Journal of General Virology (2001), **82**, 2955–2963. Printed in Great Britain

Genetic analysis of hepatitis A virus strains recovered from the environment and from patients with acute hepatitis

Sonia Pina,[1] Maria Buti,[2] Rosend Jardí,[3] Pilar Clemente-Casares,[1] Joan Jofre[1] and Rosina Girones[1]

[1] Department of Microbiology, University of Barcelona, Avd. Diagonal 645, 08028 Barcelona, Spain

[2, 3] Liver Unit[2] and Department of Biochemistry[3], Hospital General Universitario Valle Hebron, 08035 Barcelona, Spain

The molecular epidemiology of hepatitis A virus (HAV) was studied by analysing HAV strains recovered from environmental water samples over a 7 year period and strains recovered from patients with acute hepatitis over a 5 year period. A total of 54 samples of raw domestic sewage and 66 samples of river water were collected. HAV particles were concentrated and detected by nested RT–PCR. HAV infection in patients with acute hepatitis was serologically diagnosed in 26 of 74 serum samples, which were also analysed by nested RT–PCR. HAV RNA was detected in 57·4% of sewage samples, 39·2% of Llobregat river water samples, 20% of Ter river water samples and 61·6% of serum samples. The HAV genomes detected were characterized further by directly sequencing a region of the 5′ non-translated region, the VP1/2A junction region and, in some samples, the 2B region. Results showed a 95% prevalence of genotype I, with nearly 50% being either subgenotype IA or subgenotype IB. Various strains were found simultaneously in both environmental and clinical samples. These strains were closely related to those described in distant geographical areas. Genotype IIIA was also found in 5% of sewage samples and in 12·5% of serum samples. Strains belonging to a common endemic genotype were not identified. The abundance of HAV in the environment produces a situation of sanitary risk, especially considering the low prevalence of antibodies in the young population.

Introduction

Hepatitis A virus (HAV) has been classified as the only member of the genus *Hepatovirus* within the family *Picornaviridae*. As in other picornaviruses, the large open reading frame present in the HAV genome 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 VP4, and the P2 and P3 regions encode the non-structural polypeptides necessary for virus replication. The lack of antigenic diversity in HAV strains is related to a high degree of conservation in the amino acid sequences of the capsid proteins (Ticehurst *et al.*, 1989).

HAV is an enterically transmitted picornavirus that causes acute hepatitis in humans and some primates. Transmission of the virus by the faecal–oral route results in the widespread endemic asymptomatic infection of infants and children in developing countries. Exposure of non-immune adolescent and adult populations within industrialized regions of the world results in clinical disease. HAV has a single antigenic serotype and a single infection is considered to confer lifelong immunity (Gellis *et al.*, 1945; Lemon & Binn, 1983; Neefe *et al.*, 1964; Rakela *et al.*, 1976). The limited antigenic variability of HAV has precluded the use of serological approaches to differentiate precisely the sources of HAV infection in various epidemiologically defined settings of endemic and epidemic disease. Comparison of nucleotide sequences within a limited region of the genome is an alternative approach to define the genetic relatedness of different strains. Epidemiologically significant genomic analysis requires the study of the wild-type virus to circumvent any genetic changes associated with cell culture or animal adaptation. The use of nested PCR amplification and partial sequencing described here to study HAV strains detected in raw sewage, in other environmental samples and in clinical samples provides a framework for improving molecular epidemiology studies of HAV infection in a population.

Comparative studies of the nucleotide sequences of different HAV strains have suggested that sequence relatedness can be correlated with the geographical origin of the virus (Jansen *et al.*, 1990; Robertson *et al.*, 1991). Analysis

Author for correspondence: Rosina Girones.
Fax +34 93 411 05 92. e-mail rosina@bio.ub.es

of sequence data derived from the VP1/2A junction region allowed the description of seven distinct genotypes of HAV (Robertson *et al.*, 1992).

Here, we depict HAV strains, identified in Barcelona (Spain), that have been recovered from both environmental water samples analysed over a 7 year period and clinical cases of acute hepatitis over a 5 year period. The study of these HAV strains will be useful for the design of improved strategies for the control of HAV infections.

Methods

■ **Water samples.** A total of 54 raw domestic sewage samples was collected from the sewage network of Barcelona (Spain) from 1994 to 2000 (100% positive for faecal coliforms, with a mean concentration of $1{\cdot}7 \times 10^6$ c.f.u./100 ml). Samples were taken at the entry of a treatment plant receiving sewage from a population of about 1·8 million people. Each sample was harvested in a sterile 500 ml polyethylene container and kept at 4 °C for less than 8 h until the virus particles were concentrated. The method applied for the recovery of virus particles from sewage samples was chosen on the basis of previous studies (Puig *et al.*, 1994; Pina *et al.*, 1998). Briefly, samples of 40 ml were ultracentrifuged at 229600 ***g*** for 1 h at 4 °C to sediment the virus particles together with any suspended material. Viruses retained in the sediment were eluted by mixing with 4 ml 0·25 M glycine buffer (pH 9·5) on ice for 30 min. The suspended solids were then removed by centrifugation at 12000 ***g*** for 15 min after the addition of 4 ml 2 × PBS. Viruses in the supernatant were pelleted by ultracentrifugation at 229600 ***g*** for 1 h at 4 °C, resuspended in 0·1 ml PBS and stored at −80 °C.

Water samples ($n = 56$) from the Llobregat (50–250 l depending on the turbidity, 10^3–10^4 c.f.u./100 ml faecal coliforms) and Ter ($n = 10$) (250 l, < 10 c.f.u./100 ml faecal coliforms) rivers were collected from 1994 to 1998. The Llobregat and Ter rivers are located south and north of Barcelona, respectively. Both rivers receive large quantities of urban and industrial raw waste water, which contribute a high level of chemical and faecal contamination. Samples were concentrated primarily *in situ* using positively charged filters (Zeta-plus, CUNO) (Sobsey & Jones, 1979; Sobsey & Glass, 1980). Filter cartridges were transported to the laboratory and immersed in tryptone soy broth. The viruses retained were eluted with 1000 ml of 0·25 M glycine buffer (pH 9·5–10·5), concentrated by organic flocculation with 3% beef extract (pH 3·5–4·0) (Becton Dickinson) and recovered in 50 ml PBS (APHA, 1995). Finally, virus particles were concentrated by ultracentrifugation, according to the method described for the sewage samples.

■ **Human serum samples.** Serum samples ($n = 74$) from patients with acute hepatitis were randomly selected. Acute viral hepatitis was defined by the presence of specific clinical symptoms, which are as follows: serum alanine aminotransferase levels 10-fold higher than the normal upper values and exclusion of other causes of hepatocellular injury, including medications, alcohol, congestive heart failure, biliary disease, autoimmune hepatitis and other infections (e.g. cytomegalovirus and Epstein–Barr virus diseases). All serum samples were collected between October 1989 and December 2000 at the Hospital General Valle Hebron (located in Barcelona, Spain) from patients with symptoms of acute hepatitis. Samples were collected during the first month after the presentation of symptoms. Commercially available ELISA kits (Abbott) were used for the detection of IgM anti-HAV, HBV surface antigen, IgM anti-HBV, IgG anti-HEV and anti-HCV. HCV RNA was analysed by PCR (Amplicor HCV Monitor Test, version 2.0; Roche). Acute hepatitis A was diagnosed by the presence of IgM anti-HAV antibodies. Serum samples were stored at −80 °C until use for detecting genomic sequences.

■ **Viruses.** HAV strain pHM-175 was propagated in FRhK-4 cells growing in Eagle's minimal essential medium (MEM) (Auto-pow; ICN) containing 1·5% sodium bicarbonate, 2 mM L-glutamine, non-essential amino acids, 100 U/ml penicillin, 100 μg/l streptomycin and 15% foetal bovine serum. Virus suspensions were stored at −80 °C until use.

■ **Nucleic acid extraction and enzymatic amplification.** Viral nucleic acid from serum samples or virus stocks was extracted using guanidinium isothiocyanate, adsorbed onto silica particles and eluted in elution buffer containing 10 mM Tris–HCl (pH 8·0), 0·1 mM EDTA, 1 mM DTT and 0·24 U/μl ribonuclease inhibitor (Boom *et al.*, 1990). RNA solutions were diluted 1:10 in elution buffer.

For the detection of RNA genomes, 5 μl aliquots of the extracted nucleic acid (corresponding to 0·5 and 5 μl serum, 0·2 and 2 ml sewage or 1 and 5 l river water), 1·5 mM $MgCl_2$ and 1 × PCR buffer II (Perkin-Elmer), containing 10 mM Tris–HCl (pH 8·3 at 25 °C), 50 mM KCl, 0·01 M DTT, 10 nmol of each dNTP and 25 pmol of the antisense primer (Table 1) in a total volume of 10 μl were used for reverse transcription. The reaction mixture was incubated at 95 °C for 5 min before the addition of 10 units of ribonuclease inhibitor (Applied Biosystems) and 50 units of reverse transcriptase from Moloney murine leukaemia virus (Applied Biosystems). After 30 min at 42 °C, the reaction was heated for 5 min at 95 °C.

For a typical one-step PCR, a 10 μl sample of cDNA was used. Amplification was carried out in a 50 μl reaction mixture containing 10 mM Tris–HCl (pH 8·3 at 25 °C), 50 mM KCl, 1·5 mM $MgCl_2$, 200 μM of each dNTP, 25 pmol of forward primer (Table 1) and 2 units of Ampli*Taq* DNA polymerase (Applied Biosystems). The amplification mixture was incubated in a GeneAmp 2400 PCR thermocycler (Applied Biosystems). The first cycle of denaturation was carried out for 3 min at 95 °C, followed by 30 cycles of denaturation at 95 °C for 60 s, annealing at 55 °C (for non-degenerate 5′ NTR primers), 42 °C (for degenerate VP1/2A primers) or 48 °C (for the primers of the 2B-region and the degenerate 5′ NTR primers) for 60 s and extension at 72 °C for 60 s. The final extension step was done at 72 °C for 5 min. Then, 1 μl (1/50) of the reaction mixture was added to a new batch of 50 μl PCR mixture, containing 25 pmol of each nested primer (Table 1) for a new PCR amplification cycle following the same procedure. PCR products were analysed by agarose gel electrophoresis using ethidium bromide as a stain.

To reduce the probability of sample contamination by amplified DNA molecules, standard precautions were applied in all manipulations. Separate areas were used for reagents, treatment of samples and manipulation of amplified samples. All samples were analysed twice in independent experiments and a negative control was added for each sample.

■ **Sensitivity of detection.** Sensitivity was evaluated by a limiting dilution experiment, as described previously (Puig *et al.*, 1994; Pina *et al.*, 2000). Serial 10-fold dilutions of HAV strain HM-175 RNA were tested by nested RT–PCR. Sewage samples were spiked with 2×10^3 genome equivalents per ml (GE/ml). One GE, as determined by RT–PCR, is defined as the number of genomes present in the highest serial dilution shown to be positive by RT–PCR. Virus particles were recovered by the method described above. Serial 10-fold dilutions of the RNA solution were reverse transcribed, amplified by the two-step PCR procedure described above and the limits of detection were then compared.

■ **Sequencing and analysis of viral genomes.** Nested PCR products were purified using the QIAquick PCR Purification kit (Qiagen).

Table 1. Oligonucleotide primers used for PCR amplification of HAV

The sequence position refers to the HM-175 strain of HAV (Cohen *et al.*, 1987*b*). For PCR, the melting temperature (T_m) was calculated according to the equation $T_m = 4 \times$ (number of pairs GC) $+ 2 \times$ (number of pairs AT).

Region	Position	Name	Polarity	T_m (°C)	Product size (bp)	Sequence (5′ → 3′)*
5′ NTR	332–352	HAV1	Sense	64	368	TTGGAACGTCACCTTGCAGTG
	680–700	HAV2	Antisense	64		CTGAGTACCTCAGAGGCAAAC
	371–391	neHAV1	Sense	58	290	ATCTCTTTGATCTTCCACAAG
	641–661	neHAV2	Antisense	64		GAACAGTCCAGCTGTCAATGG
	332–351	HAV1d	Sense	62	376	TTGGRACGTCDCCTTGCAGT
	689–708	HAV2d	Antisense	65		AAATGCCCCTGRGTACCTCAG
	644–661	neHAV2d	Antisense	57	290	GWAMWGTCCAGCWDYHAATGG
VP1/2A	2906–2929	HHA1	Sense	59	532	TGCAAATTAYAAYCAYTCTGATGA
	3416–3438	HHA2	Antisense	61		TTTCTGTCCATTTYTCATCATTC
	2940–2960	HHA3	Sense	48	436	TTYAGTTGYTAYTTGTCTGT
	3357–3376	HHA4	Antisense	55		TCAAGAGTCCACACACTTC
2B	3627–3646	2B1	Sense	58	343	GCAGATAGAATGCTTGGATT
	3951–3970	2B2	Antisense	54		GAGTCCATYTGCCAATTAAA
	3750–3769	2B3	Sense	59	220	TTTCAYCATTCTGTGACTG

* R = A or G; D = G, A or T; W = A or T; Y = C or T; M = A or C; H = A, T or C.

Table 2. HAV strains used for phylogenetic analysis

Strain	Genotype	GenBank accession no.	Reference
HM-175	IB	M14707	Cohen *et al.* (1987*b*)
HM-175/24a	IB	M59810	Lemon *et al.* (1991)
HM-175/43c	IB	M59809	Lemon *et al.* (1991)
HM-175/18f	IB	M59808	Lemon *et al.* (1991)
HM-175/7MK5	IB	M16632	Cohen *et al.* (1987*a*)
HAF-203	IB	AF268396	Unpublished
MBB	IB	M20273	Paul *et al.* (1987)
L-A-1	IB	AF314208	Unpublished
LA	IA	K02990	Najarian *et al.* (1985)
HAS-15	IA	X15464	Sverdlov *et al.* (1987)
GBM	IA	X75215	Graff *et al.* (1994)
CR326	IA	M10033	Linemeyer *et al.* (1985)
FG	IA	X83302	Beneduce *et al.* (1995)
FH1	IA	AB020567	Unpublished
FH2	IA	AB020568	Unpublished
FH3	IA	AB020569	Unpublished
AH1	IA	AB020564	Unpublished
AH2	IA	AB020565	Unpublished
AH3	IA	AB020566	Unpublished
CF53	II	M63025	Brown *et al.* (1991)
PA21	IIIA	M63026	Brown *et al.* (1991)
Nor-24	IIIA	AJ299467	Stene-Johansen *et al.* (1999)
Nor-20	IIIA	AJ299463	Stene-Johansen *et al.* (1999)
CY145	IV	M59286	Nainan *et al.* (1991)
AGM-27	V	D00924	Tsarev *et al.* (1991)

Table 3. Sequence comparison of the conserved 5′ NTR region (248 nt) of HAV strains detected in environmental samples from 1994 to 2000

	Comparison with the control strain		Comparison with the identified strain		
Sample*	No. of identical nucleotides	Identity (%)	No. of identical nucleotides	Identity (%)	Strain
Ll-05/09/94	246	99·1	248	100	HM-175, HAF-203
Ll-15/11/94	246	99·1	248	100	HM-175, HAF-203
SA-30/06/95	236	95·1	248	100	GBM, LA
SA-28/07/95	235	94·7	247	99·6	GBM
Ll-04/09/95	246	99·1	248	100	HM-175, HAF-203
Ll-18/12/95	246	99·1	248	100	HM-175, HAF-203
T-17/01/96	246	99·1	248	100	HM-175, HAF-203
Ll-24/02/97	244	98·3	246	99·1	MBB
Ll-21/05/97	235	94·7	246	99·1	FH3
T-18/06/97	238	95·9	240	96·7	MBB
Ll-06/08/97	246	99·1	248	100	HM-175, HAF-203
SA-07/10/97	238	95·9	245	98·7	FG
SA-06/11/97	238	95·9	245	98·7	FG
Ll-31/12/97	246	99·1	248	100	HM-175, HAF-203
SA-09/01/98	245	98·7	247	99·6	HM-175, HAF-203
SA-16/01/98	237	95·5	244	98·3	FG
SA-23/01/98	238	95·9	245	98·7	FG
SA-02/02/98	235	94·7	243	97·9	GBM, LA
SA-11/02/98	237	95·5	244	98·3	FG
Ll-25/02/98	243	97·9	247	99·5	MBB
SA-22/03/99	244	98·3	246	99·1	MBB
SA-26/03/99	241	97·1	243	97·9	GBM, LA†
SA-09/06/00	242	97·6	246	99·1	MBB
SA-23/06/00	242	97·6	238	95·9	GBM, LA†
SA-11/07/00	235	94·7	235	94·7	L-A-1†
SA-03/08/00	235	94·7	241	97·1	GBM, LA
SA-16/08/00	244	98·3	238	95·9	GBM, LA†
SA-01/09/00	236	95·1	242	97·5	GBM, LA
SA-05/10/00	234	94·4	246	99·1	GBM, LA
SA-26/10/00	244	98·3	228	91·9	PA21†
SA-06/11/00	244	98·3	246	99·1	HM-175, MBB
SA-04/12/00	227	91·5	243	98·3	PA21
SA-29/12/00	236	95·1	248	100	GBM, LA

* SA-dd/mm/yy, sewage sample; Ll-dd/mm/yy, Llobregat river water sample, T-dd/mm/yy, Ter river water sample.
† Data correspond to the strain identified according to hypervariable region analysis.

Both strands of purified DNA were sequenced with the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit with Ampli*Taq* DNA polymerase (Applied Biosystems), according to the manufacturer's instructions. Results were checked using the ABI PRISM 3700 automated sequencer (Applied Biosystems). Sequences were then compared with the HAV sequences present in GenBank and EMBL using the BLAST N program, version 2.0.8, and aligned with the corresponding sequence fragments of the database using the CLUSTAL W program, version 1.8 (Thompson *et al*., 1994). Shading was carried out with the GeneDoc program, version 2.5.000.

The transition/transversion parameter was estimated from data sets using the PUZZLE program, version 4.0.2. Phylogenetic distances were determined by the maximum-likelihood method using the DNADIST program of the PHYLIP package, version 3.5c (Felsenstein, 1993). Phylograms were generated by the UPGMA algorithm using the NEIGHBOR program. The robustness of the grouping was determined by bootstrap resampling of the multiple sequence alignments (100 sets) with the programs SEQBOOT, DNADIST, NEIGHBOR and CONSENSE. Bootstrap values greater than 70% provide statistical evidence of phylogenetic association (Felsenstein, 1985). The output graphics of the trees were created with the TREEVIEW package, version 1.5 (Page, 1996).

■ **Nucleotide sequence accession numbers.** GenBank accession numbers of the HAV strain sequences characterized previously and used for phylogenetic studies are shown in Table 2. The sequences reported in this paper have been deposited into GenBank under the following accession numbers: AF386798–AF386845 for the 5′ NTR sequences, AF386846–AF386889 for the VP1/2A junction sequences and AF386890–AF386911 for the 2B region sequences.

Results

Sensitivity

RNA extracted from the HAV strain HM-175 stock was amplified using primers specific for the 5′ NTR and VP1/2A region. Visible bands (obtained after agarose gel electrophoresis) were evident up to a 10^{-6} dilution after nested PCR, which corresponds to 1 GE. The estimated sensitivity for the detection of viral RNA in sewage samples spiked with HAV strain HM-175 was shown to be 10 GE, which is equivalent to the sensitivity of 1–10 virus particles, described previously when sewage samples spiked with enterovirus were analysed (Puig *et al.*, 1994; Pina *et al.*, 1998).

The sensitivity estimated for the detection of viral RNA in serum was shown in previous studies to be of the same order as the sensitivity observed for the virus stock when nested PCR amplification was applied (Pina *et al.*, 2000); this demonstrated the absence of inhibitors of the amplification reaction.

Presence of HAV in environmental samples

HAV RNA was detected in 31 of 54 raw sewage samples tested (57·4%) after nested RT–PCR using primers specific for the conserved 5′ NTR. RNA concentration was estimated by using a limiting dilution experiment and was found to range from 10 to 10^2 GE/ml. Samples with concentrations of RNA $<$ 10 GE/ml were deemed to be negative.

Of the 56 Llobregat river water samples, 22 (39·2%) showed detectable levels of HAV, estimated to be between 10 and 10^2 GE/l. Viral RNA was also detected in 2 of 10 Ter river water samples tested (20%) and the concentration was estimated to be 10 GE/l.

HAV in clinical samples

The diagnosis of acute HAV infection was confirmed by the presence of IgM anti-HAV in 26 of 74 serum samples taken from patients with acute hepatitis. Of the 26 patients, 16 (61·6%) presented detectable levels of HAV RNA in the serum. These patients comprised 18 males and 8 females, with a mean age of 42 years (ranging from 24 to 79 years of age). Three patients were intravenous drug addicts and one had travelled to a developing country. Serum samples were obtained during the first week of symptoms in 10 cases, during the second week in 10 cases and between the third and the fourth week in six cases. All HAV RNA-positive samples were from patients whose serum was obtained between the first and the second week of presenting symptoms.

Sequence analysis of HAV RNA genomes

In order to characterize the HAV strains detected in environmental and clinical samples, nested PCR products from the 5′ NTR and the VP1/2A junction region were sequenced. Sequence analysis of 248 nt of the 290 bp nested PCR amplicon from the 5′ NTR of 33 environmental strains and 15 clinical strains showed a high degree of identity, ranging from 90 to 100% among samples isolated between 1990 and 2000. Comparison of the nucleotide sequences with that of the control strain, HM-175, showed 91·5–99·1% identity with the environmental strains and 90·7–98·4% identity with the clinical strains (Tables 3 and 4). Comparative analysis with

Table 4. Sequence comparison of the conserved 5′ NTR region (248 nt) of HAV strains detected in serum samples from 1990 to 2000

	Comparison with the control strain		Comparison with the identified strain		
Sample	No. of identical nucleotides	Identity (%)	No. of identical nucleotides	Identity (%)	Strain
HS-05/03/90	238	96·0	244	98·4	GBM, LA
HS-10/11/90	234	94·4	247	99·5	MBB
HS-17/01/91	237	95·6	243	97·9	GBM, LA
HS-27/02/91	243	98·0	247	99·5	MBB
HS-26/08/91	242	97·6	246	99·1	MBB
HS-17/09/91	236	95·2	248	100	GBM, LA
HS-25/09/91	241	97·2	245	98·7	MBB
HS-20/05/92	234	94·4	246	99·1	GBM, LA
HS-08/09/92	232	93·6	242	97·5	GBM, LA
HS-07/10/92	243	98·0	247	99·5	MBB
HS-06/04/99	244	98·4	246	99·1	HM-175, MBB
HS-27/04/00	233	93·9	245	98·7	GBM, LA
HS-16/10/00	234	94·4	246	99·1	GBM, LA
HS-13/11/00	225	90·7	242	98·3	PA21
HS-21/11/00	243	97·9	247	99·5	MBB

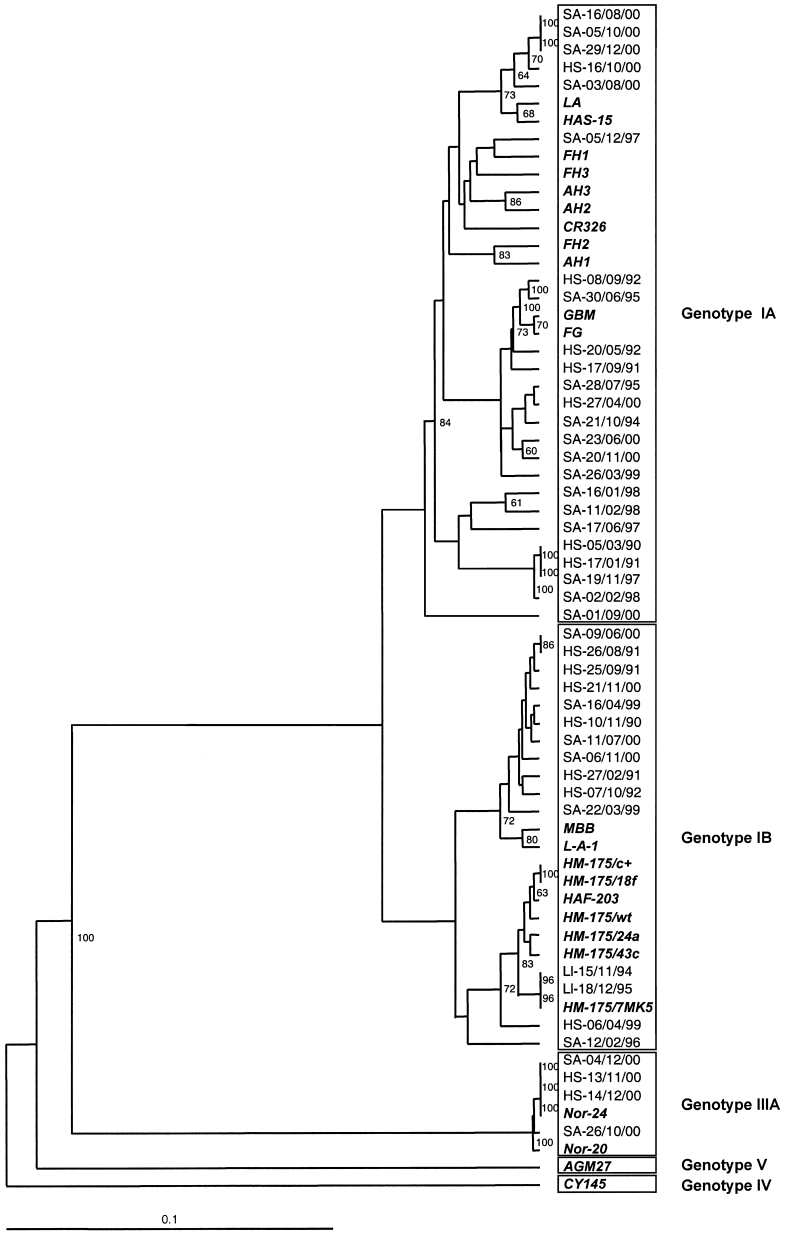

Fig. 1. Phylogram depicting the relationship between the environmental and clinical HAV strains in relation to other isolates characterized previously over a 296 bp fragment of the VP1/2A junction region. Scale represents genetic distance. The internal node numbers represent bootstrap values (100 replicates) expressed as the percentage of all trees. Values greater than 70% indicate statistically significant grouping (Felsenstein, 1985).

sequences from genetic databases indicated the presence of HAV variants that were related closely to genotypes IB (HM-175, MBB and L-A-1) and IA (GBM, LA, FG and FH), showing 94·7–100% identity with the environmental samples and 97·5–100% identity with the serum samples. Strains that cluster with strain PA21 (genotype IIIA) were identified in two sewage samples and in one serum sample, with 97–98% identity (Tables 3 and 4). Two A residues at positions 551 and 591 in the 5′ NTR fragment are characteristic of the wild-type HM-175 strain. These nucleotides change to a G in cell culture-adapted strains after long-term passage (Lemon *et al.*, 1991). All of the environmental and clinical strains detected contain an A at positions 551 and 591, as described for wild-type viruses (Lemon *et al.*, 1991).

The VP1/2A junction region is recognized as one of the most variable regions of the HAV genome (Robertson *et al.*, 1992) and was therefore chosen for genotyping and phylogenetic analysis. The sequence of 398 nt of the 436 bp nested PCR product from the VP1/2A region was determined for 27 strains detected in environmental water samples and in 16

clinical samples. Sequence comparison among the environmental and clinical samples showed 76–100% identity and 76·8–99·2% identity with the control strain, respectively. Comparative analysis with other isolates described previously confirms the presence of HAV strains belonging to genotypes IA and IB, with 93·7–100% identity. Two of the sewage samples and two serum samples collected from October to December 2000 (SA-26/10/00, HS-13/11/00, SA-04/12/00 and HS-14/12/00) showed the presence of HAV strains related closely to strains PA21 and Nor-24 (genotype IIIA), with more than 99% identity. The phylogenetic tree based on the alignment of the VP1/2A fragment is shown in Fig. 1.

Point mutations were observed in the nucleotide sequence of the VP1/2A region of the environmental and clinical strains, which differ from the control strain. The C at position 3018 of the control strain is changed to a T in all of the environmental and clinical strains. Moreover, all clinical isolates showed additional substitutions: the C at position 3002 in the HM-175 control strain is changed to a T in all strains detected in serum samples, the T at position 3248 is changed to an A, the G at position 3281 is changed to an A or a C, the G at position 3287 is changed to an A and the A at position 3353 is changed to a T. These mutations demonstrate the presence of wild-type strains in environmental water and clinical samples. Within the 2A region, the mutations described at positions 3248 and 3281 result in the substitution of the lysine and isoleucine residues at positions 47 and 58 to asparagine and methionine residues, respectively: these mutations are related to adaptation to cell culture growth (Lemon *et al*., 1991).

The results showed a similar prevalence of strains identified as genotypes IA and IB (55 and 40%, respectively) in the environmental samples and 43·7% of each subgenotype in serum samples. The strains related closely to genotype IIIA were identified in two environmental samples (5%) and in two clinical samples (12·5%).

We sequenced 182 nt of a 220 bp fragment within the 2B region in some of the environmental and clinical HAV strains. A mutation at position 3889 (C changed to a U), which has been described in cell culture-adapted strains, is responsible for an alanine to a valine substitution at amino acid position 72 of the 2B protein (Cohen *et al*., 1987*a*; Emerson *et al*., 1993). All strain sequences showed a C in that position (data not shown), as described for wild-type strains, which eliminates the possibility of cross-contamination with laboratory strains.

Discussion

The study presented here provides a framework for global analysis of the HAV strains that circulate among the population in Barcelona and includes a description of the most prevalent HAV strains present in the environment and also those strains that resulted in cases of acute viral hepatitis. Results based on virus concentration from raw sewage or other environmental water samples, nested RT–PCR amplification and partial sequencing yields reliable information on the molecular epidemiology of HAV infection in an area located in the north-eastern region of Spain.

Sequence analysis of the environmental and clinical samples showed a high prevalence of genotype I strains. Genotype I comprise 80% of HAV strains isolated worldwide, as described by Robertson *et al*. (1992). Genomic analysis shows that more than one HAV genotype could be identified within the same year in the same location; over a period of a few years, the most prevalent strains may differ. Genotypes IA and IB were found in the clinical samples collected from 1990 to 1992 at a proportion of 50% each.

Genotype IB strains, related closely to strains HM-175, MBB and L-A-1, were identified in the environment throughout the period from 1994 to 2000. Genotype IA strains were also abundant and identified as GBM variants in 52·5% of the isolates. LA- and HAS-15-related strains were the most prevalent during the year 2000 (in 5 of up to 12 isolates), together with MBB- and L-A-1-like strains (in 3 of up to 12 samples). Genotype IIIA variants were also detected in two samples.

A good concordance was observed between the strains detected in the environment and in clinical serum samples collected from patients with acute hepatitis during the year 2000, with genotypes IA, IB and IIIA circulating in the environment and among hepatitis patients. However, 100% nucleotide identity in closely related strains detected in the clinical and environmental strains was never observed. This fact indicates that a high number of strains is currently circulating in the population, since it is accepted that, during infection, no sequence differences are usually detected, even between different patients involved in the same outbreak (Robertson *et al*., 1992).

Genotype III strains were detected in the environment and in hepatitis patients during the same period of time. This genotype has been demonstrated to be highly prevalent among intravenous drug users in Northern Europe (Robertson *et al*., 1992; Stene-Johansen *et al*., 1999). This is in agreement with the fact that one of the patients that presented a virus strain similar to strain Nor-24 belonged to that group, although in the second patient, no specific risk factor was identified.

HAV infection is considered to be endemic in the Mediterranean region and, as has been described previously in other countries, is characterized in Spain by diminishing HAV seroprevalence in the population (Papaevangelou, 1992; Dal-Ré *et al*., 2000; Lopalco *et al*., 2001). Within the specific area studied (Catalonia, Spain), a seroprevalence of 67·8% in the general population has been documented; less than 5% of a group of patients who were between 5 and 14 years of age were seropositive (Bruguera *et al*., 1999). Therefore, a large percentage of the Spanish population under 30 years of age is not protected against hepatitis. The results presented here show the circulation of several HAV strains identical to, or closely related to, strains isolated previously in highly

divergent geographical areas worldwide. In this study, 79 HAV-positive samples collected over a period of 10 years were analysed. It was not possible to identify strains belonging to a common endemic genotype, as proposed by Robertson *et al.* (1992), suggesting the circulation of endemic strains in the USA between 1976 and 1990. Two hypotheses may explain this fact: (i) the higher frequency of travelling in the area and (ii) the higher prevalence of the infection in an endemic area, together with the fact that HAV has a high level of stability in the environment (Sobsey *et al.*, 1988). During the year 2000, although the number of cases of hepatitis A has been estimated to be less than 15–30 per 100000 habitants, 80% of urban sewage samples studied showed the presence of HAV. Since faecal shedding of HAV can, in some cases, last for months after the resolution of symptoms (Yotsuyanagi *et al.*, 1996), such patients could be a source of further virus spreading within the community. It is important that this factor is considered as a possible reason for the abundance of HAV observed in the environment. Moreover, low levels of HAV multiplication in the intestinal epithelial cells has been suggested (Blank *et al.*, 2000). This could represent a mechanism of virus amplification. Further monitoring will be needed to evaluate if there is a reduction in the presence of HAV in the environment over the next few years. Due to the low prevalence of antibodies in the young and the simultaneous abundance of HAV circulating in the environment, it may be important to consider the availability of a HAV vaccine in areas where sanitary is poor.

This work was partially supported by the research grants SGR/00023 from the Catalan Government and FIS 00/1123 from the 'Ministerio de Sanidad y Consumo' from the Spanish Government. Sonia Pina was in receipt of a fellowship from the Catalan Government. The authors thank the staff of the Servicios Científico-Técnicos of the University of Barcelona for their help in the sequencing of the PCR products. We also thank Drs S. U. Emerson and R. H. Purcell of the National Institutes of Health (NIAID, NIH, Bethesda, MD, USA) for their useful advice.

References

APHA (1995). Standard methods for the examination of water and wastewater, 19th edn. *American Public Health Association, Washington DC, WA, USA.*

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

Blank, C. A., Anderson, D. A., Beard, M. & Lemon, S. M. (2000). Infection of polarized cultures of human intestinal epithelial cells with hepatitis A virus: vectorial release of progeny virions through apical cellular membranes. *Journal of Virology* **74**, 6476–6484.

Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. L., Wertheim-van Dillen, P. M. E. & van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology* **28**, 495–503.

Brown, E. A., Day, S. P., Jansen, R. W. & Lemon, S. M. (1991). The 5′ nontranslated region of hepatitis A virus RNA: secondary structure and elements required for translation *in vitro*. *Journal of Virology* **65**, 5828–5838.

Bruguera, M., Salleras, L., Plans, P., Vidal, J., Navas, E., Domínguez, A., Batalla, J., Taberner, J. L. & Espuñes, J. (1999). Changes in seroepidemiology of hepatitis A virus infection in Catalonia in the period 1989–1996. Implications for a new vaccination strategy. *Medicina Clínica (Barcelona, Spain)* **112**, 406–408.

Cohen, J. I., Rosenblum, B., Ticehurst, J. R., Daemer, R. J., Feinstone, S. M. & Purcell, R. H. (1987 *a*). Complete nucleotide sequence of an attenuated hepatitis A virus: comparison with wild-type virus. *Proceedings of the National Academy of Sciences, USA* **84**, 2497–2501.

Cohen, J. I., Ticehurst, R., Purcell, R. H., Bucker-White, A. & Baroudy, B. M. (1987 *b*). 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.

Dal-Ré, R., García-Corbeira, P. & García-de-Lomas, J. (2000). A large percentage of the Spanish population under 30 years of age is not protected against hepatitis A. *Journal of Medical Virology* **60**, 363–366.

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* **184**, 475–480.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

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

Gellis, S. S., Stokes, J., Brother, G. M., Hall, W. M., Gilmore, H. R., Beyer, E. & Morrisey, 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.

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

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, USA* **87**, 2867–2871.

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.

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. *Journal of Virology* **65**, 2056–2065.

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. *Journal of Virology* **54**, 247–255.

Lopalco, P. L., Salleras, L., Barbuti, S., Germinario, C., Bruguera, M., Buti, M. & Domínguez, A. (2001). Hepatitis A and B in children and adolescents: what can we learn from Puglia (Italy) and Catalonia (Spain)? *Vaccine* **19**, 470–474.

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., 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. *Proceedings of the National Academy of Sciences, USA* **82**, 2627–2631.

Neefe, J. R., Gellis, S. S. & Stokes, J. (1964). Homologous serum hepatitis and infectious (epidemic) hepatitis. *Journal of Medical Virology* **1**, 3–22.

Page, R. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357–358.

Papaevangelou, G. (1992). Epidemiology of hepatitis A in Mediterranean countries. *Vaccine* **10** (Suppl. 1), S63–S66.

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.

Pina, S., Puig, M., Lucena, F., Jofre, J. & Girones, R. (1998). Viral pollution in the environment and in shellfish: human adenovirus detected by PCR as an index of human viruses. *Applied and Environmental Microbiology* **64**, 3376–3382.

Pina, S., Buti, M., Cotrina, M., Piella, J. & Girones, R. (2000). HEV identified in serum from humans with acute hepatitis and in sewage of animal origin in Spain. *Journal of Hepatology* **33**, 826–833.

Puig, M., Jofre, J., Lucena, F., Allard, A., Wadell, G. & Girones, R. (1994). Detection of adenoviruses and enteroviruses in polluted waters by nested PCR amplification. *Applied and Environmental Microbiology* **60**, 2963–2970.

Rakela, J., Fay, O. H., Stevenson, D., Gordon, I. & Mosley, J. W. (1976). Similarities of two hepatitis A virus strains. *Bulletin of the World Health Organization* **54**, 561–564.

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

Robertson, B. H., Jansen, R. W., Khanna, B., Totsuka, A., Nainan, O. V., Siegl, G., Widell, A., Margolis, H. S., Isomura, S., Ito, K., Ishizu, T., Moritsugu, Y. & Lemon, S. M. (1992). Genetic relatedness of hepatitis A virus strains recovered from different geographical regions. *Journal of General Virology* **73**, 1365–1377.

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

Sobsey, M. D. & Glass, J. S. (1980). Poliovirus concentration from tap water with electropositive adsorbent filters. *Applied and Environmental Microbiology* **40**, 201–210.

Sobsey, M. D. & Jones, B. L. (1979). Concentration of poliovirus from tap water using positively charged microporous filters. *Applied and Environmental Microbiology* **37**, 588–595.

Sobsey, M. D., Shields, P. A., Hauchman, F. S., Davis, A. L., Rullman, V. A. & Bosch, A. (1988). Survival and persistence of hepatitis A virus in environmental samples. In *Viral Hepatitis and Liver Diseases*, pp. 121–124. Edited by A. J. Zuckerman. New York: Alan R. Liss.

Stene-Johansen, K., Skaug, K. & Blystad, H. (1999). Surveillance of hepatitis A by molecular epidemiologic studies. *Tidsskrift for den Norske Laegeforening* **119**, 3725–3729.

Sverdlov, S. D., Tsarev, S. A., Markova, S. V., Vasilenko, S. K., Chizhikov, V. E., Petrov, N. A., Kusov, Y. Y., Natashenko, T. A. & Balayan, M. S. (1987). Cloning and expression of hepatitis A virus genome in *E. coli* cells. *Molekulyarnaya Genetika Mikrobiologiya i Virusologiya* **6**, 129–133.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

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 DC: 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.

Yotsuyanagi, H., Koike, K., Yasuda, K., Moriya, K., Shintani, Y., Fujie, H., Kurokawa, K. & Lino, S. (1996). Prolonged fecal excretion of hepatitis A virus in adult patients with hepatitis A as determined by polymerase chain reaction. *Hepatology* **24**, 10–13.

Received 11 June 2001; Accepted 31 August 2001
Published ahead of print (13 September 2001) in JGV Direct as DOI 10.1099/vir.0.17915-0