A distinct picornavirus group identified by sequence analysis

TIMO HYYPIÄ*[†], CHRISTINE HORSNELL[‡], MARITA MAARONEN^{*}, MAHBOOB KHAN[‡], NISSE KALKKINEN[§], PETRI AUVINEN^{*}, LEENA KINNUNEN[¶], AND GLYN STANWAY[‡]

Department of *Virology, University of Turku, Kiinamyllynkatu 13, SF-20520 Turku, Finland; [‡]Department of Biology, University of Essex, Wivenhoe Park, Colchester C04 3SQ, United Kingdom; §Institute of Biotechnology, University of Helsinki, Karvaamonkuja 3, SF 00380 Helsinki, Finland; and NMolecular 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. \parallel

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, PRErrY, 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.

To whom reprint requests should be addressed.

IThe 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. $(X300,000.)$

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-

50
AGAGCTTGCCTAGGGCCTTATAACCCGACTTGCTGA 230
GTGTAAATGATCTTTTCTA 410 CTAGTOCCAGCGGAACA 590 590
TAGTTGTAAGGCCCACGAAGGATGCCCAGAAGGTACCCGCAGGTAACAAGACAO 770 TAAGAGTATTGCAGATATGGCTACTGGTGTTGTAAGTTCTGTTGATTCTACCATCA ^S V ^D ^S T ^I 950 'ACTACCAACTTGACTCAA T T N L T Q
1130 'ACACATGTAGAATTACC'I V E L P 1310 **CTGGAATT** K L E F 1490 TCTGCTAACCAAGTTGAI
S A N Q V D
1670 rATAGCAGAAGGTCCTG ^I A ^E ^C ^P ^C 1650 VCCATCAGAAAATCATGG P ^S E N G 2030 **TTTCAATAG** S R G S L V L R L S V Y A S T F N R G R L R N G F F
2170 2190 2210 2210 2230 MATTC<mark>aattcu</mark>
Eqiie C L F Q ^I E 2390 GAAGATGACAC
E D D T ^I E D D S 2570 WSCATGCTTTTTTATGGA A ^F ^F ^E 2750 ACTAGTAATGATCGAGTO ^S N D R V 2930 GTTTATCTT
V Y L 70 GCTTCTATAGGAAAAAACCCT 250 TAACCTGTATTATGTA<mark>CAG</mark>G 430 ACATCTGGTAACAGATGCCTCT 610 **ATCTGATCTGG** 790 CAATGCAGTCAATGAAAA
x a v n e k N A V N E V 970 **ACATCCATCGGCACCTACGATGCC** ^H ^P S A ^P T 1150 TAAACTTTTTTGGOATCACCAJ **X V F W D** 1330 **MICTTCC** A F T ^N L ^P 1510 TACTGGGTAG
LGS V ^T ^I L G ^S ^L 1690 TTCAATGAACATGGCAAATGTC S N N H A N V 1870
Act**tgatgCaaaaggatatttc**
V D A K G Y F
2050 GATGGGCTTCTT 2230 **GCTCACATACAACAC** $N = R - L$ 2410 CAAACAATGT N C K Q T H S
2590 CCATACTTTCACCAATGACGJ ^T ^F ^T ^N ^E G 2770 CAATTTTCTGTCATCGAACGG' N F L ^S ^S N 2950 CCAAATTTCT 90 TTCCCAGCCCT0G00CTGTCAA 270 OGCAGATGSCGTSCCATAAATCTATT 450 GGGGCCAAAGCCAAGGTTGACA 630 **CTATCAGGTG** 810
TCGGTAATGAJ E S V G N E I G G 1
990 ;
S P D F S **P S P D F S**
1170 V GACaAGCccGCATATGGACAATCACO GACAAGCCCGCATATGGACAATCACL
D K P A Y G Q S R
1350 at<mark>caattt</mark>g $N - L$ 1530
T**TAGACTT**CI ;TTGCAATTAGACTTCCAAAATCCAAG L ^Q L ^D F ^Q ^N ^P ^A 1710 CTTTGTACAACTOGTGCTCAATCAGI L C ^T ^T ^C 1690 ACAACTGGTGCTCAAT :AAGTOGGCACAACAACAOCCCCCAJ OSA T A ^P ^Q 2070 CCCAATGCCACAACGGACTCAACTTC P N A T T D S T ^S 2250 AACAGTTCTAGTCCATCGGAGGTTTATTGTATAG
N S S S P S E V Y C I V
2430 2450 **:AACTAGGACTCACTTCAGCCCAAGATGATG** P N E L G L T S A
2610 LCAATGGAGAGTGCCATTGGAATTTCC Q W R V P L E F P
2790 IGTAATAACTGTACCAGCCGGAGAGCC V ^I ^T ^V ^P ^A ^C ^E ^Q 2970 TCCTGCCCC
P A P
3150 110 rAAAAACCCCCATAG 290 To 470
CCATTAGGATTGGTT
650
<mark>AAAAA</mark>CGTCTAGTGG 630 AGAAAGCGTCGGTAATTAAATTGGAGGTAATTIGTAAGTAAGTAGCTGATGA'
E S V G N E I G G N L L T K V A D D
990 1010 103 CONCAATTTTCACTC $\begin{array}{c} 0.1190 \\ 1190 \end{array}$ TACTTTGCGGCTGTA ACTTROOM 1370 LCAGGCTGACTTATG3 aw
Q A ^D L ^C 57ATTTG GTATTTGCTCAAGAT
V F A Q D ¹
1730 Q OGCTTTAGTAGGTGAM A L 1910 S I V B R
2090 :ACCTTOGACAATGCI ^T L ^D ^N A 2270 **TTATTGTATAGTCCA** 0 G X N CAAGATGATGGCCC
Q D D G P
2630 P L C Q LAACAAGGTCATGG
(Q G HG
2810 ;ATGACACTTTCAG C I T L S A 2990 $\begin{array}{cc} 5 & R \\ 3170 \end{array}$ 130 **MACCAACACCTAAGAC** 310 CTTATG
490 TCAAAACCTGAATTGT
670 ;OCCAAACCCAGGGO 650 WTAGCTGATGATGCI A 1030 rATGGCTTATGACATC 1210 AAGATGTGGCTTTCAC $\begin{smallmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{smallmatrix}$ 1390 FATCCCCTATGTTGCT ^I ^P ^Y V ^A 1570 TGTC**AACATATATGAT**
V N I Y D
1750 **AAGAGCTTTCTATGAT(** $\mathbf D$ 1930 LAAIAIIGTCTACTT IAAGACTATTTCCAAATTTGAA R ^L ^F P ^N L ^N N ^I V ^Y ^L 2110 ATACACAATATGT
YTIC $\frac{1}{2290}$ 'AAATOGGACAAGN 2470 TCAAGAAAAG E K
2650 **TATCACTGTTGTTTG**
, **s** l l **f a** L \sim S 2830 CCTACTATTCAAAC P Y Y S N 3010 CGGGGTGATATGGCA ر
ACAATTTGATCAACCCTATGC
330 CCT<mark>CTAGT</mark>
510 **ATTCAGT** GTTGTGGAAGATATTCAGTACC
690 GATCCCTGGTTTCCTTTTATTGTTAATATTGACATTATGGAGATTACTAGGACCAATTGTTTTGCAACGACTGCTGATCGACCAGA 670 TTCCAATATCCTAGGACCAAA ^S ^N ^I L ^G P ^N 1050 ACAACTGGGACAAAAATCCC
T T G D K N P
1230 N_P 6CTTCAAGTACAAGTTAATGT ^F Q ^V V ^N 1410 V TGACACAAACTATGTCAAGAC D T N Y V I 1590 NATION CARDIOIRANA kTCCTAGAACTGCAGGCAGTAA P ^R ^T ^A ^G S ^X 1950 2130
TGATATAGGTAGTGACAATAGI
D I G S D N S
2310 0 TGCCAGGTTCTTCTGCCCAACT
A R F F C P T
2490 TTATTTTCTCAATTTT CCAAATTATTTTCTCAA
P N Y F L N
2670 ATTTTACTGGTCAAC
FTGEL F A Y F T .
2850 kCAAACCATTAAGAACTGTCAG CCATTAAG
P L R ACIGICAL
T V R
3030 $\begin{array}{c}\nM \\
N\n\end{array}$ 170
CTGGTCCCCACTACTT
350
CHARLOT CHARGE TGTCTGGTGAGATGTG
530 530 :CTATCAATCTGGTAGTGGTGCAAACAC 710
ATTATGGAGACAAT CAGTAAATTGGTTCGTTGAAACCCA

S K L V R L E T B

1250 TCAAGGTACTGCAGGCAGTGCCTT
Q G T A G S A L G S A L A C ₁₄₃₀ .CGATTCGTCAGACTTAGGGCAGTTAAA G Q L R L_G
1610 GAAAAATTGGAAGAAAATTATGACCAT $\ddot{}$ 0 $\ddot{}$ $\mathbf{X}=\mathbf{I}$ \tilde{r} 1790
ATCAAGATTTGATGATCTTGTTAAGAT
S R F D D L V K I S R F D D 1970 TAACAGTTATTCATACTT ^V F ^V ^N ^S ^Y ^S ^Y 2150 Y S Y F TACAATCCCTTACTCTTT ^F ^E ^I ^T ^I P ^Y 2330 S ^F TGGTTCTGTTGTTACATTCCAGAATTC

G S V V T F Q N V P AGGTCGATGAATGTGGACATTITTA
R S M N V D I F T
2690 **EXTCATGTTCTGTTCCTAAG**
I H V L F L S $L = H - V$ L S 2670 TAGTCTTGGTTATTTGAT
S L G Y L M N N S L G Y L M
3050 -
- ATATGGTCA
- ATATGAAGAACCAGAATAGACTA
-
AGAAGAGTCTGAATTAGACTA CARP ^F L-T GT ^M ^T ⁰¹ ^I ^B ^V ^I ^L ^S ^L RC PMNF FF ^P ^L ^P ^A ^P VI SS^O R AL ROGD ^M ^A NML TO^N ^O ^P IYGQ QP 0009R $\frac{1}{2000}$ 3090 $\frac{1}{2010}$ 3110 $\frac{1}{2010}$ 3210 $\frac{1$ MRX ^L AO ^L DRO ^R ^H 0000^I ^V ^O ^D VIOQL D10111 ^F KAlA L TOGRARXF T RIB ^L ^T ⁵ DM VI BBB COE LIDI 3250O 3270 3290 3310 3330 3350 3370 3390 3410 TTINGANTONITHITORITORIGING CONTRIBUTITHITORITAGINAL ANG DISTRONCO ANG ANG DISTRONCO ANG DISTRONCO ANG DISTRONC
FRIKYLBES A V DS BBI PS V D K N C BI LA K D I BIO A I G D A I G L V G T I L L T A G L M S T I K T P A G L M S
3 **SET AN ALL PROPERTY AN ARCHITERAL EXECUTIVE AND DESCRIPTION AND DESCRIPTION** $\begin{array}{cccccccccccccccc} 3790 & 3990 & 3910 & 3990 & 3990 & 3990 & 3990 & 3990 & 4390 & 4390 & 4390 & 4310 & 4390 & 4310 &$ SIRA ^I 00V LBE ROA^R ^H ^V ^C ^O IL DIIA ^M DII V ESA ^I ^Q SAMIXTOIF^P 100R SIDC ^L AIKF ^A PI ^M ^A ^I ^C ^F RO 41 50 4170 4190 4210 4230 4250 4270 4290 ⁴310 CD MOOSIS MIT VI ALF QOB ^L ^A ^R ⁰ P0000I1001^N ^L ^I ^R ^I ^B ^P 00000 QOB ^P ^G QOK ^S ^F ^L TBH ^T ^L 00R0 ^L ^Q 4330 4350 4370 4390 4410 4430 4450 4470 4490 ^O ^C LON ^G VF TIMPITA ^S ^B ^F MI 01G D NQI OS9LO ^IDD LOOQT RI RIOKD EMN ^L ^C ^M CI SOS ^V ^P ^F ^I ^V ^P M AO LBE 4510 4530 4550 4570 4590 4610 4630 4650 4670 BIKGIXF 1101S ^L ^V VA 11010 OFDP SO TYVLOI ^S ^G AL R R FP ¹⁰rMOOI AAA IA ¹⁰ ^I ^A GAK LMNV ^S ^Q AMNA 4690 4710 4730 4750 4770 4790 4810 4830 4850 I MONATOTO ACTO ACTO COMO TO BANDA I MOS TO BRACINAL I LA LIDI VII A LIDI VII A LI I LI LI LII LII LII LII LI
THIS TIG BIC WEVIS KNIG RD WET LIK D LIVID KITTID YI WER V KNYN A WKQQ LE NO TI LII D A V ETCHARGESCHAPEN EXAMPLE AND THE RELEVANT THAN A STATEMENT AND THE RELEVANT THAN THE RELEVANT THA YII SONF ^P IA ^I POYV ^D ^E L0N IBM SITLO EOMO ^B ^A ^F ^I EIPIRP ^S ^V ^F ^R CF AM 10001001K^S ^A ^S RABV VIM 5050o 5070 5090 5110 5130 5150 5170 5190 5210 ^F 011010 NL^M ^S ^F VBERMIRAM LITV VO A VI ^S ^A ⁰⁰⁰ L L L VII ^I FIIBBE SIXD BR ^A ¹⁰ ^I - ^L ^P ^V ^A ^A PR ⁵²³⁰ ⁵²⁵⁰ ⁵²⁷⁰ ⁵²⁹⁰ ⁵³¹⁰ ⁵³³⁰ ⁵³⁵⁰ i....3 Vi') ⁵³⁹⁰ GGCACTITITITITICOLOGICOGIOSTICOLOGICOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOSTICOLOGIOST QOBIDE LITLSB 1RMR V FP ⁰ E QPOSVITQ VITL GO PMIDL A ^I LAKCIK ^L PF OF 110011S ^I ^T MAO GIB ^S ^M 5590 5610 5630 5050 5670 5690 5710 5730 5750 ^L ^I ⁰ MI 00000GI TAB VOARV HHSG 0000ARBOG ^B SKT0111000TIV AS ^C ^A ^G ^M ^C GOG ^L ^L ^I ⁰ AVE GON FRI CCTARGEDIAL CONSTANTION CONSTANT TRACENTIC CONSTANT AND THE STREET IS TO SET UP A SO SET CTGGGTATGCTGGTAATGGTGAAATGGGAGTAGCGTTAAATTTTCTTAAAAATGACATGTGAGGGAGTGTAGGATGCAGGCGAGGAGGAGAGAGGAGAGGAGAGGAGGAG
LGHHIAGNGEHGYAGERAYGGAGTAGGTTTAATTTCTTAAAAATGACATGTGAGGAGTGTTACTGAGATGCAGGCCGATGCAATGAGACAACAACACAACAGCCCTGTCTG INIUSTRATION CONSIDERATION CONTRIBUITIES AND RESIDENT CONSIDERED TRANSPORT CONSIDERED AND RESIDENCE ON THE RESIDENCE OF RESIDENCE ON THE RESIDENCE OF RESIDENCE OF RESIDENCE OF RESIDENCE OF RESIDENCE OF RESIDENCE OF RESIDEN GTCCAGCTGTCCTARGTARGTCAGCAGCCAGCTAGGARGCTGTAGARTGTCTHARAARTGCCGTTCCAGTATAGARTGTAGARGARTGTGTAGARTGARTGARTGGARTG
| PAVLSK SDTRLEBPVECLIKKS & 1909 6210
| 6130 6250 6250 6270 6290 DMA TAOI LOITS ⁰ VS ⁰ MI LOST ^S ^A 010 ^F V 0011111 KK ^L ^I ^C LBE ^P ^F ^O VA ^P ^L ^L EARL VQ OAXF DON ^L LAX 6310 6330 6350 6370 6390 6410 6430 6450 6470 0W00011T ^I ^F MIT ^C LI ^D EL RIR LIIRI ^A ⁰⁰¹ TARCO ^B ^A CE VIDI ^C ^I VI ^A ^M ^I MB 01110101KI YQ ^P ^C ¹¹⁰ $\begin{array}{cccccccccccc} \textbf{GAL} & \textbf{GSD} & \textbf$ GATGATGTATTCHTTCATTGGACAAAGAAAATTGGAGAAACTGGAAGATTGATGATGACGGAAAGATGGGAGAGAAGAAGAAGAAGGAGGAAGAATGGAGAAACCGGAAACCTGGTTATTTCCCAGAGTCTATTTCCCAGAGTCTATTTCCCAGAGTCTATTTCCCAGAGTCTATTTCCCAGAGTCTATTTCCCAGAGTCTATTTCCCAGAGTCTATTTCCC INGTAGTANATTAGACACIGANAN INGAACATTTAATGTGGATGANAMACTTTAGCACATTTAAGGAGCACCONTCCATTTAATGGCACAGCANAGACACTTATCAGACACTTATCAGGAGCATTCTAGGAGCATTATCAGGGAATGGAGCATTATCAGGGAATGGAATATCACTGGATTCTAGGAGGAATGGAATATCACTGGAATATCACTG
7210 7 10 TTTGAAAGGGGTCTCCTAC I 190
G<mark>CAATAAGAAGAGTGGAAC</mark> 370
G<mark>GAACTTATTGGAAACAAC</mark> 730 X S I A D H A
910 GAACAAGAATGTAGTTCJ
N K N V V Q AACAAGAATGTAGTTC
N K N V V Q
1090 TGAGTGGACGCCATCTTGO E0 T P ^S M 1270 OGTAGTATATGAACCTAA V V Y E P X 1450 AGTCTATGTCTGGACTCCATTGTCAATCCCCACTGGAT \mathbf{r} 1630
Cocacabartaca GAGCACAAAATACAAATGO
5 T K Y K W
1810 AGCCCAGTTATTCTCACT A O ^L ^F ^S ^V 1990 TAGGGGTTCATTAGTTTT.
R G S L V L TTCCACTTGGATGAGGGERGAG TCCACTTGGATGAGGAAAACAAATGGCCACCCAA
S T W M R K T N G E P I $\frac{M}{2350}$ ATGGGGTTCACAGA! Q M D L T N G S
2530 TGTATCACATACTA
V S B T D V ^S ^H T X V 2710 TGIGAGGGGGTTTCTGAG E R C ^F ^L R 2690 STGCAAGCCCTTCTTGA
C K P F L -T 210
AAGGATGCTTAAAGCATA
390
AATTTGCTTAATAGCATC 570 750 A T C VPO 5 S
GCCACTACTACTOTTAATAG GCCACTACTACTGTTAATA
A T T T V N 1 ;oCAAGAOGATACCAGAT A ^R ^G ^Y¹ ⁰ 1290 **CTTGTCACTTA** P V V T ^Y ^D 1470 ^L ^S ^I ^P ^T G 1650
GAACAAAAATC ECCAGAACAAAAATCGA
The the ID
1830 ;ATGGCAGITTCAAOCACI ^M A ^D ^S ^T ^T 2010 gattgagcgtcta **MATGGCCACCC** 2370 rTAACCOaCCCTCTFTT L T D P L C 2550 WATAMICTATTTGGC
D N L F G R
2730 OOTTGCACACACATATGA V A S ^T ^Y ^D 2910 CCTCTACTGG
' \$ T G M rc ra ($\overline{}$ \cdots ITI M rim NO NU **A GCTTCTATAGGAAAAAAACCCTT'** A .AC Icc **GTAATTTGTTAA** c rG4 P. LAG E LTG ly n.
L \cdots A a w I IA
T
ITK I I M ukc i ;GC i um I ;Q GATGGCAGATTCAACCACTCCATCAGAAAATCATGGAGTTGATGCAAAAGGATATTTCAAGTGGTCAGCAACAACAGCCCCCCCAGAGTATTGTCCACAGAAATATTGTCTACTTA aagattgagcgtctacgccagtactttcaatagaggtcgattgaggatgggcttcttccccaatgccacaacggactcaacttccaccttggacaatgctatatacacaatatgt UUCAAATGGCCACCCAATTGGATTGTTCAAATTGAAGTCTTAAACAGGCTCACATACAGATTCTAGTCCATGGGAGGTTTATTGTATAGTCCAAGGTAAAATGGGAGAGA Lr ;Ts LA ITs rcl AC kc RGATAACCTATTTGGGCGGGCATGGTTTTTTATGGAGCATACTTTCACCAATGAGGGACAATGGAGAGTGCCATTGGAAATTTCCAAAACAAGGTCATGGGTCCTTATCACTGTTG GGTTGCACACATATGACACTAGTAATGATGAGTCAATTTTCTGTCATCGAACGGTGTAATAACTGTACCAGCCGGAGAGCAGATGACACTTTCAGCTCCCTACTATTCAAC TGGAACCTCTACTGGTAAAA1TGAGGTTTATCTTAGCCTGAGATGTCCAAATTTCTTTTTCCCTCTTCCTGCCCCTAAGGTTACGAGTAGTCGTGCACTACGGGGTGATATGGCA

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	3Α	3Β	3С	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		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 O₁K; 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 O₁K; 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 $2-(\alpha-hydroxyben$ zyl)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 β -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 butjuxtaposed 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. & Hyypia, T. (1989) Arch. Virol. 104, 175-186.
- 2. Auvinen, P. & Hyypiä, T. (1990) J. Gen. Virol. 71, 2133–2139.
3. Wigand, R. & Sabin, A. B. (1961) Arch. Gesamte Virusforsch.
- 3. Wigand, R. & Sabin, A. B. (1961) Arch. Gesamte Virusforsch. 11, 224-247.
- 4. Shaver, D. N., Barron, A. L. & Karzon, D. T. (1961) Proc. Soc. Exp. Biol. Med. 106, 648-652.
- 5. 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.
- 7. 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. 8. Stanway, G. (1990) J. Gen. Virol. 71, 2483-2501.
- 9. Cohen, J. I., Ticehurst, J. R., Purcell, R. H., Buckler-White, A. & Baroudy, B. M. (1987) J. Virol. 61, 50-59.
- 10. Hyypid, T., Auvinen, P. & Maaronen, M. (1989) J. Gen. Virol. 70, 3261-3268.
- 11. Murphy, G. & Kavanagh, T. A. (1988) Nucleic Acids Res. 16, 5198.
12. Herikoff. S. (1984) Gene 28, 351–359.
- 12. Herikoff, S. (1984) Gene 28, 351-359.
13. Zagursky, R. J., Baumeister, K., Lo Zagursky, R. J., Baumeister, K., Lomax, N. & Berman, M. L. (1985) Gene Anal. Tech. 2, 89-94.
- 14. Huovilainen, A., Kinnunen, L., Ferguson, M. & Hovi, T. (1988) J. Gen. Virol. 69, 1941-1948.
- 15. Speicher, D. W. (1989) in Techniques in Protein Chemistry, ed.
- Hugli, T. E. (Academic, San Diego), pp. 24-35. 16. Kalkkinen, N. & Tilgmann, C. (1988) J. Protein Chem. 7, 242-243. 17. Devereux, J., Haeberli, P. & Smithies, 0. (1984) Nucleic Acids Res.
- 12, 387-395.
- 18. Felsenstein, J. (1985) Evolution 39, 783-791.
19. Palmemberg, A. C. (1989) in Molecular Asi 19. 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.
- 20. 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.
22. Melnick, J. L. (1965) in Viral and Rickettsial Infections of Man. eds.
- 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. 24. Jamison, R. M. (1974) *Arch. Gesamte Virusforsch.* 44, 184–194.
25. Fegers. H. J. & Tamm. J. (1961) *J. Exp. Med.* 113. 657–682.
- 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. Virc
- 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. 28. Chang, K. H., Auvinen, P., Hyypia, T. & Stanway, G. (1989) J.
- Gen. Virol. 70, 3269-3280. 29. Roivainen, M., Hyypia, T., Piirainen, L., Kalkkinen, N., Stanway,
- G. & Hovi, T. (1991) J. Virol. 65, 4735-4740. 30. 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.