACAAGAGACACTGT TCTGATCTGG CTATCAGGTG 810 830 7 30 750 770 790 TAAGAGTATTGCAGATATGGCTACTGGTGTTGTAAGTTCTGTTGATTCTACCATCAA CAATGAAAA TCGGTAATGAJ GTAATTTGTTAM AGTAGCTGATGATGC VPO TACTOTIANTA KSIADHATG 910 V D S T I 950 NAV N E K V E S V G N E I G G N L L T K V A D 970 990 1010 D A 1030 SNILGP 1050 GAACAAGAATGTAGTTCI CTACTA CTTGACTCAA GCACCTACGATGC TCACCAGATTTCT GTCGACAATTTTCACTCTATG TTATGACATCACAACTG T G D K N 1230 GGACAAAAATCCCAG N V 1090 Q A T T T v v D I L T Q РТ 1150 P D 1170 F D N F 1190 8 мач т N P 1210 CAAGTTAATGTG GGTACTGCAGGCAGTGCCTT Q v Q GΤ A G 1430 GSAL E W T P S W A R G Y Q I 1270 1290 ¥ E L 1310 B ٥ A Y 1350 1410 1330 ACTTAGGGCAGTTAAA GGTAGTATATGAACCTAA GTTGTCACTTA CTGGAATT ATCTTC ATGAATTT GQLK v D L G 1610 Y E P 1450 KLEF NL ٩ 1590 1470 1490 1510 1530 GAAAATTATGACCAT AGTCTATGTCTGGACTCCATTGTCAATCCCCACTGGATC TACTOGOTAC TTAGACTTO STATTO T K I 1750 NA VP3 1770 T LSIP т G N Q V 1670 L GSLLQLDP 1690 1710 Q V F A 1730 Q N I 1630 1650 1690 TGATCTTGTT GCACAAAATACA GAACAAAAATC MANGG CCTTTGIAL ACAACTGGTGCTC GGCTTTAGTAGG TTTCTATGATCCTAGA GCAGT KYKWTRTK 1810 1830 D I A E G P 1850 V A L V 1910 TAGS . D D L 1970 Q \$ G ₽ R D 1870 1930 1950 1990 TGGAGTTGATGCAAAAGGATATTTCAAGTGGTCAGCAACAA G V D A K G Y F K W S A T T 2050 2070 TTGTCTACTTAAGACTATTTCCAAATTTGAATG TAACAGTTATTCATACTT AGCCCAGTTATTCTCAGTGATGGCAGATTCAA A Q L F S V M A D S T 1990 2010 TAPQSIV 2090 TTPSENE 2030 N S Y 2150 RNIV L L F P N L 2130 YSYF 2110 AGGTAGTUM TACAATCCCTTACTCTTT CTTCTTCCCCAATGCCACAACGGACTCAACTTCCACCTTGG FFPNATTDSTSTLD 30 2250 2270 ACAATATGT GTTCATTAGTTTT GATTGAGCGTCTA TTTCAATA GATGGGCTTCTT I T I P 2330 RG SLVLRLSVYASTFNRGRLRM 2170 2190 2210 N A Y T I C 2290 22 30 GTTGTTACATTCCAGAATTC V V T F Q N SP1 2510 VP1 TTCCACTTGGATGAG AATGGCCACCC TGTTTCAAATTGJ GCTCACATACAACAC TCTAGTCCATC TATTGTATAGTCC ARFFCP 2490 Q G K M G Q D 2470 S T MRKTNGBP IGLFQI 2390 E NRL TYNSSSPSEVYCIV 0 2430 2450 W 1 2350 2 370 2410 TCAAGAAAAG ATGGGGTTCACAGA AACTAGGACTCA TATTTTCTCAATTT GATGACAC CARACANTOT CCANGATGATG R S M N V D I F 2690 W G S 2530 QMDLT D P 2550 L E D D T 2570 Q T N 2590 E L G L 2610 S A Q D D 2630 GPLGQ E K 2650 Y F L N 2670 1 FFME 2750 TATCCATGTTCTGTTCCTAAG TGTATCACATACTA CANTGAGO TATCACTGTTGTTTG ATTTIN F T G 2050 ATTTTACTGGTGAAG ACCTATTO N E 2770 т в у LS т D N L F 2730 2790 Q G 2810 G LS L L L F 2710 2870 2830 TAGTCTTGGTTATTTGAT TGAGAGGGGGTTTCTGAGGGTTGCACACAC E R G F L R V A H T 2090 2910 LGYLM 3050 s S N D R 2930 3010 30 30 P Q N R M AGCCCTTCTT CCTCTACTG GTTTATCTT CANATTTCT TCCTGCCCC N Q 32 2A 32 30 CKP FL-T G L A P 3150 S R 3170 3070 3090 3110 31 30 GATGAAACTAGCATATT ATAGAGGTTTCT TTATGGCATTATAG GTACCAACT TGACATTT ATTGACAGG TACTAAAAC S D W 3390 MKL KTALTGKA 3350 IEEE C E L D Y 3410 A Y L D R G F 3270 Y G I I 3290 Y Q 3310 L D D т 3370 32 50 3330 CTTTAGAATCAAATAT TAGAATCAAATATCTTGAATLANGAAT R I K Y L E S A V D S 3430 3450 TAATGCTGTGACAATAAAAGAATTCTTTAATCA GGATTCAGAGCACATTTTTTCAG D S E B I F S V 3470 TACACTAAGCCAACATC AGGTTAGTGO TTATTGACTGC TATGTCAACTATAAAAACCCC ATATTTTTGG E T 3490 т т IKTP N C IAKDI G THTL S Q 3530 A I G L LL A s 3510 3550 3570 3590 3430 AGTTAATGCTGTGACAATAAAAGAATTCTT V N A V T I K E P P 3630 TAATCATGCCATAGATGGCGA N B A I D G D 3650 ATTCTTTTCTTCTGCTGCAACTO FFSSAAT 3710 ATACTITGCTACATGGT I L C Y M V 3770 TGAGATTTTTGGATAATGATT E I L D N D I CTTGCTTGTACAAAAATGTACTACA L L V Q K C T T 1 3670 3690 ATAGTGAAA A I D G D E Q G L S L L V 3650 2B 3670 TIGCTTATCCACICTITTAATAATGGAIGTTACTICCT KFIV I 3750 GCTGANATT TGTGGCAGAATCAATGTCCAACACTGATGA V A E S M S N T D D 3950 ACTITATTGCCATAA TAAGAAACTGGC.... K K L A E I 3930 CTAACATTTTGACTAC LYCHKPNILT 3790 3810 3950 IGTITICAAACCCAGCATAGA V F K P S I E 4130 ATCAAAGATATG ATGAAGTTTC TGATGAAGTTAAG AT A T G T G A C A C A G T A A A G T A T A C T A A G A C C A T C A A A T C A A G G A C C T T T T A A TAGOCATATAG GATACATACATTACTC DEVKEQICDTVKYTKTI 3970 3990 4010 KGFNEVSTAFRHI 4050 4070 D W W I B T L L K I K D M 4090 4110 LS GCAAGGCTATTC GARATA TTGTAGTATCCT CAAAAGACCAAT AGACTCAAGE AAGGTATTCAG GCCAAGTTCA CATGGCAATTTGCTTTAGAAG Y Q R Y 4270 H A I C 4310 I Q W L E R N K E H V C S I 50 4170 4190 D I 4210 K T Q D F 4250 FRS S K A 4150 L D A S D Q D CLAKF 4290 42 30 TCTTTCTTGACCC TCTCCAGACAATTACAAAA CTGCCACAACAGTA AAAA101. KSFL 4470 G I 4430 LS R Q L Q K 4490 s V Y 4350 QE L A 4370 Q G E G 4330 ..... 4410 4450 ATCATGTAAGCT S C K L 4510 LE T N 4530 D D 4590 K E 4610 4650 M A 4670 H S E F M 4550 L 4630 ANGCCAGGCCATGGC S Q A M A 4850 AGTTA T S K L 4710 E K G K F 4690 BIR N K 4730 A 4770 L P Y 4790 IM AACAATGTCAACTO GAAGTATO AATTACAAT CTAGATGATGCTGTATC T M S T G E C W E V 4870 4890 5 K N G R D W 910 K D K I I E R V 4970 K N Y N A LENQT D L D D A V S 5030 4930 4950 4990 TTCAAAATTTCTAA TAGGGAAGTTGTGGACTG AACATTAA ATACATTAAGCACA P D A I 5070 GATGCCATACCATAT TGAGTATCTTAA ATCOLOCTT CTTTGCT T L 5110 R E V 5210 H N DEYL 5090 M E A 5130 ĸ P \$ с S K A \$ I Q IE 50 50 5190 5170 CACCAGTGCTATTAGTA T S A I S I 5310 CCGGTTGCCJ PVAJ 5390 TTTCAGATAAAATTA AGAGCTTATAATCCTA T K 5330 I F K K E E S 5350 S D K I K S M L S 5230 5250 V E R N K A 5270 5 A 52 90 ILLLV CACTTTCCCM GGAGTTCA CATCATT CATATATTA TCCTTATGAT G T F P V S Q R E F K N E A P Y D 5410 5450 5450 5450 GQLEE IISQMAYI 5470 5490 TGS тт G H 5510 м ACAAGAAGATGAATTGACACT TCCCAATTGAAC CCAAGTG GTAAACCAAT ATTCTCAAGTGTA NGTTGCCTTTCAGGTTCAAGAAAAACTCTAAATATTACACCAACAAGATTGGAACTGAAAGTAT CAAGAAAAALILII KKNSKY 5730 Y T N K I G T 5750 ELTLHYK 90 5610 Q E F P I E 5630 QVTLGGKPM 5650 5670 A I CKLPF ESM D 5590 Q νт D L K 5690 R F 5710 GCTAATCTGGATGAC GCATTATTA TCTTGCAAAGGAATGTG GTGGCCTACTTATT GAAGGTAACTTCAAAAT CCAAAGAGT GCATCA GAACTGAAAG ATCAGTTATA M T E Q G I I 5790 W 1 5770 T E 5850 G LI Q R V 5810 G S Y 5870 с K 5890 GGLL 5910 Е G N 5930 F 5830 D M S AGGCATTGTTAC CTGTTGAGGT CCTGGGTATGCATA ACA AGAGO M H 5950 D Q G IA I CAGCTGTCCT GTCCTAAGIANGE GAATGAATGGTATTGT ITU. 5 K ± 6270 G P A 1 6130 R Е Р V E 6170 с SASKYR 6190 V N 6210 N N E 6230 LW Õ C K A 6250 с G M N G I 6290 TGACATGAAAACTGCCATTTTAGGAACTTCTCAT ATTCTATGGATTTGAG GGTTATAGTT ATCTAATTTGCTT TTCTCAGTAGC GAGAGGCTTGT AATTCCACAACTTACTAAA TAILGT M D L 6350 S F Y K H N L L K 6470 8 s G D L I C 6410 L E R L 6450 6310 6330 6370 6390 6430 AATCAAATA T T T F N T 6510 TTTATGAC ACTCCATGTTACTACTC T P C Y Y S MEIYD 6630 Q I 6490 LKDE L R 6530 A S D Y 6590 ΥΩΤ P C 6650 LDKI K T R C I 6570 A C E V C I V Y R G 6610 6610 6630 ANTGTTGCTATGGGAGCTGTGGAGCTTACTGCTACTG N L L W E A V E V L A Y C E D 6790 6610 ITTGATGGTGTTTTTACCTACAATTTAATCAGTCCTGGAA L N H C I Y T N L I S P G I 6970 5970 5990 6590 RGACTATTTCTCAATATGATGGCTCCCCTTAGTTCAATG D Y S Q Y D G S L S S H 1 6750 6770 6490 5310 Aggactagctgttggaatcaaccaatacaagat G L A V G I N P Y K D 6670 6690 6650 TTCACCTGATCTCG S P D L V 6830 GCACTTCATGATTAATG E F M I N A 6710 TGATTACAATTATGAAAT D Y N Y E M 6730 M Q L H L L W E A 1 6790 A L GCACAAACCAGTCATTGACTCTGACCATGTTGT H K P V I D S D H V V 6850 6870 TCACTATGCAAT GCACAAACCAGTCATT GATGGTTGA TGATTGTCTACCAATCGTATATGG TATOCCATCA S G S P C T T 6930 CCATGTACTACTGTC V L N S L C N 6950 VFNER I B G G M P 6910 GIDCLPI 7010 VYG W L 6890 CGATGATGTTATTCTT CATTGGACAAAGAAATTG GAATTCTTAAAGCGGAAACCTGGTTATTTCCCAGAGTCTACTTT AGAAACTGCAAAGT GATTCATTT GTGACTOGGTC. ATGAACCTCCTTC CAAGAATGGA M A D S F 7090 D D V I L S L D K E I E P E K L Q S I 7030 7050 7070 G A E V T G S R K D E P P 7110 7130 SLKPRMEVEFLKRKPGYFPESTF 7150 7170 7190 TATAGTAGTAAATTAGACACTGAAAATATGATACAACATTTAATGGGGAAAAACTTTAGCAC I V G K L D T E N M I Q H L N W M K N F S T 7210 7250 7250 ATTANGCAGCAGCTOCATCTATTAATGGAACTATGCCTCCATGGAAAAGACACTATCAACAACTATCAATAAGATCTAGGAACCATATCAACAGCATGGAACAATATCACTGG PKQQLQSYLNELCLBGKDYYQBYIKILEPYLQENNITV 7270 7290 7330 TAGCACATTTAAGC 

FIG. 2. Nucleotide sequence of cDNA representing echovirus 22 genomic RNA. The predicted amino acid sequence of the polyprotein and suggested cleavage sites, determined by alignment with known picornavirus polypeptides, are indicated. The amino-terminal sequences of VP3 and VP1 were verified by directly sequencing the polypeptides.

Table 1. Comparison of nucleotide and amino acid (coding region) identities of echovirus 22 with representatives of other picornavirus groups

Virus	Identity with echovirus 22, %												
	5'UTR	VP4	VP2	VP3	VP1	2A	2B	2C	3A	3B	3C	3D	3'UTR
Polio 1	38	24	22	19	22	14	21	30	15	35	24	27	36
HRV1B	35	20	22	17	15	15	19	29	19	30	23	26	48
FMDV	46	20	23	20	23		15	29	14	25	18	25	29
EMCV	43	20	24	22	20	18	23	29	20	30	18	25	34
HAV	36	19	18	23	21	17	27	25	29	35	19	27	44

These identities were calculated by using the GAP program of the Genetics Computer Group package. Polio 1, poliovirus 1; HRV1B, human rhinovirus 1B; FMDV, FMDV strain  $O_1K$ ; EMCV, encephalomyocarditis virus. In the 5' UTR optimal alignments were used for comparison.

parison (Fig. 3). This is clearly a lower degree of relatedness than, for example, between the major enterovirus/rhinovirus group (51%), and this value also slightly exceeds the value seen between these two groups, and aphtho- and cardioviruses (68%). Very similar results are obtained when the region, corresponding to the VP2 polypeptide in other picornaviruses, is used for comparison (data not shown).

## DISCUSSION

Classification of viruses is mainly based on virion morphology, the nature of the nucleic acid, and replication strategy. Other biochemical and antigenic properties are also used, and they are important in identification and grouping of individual virus types and strains. Optimal classification should combine the essential biological properties of the group and, in addition, give an adequate tool for diagnosis and treatment of clinical diseases. Introduction of diagnostic procedures based on nucleic acid sequences has recently increased the need for precise classification of viruses according to their molecular characteristics.

Echoviruses were discovered soon after the first tissueculture techniques were introduced into laboratories (22). These viruses were isolated from human intestinal tract, as are polioviruses and coxsackieviruses, but they did not share the major pathogenic properties of these groups. Because the association with human disease was, at the time of their discovery, unknown, they were given the name ECHO (enteric, cytopathogenic, human, orphan) viruses. It was originally postulated that whenever an echovirus was estab-



FIG. 3. Dendrogram of the relationships between picornavirus groups based on comparisons of amino acid differences of 2C polypeptides. FMDVO, FMDV strain  $O_1K$ ; FMDVA, FMDV strain A-10-61; TMEV, Theiler's murine encephalomyocarditis virus; EMCV, encephalomyocarditis virus; HRV, human rhinovirus; Polio, poliovirus; CAV, coxsackie A virus; CBV, coxsackie B virus; SVDV, swine vesicular disease virus; EV, echovirus; eV, enterovirus; BEV, bovine enterovirus.

lished as the etiologic agent of a clinically distinct disease, it would be removed from the group (23). Later on, it became evident that individual echovirus types cannot be directly associated with individual illnesses but rather with a wide spectrum of clinical manifestations. This is why they still are classified together, forming the largest of the enterovirus subgroups. However, from the original 33 echovirus serotypes, types 1 and 8 were closely enough related to be combined, type 10 was shown to be a reovirus, and type 28 appeared to be a rhinovirus.

Echovirus serotypes 22 (including the prototype 101 Harris virus) and 23 were isolated during a study of summer diarrhea in 1956 (3). Although these viruses shared the general properties found in the group, the cytopathic destruction of cell cultures was incomplete when compared with previously characterized echoviruses. It has been reported that cytologic changes caused by echovirus types 22 and 23 involve distinctive nuclear manifestations (4, 24) and that the replication of these viruses is not inhibited by 2-( $\alpha$ -hydroxybenzyl)benzimidazole, in contrast to other members of the subgroup (25). There is also evidence of exceptional secondary structure of EV22 RNA (6). Recently EV22 infection has been shown not to cause the host-cell protein synthesis shut-off (7, 26) characteristic of enteroviruses.

In addition to the biological characteristics, the epidemiology and pathogenesis of EV22 infections also differ when compared with the other echoviruses. Perhaps the most striking difference is that 60% of the EV22 isolations are from children of <1 yr of age (17% in all the echoviruses), and altogether 92% of the isolations are from children under 5 yr of age (5). From the isolation-positive cases, 29% had gastrointestinal symptoms (11% in all the echoviruses), and only 11% had symptoms indicating involvement of the central nervous system (56% in the entire echovirus subgroup).

Previous work has, thus, given evidence that EV22 differs biologically from other echoviruses. Our data indicate the molecular basis for these characteristics. Although clearly a picornavirus in terms of electron microscopy and genome organization, EV22 has a very divergent nucleotide and amino acid sequence. These characteristics locate it well outside the enterovirus group and, together with the other biological data, suggest that EV22 belongs to another picornavirus genus.

The diverse nature of the capsid proteins, when compared with other picornaviruses, made prediction of some cleavage sites relatively difficult, and, therefore, the VP2/VP3 and VP3/VP1 cleavages were established by directly analyzing the amino-terminal sequence of polypeptides VP3 and VP1. The amino-terminal sequence of VP2 protein could not be determined, and because VP4 can be myristoylated in picornaviruses, one explanation for lack of a signal is that no cleavage occurs between VP4/VP2 during maturation of EV22 virions. This alternative is strongly supported by the results from direct amino acid analysis of the tryptic peptides from the 38-kDa band, and these data suggest that capsid

structure of EV22 differs from other picornaviruses in this respect. Sequence comparison gives some evidence of conserved amino acids in the capsid proteins of EV22 (data not shown), indicating that they probably share the typical threedimensional structure found in picornaviruses, an eightstranded antiparallel  $\beta$ -barrel. The unique amino-terminal extension of  $\approx 20$  residues in VP3, contains eight positively charged amino acids and may have important functions in the virus structure. The observation of an RGD motif near the carboxyl terminus of EV22 VP1 may give some indication of the type of the cell receptor used by the virus. In aphthoviruses (FMDV) the carboxyl terminus of VP1 is known to project onto the virus surface and participate in receptor binding in conjunction with an RGD motif located elsewhere in the VP1 primary sequence but juxtaposed structurally (27). An RGD at the carboxyl terminus of VP1 has recently been implicated in receptor binding of coxsackievirus A9 (28, 29), and a comparison with strains of this virus and FMDV shows that the EV22 motif (RGDMANL) conforms well to a consensus that can be derived (RGDM/LXXL; ref. 30).

The 2A protein, known to be associated with a cis-acting proteolytic activity in other picornaviruses, does not conform in sequence to either of the two well-characterized types (31). In entero- and rhinoviruses, the 2A protein acts autocatalytically at its amino terminus and is also involved in cleaving the p220 component of the cap-binding complex, shutting off host-cell protein synthesis. This protein is similar to the picornaviral 3C proteases, as it is also a member of the trypsin-like proteases in which the usual active-site serine has been replaced by cysteine. Such proteases have a catalytic triad made up of histidine, aspartate/glutamate, and serine/ cysteine residues, the latter being found in the context GXCG, which is well-conserved between these proteins from a wide range of sources. EV22 2A is similar in size to that of entero- and rhinoviruses, and there are a few short regions of amino acid homology. However, the critical GXCG motif is absent, and the protein is unlikely to function exactly in the same manner. This protein is also dissimilar in sequence from the 2A protein of aphthoviruses and cardioviruses. In these viruses, 2A protein cleaves at its carboxyl terminus and does not participate in host-cell shut-off, which, in aphthovirusinfected cells, is accomplished by the proteolytic activity of the leader. The cardiovirus leader protein has no known proteolytic function, and this virus does not shut-off host-cell synthesis by inactivating the cap-binding complex, which also happens in EV22 infection (7, 26). Further insights into the role of 2A protein in EV22 replication must await direct analysis of its proteolytic capability.

In view of the highly divergent nature of the rest of the genome, the similarity of the 5' UTR to cardio- and aphthoviruses is somewhat unexpected. The observation raises the possibility that EV22 is a recombinant virus or, alternatively, that the 5' UTR evolves more slowly than other genome regions. Predictions suggest that similarity of this region is reflected in secondary-structural terms, and this may play a role in the observed viral properties. For instance, because the 5' UTR is vital in translating the virus polyprotein, this similarity may explain the efficient translation of EV22 RNA in rabbit reticulocyte lysate (7) also seen in cardio- and aphthoviruses, and contrasting with the poor translation of entero- and rhinovirus RNA in this system. As suggested previously by hybridization experiments (7), the EV22 5' UTR lacks the poly(C) tract seen in aphthoviruses and encaphalomyocarditis virus but absent from Theiler's murine encephalitis virus, another virus that shares the common 5' UTR structure. The 3' UTR has been shown (2) to have an exceptional predicted secondary structure when compared with enteroviruses.

In conclusion, our data demonstrate that EV22 has a structure and genome organization corresponding to other picornaviruses. However, its similarity to other sequenced members of this virus family is low. This similarity is approximately at the same level as that observed for HAV (9), originally classified as an enterovirus but now assigned to the hepatovirus genus of picornaviridae. The molecular data presented here suggest that classification of EV22 should also be reconsidered. Although this virus is already known to be found in young children and is often associated with gastroenteritis, the role of EV22 virus in clinical disease needs careful reevaluation against the background of its distinct molecular properties. As soon as these data and the identity and characteristics of possible relatives of EV22 are known, it would be possible to give the putative additional genus a nomenclature that corresponds to that of other picornavirus groups.

We thank Dr. Carl-Henrik von Bonsdorff for his advice in electron microscopy and Drs. Pekka Halonen and Tuija Pöyry for their comments during manuscript preparation. The study was supported by grants from the Sigrid Juselius Foundation, the Turku University Foundation, and the Medical Research Council.

- 1. Auvinen, P., Stanway, G. & Hyypiä, T. (1989) Arch. Virol. 104, 175-186.
- 2. Auvinen, P. & Hyypiä, T. (1990) J. Gen. Virol. 71, 2133-2139.
- Wigand, R. & Sabin, A. B. (1961) Arch. Gesamte Virusforsch. 11, 224-247.
- Shaver, D. N., Barron, A. L. & Karzon, D. T. (1961) Proc. Soc. Exp. Biol. Med. 106, 648-652.
- Grist, N. R., Bell, E. J. & Assaad, F. (1978) Prog. Med. Virol. 24, 114-157.
- 6. Seal, L. A. & Jamison, R. M. (1984) J. Virol. 50, 641-644.
- Coller, B.-A. G., Chapman, N. M., Beck, M. A., Pallansch, M. A., Gauntt, C. J. & Tracy, S. M. (1990) J. Virol. 64, 2692–2701.
   Stanway, G. (1990) J. Gen. Virol. 71, 2483–2501.
- Cohen, J. I., Ticchurst, J. R., Purcell, R. H., Buckler-White, A. & Baroudy, B. M. (1987) J. Virol. 61, 50–59.
- Hyypiä, T., Auvinen, P. & Maaronen, M. (1989) J. Gen. Virol. 70, 3261-3268.
- Murphy, G. & Kavanagh, T. A. (1988) Nucleic Acids Res. 16, 5198.
   Herikoff, S. (1984) Gene 28, 351-359.
- Herikoff, S. (1984) Gene 28, 351-359.
   Zagursky, R. J., Baumeister, K., Lomax, N. & Berman, M. L. (1985) Gene Anal. Tech. 2, 89-94.
- Huovilainen, A., Kinnunen, L., Ferguson, M. & Hovi, T. (1988) J. Gen. Virol. 69, 1941–1948.
- Speicher, D. W. (1989) in Techniques in Protein Chemistry, ed. Hugli, T. E. (Academic, San Diego), pp. 24-35.
   Kalkkinen, N. & Tilgmann, C. (1988) J. Protein Chem. 7, 242-243.
- Kalkkinen, N. & Tilgmann, C. (1988) J. Protein Chem. 7, 242–243.
   Devereux, J., Haeberli, P. & Smithies, O. (1984) Nucleic Acids Res.
- 12, 387–395.
   18. Felsenstein, J. (1985) Evolution 39, 783–791.
- Palmemberg, A. C. (1989) in Molecular Aspects of Picornavirus Infection and Detection, eds. Semler, B. L. & Ehrenfeld, E. (Am. Soc. Microbiol., Washington), pp. 211-241.
- Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H.-J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., Rueckert, R. R., Sherry, B. & Vriend, G. (1985) *Nature (London)* 317, 145-153.
- 21. Ruoslahti, E. & Pierschbacher, M. D. (1987) Science 238, 491-493.
- Melnick, J. L. (1965) in Viral and Rickettsial Infections of Man, eds. Horsfall, F. L., Jr., & Tamm, I. (Lippincott, Philadelphia), 4th Ed., pp. 513-545.
- 23. Committee on the ECHO viruses (1955) Science 122, 1187–1188.
- 24. Jamison, R. M. (1974) Arch. Gesamte Virusforsch. 44, 184-194.
- 25. Eggers, H. J. & Tamm, I. (1961) J. Exp. Med. 113, 657-682.
- 26. Coller, B.-A. G., Tracy, S. M. & Etchison, D. (1991) J. Virol. 65,
- 3903–3905.
  27. Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D. & Brown,
- F. (1989) Nature (London) 337, 709–716.
   Chang, K. H., Auvinen, P., Hyypiä, T. & Stanway, G. (1989) J.
- Chang, K. H., Auvinen, F., Hyypia, T. & Staliway, G. (1967) J. Gen. Virol. 70, 3269–3280.
   Devicing M. Hurging T. Birgingan I. Kolkking N. Stanway.
- Roivainen, M., Hyypiä, T., Piirainen, L., Kalkkinen, N., Stanway, G. & Hovi, T. (1991) J. Virol. 65, 4735–4740.
- Chang, K. H., Day, C., Walker, J., Hyypiä, T. & Stanway, G. (1992) J. Gen. Virol. 73, 621–626.
- 31. Harris, K. S., Hellen, C. U. T. & Wimmer, E. (1990) Sem. Virol. 1, 323-333.