Genetic relatedness of hepatitis A virus strains recovered from different geographical regions

Omana V. Nainan,[1] Gunter Siegl,[4] Anders Widell,[5] Harold S. Margolis,[1] Shin Isomura,[6] Kiichi Ito,[7] Tetsuya Ishizu,[8] Yasuo Moritsugu[3] and Stanley M. Lemon[2]

[1]*Hepatitis Branch (A33), Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control (World Health Organization Collaborating Centre for Research and Reference in Viral Hepatitis), Atlanta, Georgia 30333,* [2]*Division of Infectious Diseases, Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599–7030, U.S.A.,* [3]*National Institute of Health, Murayama Annex, 4-7-1 Gakuen, Musashimurayama, Tokyo 208, Japan,* [4]*Institute for Clinical Microbiology and Immunology, Frohbergstrasse 3, CH 9000 St Gallen, Switzerland,* [5]*Department of Medical Microbiology, University of Lund, Malmo General Hospital, S-21401, Malmo, Sweden,* [6]*Aichi Prefectural Institute of Public Health, 7-6 Nagare, Tsujimachi, Kita-Ku, Nagoya 462,* [7]*Shizuoka Prefectural Institute of Public and Environmental Health, 4-27-2 Kita-Ando, Shizuoka 420 and* [8]*Mie Prefectural Institute of Public Health, 3-446-34 Sakurabashi, Tsu 514, Japan*

A pairwise comparison of the nucleic acid sequence of 168 bases from 152 wild-type or unique cell culture-adapted strains of hepatitis A virus (HAV) revealed that HAV strains can be differentiated genetically into seven unique genotypes (I to VII). In general, the nucleotide sequence of viruses in different genotypes differs at 15 to 25% of positions within this segment of the genome. Viruses from four of the genotypes (I, II, III and VII) were recovered from cases of hepatitis A in humans, whereas viruses from the other three genotypes (IV, V and VI) were isolated only from simian species developing a hepatitis A-like illness during captivity. Among non-epidemiologically related human HAV strains, 81 were characterized as genotype I, and 19 as genotype III. Within each of these major genotypes, there were two distinct groups (sub-genotypes), which differed in sequence at approximately 7·5% of base positions. Each genotype and sub-genotype has a characteristic amino acid sequence in this region of the polyprotein, with the most divergent genotypes differing at 10 of 56 residues. Strains recovered from some geographical regions belonged to a common (endemic) genotype, whereas strains from other regions belonged to several, probably imported, genotypes. Thus, HAV strains recovered in North America were for the most part closely related at the nucleotide sequence level, whereas in other regions, such as Japan and Western Europe, HAV strains were derived from multiple genotypes or sub-genotypes. These data indicate that patterns of endemic transmission can be differentiated from situations in which infections are imported due to travel.

Introduction

Human hepatitis A virus (HAV), a hepatotropic member of the picornavirus family (Melnick 1982; Gust *et al.*, 1983), has recently been reclassified as the type species of a new genus, *Hepatovirus* (Minor, 1991). HAV was previously classified as enterovirus 72 based upon its physical properties, which are similar to those of other enteroviruses, and its faecal–oral route of transmission. Despite its overall physical and epidemiological similarity to enteroviruses, the unique structural composition of HAV, its stability characteristics, tissue tropism and genetic distance from members of other picornavirus genera indicate that HAV is unique within the picornavirus family, and have provided the impetus for the establishment of a separate genus for HAVs (Palmenberg, 1989). As in other picornaviruses, the large open reading frame present in the genome of HAV may be divided into three functional regions, termed P1, P2 and P3 (Rueckert & Wimmer, 1984). The P1 genomic region encodes the capsid polypeptides (VP1, VP2, VP3 and a putative VP4), and the P2 and P3 regions encode the non-structural polypeptides necessary for virus replication. The genome itself is 7·5 kb in length and of positive sense.

Antibodies to human HAV elicited by either natural or

experimental infection do not distinguish between individual strains of HAV, and thus there is only a single serotype of human HAV (Gellis *et al.*, 1945; Neefe *et al.*, 1946; Provost *et al.*, 1973; Rakela *et al.*, 1976; Lemon & Binn, 1983). Furthermore, monoclonal antibodies raised to various strains of human HAV have failed to differentiate more than one antigenic type (MacGregor *et al.*, 1983; Dawson *et al.*, 1984; Crevat *et al.*, 1990). Owing to this high degree of antigenic conservation among human HAV strains, infection with HAV is likely to confer life-long immunity that protects against subsequent symptomatic reinfection. As expected, the lack of antigenic diversity is mirrored by a high degree of conservation in the amino acid sequences of capsid proteins (Ticehurst *et al.*, 1989).

Early comparative studies of the nucleotide sequences of different human HAV strains suggested that isolates of diverse epidemiological origin were remarkably closely related (Ticehurst *et al.*, 1989). However, more recent nucleotide sequencing of selected genome regions encoding the C terminus of VP3, the N terminus of VP1, or the putative VP1/2A junction region of wild-type HAV strains present in human specimens has demonstrated substantial sequence heterogeneity. These studies have suggested that sequence relatedness can be correlated with the geographical origin of viruses (Jansen *et al.*, 1990; Robertson *et al.*, 1991*a*). In this manuscript we present a comprehensive analysis of sequence data derived from the VP1/2A junction region, resulting in the definition of seven distinct genotypes of HAV. The 107 unique virus strains included in this analysis represent the majority of HAV strains available worldwide. This analysis, which includes virus strains of both human and simian origin, provides a framework for the future use of partial genomic sequencing in defining the molecular epidemiology of HAV infection, and in the design of improved strategies for control of hepatitis A.

Methods

Virus strains. One-hundred and seventy-one individual virus isolates or clinical specimens containing HAV (stools, liver suspensions and serum) were assembled from a variety of sources, including the HAV Strain Bank organized by the Programme for Vaccine Development of the WHO. For each virus, the genomic sequence in the region of the VP1/2A cleavage site was determined, as described below. However, only wild-type viruses from primary clinical material, or cell culture isolates of HAV which had unique nucleotide sequences were included in the analysis of the genetic relatedness among HAV strains (total 152 strains), as preliminary studies have indicated that some cell culture isolates are in fact laboratory virus contaminants (Jansen *et al.*, 1990). Virus strains included in the analysis of genetic relatedness are listed in Table 1 according to their presumed geographical source and date of recovery (only a single virus is listed where multiple identical viruses were recovered from patients involved in a common outbreak of disease). The 19 cell culture isolates found to have VP1/2A sequences identical to those of common laboratory strains of HAV are listed in Table 2.

Antigen capture, reverse transcription and polymerase chain reaction (PCR). Virus in stool or liver suspensions (2 to 20%), or from cell culture was affinity-captured using anti-HAV-specific antibodies, followed by heat denaturation of the viral capsid and RNA, reverse transcription with an appropriate negative-sense primer and subsequent amplification of cDNA by PCR as described previously (Jansen *et al.*, 1990; Robertson *et al.*, 1991*a*). Viral RNA was recovered from serum by proteinase K digestion and phenol–chloroform extraction, followed by ethanol precipitation prior to reverse transcription and PCR amplification (Robertson *et al.*, 1991*a*). The genome region selected for comparison was located at the putative junction between the C terminus of VP1 and the N terminus of P2A (Cohen *et al.*, 1987). This region has been shown previously to have a relatively high degree of sequence variability in comparison with other regions of the HAV genome (Brown *et al.*, 1989). It is also analogous to the region utilized for a similar comparative study of nucleotide sequences of poliovirus strains (Rico-Hesse *et al.*, 1987). However, it is important to note that the precise HAV VP1/2A cleavage site has yet to be determined. The oligonucleotide primer pairs used for PCR of HAV cDNA included: +2984 and −3285, +2984 and −3265, +2949 and −3285, +2934 and −3285, +2799 and −3273 (Table 3).

DNA purification and nucleic acid sequencing. Amplified cDNA was purified by agarose or acrylamide gel electrophoresis and sequenced either by extension of a ^{32}P-5-end-labelled primer (Jansen *et al.*, 1990), or by incorporative [^{35}S]- (Robertson *et al.*, 1991*a*) or [^{32}P]dideoxynucleotide sequencing in the presence of 10% DMSO (Winship, 1989) with Sequenase (United States Biochemicals).

Computer analysis of sequences. Nucleotide sequences were analysed using computer algorithms designed for the VAX computer and provided by the University of Wisconsin (Devereux *et al.*, 1984). Dendrogram analyses were carried out using a program written by Mark Pallansch (CDC) and used previously for the analysis of poliovirus genomic segments (Rico-Hesse *et al.*, 1987).

Results

Genetic analysis of wild-type HAV strains

We compared the partial nucleotide sequences of 152 individual HAVs recovered from stool, liver, serum or cell culture materials. Most of these strains were studied directly in human clinical materials, although we also included cell culture isolates of HAV when their sequences were unique compared to common laboratory strains of HAV (HM175, MBB or HAS15) (see Methods). Together, these strains originated in 29 countries in seven continents and subcontinents and thus include HAV strains causing hepatitis in all regions of the world. The earliest strains to be recovered were CR326 and MS-1, which were identified in human faecal specimens collected approximately 30 years ago (Table 1). However, most of the strains we studied were collected between 1978 and 1990. Four of the HAV strains were recovered from non-human primates held in captivity: PA21 from owl monkeys, JM55 and Cy145 from cynomolgus monkeys, and AGM27 from African

Table 1. *Genotype, source and year of recovery of HAV strains**

Strain	Genotype	Location	Year	Contributor
North America				
MS-1	IA	Willowbrook, N.Y.	1964	S. M. Krugman
SD-11	IA	San Diego, Ca.	1974	R. H. Purcell
LA	IA	Los Angeles, Ca.	1975	J. Rakela†
AK76	IA	Alaska, U.S.A.	1976	Arctic Investigation Laboratory
FR-AL	IA	Fort Richardson, Alaska	1976	W. Bancroft
GA76	IIIA	Georgia, U.S.A.	1976	CDC Archive sample
HAS15	IA	Phoenix, Ariz.	1979	CDC Archive sample
LV-BE	IA	Levenworth, Kan.	1982	R. Redfield
CA85	IA	Berkley, Ca.	1985	L. S. Oshiro
HA88	IA	Honolulu, Haw.	1988	M. E. Melish
UT88	IA	Utah, U.S.A.	1988	A. Pavia
GA88	IA	Atlanta, Ga.	1988	C. H. Cowart
NC-1	IA	Chapel Hill, N.C.	1989	E. Brown
CA89	IA	Los Angeles, Ca.	1989	J. Vargas
GA90	IA	Georgia, U.S.A.	1990	M. Niu
MT90	IA	Montana, U.S.A.	1990	L. Polish
Central and South America				
CR326	IA	Costa Rica	1960	V. Villarejos‡
RDJ	IA	Brasilia, Brazil	1980	J. V. Parry
Ep-13·006	IA	Belem, Brazil	1981	G. Bensabath
H-201	IB	S. America (Sweden)	1982	A. Widell
Ep-35·730	IA	Maraba, Brazil	1985	G. Bensabath
MEX-1	IA	Mexico (U.S.A.)	1986	S. Lemon
Ep-49·674	IA	Belem, Brazil	1988	G. Bensabath
Arg90	IA	Argentina	1990	J. Pintera
Europe				
H-122	IIIA	Malmo, Sweden	1979	A. Widell
H-132	IIIA	Malmo, Sweden	1979	A. Widell
H-141	IIIA	Malmo, Sweden	1979	A. Widell
KPH	IIIB	Copenhagen, Denmark	1979	A. Widell
KMW-1§	IA	Switzerland	1979	G. Siegl
CP	IA	England	1979	J. V. Parry
CF-53	II	Clermont-Ferrand, France	1979	M. Deloince
H-148	IA	Vasteras, Sweden	1980	A. Widell
H-152	IB	Germany (Sweden)	1981	A. Widell
H-174	IIIA	Lund, Sweden	1978	A. Widell
GR7	IA	Grafenwohr, Germany	1982	S. Lemon
RM 238	IA	Czechoslovakia	1982	R. Scheid
JF 136	IA	Czechoslovakia	1982	R. Scheid
Ag11	IB	Athens, Greece	1983	G. Papaevangelou
Ag5978	IB	Thiva, Greece	1983	G. Papaevangelou
Ag6084	IB	Athens, Greece	1983	G. Papaevangelou
Ag6014	IB	Chalkis, Greece	1983	G. Papaevangelou
H-209	IIIA	Malmo, Sweden	1984	A. Widell
H-213	IIIA	Malmo, Sweden	1984	A. Widell
H-217	IIIA	Malmo, Sweden	1984	A. Widell
H-228	IIIA	Malmo, Sweden	1986	A. Widell
Carina	IA	Italy	1986	A. Pana
U.S.S.R.				
1406	IA	East Siberia, U.S.S.R.	1983	M. Balayan
2424	IA	Lithuania	1985	M. Balayan
406-I	IA	Gorky, U.S.S.R.	1987	M. Balayan
China				
China 81	IA	Shanghai, P.R.C.	1981	Z. Y. Xu
China 83	IA	Jiansu, P.R.C.	1983	Z. Y. Xu
LCDC-1§	IA	Shanghai, P.R.C.	1984	R. Chaudhury
S85-1§	IA	Sichuan, P.R.C.	1985	Anti epidemic station
PRC16	IA	Shanghai, P.R.C.	1988	M. Hu
PRC37	IA	Shanghai, P.R.C.	1988	M. Hu

Table 1.—(*continued*)

Strain	Genotype	Location	Year	Contributor
Southeast Asia, Japan and India				
KRM031	IA	Saga, Japan	1977	Y. Moritsugu
KRM016	IIIB	Saga, Japan	1977	Y. Moritsugu
KRM003	IIIB	Fukuoka, Japan	1979	Y. Moritsugu
H-229	IIIA	Sri Lanka (Sweden)	1984	A. Widell
SMA 292	IIIA	Senang, Malaysia	1986	I. Gust
M-136	IA	Mie, Japan	1986	T. Ishizu
M-138	IA	Mie, Japan	1986	T. Ishizu
SR082	IA	Bangkok, Thailand	1987	B. Innis
SR035	IA	Bangkok, Thailand	1987	B. Innis
SR102	IA	Bangkok, Thailand	1987	B. Innis
PMA 32	IA	Penang, Malaysia	1988	I. Gust
TK023	IIIA	Kathmandu, Nepal	1989	B. Innis
India90	IIIA	India	1990	B. Khanna
PH074	IA	Nan, Thailand	1990	B. Innis
A-4	IA	Aichi, Japan	1990	S. Isomura
A-12	IA	Aichi, Japan	1990	S. Isomura
A-113	IA	Aichi, Japan	1990	S. Isomura
A-127	IA	Aichi, Japan	1990	S. Isomura
A-129	IA	Aichi, Japan	1990	S. Isomura
A-148	IA	Aichi, Japan	1990	S. Isomura
A-149	IA	Aichi, Japan	1990	S. Isomura
A-162	IIIB	Aichi, Japan	1990	S. Isomura
A-167	IA	Aichi, Japan	1990	S. Isomura
A-177	IIIB	Aichi, Japan	1990	S. Isomura
A-185	IA	Aichi, Japan	1990	S. Isomura
A-261	IA	Aichi, Japan	1990	S. Isomura
TJ-D-1	IA	Gifu, Japan	1990	Y. Moritsugu
S-3	IA	Shizuoka, Japan	1990	K. Ito
S-6	IA	Shizuoka, Japan	1990	K. Ito
S-13	IA	Shizuoka, Japan	1990	K. Ito
M-25	IIIB	Mie, Japan	1990	T. Ishizu
M-30	IA	Mie, Japan	1990	T. Ishizu
M-53	IA	Mie, Japan	1990	T. Ishizu
M-63	IA	Mie, Japan	1990	T. Ishizu
M-74	IA	Mie, Japan	1990	T. Ishizu
M-91	IA	Mie, Japan	1990	T. Ishizu
M-100	IA	Mie, Japan	1990	T. Ishizu
M-102	IA	Mie, Japan	1990	T. Ishizu
M-116	IA	Mie, Japan	1990	T. Ishizu
M-118	IA	Mie, Japan	1990	T. Ishizu
KS-T-1	IA	Tokyo, Japan	1990	Y. Moritsugu
Australia, Africa and the Middle East				
HM175	IB	Australia	1976	I. Gust
MBB	IB	N. Africa (Germany)	1978	G. Frosner
TKM005	IB	Iraq (Japan)	1981	Y. Moritsugu
TKM002	IA	Nigeria (Japan)	1981	Y. Moritsugu
H153	IB	Tunisia (Sweden)	1981	A. Widell
No. 4	IB	N. Africa	?	H. Garelick
Jor88	IB	Amman, Jordan	1988	A. Toukan
SLF88	VII	Sierra Leone	1988	L. Chapman
Non-human primates				
PA21	IIIA	Panama	1980	*Aotus trivirgatus*
JM55	VI	Indonesia (U.S.S.R.)	1985	*Macaca fascicularis*
AGM27	V	Kenya (U.S.S.R.)	1985	*Cercopithecus aethiops*
Cy145	IV	Philippines (U.S.A.)	1988	*Macaca fascicularis*

* When the source of infection was strongly considered to be travel-associated, the strain was classified by the country in which the infection was presumed to have been acquired (the country within which the strain was actually recovered is indicated in parentheses). Some dates may represent year of publication rather than actual year of recovery. When strains were recovered from patients in a common outbreak, only one strain is listed (see Table 4).

† Najarian *et al.* (1985). ‡ Linemeyer *et al.* (1985). § Probably a contaminant.

Table 2. *Cell culture isolates of HAV excluded from analysis*

Isolate	Location	Year	Passage	Cells
MBB-like isolates				
CLF	Switzerland	1983	1	MRC-5
GBM	Germany	1976	1	MRC-5
LSH/S	England	1983	1	MRC-5
KMW-1*	Switzerland	1979	7	MRC-5
Sy77	China	1978	5	MRC-5
BR 51/7	Germany	1982	1	PRC/PRF/5
CF979	U.S.A.	1979	1	PRC/PRF/5
SM1	China	1978	6	PRC/PRF/5
190/1	Greece	1987	2	FRhK4
HEL2	Germany	1986	3	?
HM175-like isolates				
MD1	U.S.A.	1981	?	AGMK?
S84-1	China	1984	15	Primary human lung
S85-1*	China	1985	3	Primary human lung
HLJA	China	?	5	Primary human lung
S7	China	?	2	?
HAS15-like isolates				
LCDC-1*	China	1984	20	FRhK4

* Those isolates strongly suspected of being contaminants.

Table 3. *Primers used for amplification and sequencing*

Positive sense primers	
+2799	5′ ATTCAGATTAGACTGCCTTGGTA 3′
+2891	5′ GGTTTCTATTCAGATTGCAAATTA 3′
+2933	5′ TTTGTCTTTTAGTTGTTATTTGTCTGT 3′
+2934	5′ TTGTCTTTTAGTTGTTATTTGTCTG 3′
+2949	5′ TATTTGTCTGTCACAGAACAATCAG 3′
Negative sense primers	
−3192	5′ AGGAGGTGGAAGCACTTCATTTGA 3′
−3265	5′ CATTATTTCATGCTCCTCAG 3′
−3273	5′ CCAAGAAACCTTCATTATTTCATG 3′
−3285	5′ AGTCACACCTCTCCAGGAAAACTT 3′
−3375	5′ AGTAAAAACTCCAGCATCCATTTC 3′

green monkeys. The remainder of the strains were from human cases of hepatitis A. A detailed list of the strains is shown in Table 1.

The extent of nucleotide variation among these 152 strains within a 168 base region encoding the putative VP1/2A junction provided a measure of the relatedness of individual human strains to each other, and to recently characterized simian HAV strains (Nainan *et al.*, 1991; Tsarev *et al.*, 1991). A pairwise analysis of the nucleotide sequences of cDNA amplified by a PCR-based procedure resulted in the dendrogram shown in Fig. 1. This figure includes only 93 of the 152 strains we evaluated, as a number of strains with sequences identical to those of others (Table 4) were excluded to simplify the graphic comparison. These data confirm the relatively high degree of genomic conservation found among HAV strains in general, but they also demonstrate the existence of genetically related groups which are often (but not always) associated with certain geographical regions.

Based upon percentage nucleotide identity, we have grouped strains of HAV into seven major genotypes (Fig. 1 and Table 1). Adopting criteria applied to the comparison of poliovirus sequences (Rico-Hesse *et al.*, 1987), we have defined an HAV genotype as a group of viruses having nucleotide sequences which differ from each other at no more than 15% of base positions (Jansen *et al.*, 1991). Four of the genotypes shown in Fig. 1 (I, II, III and VII; original genotype numbering from Jansen *et al.*, 1990) include strains of HAV that have been associated with human disease, whereas the remaining three genotypes (IV, V and VI) each include a single unique simian HAV strain. Genotype III is unique in that it includes strains that have been recovered from both human and non-human primate sources.

Genotype I

This genotype includes 82 of the 104 (80%) human HAV strains studied (Table 1 and Fig. 1). We further divided this genotype into two sub-genotypes (IA and IB) differing from each other at approximately 7·5% of base positions (Jansen *et al.*, 1991), as the dendrogram (Fig. 1) and related amino acid sequences (Fig. 2) suggest a natural division. Sub-genotype IA comprises the majority of the human strains studied (69 of 104, 67%), and includes strains found world-wide. It constitutes the major virus population in both North and South America, China, Japan, the former U.S.S.R. and Thailand. However, three geographically related clusters of viruses with closely related sequences are found in this genotype. These include one group of strains from the U.S.A., another from Japan and a third group recovered both in Japan and China (Fig. 1, shaded areas). The remaining strains within the IA sub-genotype appear to be more randomly dispersed with respect to geographical origin. Sub-genotype IB contains strains from Jordan, North Africa, Australia, Europe, Japan and South America. However, the majority of these strains were recovered from locations near the Mediterranean (Table 1).

Genotype III

Most of the remaining human HAV strains segregate into a single genotype that can also be divided into two sub-genotypes, IIIA and IIIB. The prototype virus strain of the IIIA sub-genotype, PA21, was originally isolated from recently captured Panamanian owl monkeys

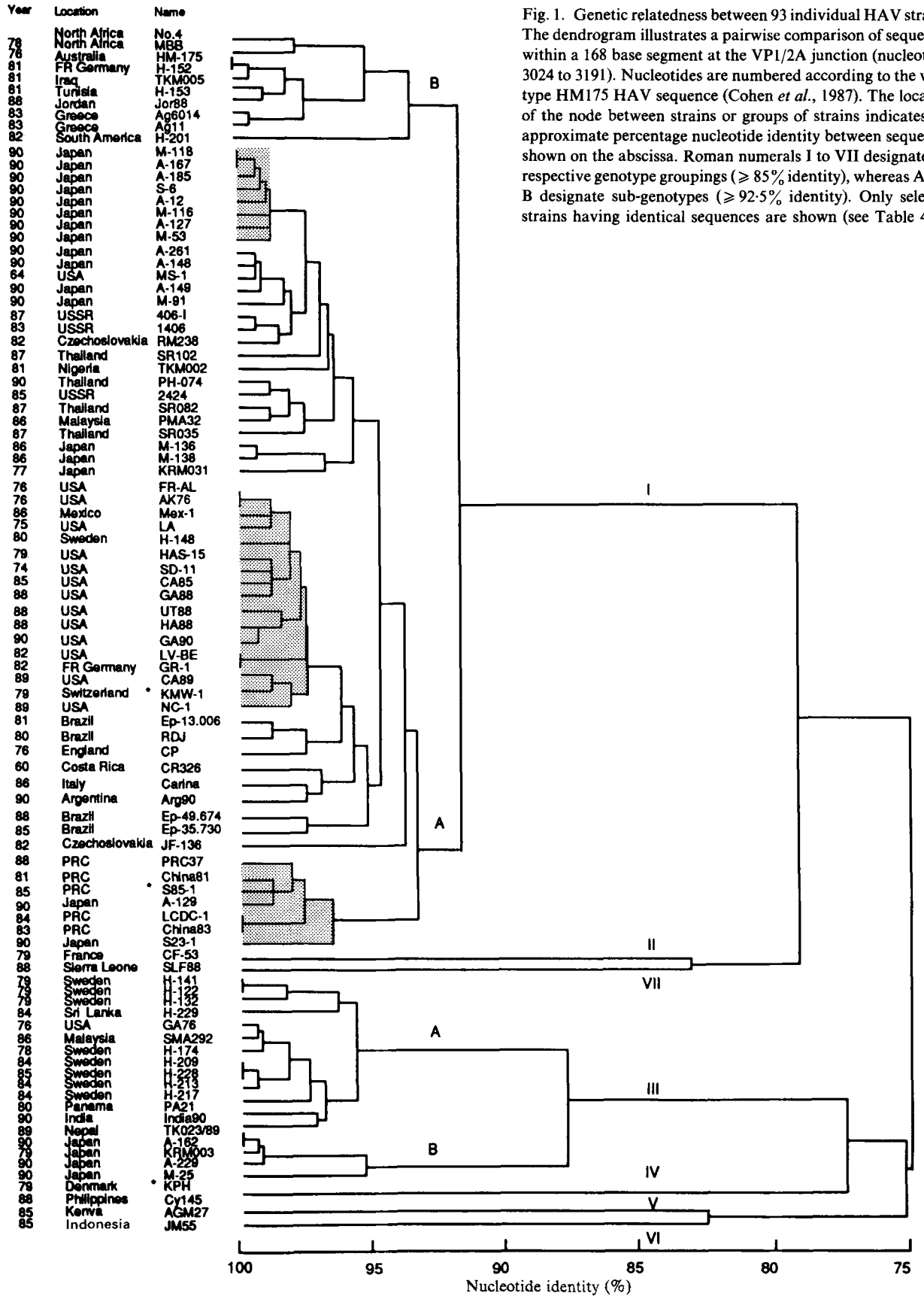

Fig. 1. Genetic relatedness between 93 individual HAV strains. The dendrogram illustrates a pairwise comparison of sequences within a 168 base segment at the VP1/2A junction (nucleotides 3024 to 3191). Nucleotides are numbered according to the wild-type HM175 HAV sequence (Cohen *et al.*, 1987). The location of the node between strains or groups of strains indicates the approximate percentage nucleotide identity between sequences shown on the abscissa. Roman numerals I to VII designate the respective genotype groupings ($\geq 85\%$ identity), whereas A and B designate sub-genotypes ($\geq 92{\cdot}5\%$ identity). Only selected strains having identical sequences are shown (see Table 4).

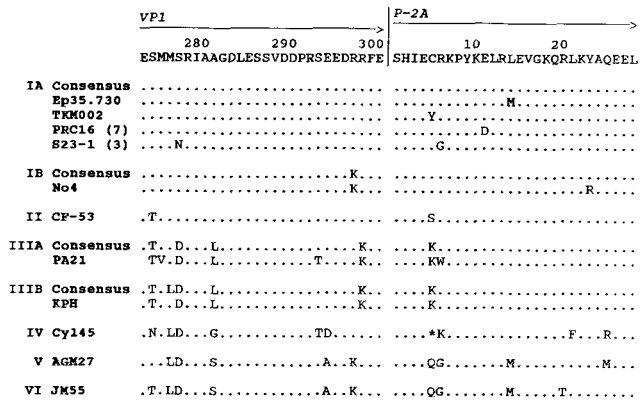

Fig. 2. Comparison of the predicted amino acid sequences near the putative VP1/2A junction. The consensus amino acid sequence for the predominant sub-genotype IA is shown on the top line, with the corresponding amino acid residue numbers for VP1 and P2A based on the cleavage site proposed for HAV strain HM175 (Cohen *et al.*, 1987). The consensus for each genotype is shown, along with the sequence of strains which differ from the consensus (numbers in parentheses indicate the number of strains with an identical divergent sequence). Dots indicate conserved amino acids; differences are shown by the appropriate single letter amino acid code. An asterisk indicates a deletion.

(Brown *et al.*, 1989). However, closely related viruses have subsequently been identified in specimens collected from humans with hepatitis A in India, Sri Lanka, Nepal, Malaysia, Sweden and the U.S.A. (Jansen *et al.*, 1991; Khanna *et al.*, 1991; Robertson *et al.*, 1991*b*). The IIIB sub-genotype is responsible for cases of HAV infection in Japan and Denmark.

Genotypes II and VII

These two genotypes each contain only a single strain. Genotype II is represented by a strain isolated in France in 1979, CF-53, whereas genotype VII contains a strain isolated in Sierra Leone in 1988, SLF88. These strains differ from each other at approximately 17% of base positions, and from other HAV strains by 20 to 25%. The CF-53 strain was recovered from a patient with typical disease in an otherwise unremarkable outbreak of hepatitis A in central France. The SLF88 strain was recovered from the liver of a patient who died with fulminant hepatitis A.

Genotypes IV, V and VI

The remaining three genotypes are each represented by a single unique virus strain recovered from an Old World monkey species. Two of these viruses were recovered at a single primate holding facility in the Soviet Union, one from a cynomolgus macaque (*Macaca fasicularis*) recently imported from Indonesia (genotype VI, JM55 virus), and the other from an African green monkey (*Cercopithecus aethiops*) imported from Kenya (genotype V, AGM27 virus). The third of these simian strains (genotype IV, Cy145 virus) was recovered in Atlanta, Ga., U.S.A. from a cynomolgus macaque imported from the Philippines. These three simian strains tend to differ genetically as much between themselves as they do from the human viruses. The AGM27 and JM55 strains appear to be more closely related to each other than to the Cy145 strain, even though they were recovered from different primate species.

Amino acid sequence diversity

The nucleotide sequence heterogeneity in this region of the HAV genome of these 108 virus strains results in only limited differences in the amino acid sequence (Fig. 2). Within sub-genotype IA, the amino acid sequence is conserved with the exception of a cluster of strains from China and Japan that have an Asn substitution at Met 277 of VP1 (residue 1-277) and a Gly substitution at Arg 2A-6, or an Asp substitution at Glu 2A-10. The amino acid sequences of sub-genotype IB strains are also highly conserved, and differ from most sub-genotype IA strains in that they have a Lys substitution at Arg 1-297. Although the nucleotide sequences of the CF-53 and SLF88 viruses (genotypes II and VII) differ from each other and those of genotype I and III viruses at about 15% of base positions, their predicted amino acid sequences are identical in this region, and include a Ser substitution at Thr 1-274 and a Cys substitution at Ser 2A-5. Greater differences are evident in the amino acid sequences of the remaining four genotypes (Fig. 2). The amino acid substitutions that distinguish the genotype III viruses from the consensus sequence derived from genotype I viruses include a Thr residue at Ser 1-274, Asp at Ser 1-277, Leu at Ala 1-281, Lys at Arg 1-298 and Lys at Cys 2A-5. The common sites of amino acid substitutions within simian genotypes IV to VI include Leu at Met 1-276, Asp at Ser 1-277, Ser or Gly at Ala 1-281, Ala or Asp at Glu 1-294, Glu or a deletion at Cys 2A-5, and Gly or Lys at Arg 2A-6. In addition, there are several other substitutions in the amino acid sequences of these three strains (Fig. 2). Overall, certain residues are more likely to be sites of subsitutions in the different genotypes than others. These hypervariable residues include 1-273, 1-277, 1-281, 1-294, 2A-5, 2A-6 and 2A-24, at each of which at least three different amino acids are found in different genotypes.

Molecular epidemiology of HAV

The ability to determine the genotype of HAV strains during epidemic or endemic periods offers the potential

Table 4. *HAV strains with identical sequences in the 168 base VP1/2A region*

Strain	Genotype	Location	Year	Contributor
GR1*	IA	Grafenwohr, Germany	1982	S. Lemon
LV-BE		Levenworth, Kan.	1982	R. Redfield
GA90	IA	Georgia, U.S.A.	1990	M. Niu
MT90		Montana, U.S.A.	1990	L. Polish
HM175	IB	Australia	1976	I. Gust
H-152†		Germany (Sweden)	1981	A. Widell
FR-AL	IA	Fort Richardson, Alaska	1976	W. Bancroft
AK76*		Alaska, U.S.A.	1976	Arctic Investigation Laboratory
China83	IA	Jiansu, PRC	1983	Z. Y. Xu
LCDC-1		Shanghai, PRC	1984	R. Chaudhury
H-122	IIIA	Malmo, Sweden	1979	A. Widell
H-141		Malmo, Sweden	1979	A. Widell
H-209	IIIA	Malmo, Sweden	1984	A. Widell
H-213		Malmo, Sweden	1984	A. Widell
H-228		Malmo, Sweden	1986	A. Widell
Ag11	IB	Athens, Greece	1983	G. Papaevangelou
Ag5978		Thiva, Greece	1983	G. Papaevangelou
Ag6084	IB	Athens, Greece	1983	G. Papaevangelou
Ag6014		Chalkis, Greece	1983	G. Papaevangelou
KRM016*	IIIB	Saga, Japan	1977	Y. Moritsugu
KRM003		Kyushu, Japan	1979	Y. Moritsugu
A-162†		Aichi, Japan	1990	S. Isomura
A-177	IIIB	Aichi, Japan	1990	S. Isomura
A-229†		Aichi, Japan	1990	S. Isomura
A-149*	IA	Aichi, Japan	1990	S. Isomura
A-204		Aichi, Japan	1990	S. Isomura
A-211		Aichi, Japan	1990	S. Isomura
A-322		Aichi, Japan	1990	S. Isomura
M-87		Aichi, Japan	1990	S. Isomura
M-63†		Mie, Japan	1990	T. Ishizu
A-167	IA	Aichi, Japan	1990	S. Isomura
A-181		Aichi, Japan	1990	S. Isomura
M-118†		Mie, Japan	1990	T. Ishizu
M-119		Mie, Japan	1990	T. Ishizu
M-121		Mie, Japan	1990	T. Ishizu
TJ-D-1*		Gifu, Japan	1990	Y. Moritsugu
KS-T-1		Tokyo, Japan	1990	Y. Moritsugu
M-74		Mie, Japan	1990	T. Ishizu
M-75		Mie, Japan	1990	T. Ishizu
S-6		Shizuoka, Japan	1990	K. Ito
A-127	IA	Aichi, Japan	1990	S. Isomura
M-100†		Mie, Japan	1990	T. Ishizu
S-32[2]		Shizuoka, Japan	1990	K. Ito
A-4	IA	Aichi, Japan	1990	S. Isomura
A-8		Aichi, Japan	1990	S. Isomura
A-37		Aichi, Japan	1990	S. Isomura
A-192		Aichi, Japan	1990	S. Isomura
A-226		Aichi, Japan	1990	S. Isomura
A-231		Aichi, Japan	1990	S. Isomura
A-302		Aichi, Japan	1990	S. Isomura
A-43†		Aichi, Japan	1990	S. Isomura
A-174†		Aichi, Japan	1990	S. Isomura
M-30†		Mie, Japan	1990	T. Ishizu
M-49†		Mie, Japan	1990	T. Ishizu
M-54†		Mie, Japan	1990	T. Ishizu
M-61†		Mie, Japan	1990	T. Ishizu
S-13†		Shizuoka, Japan	1990	K. Ito
S-18†		Shizuoka, Japan	1990	K. Ito
S-23-1-1*†		Shizuoka, Japan	1990	K. Ito
M-102	IA	Mie, Japan	1990	T. Ishizu
A-113		Aichi, Japan	1990	S. Isomura
A-188		Aichi, Japan	1990	S. Isomura

Table 4.—(*continued*)

Strain	Genotype	Location	Year	Contributor
A-193	IA	Aichi, Japan	1990	S. Isomura
A-249		Aichi, Japan	1990	S. Isomura
A-300		Aichi, Japan	1990	S. Isomura
A-251		Aichi, Japan	1990	S. Isomura

* Multiple identical strains collected from different patients involved in a school outbreak or common source outbreak of infection, or from family contact.

† May be distinguished from other strains within the group by sequencing in other genome regions.

to define various aspects of HAV infection and transmission at the level of molecular epidemiology. We found several different patterns when we examined the geographical distribution of genotypes. One pattern is that seen within the U.S.A., for which 15 of the 16 strains evaluated were of sub-genotype IA. Moreover, 13 of these strains formed a discrete, closely related geographical cluster within this genotype (Fig. 1). These data suggest the existence of a circulating endemic population of HAV. In contrast, strains collected from Western Europe (Sweden, Germany, Denmark and Switzerland) differ widely in their respective genotypes, suggesting that there is little endemic transmission of HAV in this geographical region and that most cases are related to importation from other regions in which endemic transmission occurs. Although the origin of some of these strains is known for certain, it can only be surmised in most cases (MBB, for example, was recovered in Germany from a traveller to Africa) (Table 1).

Fig. 3(*a*) illustrates the genetic relatedness of viruses recovered in Sweden, a country with excellent public health sanitation services. Among these viruses, the sub-genotype IA and IB strains (H148, H153, H152 and H201) were generally associated with a history of recent travel (Tunisia, West Germany or South America) during 1980 to 1982. In contrast, genotype IIIA viruses were recovered from cases of HAV predominantly associated with parenteral drug abuse within the Malmo region during the interval from 1979 to 1985. In only one of these sub-genotype IIIA cases was there a history of travel (Sri Lanka). Thus, in Sweden, there appears to have been endemic circulation of a particularly closely related group of viruses among parenteral drug users in one city, whereas most of the other cases involved quite distantly related viruses imported from other geographical regions.

The Japanese strains studied included viruses causing sporadic cases as well as outbreaks of hepatitis A (Fig. 3*b*). This relatively small geographical region has a low endemic rate of HAV. However, more than one genotype of HAV was identified within the same year and even within the same location. Remarkably, two unique strains of HAV were recovered from different children involved in an HAV outbreak which occurred in Kiyama Elementary School, Saga Prefecture during 1977. Of 11 children involved in this outbreak, nine were infected with KRM031, a sub-genotype IA strain. However, two children were infected with a distantly related sub-genotype IIIB virus, KRM016, which was identical in sequence to KRM003 (Table 4). KRM031 and KRM003 differ in sequence at approximately 25% of base positions (Fig 3*b*). KRM003 was recovered 2 years later, from a sporadic hepatitis A case which occurred in Fukuoka Prefecture in 1979. In addition, a very closely related virus (A-162) was identified 11 years later in a sporadic case occurring in Aichi Prefecture during 1990. A-162 is identical to KRM003 within the genomic region analysed in Fig. 1, but these two strains can be distinguished by sequencing a larger 360 base segment in this region (data not shown).

Also in Japan, a cluster of seven strains having sequences identical to that of strain A-167 was identified among cases occurring in four different locations (Table 4). Two of these viruses, as well as other viruses having sequences identical to those of the A-149, M-91 or M-100 strains, were associated with transmission within families (data not shown). Thus, the strains of HAV studied in Japan probably represent a combination of travel-associated and endemic transmission patterns, with local spread of otherwise imported viruses. However, TKM002 and TKM005 are the only two strains for which a history of travel was known (Fig. 3*b*). These were recovered from patients after travel to Nigeria and Iraq, respectively.

Cell culture isolates

The growth of HAV in cell culture generally requires a long adaptation period and generally does not result in a c.p.e. Thus the contamination of cultured cells with laboratory-passaged strains is a constant possibility. For this reason we included in the analysis described above only those sequences derived from cell culture-adapted viruses that were unique in comparison to common

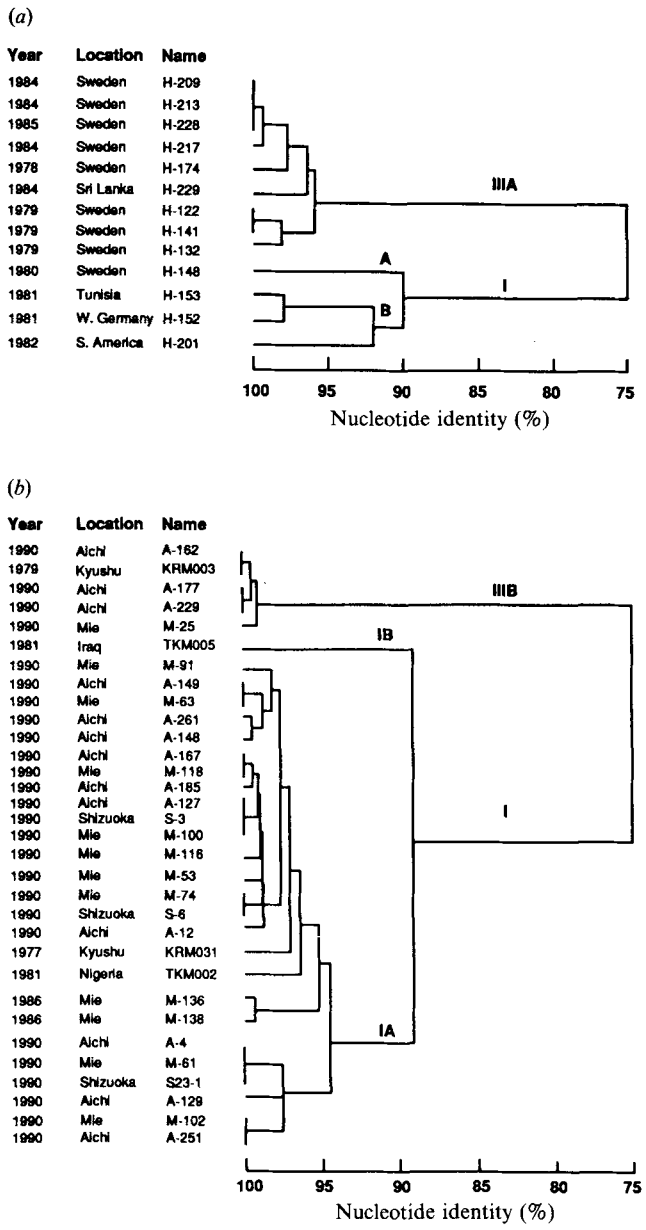

Fig. 3. Dendrograms depicting the genetic relatedness of human HAV strains recovered in Sweden (*a*) and Japan (*b*). See legend to Fig. 1.

laboratory isolates of HAV. This resulted in the exclusion of 19 virus strains with sequences in this region of the genome identical to those of MBB, HM175 or HAS15, and which therefore might be contaminants (Table 2). MBB, HM175 and HAS15 are well characterized cell culture-adapted isolates (Baroudy *et al.*, 1985; Ovchinnikov *et al.*, 1985; Paul *et al.*, 1987). These strains were adapted to cell culture relatively early in the history of HAV propagation, and each has been widely distributed among laboratories engaged in HAV research. The cell culture-adapted strains listed in Table 2 were isolated in a number of independent laboratories, suggesting that cross-contamination of cell cultures might be a generic problem in the isolation of HAV strains from clinical materials. In several instances (strains marked by the symbol § in Table 1 and an asterisk in Table 2 and Fig. 1), the cell culture 'isolate' is strongly suspected to be a contaminant because virus with a substantially different sequence has been identified in the primary clinical specimen (see Fig. 1) (Jansen *et al.*, 1990). Although the other isolates listed in Table 2 must be considered possible contaminants, it should be noted that a number of apparently unrelated virus strains, studied after direct isolation from primary clinical materials, have been found to have identical nucleotide sequences in the VP1/2A genomic region (see Table 4).

Stability of HAV genotypes

To assess the stability of the VP1/2A sequence of HAV within infected individuals over a period of time, we studied viruses identified in 10 serial specimens collected from four different persons. The nucleotide sequence of 360 bases was determined for each virus. In one case, a virus-positive specimen was collected 41 days after initial recovery of the virus, whereas in the other cases this interval ranged from 11 to 26 days. No sequence differences were identified in any of the viruses isolated from each of these patients (data not shown). These data are consistent with the fact that viruses isolated from different patients involved in several different outbreaks of hepatitis A always had a common sequence (Table 4). In addition, no change was found in the sequence of the VP1/2A region after six passages of the KRM003 strain in non-human primates (marmosets) (data not shown). A single nucleotide substitution (C to T at base 3002, HM175 numbering) was identified in this strain following 72 passages in cell culture.

Discussion

Limited genomic sequencing of cDNA amplified by PCR from viral RNA present within infected stool, liver or serum provides a valuable tool for defining populations of wild-type HAV strains without the potentially confounding factors of extended cell culture adaptation or alternative animal passage. Four genotypes of HAV associated with human disease have been identified and a framework has been constructed for epidemiological tracking of wild-type HAV in both endemic and epidemic settings, and for differentiating wild-type strains from potential vaccine-derived virus in future vaccine trials.

Overall, the seven HAV genotypes recovered from

humans and non-human primates differ from each other at approximately 15 to 25% of base positions in the VP1/2A region. However, HAV does not appear to undergo the rapid accumulation of genetic changes seen in many RNA viruses. If one assumes a common ancestral root for viruses in the IA sub-genotype, and accepts the caveat that the region of the genome we have examined reflects the overall pattern of genetic relatedness among different strains, an average of 10 substitutions/year would be predicted to accumulate within the entire genome. This may actually be an overestimate, because we chose the VP1/2A region for its relative variability compared with the VP3 and 5′ non-translated regions (Brown *et al.*, 1989). In contrast, poliovirus averages one to two substitutions/week based upon nucleic acid fingerprinting of serial samples examined during an epidemic (Nottay *et al.*, 1981; Kew *et al.*, 1984). Within sub-genotypes of HAV (IA, IB, IIIA or IIIB), there is less than 7·5% nucleotide diversity. Moreover, within a geographical cluster such as that seen in North America (Fig. 1), there is less than 3% diversity between individual strains, even when strains collected over a period of 15 years are examined. These observations lead one to speculate that HAV strains have maintained their low rate of accumulating mutations over a long period of time and may have developed specific ecological niches.

Although the overall sequence diversity evident among the HAV strains included in Fig. 1 is reminiscent of that found among individual serotypes of other picornaviruses (Palmenberg, 1989), there are important differences. We found only four genotypes among 89 non-epidemiologically related individual human strains (Fig. 1). In contrast, Rico-Hesse *et al.* (1987) found 19 genotypes of human poliovirus type 1 strains, using identical criteria for genotypic differentiation in a similar region of the genome. Moreover, unlike HAV there was no correlation between amino acid changes and genotype grouping among the polioviruses. Within HAV genotypes, there are only limited amino acid differences (one or two substitutions in 56 possible residues), whereas between genotypes the number of amino acid substitutions varies from approximately 3% (2/56) to about 18% (10/56). Thus, the amino acid sequence in the VP1/2A region correlates closely with genotype and even sub-genotype, and possibly even selected geographical clusters (Fig. 2). The exceptions to this observation were strains JF-136 and KPH, which grouped genetically as sub-genotypes IA and IIIB, but the predicted amino acid sequences of which sorted with sub-genotypes IB and IIIA, respectively.

There is evidence that viruses from different genotypes may have different biological phenotypes, particularly when one contrasts the viruses recovered from Old World monkeys (genotypes IV, V and VI) with the human strains. These simian-derived HAV agents are virulent for the primate species from which they were originally recovered, but they do not appear to replicate or cause disease when inoculated into chimpanzees, which are generally susceptible to human strains (O. Nainan *et al.*, unpublished results). In addition, there is no evidence of human disease due to the transmission of these viruses to animal handlers. These three simian strains also have important amino acid substitutions (Nainan *et al.*, 1991; Tsarev *et al.*, 1991) within the immunodominant antigenic site of HAV (Ping *et al.*, 1988), and thus they may represent serological variants of HAV. On the other hand, there is no obvious difference in disease expression between genotype I and III viruses. Animal transmission studies in Panamanian owl monkeys and chimpanzees suggest that both genotypes cause classical signs and symptoms of HAV, and that representative members of these genotypes confer cross-protection against each other (Brown *et al.*, 1989; Khanna *et al.*, 1991). Moreover, only very minor antigenic differences exist between these genotypes (Brown *et al.*, 1989).

Epidemiological tracking of various strains is another potential use of the information derived from these studies. For example, viruses (H209, H213 and H228) recovered from intravenous drug abusers in Malmo, Sweden between March 1984 and January 1985 are identical in this region, implying sequence conservation over a period of approximately 1 year. However, the strains we have studied also include several examples of viruses which are not closely related epidemiologically, but which nonetheless have identical base sequences in the VP1/P2A region (Table 4). For example, KRM003 and A-162 viruses, collected in Kyushu and Aichi, Japan during 1989 and 1990, are identical in this region, as are MT90 and GA90, which were collected in Montana and Georgia, U.S.A. during 1990. Another example is provided by LV-BE and GR-1, viruses recovered from American soldiers in the U.S.A. and Europe during 1982 (Jansen et al., 1990). Such data provide at least presumptive evidence for epidemiological linkage.

However, the stability of the HAV genome complicates an absolute assignment of direct genetic linkage based upon the sequence of a single region. For example, H152 (West Germany, 1981) and HM175 (Australia, 1976) viruses also share identical sequences in the VP1/2A region. Yet, even though these two viruses have identical nucleotide sequences in the VP3 coding region as well (altogether forming about 5% of the genome), we have found significant differences in the sequence of the 5′ non-translated regions of these viruses. Thus, care must be taken in reaching strong conclusions about the epidemiological relatedness of strains. Similar caution

must be exercised in considering that each of the cell culture-adapted strains listed in Table 2 is a contaminant (although it is likely that most are).

The sample population used in this analysis contains a disproportionate number of strains (>47%) from the U.S.A., Sweden and Japan. These three countries all have low endemic rates of hepatitis A compared to many less developed parts of the world. Consequently, these regions are substantially over-represented in this analysis. Nevertheless, the wide variety of genotypes found in both Sweden and Japan suggests that the strains recovered within these countries represent a mixture of imported strains as well as a continued low level of transmission of endemic strain(s) in certain high risk groups, such as drug addicts in Sweden (Widell *et al.*, 1983) or elementary school children in Saga, Japan. By contrast, most of the strains from the U.S.A. are tightly clustered genetically, suggesting that they represent almost exclusive endemic transmission of a predominant genotype which continues to circulate in that country. Samples from regions of the world which have a hyperendemic pattern of HAV (Gust, 1984), such as India, central and south Africa, and parts of South America, are particularly under-represented in the sample of viruses studied, although the strains sequenced represent the vast majority of HAV strains available for study world-wide. The strains examined were recovered from symptomatic cases of hepatitis A, but residents of hyperendemic regions are generally infected with HAV at a very early age and do not develop symptomatic infection. This may have influenced the number of possible clinical specimens available for analysis.

We would like to acknowledge Jeff Moyer who provided valuable assistance as a summer student at the Centers for Disease Control. We thank S. Tsarev for providing the 168 base sequence of the VP1/2A region of AGM27, and L. Binn, B. Innis, G. Papaevangelou, M. Balayan and our many other colleagues whose past efforts have made this work possible. This work was supported in part by a grant (S.M.L.) from the U.S. Army Medical Research and Development Command (DAMD 17-89-Z-9022) and Technical Services Agreements (S.M.L. and G.S.) with the Programme for Vaccine Development of the World Health Organization. The studies in Japan were supported by a grant-in-aid from the Ministry of Health and Welfare.

References

BAROUDY, B. M., TICEHURST, J. R., MIELE, T. A., MAIZEL, J. V., JR, PURCELL, R. H. & FEINSTONE, S. M. (1985). Molecular cloning and partial sequencing of hepatitis A virus cDNA coding for capsid proteins and RNA polymerase. *Proceedings of the National Academy of Sciences, U.S.A.* **82**, 2134–2147.

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. *Journal of Virology* **63**, 4932–4937.

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. *Journal of Virology* **61**, 50–59.

CREVAT, D., CRANCE, J. M., CHERVINAIS, A. M., PASSAGOT, J., BIZIAGOS, E., SOMME, G. & DELOINCE, R. (1990). Monoclonal antibodies against an immunodominant and neutralizing epitope on hepatitis A virus antigen. *Archives of Virology* **113**, 95–98.

DAWSON, G. J., DECKER, R. H., NORTON, D. K., BRYCE, W. H., WHITTINGON, R. O., TRIBBY, I. I. & MUSHAHWAR, I. K. (1984). Monoclonal antibodies to hepatitis A virus. *Journal of Medical Virology* **14**, 1–8.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the Vax. *Nucleic Acids Research* **12**, 387–395.

GELLIS, S. S., STOKES, J., BROTHER, G. M., HALL, W. M., GILLMORE, H. R., BEYER, E. & MORRISSEY, R. A. (1945). The use of human immune serum globulin (gamma globulin) in infectious (epidemic) hepatitis in the Mediterranean theatre of operations. *Journal of the American Medical Association* **128**, 1062–1063.

GUST, I. D. (1984). The epidemiology of viral hepatitis. In *Viral Hepatitis and Liver Disease*, pp. 415–421. Edited by G. N. Vyas, J. L. Dienstag & J. H. Hoofnagle. Orlando: Harcourt Brace Jovanovich.

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.

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. *Proceedings of the National Academy* of *Sciences U.S.A.* **87**, 2867–2871.

JANSEN, R. W., SIEGL, G. & LEMON, S. N. (1991). Molecular epidemiology of human hepatitis A virus (HAV). In *Viral Hepatitis and Liver Disease*, pp. 58–62. Edited by B. N. Hollinger, S. M. Lemon & H. S. Margolis. Baltimore: Williams & Wilkins.

KEW, O. M., NOTTAY, B. K. & OBIJESKI, J. F. (1984). Applications of oligonucleotide fingerprinting to the identification of viruses. In *Methods in Virology* vol. 8, pp. 41–84. Edited by K. Maramorosch & H. Koprowski. New York: Academic Press.

KHANNA, B., SPELBRING, J. E., INNIS, B. L. & ROBERTSON, B. H. (1991). Characterization of a genetic variant of human hepatitis A virus. *Journal of Medical Virology* **36**, 118–124.

LEMON, S. M. & BINN, L. N. (1983). Antigenic relatedness of two strains of hepatitis A virus determined by cross-neutralization. *Infection and Immunity* **42**, 418–420.

LINEMEYER, D. L., MENKE, J. G., MARTIN-GALLARADO, A., HUGHES, J. V., YOUNG, A. & MITRA, S. W. (1985). Molecular cloning and partial sequencing of hepatitis A viral cDNA. *Journal of Virology* **54**, 247–255.

MACGREGOR, A., KORNITSCHUK, M., HURRELL, J. G. R., LEHMANN, N. I., COULEPIS, A. G., LOCARNINI, S. A. & GUST, I. D. (1983). Monoclonal antibodies against hepatitis A virus. *Journal of Clinical Microbiology* **18**, 1237–1243.

MELNICK, J. L. (1982). Classification of hepatitis A virus as enterovirus type 72 and of hepatitis B virus as hepadnavirus type I. *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, Archives of Virology, Supplementum 2*, pp. 320–326. Edited by R. I. B. Francki, C. M. Fauquet, D. L. Knudson & F. Brown. Wien: 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*). *Journal of General Virology* **72**, 1685–1689.

NAJARIAN, R. D., CAPUT, D., GEE, W., POTTER, S. J., RENARD, A., MERRYWEATHER, J., NEST, G. V. & DINA, D. (1985). Primary structure and gene organization of human hepatitis A virus. *Proceedings of the National Academy of Sciences, U.S.A.* **82**, 2627–2631.

NEEFE, J. R., GELLIS, S. S. & STOKES, J. (1946). Homologous serum hepatitis and infectious (epidemic) hepatitis. *Journal of Medical Virology* **1**, 3–22.

NOTTAY, B. K., KEW, O. M., HATCH, M. H., HEYWARD, J. T. & OBIJESKI, J. F. (1981). Molecular variation of type 1 vaccine-related and wild polioviruses during replication in humans. *Virology* **108**, 405–423.

OVCHINNIKOV, I. A., SVERDLOV, E. D., TSAREV, S. A., ARSENIAN, S. G., ROKHLINA, T. O., CHIZHIKOV, V. E., PETROV, N. A., PRIKHOD'KO, G. G., BLINOV, V. M., BASILENKO, S. K, SANDAKHCHIEV, L. S., KUSOV, I. I., GRABKO, V. I., FLEER, G. P., BALAYAN, M. S. & DROZDOV, S. G. (1985). Sequence of 3372 nucleotide units of RNA of the hepatitis A virus, coding the capsids VP4-VP1 and some nonstructural proteins. *Doklady Akademii nauk SSSR* **285**, 1014–1018.

PALMENBERG, A. C. (1989). Sequence alignments of picornaviral capsid proteins. In *Molecular Aspects of Picornaviral Infection and Detection*, pp. 211–241. Edited by B. L. Semler & E. Ehrenfeld. Washington, D.C.: American Society for Microbiology.

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 Research* **8**, 153–171.

PING, L.-H., JANSEN, R. W., STAPLETON, J. T., COHEN, J. I. & LEMON, S. M. (1988). Identification of an immunodominant antigenic site involving the capsid protein VP3 of hepatitis A virus. *Proceedings of the National Academy of Sciences, U.S.A.* **85**, 8281–8285.

PROVOST, P. J., ITTENSOHN, O. L., VILLAREJOS, V. M., ARGUEDAS, J. A. & HILLEMAN, M. R. (1973). Etiologic relationship of marmoset-propagated CR326 hepatitis A to hepatitis in man. *Proceedings of the Society for Experimental Biology and Medicine* **142**, 1257–1267.

RAKELA, J., FAY, O. H., STEVENSON, K., GORDON, I. & MOSLEY, J. W. (1976). Similarities of two hepatitis A virus strains. *WHO Bulletin* **54**, 561–564.

RICO-HESSE, R., PALLANSCH, M. A., NOTTAY, B. K. & KEW, O. M. (1987). Geographic distribution of wild poliovirus type 1 genotypes. *Virology* **160**, 311–322.

ROBERTSON, B. H., KHANNA, B., NAINAN, O. V. & MARGOLIS, H. S. (1991*a*). Epidemiologic patterns of wild-type hepatitis A virus determined by genetic variation. *Journal of Infectious Diseases* **163**, 286–292.

ROBERTSON, B. H., KHANNA, B., NAINAN, O. V. & MARGOLIS, H. S. (1991*b*). Genetic variation of wild-type hepatitis A isolates. In *Viral Hepatitis and Liver Disease*, pp. 54–58. Edited by B. N. Hollinger, S. M. Lemon & H. S. Margolis. Baltimore: Williams & Wilkins.

RUECKERT, R. R. & WIMMER, E. (1984). Systematic nomenclature of picornavirus proteins. *Journal of Virology* **50**, 957–959.

TICEHURST, J., COHEN, J. I., FEINSTONE, S. M., PURCELL, R. H., JANSEN, R. W. & LEMON, S. M. (1989). Replication of hepatitis A virus: new ideas from studies with cloned cDNA In *Molecular Aspects of Picornaviral Infection and Detection*, pp. 27–50. Edited by B. L. Semler & E. Ehrenfeld. Washington, D. C.: American Society for Microbiology.

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. *Journal of General Virology* **72**, 1677–1683.

WIDELL, A., HANSSON, B. G., MOESTRUP, T. & NORDENFELDT, E. (1983). Increased occurrence of hepatitis A with cyclic outbreaks among drug addicts in a Swedish community. *Infection* **11**, 198–200.

WINSHIP, P. (1989). An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Research* **17**, 1266.

(*Received 18 December 1991; Accepted 21 February 1992*)