Characterization of the complete genomic sequence of genotype II hepatitis A virus (CF53/Berne isolate)

Ling Lu,[1]† Karen Z. Ching,[1]‡ Vanessa Salete de Paula,[1]§ Tatsunori Nakano,[1]|| Gunter Siegl,[2]¶ Manfred Weitz[2]# and Betty H. Robertson[1]

[1]Laboratory Branch, Division of Viral Hepatitis, National Center for Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road NE, MS C12, Atlanta, GA 30333, USA

[2]Institut für Klinische Mikrobiologie und Immunologie, St Gallen, Switzerland

Correspondence
Betty H. Robertson
bjr1@cdc.gov

Received 21 May 2004
Accepted 1 June 2004

The complete genomic sequence of hepatitis A virus (HAV) CF53/Berne strain was determined. Pairwise comparison with other complete HAV genomic sequences demonstrated that the CF53/Berne isolate is most closely related to the single genotype VII strain, SLF88. This close relationship was confirmed by phylogenetic analyses of different genomic regions, and was most pronounced within the capsid region. These data indicated that CF53/Berne and SLF88 isolates are related more closely to each other than are subtypes IA and IB. A histogram of the genetic differences between HAV strains revealed four separate peaks. The distance values for CF53/Berne and SLF88 isolates fell within the peak that contained strains of the same subtype, showing that they should be subtypes within a single genotype. The complete genomic data indicated that genotypes II and VII should be considered a single genotype, based upon the complete VP1 sequence, and it is proposed that the CF53/Berne isolate be classified as genotype IIA and strain SLF88 as genotype IIB. The CF53/Berne isolate is cell-adapted, and therefore its sequence was compared to that of two other strains adapted to cell culture, HM-175/7 grown in MK-5 and GBM grown in FRhK-4 cells. Mutations found at nucleotides 3889, 4087 and 4222 that were associated with HAV attenuation and cell adaptation in HM175/7 and GMB strains were not present in the CF53/Berne strain. Deletions found in the 5′UTR and P3A regions of the CF53/Berne isolate that are common to cell-adapted HAV isolates were identified, however.

INTRODUCTION

Hepatitis A virus (HAV), classified as *Hepatovirus* (Minor, 1991) within the family *Picornaviridae* (Melnick 1982; Gust *et al.*, 1983), is composed of an icosahedral capsid that contains a positive-sense, single-stranded RNA genome of approximately 7·5 kbp in length. This virus is a single serotype that precludes serological approaches to differentiate virus isolates (Lemon *et al.*, 1992). Genetic variants of HAV have been documented by RT-PCR sequencing of the C terminus of the VP3 protein (Jansen *et al.*, 1990), the N terminus of the VP1 protein (Robertson *et al.*, 1991; Apaire-Marchais *et al.*, 1995; De Serres *et al.*, 1999; Robertson *et al.*, 2000; Arauz-Ruiz *et al.*, 2001; Bruisten *et al.*, 2001; Costa-Mattioli *et al.*, 2001a, b) and the junctional region of the VP1/P2A proteins (Jansen *et al.*, 1990; Robertson *et al.*, 1992, 2000; Normann *et al.*, 1995; Taylor, 1997; Bosch *et al.*, 2001; Bruisten *et al.*, 2001; Byun *et al.*, 2001; Costa-Mattioli *et al.*, 2001a; Diaz *et al.*, 2001; Pina *et al.*, 2001; de Paula *et al.*, 2002). Sequence variation within the VP1/P2A junction has defined seven genotypes that differ by at least 15 % and subtypes that differ by 7·0–7·5 % (Robertson *et al.*, 1992). Four genotypes (I, II, III and VII) are associated with human HAV infections and three were derived from simian HAV strains (IV, V and VI).

†Present address: Division of Gastroenterology/Hepatology, Kansas University Medical Center, 4035 Delp MS1023, Kansas City, KS 66160, USA.

‡Present address: Respiratory and Enterovirus Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, CDC, 1600 Clifton Road NE, MS G-10, Atlanta, GA 30333, USA.

§Present address: Oswaldo Cruz Foundation, Department of Virology, Av. Brasil 4365, 21045-900 Rio de Janeiro, Brazil.

||Present address: Department of Internal Medicine, Ichinomiya Nishi Hospital, Okucho Origuchinishi 89-1, Ichinomiya, Aichi 491-0201, Japan.

¶Present address: Seeblickstrasse 17, CH 9402 Mörschwil, Switzerland.

#Present address: Swissmedic, Schweizerisches Heilmittelinstitut, Berne, Switzerland.

The GenBank/EMBL/DDBJ accession numbers for the CF53/Berne sequences reported in this manuscript are AY644671–AY644673 and AY644676.

Genotype IA appears to be the agent responsible for the majority of hepatitis A cases worldwide (Robertson *et al.*, 1992; Costa-Mattioli *et al.*, 2002) and has been isolated from all parts of the world. Genotype IB appears to occur in the Mediterranean region, whereas genotype III viruses have been isolated from diverse sources such as Panamanian owl monkeys, drug abusers in Sweden and patients from India and Nepal. Single representatives of genotype II and VII were isolated from individual patients from Sierra Leone and France (Robertson *et al.*, 1992; Ching *et al.*, 2002).

Complete genomic information is now available for all human HAV genotypes with the exception of genotype II. In this study, we have determined the complete genomic sequence from the single representative of genotype II, the CF53/Berne isolate.

METHODS

Specimen. The CF53/Berne isolate is a cell-culture isolate of HAV that was passaged six times in PLC/PRF/5 cells (passages 1–6) prior to donation to the Berne laboratory where the virus was passaged twice in PLC/PRF/5 cells (passages 7–8) followed by four passages in FRhK-4 cells (passages 9–12). The original virus was derived in 1979 from the stool of a patient in Clermont-Ferrand, France, who had sporadic hepatitis A 3 days after the onset of jaundice (Crance *et al.*, 1985). The VP3 region and VP1/P2A junctional region of virus from passage 10 was sequenced and classified as genotype II (Jansen *et al.*, 1990; Robertson *et al.*, 1992). For reference, this cultured strain is called HAV CF53/Berne isolate. Cell-culture passages 7, 8, 10 and 12 were available for this study.

Primers. HAV primers that were available in our laboratory for the amplification of various HAV strains (Hutin *et al.*, 1999; Robertson *et al.*, 2000; Ching *et al.*, 2002) were used to amplify the complete CF53/Berne isolate genomic sequence. Sequence from the amplifiable fragments provided more information for designing additional primers specific to CF53/Berne isolate. Table 1 identifies the primers that were used for amplification of the complete genomic sequence of the CF53/Berne strain.

RT-PCR. Total RNA was extracted using Tripure (Boehringer Mannheim) from 100 µl cell-culture material that contained the CF53/Berne virus. Extracted RNA was resuspended in 8·5 µl diethylpyrocarbonate (DEPC)-treated water and cDNA was synthesized in a 20 µl reaction volume that contained 4 µl 5 × avian myeloblastosis virus reverse transcriptase (AMV RT) buffer (Promega), 1 mM dNTPs (Boehringer Mannheim), 2·0 µl DMSO (Invitrogen), 8·5 µl RNA, 25 U AMV RT (Boehringer Mannheim), 20 U RNasin (Boehringer Mannheim) and 1 µg random primers (Promega). Following denaturation at 95 °C for 3 min, the reaction mixture was placed at 4 °C for 3 min and then AMV RT and RNasin were added for reverse transcription at 43 °C for 60 min. First-round PCR was carried out in a 100 µl reaction volume that contained 1 µl cDNA, 10 pmol of each of the appropriate outer primers, 20 µl 5 × PCR buffer (Invitrogen), 2 µM dNTPs (Boehringer Mannheim), 2 µl DMSO and 2·5 U *Taq* polymerase (Boehringer Mannheim). Second-round reactions were carried out in a 100 µl volume that contained 3–10 µl first-round PCR product, 20 pmol of each of the appropriate inner primers, 10 µl 10 × PCR buffer (Boehringer Mannheim), 2 µM dNTPs (Boehringer Mannheim) and 5 U *Taq* polymerase (Boehringer Mannheim). Both rounds of PCR were performed with 1 cycle at 95 °C for 2 min followed by 35 cycles of 94 °C for 40 s, 45 °C (for the first round) or 55–63 °C (varied to different primer pair T_m for the second round) for 40 s and 72 °C for 1–4 min depending upon the amplicon length. The final PCR products were resolved on 1 % agarose gels with ethidium bromide.

Sequence strategy, determination and analysis. Amplicons with the expected molecular sizes were purified using the QIAquick PCR Purification kit (Qiagen) and the purified PCR products were then sequenced in both directions using dRhodamine terminator reagents and the appropriate primer in the ABI Prism 3100 Genetic Analyzer (PE Applied Biosystems).

Algorithms within the Genetics Computer Group (GCG) package (Wisconsin Sequence Analysis Package; Genetic Computer Group, Madison, WI, version 10.1) were used for alignment of nucleotide and amino acid sequences. Initial multiple alignments of nucleotide and amino acid sequences were performed with the PILEUP program; further adjustment to the alignment was performed manually using visual correction based on sequence comparison generated with the PRETTY program in GCG. Calculation of nucleotide and amino acid sequence identities, calculation of genetic distances between sequences and construction of phylogenetic trees were performed using the computer software MEGA2 (Kumar *et al.*, 1994).

Genetic distances for the full-length genomic nucleotide sequences, all nucleotide positions of the P1, P2 and P3 regions and amino acid sequences were calculated using the Jukes–Cantor method. Genetic distances for synonymous and non-synonymous nucleotide substitutions of coding regions were calculated using the Jukes–Cantor parameter within the modified Nei–Gojobori method included in MEGA2. All other phylogenetic trees were constructed by the neighbour-joining method, except for those in Fig. 4 (constructed by the UPMGA method). Reliability of the trees was confirmed by bootstrap resampling of 500 replications with 60 000 random seeds. Pairwise genetic distances that were used for plotting histograms to display the relationship of HAV strains, subtypes and genotypes were calculated using the Jukes–Cantor method of the PAUP software (version 4.0, Sinauer Associates, Sunderland, MA). The PLOTSIMILARITY program from the GCG package was used for pairwise comparison of sequences with a sliding window of 100 nt in order to exclude the possibility of recombination between genotypes and subtypes,

RESULTS

Genetic characterization and comparison

A total of 15 overlapping fragments, spanning the entire HAV genome (Fig. 1), were amplified from the cell-culture-passed CF53/Berne strain. A 7426-nt sequence from the passage 12 CF53/Berne isolate genome, encompassing nt 32–7465 (based upon the numbering of HM-175, M14707; Cohen *et al.*, 1987), was determined. The 5′ untranslated region (UTR) consisted of 685 nt, followed by a single, long, open reading frame and 63 nt of 3′UTR. The nucleotide composition was 28·79 % A, 15·98 % C, 22·45 % G, 32·78 % T, and the overall G + C content was 38·43 %. The CF53/Berne isolate sequence was compared to the complete available sequences of four representative HAV genotypes [genotype IA (GBM; Graff *et al.*, 1994a), genotype IB (HM-175; Cohen *et al.*, 1987), genotype III (NOR-21, unpublished GenBank accession no. AJ299464), genotype V (AGM27; Tsarev *et al.*, 1991) and genotype VII (SLF88; Ching *et al.*, 2002)] and shown to be related closest to the SLF88 strain. This close relationship was maintained when

Table 1. PCR primers

Primer sets were used to amplify fragments corresponding to the amplicons illustrated in Fig. 1. Positions refer to the nucleotide numbering of a given HAV sequence.

Primer	Sequence (5′–3′)	Position (nt)	Specificity*	Span†
Set 1				
F1	TTCAAGAGGGGTCTCCGGGGA	1–21	M20273	
R2451	ATGCCAAATCTTGCATATGTATG	2429–2451	M14707	
Set 1a				
F31	CCCTCTTGGAAGTCCATGGTGAGGGGA	32–58	M14707	
R701	GGAAAAACCTAAATGCCCCTGAG	696–718	M14707	1–669
Set 1b				
F95	GTAAATATTAATTCCTGCAGGTTCAGG	95–121	CF53/Berne	
R749	AAGACAGGATGTGGTCAAGACCACT	725–749	CF53/Berne	95–749
Set 1c				
F701	ACTCAGGGGCATTTAGGTTTTTCCT	695–719	M14707	
R1051	ACATCCAATTTTGCCACTTCAT	1030–1051	CF53/Berne	646–1002
Set 2				
F701	ACTCAGGGGCATTTAGGTTTTTCCT	695–719	M14707	
R2451	GATCTGATGTATGTCTGGATTCT	2429–2451	M14707	
F788	ATTTTCCAGACTGTTGGGAGTGG	756–778	M14707	
R2167	TAGCATGATAAAGAGGAGCAAAAC	2167–2191	M14707	707–2143
Set 3				
F2132	GTTAATGTTTATCTTTCAGCAAT	2086–2108	X75215	
R3381	CCATTTCAAGAGTCCACACACT	3360–3381	M14707	
Set 3a				
F2167	GTTTTGCTCCTCTTTATCATGCTATG	2167–2192	M14707	
R3288	AACTTCATTATTTCATGCTCCT	3268–3289	M14707	2118–3240
Set 3b				
F2132	GTTAATGTTTATCTTTCAGCAAT	2086–2108	X75215	
R2414	GGAAATGTCTCAGGTACTTTCTTTG	2389–2413	M14707	2084–2364
Set 3c				
F2167	GTTTTGCTCCTCTTTATCATGCTATG	2167–2192	M14707	
R2919	GATTGTAATTTGCAATCTGAATAG	2896–2919	M14707	2118–2870
Set 3d				
F2775	TTTAACACAAGAAGAACAGGAA	2775–2796	CF53/Berne	
R3288	AACTTCATTATTTCATGCTCCT	3268–3289	M14707	2726–3240
Set 4				
F2775	TTTAACACAAGAAGAACAGGAA	2775–2796	CF53/Berne	
R4392	CTGGTTCACATCTTGTCACCAT	4333–4354	AY032861	
F3065	TCAGTGGATGATCCTAGATCAG	3066–3087	M14707	
R3600	AATGCAGCAAGAAACACAAATC	3579–3600	M14707	3017–3551
Set 5				
F3065	TCAGTGGATGATCCTAGATCAG	3066–3087	M14707	
R4724	ACCAATTTGAAGTTGCAATTAT	4651–4672	AY032861	
F3288	AGGAGCATGAAATAATGAAGTT	3268–3289	M14707	
R4392	CTGGTTCACATCTTGTCACCAT	4333–4354	AY032861	3219–4355
Set 6				
F3288	AGGAGCATGAAATAATGAAGTT	3268–3289	M14707	
HAVRDT	TTTTTTTTTTTTTTTTATTTACTGATA	7421–7446	X75215	
F4159	TTACAGATTCAGGATGTGGAA	4147–4167	AY032861	
R5078	GGTTCACCAGATGGAAAAGACTG	5046–5068	M14707	4148–5013
Set 7				
F4159	TTACAGATTCAGGATGTGGAA	4147–4167	AY032861	
R5356	CTCACACATCCATTTTTCTCTC	5296–5318	AY032861	
F4392	ATGGTGACAAGATGTGAACCAG	4333–4354	AY032861	
R4724	ACCAATTTGAAGTTGCAATTAT	4651–4672	AY032861	4334–4673

Table 1. cont.

Primer	Sequence (5′–3′)	Position (nt)	Specificity*	Span†
Set 8				
F4392	ATGGTGACAAGATGTGAACCAG	4338–4354	AY032861	
HAVRDT	TTTTTTTTTTTTTTTTATTTACTGATA	7421–7446	X75215	
F4724	ATAATTGCAACTTCAAATTGGT	4651–4672	AY032861	
R5078	GGTTCACCAGATGGAAAAGACTG	5046–5068	M14707	4652–5013
Set 9				
R6303	TTTCCTGAACATAGGGAAATCCGGGAGAA	6275–6303	CF53/Berne	
F4906	ATAATATTTCACTGATGGATTTGCT	4906–4930	CF53/Berne	
R6274	GAATCCATATTGATTGCGTCTATTCC	6249–6274	CF53/Berne	4857–6219
Set 10				
F4724	ATAATTGCAACTTCAAATTGGT	4651–4672	AY032861	
HAVRDT	TTTTTTTTTTTTTTTTATTTACTGATA	7421–7446	X75215	
F6232	ATCTTGACATGGCCATTACAGG	6170–6191	AY032861	
R7481	AAAGAAATAAACAAACCTCAGA	7446–7467	M14707	6165–7412

*Primer sequence completely identical to the numbered positions of the given HAV sequence (isolate or GenBank accession number).
†Nucleotide range determined by the amplicon for the complete CF53/Berne sequence.

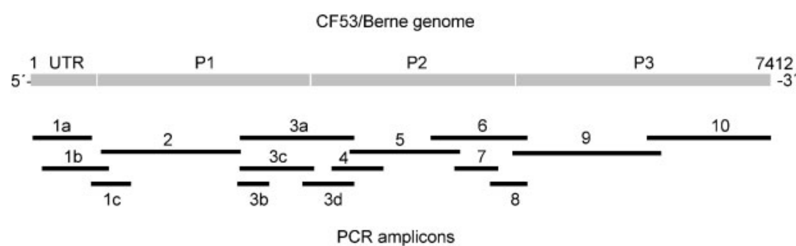

Fig. 1. Strategy for amplification of the entire CF53/Berne isolate genome. The shaded bar at the top represents the entire HAV genome divided into five regions, the 5′UTR, the P1, P2 and P3 regions and the 3′UTR. The solid bars below represent the 15 overlapping fragments amplified to cover the complete CF53/Berne isolate genome.

Table 2. Nucleotide and amino-acid sequence identity of the CF53/Berne isolate relative to representative strains of human HAV genotypes IA, IB, III and VII and simian HAV genotype V

Region	SLF88 (genotype VII)	GBM (genotype IA)	HM-175 (genotype IB)	Nor21 (genotype III)	AGM27 (genotype V)
Nucleotide sequence					
Complete genome	90·98	86·29	86·22	–	81·58
5′UTR	96·73	92·00	92·84	–	82·83
Full-length coding region	90·31	85·57	85·36	80·39	81·15
P1 region	90·83	85·89	85·72	80·50	81·75
P2 region	89·43	85·53	85·58	78·50	80·03
P3 region	90·45	85·31	84·85	81·76	81·19
P3 ABC region	89·38	84·82	84·82	80·53	80·66
P3 D region	91·14	85·62	84·87	82·55	81·53
Amino acid sequence					
Full-length coding region	98·07	96·13	96·58	93·92	92·08
P1 region	99·24	98·73	98·86	94·92	95·34
P2 region	97·62	95·09	95·72	92·55	90·65
P3 region	97·26	94·40	95·02	91·91	91·42
P3 ABC region	97·45	96·82	98·73	95·59	94·55
P3 D region	97·14	92·84	92·64	89·78	89·57

the three major protein-encoding regions (P1, P2 and P3) were analysed separately. Evaluation of the more conserved 5′UTR revealed that the CF53/Berne isolate had over 96 % identity to human HAV genotype VII, about 92 % identity to the human genotype IA and IB and approximately 83 % identity to the simian HAV genotype V sequence (Table 2).

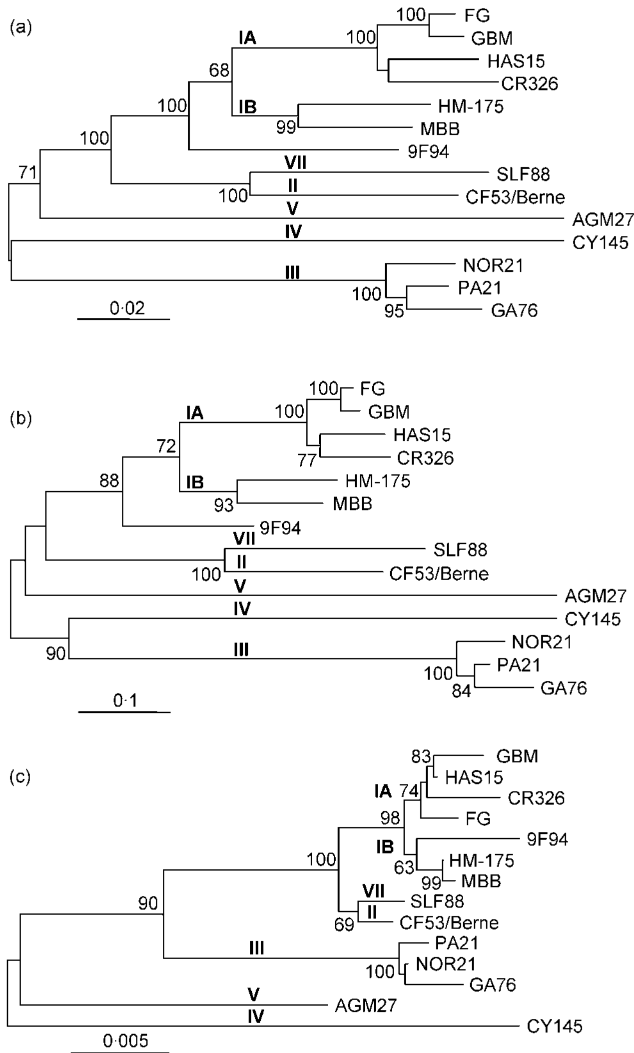

Fig. 2. Phylogenetic trees based on nucleotide sequences in the P1 region from 13 HAV isolates compared to CF53/Berne isolate based on all nucleotide positions (a), synonymous positions (b) and non-synonymous positions (c). Bootstrap values (%) are illustrated at nodes and genetic distance per nucleotide site is shown by the horizontal bar. Abbreviations and the corresponding references are: CF53/Berne (this study); FG (Beneduce *et al*., 1995); GBM/WT (Graff *et al*., 1994a); HAS15 (Sverdlov *et al*., 1987); CR326 (Linemeyer *et al*., 1985); HM-175 (Cohen *et al*., 1987); MBB (Paul *et al*., 1987); 9F94 (Costa-Mattioli *et al*., 2002, 2003); SLF88 (Ching *et al*., 2002); AGM27 (Tsarev *et al*., 1991); CY145 (Nainan *et al*., 1991); NOR-21 (Stene-Johansen *et al*., 1999); PA21 (Brown *et al*., 1989); GA76 (Khanna *et al*., 1992).

For some HAV isolates, only the capsid sequences are available. These include HAS15 (Sverdlov *et al.*, 1987) and CR326 (Linemeyer *et al.*, 1985) from genotype IA, CY145 from genotype IV (Nainan *et al.*, 1991), PA21 (Brown *et al.*, 1989) and GA76 (Khanna *et al.*, 1992) from genotype III and 9F94 (AJ437317 and AJ519487), a recombinant from genotypes I and II (Costa-Mattioli *et al.*, 2002, 2003). Fig. 2 illustrates phylogenetic analyses of the available capsid (P1) region sequences; Fig. 2(a) was based on all nucleotide positions, Fig. 2(b) on the synonymous positions and Fig. 2(c) on the non-synonymous positions. Synonymous and non-synonymous positions were evaluated independently to determine the differential, evolution-reflecting mutations resulting from the absence or presence of selective pressure. In all three trees, the CF53/Berne isolate branch clusters with the SLF88 branch, with bootstrap values of 100, 100 and 69 % respectively. In addition, the two isolates (SLF88 and CF53) compose a group adjacent to the genotype I clade. The distance between the CF53/Berne and SLF88 isolates is less than that between subtypes IA and IB, particularly in the tree based on all nucleotide positions (Fig. 2a). Analyses of available sequences for the P2 and P3 regions representing five HAV genotypes (genotypes IV and VI are not available) also resulted in phylogenies bearing topologies and bootstrap values similar to those in Fig. 2 (data not shown). Strains CF53/Berne and SLF88 therefore appear to represent one genotype based on these results and, within this genotype, they each represent subtypes, analogous to subgenotypes IA and IB.

Pairwise genetic distances were calculated using the sequences included in Fig. 2(a) and from 17 additional HAV genotype I capsid regions and the values were plotted as a histogram (Fig. 3). Four peaks are clearly defined,

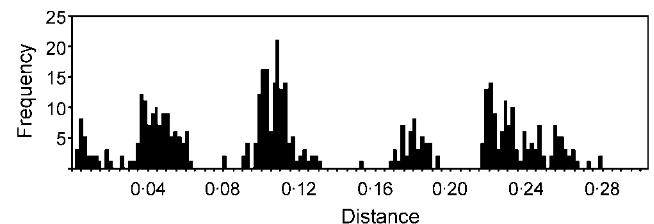

Fig. 3. Histogram reflecting the evolutionary relatedness between HAV isolates. Genetic distances per nucleotide site were derived from a pairwise comparison of sequences in P1 region between HAV isolates and are indicated by the horizontal scale. Distribution frequency is identified by the vertical scale. In addition to the 14 sequences analysed in Fig. 2, 17 other sequences were included in this analysis: LA-15 (Najarian *et al*., 1985), AH1, AH2, AH3, FH1, FH2 and FH3 (Fujiwara *et al*., 2001), LU38/WT (AF357222; unpublished), LY6 (AF485328; unpublished), GBM/FRhK and GBM/HFS (Graff *et al*., 1994a), HM-175/7 MK-5 (Cohen *et al*., 1987), HM-175/18f, HM-175/43c, HM-175/24a (Lemon *et al*., 1991), HAF-203 (AF268396; Baptista *et al*., unpublished) and L-A-1 (AF314208; unpublished).

and each represents a different phylogenetic level of relatedness. The first peak (0·003–0·065) illustrates the distances between strains. The second peak (0·08–0·135) represents distances between subtypes; the distance between CF53/Berne and SLF88 isolates resides in this peak, indicating that they are different subtypes of the same genotype. The third peak is smaller but encompasses distances between 0·165 and 0·195. This peak represents the distances of CF53/Berne and SLF88 isolates versus any one isolate in genotype I, indicating a close evolutionary relatedness between these two genotypes when compared to the other HAV genotypes. The last peak (0·215–0·28) illustrates the distances between genotypes.

The VP1/P2A junctional region (168 bp long) has been used historically to distinguish one HAV strain from another, and previous analysis resulted in strains CF53/Berne and SLF88 being assigned as representing two different HAV genotypes (Robertson *et al.*, 1992). In this paper, we have re-evaluated this approach, including additional sequences representing genotype III. The results (Fig. 4) indicate that the CF53/Berne and SLF88 strains should be reclassified as two subtypes of the same genotype, as the distances between these two strains are similar to the genetic distances between genotypes IA and IB and less than that between genotypes IIIA and IIIB.

Phenotypic characteristics: amino acid changes and cell-culture adaptation

The translated amino acid sequences of the P1, P2 and P3 regions in the CF53/Berne isolate were compared to those

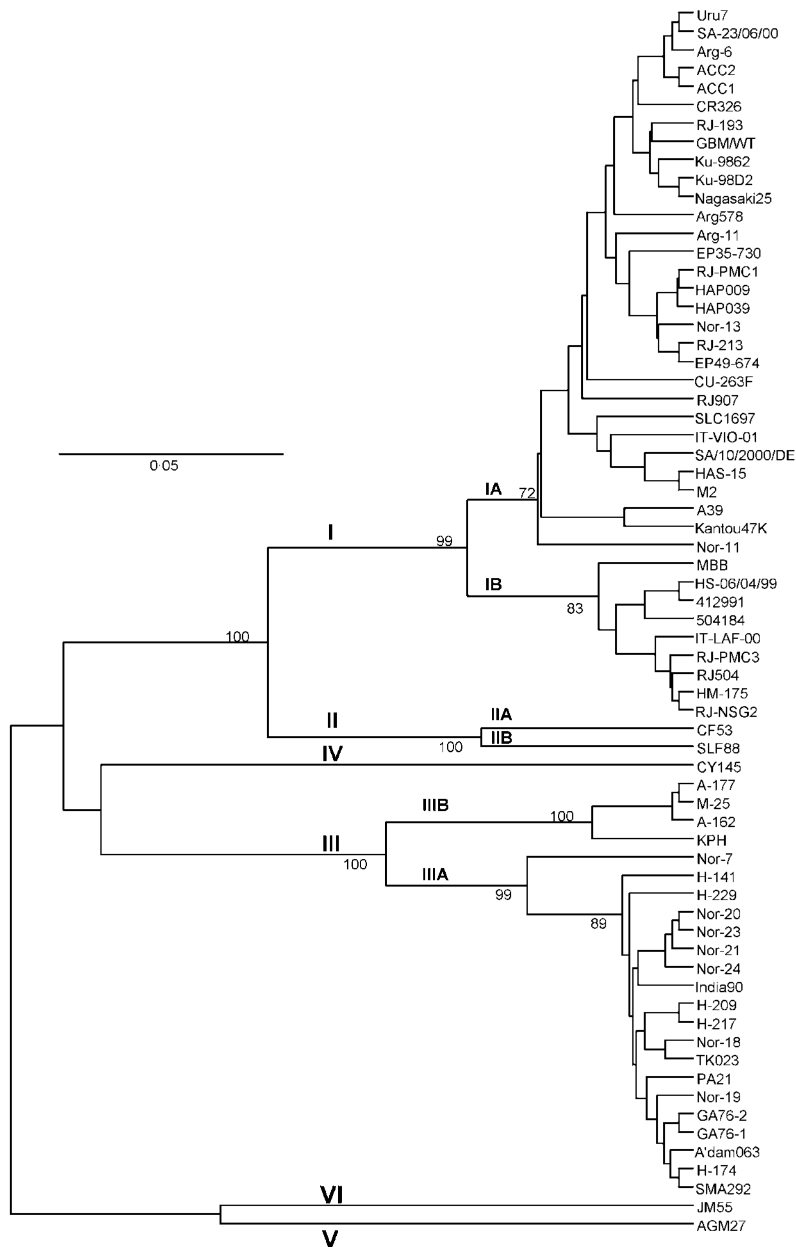

Fig. 4. Phylogenetic tree based on HAV sequences in the VP1/P2A junctional region. Statistically significant bootstrap values (%) are illustrated at nodes and genetic distance per nucleotide per site is shown by the horizontal bar. Abbreviations and the corresponding reference (or GenBank accession numbers) are: Uru7, Arg-6, Arg-11 (AJ306371, AJ306370, AJ306863; unpublished); SA-23/06/00, HS-06/04/99 (Pina *et al.*, 2001); Nagasaki25, ACC1, ACC2, Kantou47K (AB038292, AB038299, AB038300, AB038305; unpublished); CR326 (Linemeyer *et al.*, 1985); RJ-PMC3, RJ-NSG2, RJ504, RJ-193, RJ-907, RJ-PMC1 (De Paula *et al.*, 2002); GBM/WT (Graff *et al.*, 1994a); KU98-42, KU98-D2 (Byun *et al.*, 2001); Arg578 (Mbayed *et al.*, 2002); EP35-730, EP49-674, H-141, KPH, CF53/Berne, H-174, H209, H217, H229, SMA292, TK-23, PA21, JM-55, India90, A-177, M-25, A-162 (Robertson *et al.*, 1992); HAP009, HAP039 (AF538723, AF538724; unpublished); Nor-7, Nor-11, Nor-13 (Grinde *et al.*, 1997); Nor-18, Nor-19, Nor-20, Nor-21, Nor-23, Nor-24 (Stene-Johansen *et al.*, 1999); RJ-213 (De Paula *et al.*, 2003); CU236-F (Diaz *et al.*, 2001); SLC1697 (Hutin *et al.*, 1999); IT-VIO-01, IT-LAF-00 (AJ505577, AJ505624; unpublished); SA/10/2000/DE (AY028976; unpublished); HAS15 (Sverdlov *et al.*, 1987); M2 (Diaz *et al.*, 1999); A39 (AB046901; unpublished); MBB (Paul *et al.*, 1987); 412991 and 504184 (Taylor, 1997); HM-175 (Cohen *et al.*, 1987); SLF88 (Ching *et al.*, 2002); CY145 (Nainan *et al.*, 1991); A'dam063 (AY101269; unpublished); AGM27 (Tsarev *et al.*, 1991); and GA76 (Robertson *et al.*, 1992; Khanna *et al.*, 1992).

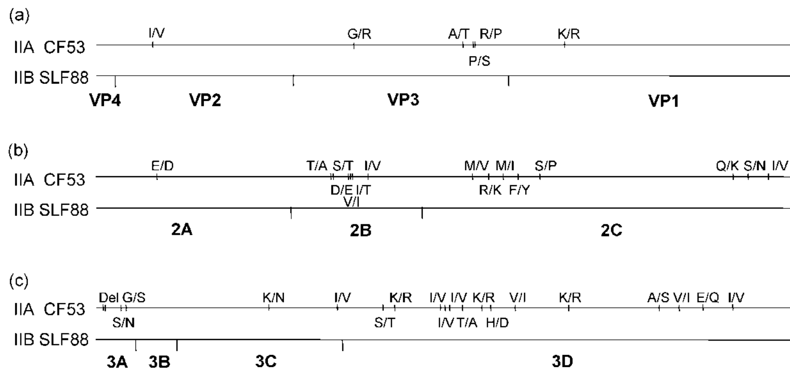

Fig. 5. Amino acid differences between the CF53/Berne and SLF88 isolates in the P1 (a), P2 (b) and P3 (c) genomic regions. The horizontal lines for CF53/Berne indicate sequence identity to strain SLF88, whereas vertical lines indicate the position of changed amino acids; these are represented by single capital letters separated by a slash giving the residue in CF53/Berne (before) and that in SLF88 (after). The SLF88 lines delimit the ranges of different HAV subgenomic regions. SLF88 is the isolate previously determined as genotype VII.

of the SLF88 strain. These results, illustrated in Fig. 5, reveal that the CF53/Berne isolate contains six amino acid differences in the P1 region, 15 amino acid differences in the P2 region and 18 amino acid substitutions and two amino acid deletions in the P3 region when compared to the SLF88 strain. Dissimilar amino acid changes in the P1, P2 and P3 regions number four, three and four, respectively. The 2C and 3D regions contain more substitutions than other regions, and the two amino acid deletions may be a result of cell adaptation (see below).

As CF53/Berne isolate is a cell-adapted strain of HAV, we evaluated this sequence for the cell-adaptive nucleotide and amino acid changes that were found in the HM-175/7 strain when grown in MK-5 cells; the HM-175/7 nucleotide changes are listed in Table 3 (column 3) (Emerson *et al.*, 1991). These positions were compared in the cell-adapted CF53/Berne isolate sequence and the HM175/7 strain sequence. A single base deletion and an 18-nt deletion within the 5′UTR was identified, whereas nucleotide substitutions were found within 10 codons in the coding

Table 3. Comparison of the nucleotide changes in line with HM-175/7 MK-5

Nucleotide positions, used to mark single or bold bases, are based on the numbering of the genome of HM-175/7 MK-5 (Cohen *et al.*, 1987). To show its relation to the amino acid sequence, nucleotide triplets are displayed. del, Deletion; −, substitution not found.

Nucleotide position(s)	Genomic region	Nucleotide sequence		Amino acid sequence	
		HM-175/7 MK-5	CF53/Berne	HM-175/7 MK-5	CF53/Berne
124	5′UTR	T→C	del (18 nt)*		
131–134	5′UTR	del	Covered above		
152	5′UTR	A→G	−		
203–207	5′UTR	del	del (1 nt)*		
964	VP2	AAG→AGG	AAG→AAA*	Lys→Arg	−
1742	VP3	ACG→ACA	ACG→TCA*	−	Thr→Ser*
3025	VP1	GAA→GTA	GAA→GAG	Glu→Val	−
3196	2A	AAT→AGT	−	Asn→Ser	−
3889	2B	GCA→GTA	GCA→GCT*	Ala→Val	−
3919	2B	GGC→GCC	−	Gly→Ala	−
4043	2C	TTT→TTC	−	−	−
4087	2C	AAG→ATG	AAG→AAA	Lys→Met	−
4185	2C	GAA→AAA	GAA→AAT	Glu→Lys	Glu→Asn*
4222	2C	TTT→TCT	−	Phe→Ser	−
4563	2C	GTT→ATT	GTT→GTG	Val→Ile	−
5204	3A	GAG→GAA	−	−	−
5323	3B	CAT→TAT	−	His→Tyr	−
6147	3D	GAT→AAT	GAT→GGC	Asp→Asn	Asp→Gly*
6522	3D	TCA→ACA	TCA→ACA	Ser→Thr	Ser→Thr*
7032	3D	TTG→CTG	TTG→CAA	−	Leu→Gln*
7430	3′UTR	A→G	−		

*Substitution shared by the CF53/Berne and SLF88 isolates.

region that resulted in five amino acid replacements, one in the VP3, one in the 2C and three in the 3D region. As the 5′UTR deletions and the amino acid replacements (indicated by asterisks in Table 3) were conserved between the CF53/Berne isolate and SLF88 (a wild-type genotype II HAV) strain, they are not a result of cell adaptation by the CF53/Berne isolate. The three changes at nt 3889, 4087 and 4222, which are the non-synonymous mutations associated with wild-type to cell-culture adaptation in HM-175 (Emerson *et al.*, 1992, 1993), are not present in the CF53/Berne isolate. The other complete sequence from a cell-adapted strain, HAV GBM/FRhK4, has identified cell-culture adaptive mutations in two regions: nt 3889 in the 2B region resulting in an alanine to valine substitution and an in-frame deletion of nine nucleotides (nt 5013–5021) in the 3A region, resulting in the deletion of three amino acids (Graff *et al.*, 1994b). The cell-adapted CF53/Berne isolate has the wild-type alanine, based upon the codon surrounding nt 3889; however, it does have an in-frame deletion of six nucleotides at position 5017–5021. This results in the deletion of two amino acids that correspond to sites deleted in the GBM strain cell-culture mutant.

In order to trace the occurrence of nucleotide mutations that are potentially related to cell-culture adaptation, single fragments corresponding to PCR amplicon 9 in Fig. 1 were also amplified from samples from cell-culture passages 7, 8 and 10. The sequences were found to be identical to each other and to the corresponding region in the complete genomic sequence that we had determined from the sample from cell-culture passage 12 (data not shown).

DISCUSSION

The original study that defined HAV genotypes was based on the comparison of 168 bases in the VP1/P2A junctional region from 152 strains (Robertson *et al.*, 1992) and identified seven unique genotypes. In this study and one of our recent reports (Ching *et al.*, 2002) complete genomic sequences were characterized from two single HAV strains (CF53/Berne and SLF88 isolates) that represented genotypes II and VII, respectively. Analysis of the complete VP1 region sequence from strain 9F94, which was related to the CF53/Berne and SLF88 isolates, suggested that these strains are subtypes of the same genotype (Costa-Mattioli *et al.*, 2002). Our analysis of the complete genomic sequence has confirmed this interpretation, and we propose that they be recognized as different subtypes of genotype II. We propose that the CF53/Berne isolate be identified as subtype IIA and the SLF88 isolate as subtype IIB. This classification is consistent with phylogenetic analyses of the sequences of the three individual protein-coding regions, P1, P2 and P3, and analysis using only synonymous or non-synonymous nucleotide substitutions. Re-analysis of longer sequences in the VP1/P2A junctional region was consistent with this conclusion (data not shown).

In the current study, the conversion of genetic distances into a histogram was used as an alternative form in order to illustrate graphically the phylogenetic relatedness of HAV isolates. We found that, in addition to peaks representing strain, subtype and genotype variation, there was an additional peak located between the subtype and genotype levels. This peak represented the genetic differences derived from CF53/Berne and SLF88 isolates versus any other isolate of genotype I. This indicates a closer evolutionary relationship between genotypes I and II than between any other genotypes. Subgenotypes IA and IB account for the majority of characterized HAV cases worldwide (Robertson *et al.*, 1992; Costa-Mattioli *et al.*, 2002). In contrast, subgenotypes IIA and IIB are each represented by a single strain from individual patients and were isolated many years ago (Robertson *et al.*, 1992; Ching *et al.*, 2002). The limited detection of these two subgenotypes (IIA and IIB) suggests that their occurrence in the human population may reflect a unique source of infection or route of transmission or their ability to maintain infections in humans.

Recombination between different virus genotypes, which was recently described for HAV (Costa-Mattioli *et al.*, 2002, 2003), is a plausible explanation for the appearance of a strain or group of viruses between established genotypes. PLOTSIMILARITY analysis of sequences of the CF53/Berne and SLF88 isolates against other genotypes (figures not shown) revealed no evidence of such recombination, confirming that they both belong to a unique HAV genotype. In contrast, inclusion of strain 9F94 (isolated in France from a sporadic case) in phylogenetic analysis (Fig. 2) repeatedly resulted in its positioning as a new subtype of genotype I. PLOTSIMILARITY analysis (not shown) also revealed that strain 9F94 is an HAV recombinant, with its upstream sequence (about 1400 nt corresponding to VP2 and VP3 regions) derived from the genotype IB (95 % identical to HM-175) and downstream sequence (about 900 nt corresponding to VP1 region) derived from IIA (95 % identical to CF53/Berne isolate).

The virus isolate of CF53/Berne that we sequenced was derived from 12 cell-culture passages. We compared its sequence with that of other cell-culture-adapted HAV strains to evaluate the presence of mutations associated with cell-culture adaptation and attenuation. The CF53/Berne strain does not contain the mutation at nt 3889 (wild-type valine to cell-culture-adapted alanine) found in the HM-175/7 MK-5 and GBM/FRhK4 isolates that is associated with their cell-culture adaptation. An in-frame deletion in the vicinity of nine deleted nucleotides (3A region, nt 5013–5021) in GBM/FRhK4 isolates was the only other mutation found that is common to cell-adapted strains. These results indicate that cell adaptation by different HAV subtypes involves mutation in different regions and that the P2 region mutation at nt 3889 is not necessary for cell adaptation in genotype II.

In conclusion, we propose that the nomenclature for HAV genotypes be revised to reflect that the SLF88 and

CF53/Berne isolates are subtypes of the same genotype, based upon complete genomic sequences. If future strains are proposed as new genotypes, we also suggest that their complete genomic sequence is used for the basis for this decision.

REFERENCES

Apaire-Marchais, V., Robertson, B. H., Aubineau-Ferre, V., Le Roux, M. G., Leveque, F., Schwartzbrod, L. & Billaudel, S. (1995). Direct sequencing of hepatitis A virus strains isolated during an epidemic in France. *Appl Environ Microbiol* **61**, 3977–3980.

Arauz-Ruiz, P., Sundqvist, L., Garcia, Z., Taylor, L., Visona, K., Norder, H. & Magnius, L. O. (2001). Presumed common source outbreaks of hepatitis A in an endemic area confirmed by limited sequencing within the VP1 region. *J Med Virol* **65**, 449–456.

Beneduce, F., Pisani, G., Divizia, M., Pana, A. & Morace, G. (1995). Complete nucleotide sequence of a cytopathic hepatitis A virus strain isolated in Italy. *Virus Res* **36**, 299–309.

Bosch, A., Sanchez, G., Le Guyader, F., Vanaclocha, H., Haugarreau, L. & Pinto, R. M. (2001). Human enteric viruses in Coquina clams associated with a large hepatitis A outbreak. *Water Sci Technol* **43**, 61–65.

Brown, E. A., Jansen, R. W. & Lemon, S. M. (1989). Characterization of a simian hepatitis A virus (HAV): antigenic and genetic comparison with human HAV. *J Virol* **63**, 4932–4937.

Bruisten, S. M., van Steenbergen, J. E., Pijl, A. S., Niesters, H. G., van Doornum, G. J. & Coutinho, R. A. (2001). Molecular epidemiology of hepatitis A virus in Amsterdam, the Netherlands. *J Med Virol* **63**, 88–95.

Byun, K. S., Kim, J. H., Song, K.-J. & 8 other authors (2001). Molecular epidemiology of hepatitis A virus in Korea. *J Gastroenterol Hepatol* **16**, 519–524.

Ching, K. Z., Nakano, T., Chapman, L. E., Demby, A. & Robertson, B. H. (2002). Genetic characterization of wild-type genotype VII hepatitis A virus. *J Gen Virol* **83**, 53–60.

Cohen, J. I., Ticehurst, J. R., Purcell, R. H., Buckler-White, A. & Baroudy, B. M. (1987). Complete nucleotide sequence of wild-type hepatitis A virus: comparison with different strains of hepatitis A virus and other picornaviruses. *J Virol* **61**, 50–59.

Costa-Mattioli, M., Ferré, V., Monpoeho, S., Garcia, L., Colina, R., Billaudel, S., Vega, I., Perez-Bercoff, R. & Cristina, J. (2001a). Genetic variability of hepatitis A virus in South America reveals heterogeneity and co-circulation during epidemic outbreaks. *J Gen Virol* **82**, 2647–2652.

Costa-Mattioli, M., Monpoeho, S., Schvoerer, C., Besse, B., Aleman, M.-H., Billaudel, S., Cristina, J. & Ferré, V. (2001b). Genetic analysis of hepatitis A virus outbreak in France confirms the co-circulation of subgenotypes Ia, Ib and reveals a new genetic lineage. *J Med Virol* **65**, 233–240.

Costa-Mattioli, M., Cristina, J., Romero, H. & 12 other authors (2002). Molecular evolution of hepatitis A virus: a new classification based on the complete VP1 protein. *J Virol* **76**, 9516–9525.

Costa-Mattioli, M., Ferré, V., Casane, D. & 7 other authors (2003). Evidence of recombination in natural populations of hepatitis A virus. *Virology* **311**, 51–59.

Crance, J. M., Passagot, J., Verwaerde, N., Biziagos, E., Gandre, H. & Deloince, R. (1985). Adaptation to the cell line PLC/PRF/5 of hepatitis A virus released in the cell culture medium. *C R Acad Sci III* **301**, 361–363 (in French).

de Paula, V. S., Baptista, M. L., Lampe, E., Niel, C. & Gaspar, A. M. C. (2002). Characterization of hepatitis A virus isolates from subgenotypes IA and IB in Rio de Janeiro, Brazil. *J Med Virol* **66**, 22–27.

de Paula, V. S., Saback, F. L., Gaspar, A. M. C. & Niel, C. (2003). Mixed infection of a child care provider with hepatitis A virus isolates from subgenotypes IA and IB revealed by heteroduplex mobility assay. *J Virol Methods* **107**, 223–228.

De Serres, G., Cromeans, T. L., Levesque, B. & 8 other authors (1999). Molecular confirmation of hepatitis A virus from well water: epidemiology and public health implications. *J Infect Dis* **179**, 37–43.

Diaz, B. I., Sariol, C. A., Martinez, R. & Rodríguez, L. (1999). Genomic classification and genetic relationships of a new variant of hepatitis A virus isolated in Cuba. *Mem Inst Oswaldo Cruz* **94**, 361–363.

Diaz, B. I., Sariol, C. A., Normann, A., Rodríguez, L. & Flehmig, B. (2001). Genetic relatedness of Cuban HAV wild-type isolates. *J Med Virol* **64**, 96–103.

Emerson, S. U., McRill, C., Rosenblum, B., Feinstone, S. & Purcell, R. H. (1991). Mutations responsible for adaptation of hepatitis A virus to efficient growth in cell culture. *J Virol* **65**, 4882–4886.

Emerson, S. U., Huang, Y. K., McRill, C., Lewis, M. & Purcell, R. H. (1992). Mutations in both the 2B and 2C genes of hepatitis A virus are involved in adaptation to growth in cell culture. *J Virol* **66**, 650–654.

Emerson, S. U., Huang, Y. K. & Purcell, R. H. (1993). 2B and 2C mutations are essential but mutations throughout the genome of HAV contribute to adaptation to cell culture. *Virology* **194**, 475–480.

Fujiwara, K., Yokosuka, O., Fukai, K., Imazeki, F., Saisho, H. & Omata, M. (2001). Analysis of full-length hepatitis A virus genome in sera from patients with fulminant and self-limited acute type A hepatitis. *J Hepatol* **35**, 112–119.

Graff, J., Normann, A., Feinstone, S. M. & Flehmig, B. (1994a). Nucleotide sequence of wild-type hepatitis A virus GBM in comparison with two cell culture-adapted variants. *J Virol* **68**, 548–554.

Graff, J., Kasang, C., Normann, A., Pfisterer-Hunt, M., Feinstone, S. M. & Flehmig, B. (1994b). Mutational events in consecutive passages of hepatitis A virus strain GBM during cell culture adaptation. *Virology* **204**, 60–68.

Grinde, B., Stene-Johansen, K., Sharma, B., Hoel, T., Jensenius, M. & Skaug, K. (1997). Characterisation of an epidemic of hepatitis A virus involving intravenous drug abusers – infection by needle sharing? *J Med Virol* **53**, 69–75.

Gust, I. D., Coulepis, A. G., Feinstone, S. M., Locarnini, S. A., Moritsugu, Y., Najera, R. & Siegl, G. (1983). Taxonomic classification of hepatitis A virus. *Intervirology* **20**, 1–7.

Hutin, Y. J. I., Pool, V., Cramer, E. H. & 9 other authors (1999). A multistate, foodborne outbreak of hepatitis A. National Hepatitis A Investigation Team. *N Engl J Med* **340**, 595–602.

Jansen, R. W., Siegl, G. & Lemon, S. M. (1990). Molecular epidemiology of human hepatitis A virus defined by an antigen-capture polymerase chain reaction method. *Proc Natl Acad Sci U S A* **87**, 2867–2871.

Khanna, B., Spelbring, J. E., Innis, B. L. & Robertson, B. H. (1992). Characterization of a genetic variant of human hepatitis A virus. *J Med Virol* **36**, 118–124.

Kumar, S., Tamura, K. & Nei, M. (1994). MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput Appl Biosci* **10**, 189–191.

Lemon, S. M., Murphy, P. C., Shields, P. A., Ping, L. H., Feinstone, S. M., Cromeans, T. & Jansen, R. W. (1991). Antigenic and genetic

variation in cytopathic hepatitis A virus variants arising during persistent infection: evidence for genetic recombination. *J Virol* **65**, 2056–2065.

Lemon, S. M., Jansen, R. W. & Brown, E. A. (1992). Genetic, antigenic and biological differences between strains of hepatitis A virus. *Vaccine* **10** (Suppl. 1), S40–S44.

Linemeyer, D. L., Menke, J. G., Martin-Gallardo, A., Hughes, J. V., Young, A. & Mitra, S. W. (1985). Molecular cloning and partial sequencing of hepatitis A viral cDNA. *J Virol* **54**, 247–255.

Mbayed, V. A., Sookoian, S., Alfonso, V. & Campos, R. H. (2002). Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina. *J Med Virol* **68**, 168–174.

Melnick, J. L. (1982). Classification of hepatitis A virus as enterovirus type 72 and of hepatitis B virus as hepadnavirus type 1. *Intervirology* **18**, 105–106.

Minor, P. D. (1991). *Picornaviridae*. In *Classification and Nomenclature of Viruses. Fifth Report of the International Committee on Taxonomy of Viruses*, pp. 320–326. Edited by R. I. B. Francki, C. M. Fauquet, D. L. Kundson & F. Brawn. Vienna & New York: Springer Verlag.

Nainan, O. V., Margolis, H. S., Robertson, B. H., Balayan, M. & Brinton, M. A. (1991). Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus macaques (*Macaca fascicularis*). *J Gen Virol* **72**, 1685–1689.

Najarian, R., Caput, D., Gee, W., Potter, S. J., Renard, A., Merryweather, J., Van Nest, G. & Dina, D. (1985). Primary structure and gene organization of human hepatitis A virus. *Proc Natl Acad Sci U S A* **82**, 2627–2631.

Normann, A., Pfisterer-Hunt, M., Schade, S., Graff, J., Chaves, R. L., Crovari, P., Icardi, G. & Flehmig, B. (1995). Molecular epidemiology of an outbreak of hepatitis A in Italy. *J Med Virol* **47**, 467–471.

Paul, A. V., Tada, H., von der Helm, K., Wissel, T., Kiehn, R., Wimmer, E. & Deinhardt, F. (1987). The entire nucleotide sequence of the genome of human hepatitis A virus (isolate MBB). *Virus Res* **8**, 153–171.

Pina, S., Buti, M., Jardí, R., Clemente-Casares, P., Jofre, J. & Girones, R. (2001). Genetic analysis of hepatitis A virus strains recovered from the environment and from patients with acute hepatitis. *J Gen Virol* **82**, 2955–2963.

Robertson, B. H., Khanna, B., Nainan, O. V. & Margolis, H. S. (1991). Epidemiologic patterns of wild-type hepatitis A virus determined by genetic variation. *J Infect Dis* **163**, 286–292.

Robertson, B. H., Jansen, R. W., Khanna, B. & 7 other authors (1992). Genetic relatedness of hepatitis A virus strains recovered from different geographical regions. *J Gen Virol* **73**, 1365–1377.

Robertson, B. H., Averhoff, F., Cromeans, T. L. & 8 other authors (2000). Genetic relatedness of hepatitis A virus isolates during a community-wide outbreak. *J Med Virol* **62**, 144–150.

Stene-Johansen, K., Skaug, K. & Blystad, H. (1999). Surveillance of hepatitis A by molecular epidemiologic studies. *Tidsskr Nor Laegeforen* **119**, 3725–3729 (in Norwegian).

Sverdlov, S. D., Tsarev, S. A., Markova, S. V., Vasilenko, S. K., Chizhikov, V. E., Petrov, N. A., Kusov, Y. Y., Nastashenko, T. A. & Balayan, M. S. (1987). Cloning and expression of hepatitis A virus genome in *E. coli* cells. *Mol Gen Microbiol Virol* **6**, 129–133 (in Russian).

Taylor, M. B. (1997). Molecular epidemiology of South African strains of hepatitis A virus: 1982-1996. *J Med Virol* **51**, 273–279.

Tsarev, S. A., Emerson, S. U., Balayan, M. S., Ticehurst, J. & Purcell, R. H. (1991). Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure and growth in cell culture with other HAV strains. *J Gen Virol* **72**, 1677–1683.