Epidemiology and Genetic Characterization of Hepatitis A Virus Genotype IIA[▽]

Delphine Desbois,[1,2,3] Elisabeth Couturier,[4] Vincent Mackiewicz,[1] Arielle Graube,[1] Marie-José Letort,[4] Elisabeth Dussaix,[1,2] and Anne-Marie Roque-Afonso[1,2,3]*

Centre National de Référence pour les Virus des Hépatites à Transmission Entérique, AP-HP, Hôpital Paul Brousse, Villejuif 94804,[1] Université Paris-Sud, UMR-S 602,[2] and INSERM, Unité 785,[3] Villejuif F-94800, and Institut de Veille Sanitaire, Département des Maladies Infectieuses, Saint-Maurice 94415,[4] France

Three hepatitis A virus (HAV) genotypes, I, II, and III, divided into subtypes A and B, infect humans. Genotype I is the most frequently reported, while genotype II is hardly ever isolated, and its genetic diversity is unknown. From 2002 to 2007, a French epidemiological survey of HAV identified 6 IIA isolates, mostly from patients who did not travel abroad. The possible African origin of IIA strains was investigated by screening the 2008 mandatory notification records of HAV infection: 171 HAV strains from travelers to West Africa and Morocco were identified. Genotyping was performed by sequencing of the VP1/2A junction in 68 available sera. Entire P1 and 5′ untranslated regions of IIA strains were compared to reference sequences of other genotypes. The screening retrieved 5 imported IIA isolates. An additional autochthonous case and 2 more African cases were identified in 2008 and 2009, respectively. A total of 14 IIA isolates (8 African and 6 autochthonous) were analyzed. IIA sequences presented lower nucleotide and amino acid variability than other genotypes. The highest variability was observed in the N-terminal region of VP1, while for other genotypes the highest variability was observed at the VP1/2A junction. Phylogenetic analysis identified 2 clusters, one gathering all African and two autochthonous cases and a second including only autochthonous isolates. In conclusion, most IIA strains isolated in France are imported by travelers returning from West Africa. However, the unexplained contamination mode of autochthonous cases suggests another, still to be discovered geographical origin or a French reservoir to be explored.

Hepatitis A virus (HAV) is a small, nonenveloped hepatotropic virus classified in the genus *Hepatovirus* within the family *Picornaviridae*. Its genome consists of a 7,500-nucleotide linear, positive-stranded RNA with a single open reading frame (ORF) flanked by 5′ and 3′ untranslated regions (UTR) (3). The 5′ UTR is the most conserved region of the genome and contains an internal ribosome entry site (IRES). The ORF encodes a polyprotein organized into three functional regions: P1, P2, and P3. P1 is secondarily cleaved into four capsid proteins, VP1 to VP4, whereas P2 and P3 encode nonstructural proteins.

Despite significant genetic variability, a single serotype has been described (13). Sequence variation of a 168-nucleotide fragment encompassing the VP1/2A junction has initially defined seven genotypes that differ by at least 15% and subtypes that differ by 7.0 to 7.5% (19). Six HAV genotypes (genotypes I to VI) are now defined based on analysis of the 900 nucleotides of the complete VP1 protein (6). Genotypes I, II, and III, divided into subtypes A and B, infect humans. The former genotype VII has been reclassified within the genotype II clade as subgenotype IIB (6, 14). Data on genotype distribution showed that genotype I was the most prevalent worldwide, with IA being reported more frequently than IB, and that subgenotype IIIA was prevalent in Central Asia (6, 16, 19). In areas of low endemicity, such as the United States and Western Europe, subgenotype IA dominates, but all genotypes and subtypes have been reported (6, 16, 23). Most HAV strains in these countries can be identified as imported strains by phylogenetic analysis thanks to growing sequence databases, which allow for the tracing of the geographic origin of a given subgenotype (16).

Among HAV genotypes, subgenotypes IIA (former genotype II) and IIB (former genotype VII) were defined based on a single isolate of each: CF-53/Berne, isolated in France in 1979, and SLF88, isolated in Sierra Leone in 1988 (19). Their complete sequences are now available (4, 14). These two subgenotypes are rarely reported. Recently, sequence analysis of the P1 region allowed the identification of a recombinant IB–IIA virus in a patient returning from Morocco (7), and a second IIA strain, identified by VP1–2A analysis, was reported in a Dutch patient returning from Spain (23).

From 2002 to 2007, six subgenotype IIA strains were isolated at the National Reference Centre for HAV in France (6/693 available strains [<1%]). Of these, only one was isolated from a patient reporting travel in an area of endemicity, in Benin, West Africa. Acute HAV infection has been a mandatory reportable disease in France since 2006. The mandatory notification records allowed us, first, to investigate the possible sub-Saharan origin of IIA strains by determining their prevalence among French travelers returning from Africa in 2008. Second, the genetic variability of all IIA isolates identified was characterized in order to gain insight into the molecular epidemiology of the IIA subgenotype and to elucidate the origin of autochthonous IIA cases.

* Corresponding author. Mailing address: Laboratoire de Virologie, Hôpital Paul Brousse, 12 Avenue Paul Vaillant-Couturier, 94804 Villejuif, France. Phone: (33) 1-45593956. Fax: (33) 1-45593724. E-mail: anne-marie.roque@pbr.aphp.fr.

TABLE 1. Primers used for amplification and sequencing of HAV isolates

Primer	Region	Position (nt)	Sequence (5′–3′)	Source or reference
IRES				
IRES3	5′ UTR	63	ATACCTCACCGCCGTTTGC	This study
IRES4	5′ UTR	785	GTCAAGRCCACTCCCAACAG	This study
P1				
First round of PCR				
+697	5′ UTR	697	TACGGGGCATTTAGGTTTTC	This study
−3444	2A	3444	CTTTCATTTCTGTCCATTTYTCATC	This study
Second round of PCR				
Set 1				
+697	5′ UTR	697	TACGGGGCATTTAGGTTTTC	This study
−1591	VP3	1591	TTTGATTCCTCCACCCTGAG	This study
Set 2				
+1458	VP2	1458	CTTTCTACACARATGATGAGAAATG	This study
−2268	VP1	2268	GGTCTGGAACATTCTGT	This study
Set 3				
+2112	VP3	2065	AAYGTTGCTTCYCATGTYAGRGT	22a
−3444	2A	3444	CTTTCATTTCTGTCCATTTYTCATC	This study
N-terminus of VP1				
−1865	VP1	2612	ACCADGCCATDCCATCHACATC	22a
+2112	VP3	2065	AAYGTTGCTTCYCATGTYAGRGT	22a
VP1/2A junction				
+2870	VP1	2870	GACAGATTCTACATTTGGATTGG	2a
−3381	2A	3381	Biotin-CCATTTCAAGAGTCCACACACT	2a

MATERIALS AND METHODS

Samples. The 2008 mandatory notification records were screened for travel to sub-Saharan Africa as risk factor for HAV. Serum samples were obtained from routine diagnostic laboratories and were frozen at −80°C until further use. Thirty samples with subgenotype IA (n, 10), IB (n = 10), or IIIA (n, 10) identified in 2008 were used as controls. They were selected from 202 sera of the 2008 collection of the CNR, based on the availability of alanine aminotransferase (ALT) levels and dates of symptom onset.

HAV RNA extraction. Viral RNA was extracted from 140 µl of serum by using the QIAmp viral RNA kit (Qiagen, les Ulis, France) according to the manufacturer's instructions and was eluted from the spin columns with 60 µl of RNase/DNase-free water.

HAV amplification. Table 1 identifies the primers that were used for the amplification of each region, and Fig. 1 shows their positions. For the IRES, the N terminus of VP1, and the VP1/2A region, 10 µl of extracted RNA was subjected to reverse transcription (RT) and PCR amplification in a final volume of 50 µl using the One-Step RT-PCR kit (Qiagen) according to the manufacturer's recommendations. The P1 segment, coding for the VP4, VP2, VP3, and VP1 capsid proteins, was amplified using the SuperScript III One-Step RT-PCR system with Platinum *Taq* DNA polymerase High Fidelity (Invitrogen, Cergy-Pontoise, France). The reaction mixture was prepared, according to the manufacturer's instructions, in a total volume of 50 µl containing 10 µl of extracted

RNA, 200 nM forward primer, and 200 nM reverse primer (Table 1). Second-round reactions were carried out with appropriate primers (Table 1) by using Platinum *Taq* DNA polymerase High Fidelity (Invitrogen) according to the manufacturer's instructions in a 50-µl volume that contained 5 µl first-round PCR products.

HAV RNA quantification. HAV RNA was quantified using the SuperScript III Platinum One-Step quantitative RT-PCR system (Invitrogen) according to the manufacturer's protocol. The set of primers and the probe, targeting the 5′ UTR as previously described, were used at 800 nM and 200 nM (final concentrations), respectively (8). Twenty microliters of the reaction mixture was added to PCR tubes containing 5 µl RNA from serum samples or the standard. HAV RNA was reverse transcribed into cDNA (30 min at 50°C), and the 87-bp fragment was amplified by PCR (15 s at 95°C and 1 min at 60°C) for 50 cycles on an Applied Biosystems (Foster City, CA) real-time PCR system, model 7500. A range of 10-fold dilutions of the cell culture-adapted HAV strain HM-175 at 10^6 PFU ml^{-1} was used to construct the standard quantification curve. Viral-load results were calibrated by using the National Institute for Biological Standards and Control (NIBSC) HAV RNA working reagent (Hertfordshire, United Kingdom), titrating 2,000 copies ml^{-1}. The sensitivity of this real-time RT-PCR assay was 125 copies ml^{-1} as assessed on serial dilutions of the NIBSC working reagent.

Sequencing of PCR amplicons. The amplification products were sequenced on both strands using the BigDye Terminator cycle-sequencing ready reaction kit,
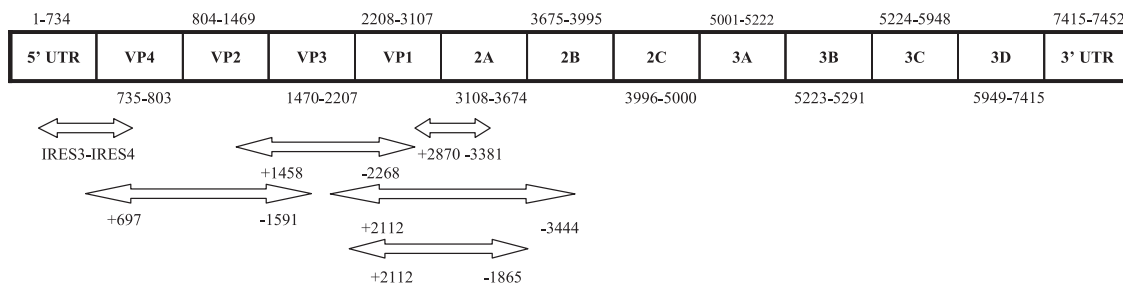

FIG. 1. Strategy for the amplification of the IRES, the VP1/2A junction, the N terminus of VP1, and the entire P1 region. The HAV genome is represented by a rectangle, in which the name of each HAV protein is shown in boldface. Double-headed open arrows below the rectangle represent the fragments of the HAV genome amplified using the primers listed in Table 1. UTR, untranslated region.

TABLE 2. HAV reference sequences used in this study

GenBank accession no.	Strain code	Subgenotype	Length of genome region (bp)	Place of isolation (infection)[a]
X75215	GBM	IA	Full length (7,431)	Germany
AF357222	LU38	IA	Full length (7,477)	China
AF485328	LY6	IA	Full length (7,477)	China
AF512536	DL3	IA	Full length (7,476)	China
AB020564	AH1	IA	Full length (7,477)	Japan
AB020565	AH2	IA	Full length (7,477)	Japan
AB020566	AH3	IA	Full length (7,445)	Japan
AB020567	FH1	IA	Full length (7,477)	Japan
AB020568	FH2	IA	Full length (7,478)	Japan
AB020569	FH3	IA	Full length (7,477)	Japan
EU251188	VBA-07	IA	Full length (7,487)	Russia
EU131373	HAV5	IA	Full length (7,452)	Uruguay
M14707	HM-175	IB	Full length (7,478)	Australia
M20273	MBB	IB	Full length (7,470)	Germany (North Africa)
AF268396	HAF-203	IB	Full length (7,468)	Brazil
AF314208	L-A-1	IB	Full length (7,420)	China
AY644676	CF53	IIA	Full length (7,426)	France
AY574059	BRAB13	IIA	VP1–2A (177)	Netherlands (Spain)
AJ519486	9F94	IIA	VP1 (900)	France (Morocco)
AY644670	SLF88	IIB	Full length (7,414)	United States (Sierra Leone)
AB279732	HA-JNG04-90F	IIIA	Full length (7,478)	Japan
AB279733	HA-JNG08-92F	IIIA	Full length (7,476)	Japan (Madagascar)
AB279734	HAJ95-8F	IIIA	Full length (7,478)	Japan (Philippines)
AY644337	HMH	IIIA	Full length (7,166)	Germany
AB279735	HAJ85-1F	IIIB	Full length (7,478)	Japan
AB258387	HA-JNG06-90F	IIIB	Full length (7,480)	Japan

[a] The place of infection is given only in cases where it is different from the place of isolation.

version 3 (Applied Biosystems, Foster City, CA) on an Applied Biosystems automatic sequencer, model 3130, according to the manufacturer's protocol.

Sequence analysis. Strain genotyping was carried out by the phylogenetic analysis of a 486-nucleotide fragment encompassing the VP1/2A junction. Sequences were aligned with Clustal X software. Phylogenetic trees were constructed with MEGA software (version 4.1) by using a Kimura 2-parameter model with the neighbor-joining algorithm. The reliability of the trees was tested by bootstrap resampling of 500 replications. Only bootstrap values above 70% are shown on the phylogenetic trees. Twenty-four reference sequences of complete genomes (16 of genotype I, 2 of genotype II, and 6 of genotype III), 1 subgenotype IIA VP1/2A sequence from strain BRAB13 (23), and 1 additional VP1 sequence of a recombinant IB–IIA virus, strain 9F94 (7), were used for the phylogenetic analysis (Table 2).

Mean percentages of nucleotide identity within and between genotypes were computed using the p-distance model included in the MEGA software. Genetic distances for synonymous and nonsynonymous nucleotide substitutions were determined using the Jukes-Cantor parameter within the modified Nei-Gojobori method.

The P1 and P2A sequences of subgenotype IIA isolates were compared to consensus sequences of genotypes I and III in order to identify polymorphic substitutions. Consensus sequences were constructed using Consensus Maker software (http://www.hiv.lanl.gov). The consensus sequences were generated from the reference sequences used for the phylogenetic analysis (Table 2).

Statistical analysis. Results are expressed as means ± standard deviations (SD). Due to the small sample sizes, only nonparametric statistical analyses were performed, with no multiple-testing corrections. The Wilcoxon test was used to compare the characteristics of patients infected with HAV subgenotype IIA with those of patients infected with other subgenotypes (IA, IB, and IIIA). A P value of <0.05 was considered significant.

Nucleotide sequence accession numbers. The GenBank accession numbers of the subgenotype IIA P1 and P2A sequences are GU390565, GU390566, GU390567, GU390568, GU390569, GU390570, GU390571, GU390572, GU390573, GU390574, GU390575, GU390576, and GU390577. The GenBank accession numbers of the subgenotype IIA IRES sequences are GU390553, GU390554, GU390555, GU390556, GU390557, GU390558, GU390559, GU390560, GU390561, GU390562, GU390563, and GU390564.

RESULTS

Epidemiological characteristics of subgenotype IIA strains. In 2008, 1,182 cases of acute hepatitis A were reported to the district health departments (incidence, 1.9/100,000). Travel outside mainland France was the main risk exposure for 459/1,182 (39%) patients. Of these 459 patients, 301 had traveled to Africa: 117 to the Maghreb (except Morocco), 13 to East Africa, and 171 to Morocco (n, 117) and West Africa (n, 54). HAV RNA amplification and strain genotyping were prospectively performed on 68/171 (40%) available sera from Morocco and West Africa; 48 (70.6%) clustered with subgenotype IB, 15 (22.0%) with subgenotype IA, and 5 (7.4%) with subgenotype IIA.

In addition to these 5 African cases, one autochthonous IIA strain and 2 additional African isolates were identified in 2008 and 2009, respectively. These strains were pooled with the 6 IIA strains identified from 2002 to 2007. Table 3 lists the geographical origins of the strains. Eight were imported from Africa. For the 6 autochthonous cases, no risk factor for HAV was identified, except living in poor sanitary conditions (for 3 individuals). Notification records were available for the 12 patients infected after 2006. Clinical symptoms reported were acute icteric hepatitis. None of these patients developed fulminant hepatitis, and only one presented with severe hepatocellular insufficiency. The IIA-infected patients were compared to 30 selected patients infected with genotype I or III in 2008 (Table 4). No significant difference in terms of age, ALT level, or the interval from the onset of symptoms to blood sampling was observed between patients infected by different subgenotypes.

TABLE 3. Origins of subgenotype IIA clinical isolates

Clinical IIA isolate[a]	Geographical origin
2004-1	Autochthonous[b]
2004-2	Autochthonous
2007-1	West Africa (Benin)
2007-2	Autochthonous
2007-3	Autochthonous
2007-4	Autochthonous
2008-1	West Africa (Cameroon)
2008-2	Autochthonous
2008-3	West Africa (Togo)
2008-4	West Africa (Mauritania)
2008-5	West Africa (Cameroon)
2008-6	West Africa (Cameroon)
2009-1	West Africa (Cameroon)
2009-2	West Africa (Cameroon)

[a] Labeled with the year of isolation and a number referring to the chronological order of isolation.
[b] Autochthonous, acquired without travel outside mainland France.

TABLE 5. Nucleotide identity within and between genotypes

Comparison	% Nucleotide identity (mean ± SD)[a]			
	IRES	P1	N terminus of VP1	VP1/2A
Strains within:				
Genotype I	97.6 ± 0.3	94.2 ± 0.4	94.0 ± 0.6	94.4 ± 1.0
Genotype II[b]	98.7 ± 0.2	96.2 ± 0.3	96.0 ± 0.5	96.6 ± 0.7
Genotype III	97.2 ± 0.5	91.6 ± 0.5	90.8 ± 0.9	91.3 ± 1.5
All genotypes together	95.4 ± 0.5	87.9 ± 0.5	88.6 ± 0.9	85.3 ± 1.6
Genotypes:				
I vs II	95.6 ± 0.7	85.9 ± 0.8	87.3 ± 1.3	82.6 ± 2.5
II vs III	92.2 ± 1.0	80.9 ± 0.9	81.8 ± 1.4	75.5 ± 2.7
I vs III	91.9 ± 0.7	81.2 ± 0.9	82.4 ± 1.4	74.9 ± 2.7

[a] Computed with MEGA software for clinical and reference sequences.
[b] For the comparison of strains within genotype II, the following numbers of sequences were used: 14 IRES, 14 P1, 16 VP1 N-terminal, and 16 VP1/2A sequences.

Viremia. HAV RNA levels in 12/14 available IIA samples and in the sera from the 30 controls were quantified by real-time RT-PCR. Our house technique failed to quantify HAV RNA levels in one IIA sample and two IB samples. The mean level of viremia was 4.0 $\log_{10}$ copies ml^{-1} (range, 2.64 to 5.31 $\log_{10}$ copies ml^{-1}) for samples collected, on average, 4.7 days after disease onset. Viral loads tended to be higher in IA samples; the difference from IIA samples was significant (Table 4).

Genetic analysis. Sequences for the N terminus of VP1 and the VP1/2A junction were available for all IIA strains. IRES and entire P1 sequences were available for 13/14 isolates. Among the 14 patients were a mother and her child with identical IRES and P1 sequences, so we excluded one of the two strains for the calculation of genetic distances. Distance calculation for genotype II strains included clinical sequences and IIA and IIB reference sequences.

As shown in Table 5, the variability of genotype II was lower than that of genotype I or III in the four regions analyzed. As with the other genotypes, the most conserved region was the 5′ UTR, with 98.7% ± 0.2% nucleotide identity over the 580 nucleotides studied encompassing the IRES. The CF53/Berne isolate presented higher diversity than the other strains, including strain SLF88. Interestingly, the highest genetic distance for genotype II was observed within the N terminus of the VP1 region, with 95.9% ± 0.5% identity over 510 nucleotides, while for other genotypes the most variable region was the VP1/2A junction. Genotypes I and II appeared closer to each other than to genotype III isolates, regardless of the genomic region.

The calculated nonsynonymous/synonymous distance ratio was very low along the P1 and P2A regions for genotype II as well as for genotype I and genotype III sequences (0.005 for genotype II versus 0.012 and 0.007 for genotypes I and III, respectively). Amino acid variability within genotype II sequences was extremely low (0.9% within the 835 amino acids analyzed). As shown in Fig. 2, only 2 clinical isolates differed from the type II consensus: 2004–2, by 2 amino acids at the N terminus of VP1 and in the P2A region, and 2009–2, by 1 amino acid in the P2A region. In addition, all IIA clinical isolates differed from the IIA isolate CF53/Berne by 3 amino acids in the VP3 region (residues 163, 177, and 180) and from the IIB isolate SLF88 by 1 amino acid in VP2 (residue 40) and 2 amino acids in VP3 (positions 45 and 145) (Fig. 2). IIA isolates differed from type IA, IB, and III consensus sequences by 5, 7, and 23 amino acids, respectively, especially at the N terminus of VP1 and at the VP1–2A junction. The amino acid variabilities of genotype I and III sequences were 4.5% and 1.4%, respectively. Analysis of the VP1 and P2A regions of all available strains revealed higher amino acid variability than that observed for the other capsid proteins (34 substitutions over 491 amino acids for VP4 and VP3 versus 43 over 344 amino acids for VP1 and P2A; P, 0.01).

TABLE 4. Patient characteristics

Group (no. of patients)	Value for the following characteristic[a] (mean ± SD):			
	Age (yr)	ALT[b] level (times the upper limit of normal)	Interval between hepatitis onset and blood sampling (days)	HAV viremia ($\log_{10}$ copies/ml)
All genotypes (44)	25.7 ± 15.7	79.0 ± 55.7	6.9 ± 6.2	4.6 ± 1.2
Subgenotype IA (10)	21.1 ± 13.5	88.1 ± 52.5	7.1 ± 8.6	5.9 ± 1.3*
Subgenotype IB (10)	20.4 ± 12.2	70.4 ± 47.2	7.0 ± 4.8	4.1 ± 0.9
Subgenotype IIIA (10)	33.3 ± 17.1	86.0 ± 41.8	8.1 ± 6.8	4.4 ± 0.8
Subgenotype IIA (14)	27.6 ± 16.9	72.8 ± 74.9	4.7 ± 2.9	4.0 ± 0.8*

[a] There was no statistical difference in these parameters between the different genotypes except for viremia, which differed between patients infected with subgenotype IA and those infected with subgenotype IIA ($P = 0.03$) (indicated by asterisks).
[b] ALT, alanine aminotransferase.

FIG. 2. Alignment of clinical IIA sequences and genotype consensus (CON) sequences for P1 and P2A. The specific consensus sequences for genotypes IA, IB, and III were constructed with available full-length reference sequences. The genotype II consensus was constructed with clinical sequences as well as with the CF53 and SLF88 reference sequences.

```
                                               VP3
           310       320       330       340       350       360       370       380       390       400
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CON-GII  SIPTLAAQFPFNASDSVGQQIKVIPVDPYFFQMTNSNPDQKCITALASICQMFCFWRGDLVFDFQVFPTKYHSGRLLFCFVPGNELIDVSGITLKQATTA
2008-6   ....................................................................................................
2007-2   ....................................................................................................
2007-1   ....................................................................................................
2008-2   ....................................................................................................
2007-3   ....................................................................................................
2008-4   ....................................................................................................
2009-1   ....................................................................................................
2004-2   ....................................................................................................
2004-1   ....................................................................................................
2008-1   ....................................................................................................
2008-3   ....................................................................................................
2008-5   ....................................................................................................
2009-2   ....................................................................................................
CF53-BERNE ..................................................................................................
SLF88-IIB  ..................................................................................................
CON-IA   ...................................T...................................................T...........
CON-IB   ...................................T................................................................
CON-III  ...................................T..E.........................................H..................
```

```
                                               VP3                                         VP1
           410       420       430       440       450       460       470       480       490       500
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CON-GII  PCAVMDITGVQSTLRFRVPWISDTPYRVNRYTKSAHQKGEYTAIGKLIVYCYNRLTSPSNVASHVRVNVYLSAINLECFAPLYHAMDVTTQVGDDSGGFS
2008-6   ....................................................................................................
2007-2   ....................................................................................................
2007-1   ....................................................................................................
2008-2   ....................................................................................................
2007-3   ....................................................................................................
2008-4   ....................................................................................................
2009-1   ....................................................................................................
2004-2   ....................................................................................................
2004-1   ....................................................................................................
2008-1   ....................................................................................................
2008-3   ....................................................................................................
2008-5   ....................................................................................................
2009-2   ....................................................................................................
CF53-BERNE .......A.............P..R...........................................................................
SLF88-IIB  ..................................................................................................
CON-IA   ....................................................................................................
CON-IB   ....................................................................................................
 CON-III ............................S.......................................................................
```

```
                                               VP1
           510       520       530       540       550       560       570       580       590       600
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CON-GII  TTVSTEQNVPDPQVGITTMRDLKGKANRGKMDVSGVQAPVGAITTIEDPVLAKKVPETFPELKPGESRHTSDHMSIYKFMGRSHFLCTFTFNSNNKEYTF
2008-6   ....................................................................................................
2007-2   ....................................................................................................
2007-1   ....................................................................................................
2008-2   ....................................................................................................
2007-3   ....................................................................................................
2008-4   ....................................................................................................
2009-1   ....................................................................................................
2004-2   ...................S................................................................................
2004-1   ....................................................................................................
2008-1   ....................................................................................................
2008-3   ....................................................................................................
2008-5   ....................................................................................................
2009-2   ....................................................................................................
CF53-BERNE ........................................................................................................
SLF88-IIB  ...................R................................................................................
CON-IA   ....................................................................................................
CON-IB   ....................K...............................................................................
CON-III  .....K...........VK....R..Q....I....................................................................
```

FIG. 2—*Continued.*

```
                                              VP1
          ───────────────────────────────────────────────────────────────────────────────
              610       620       630       640       650       660       670       680       690       700
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CON-GII   PITLSSTSNPPHGLPSTLRWFFNLFQLYRGPLDLTIIITGATDVDGMAWFTPVGLAVDTPWVEKESALSIDYKTALGAVRFNTRRTGNIQIRLPWYSYLY
2008-6    ....................................................................................................
2007-2    ....................................................................................................
2007-1    ....................................................................................................
2008-2    ....................................................................................................
2007-3    ....................................................................................................
2008-4    ....................................................................................................
2009-1    ....................................................................................................
2004-2    ....................................................................................................
2004-1    ....................................................................................................
2008-1    ....................................................................................................
2008-3    ....................................................................................................
2008-5    ....................................................................................................
2009-2    ....................................................................................................
CF53-BERNE ...................................................................................................
SLF88-IIB ....................................................................................................
CON-IA    ....................................................................................................
CON-IB    ....................................................................................................
CON-III   ....................................................................................................

                                       VP1                                              2A
          ─────────────────────────────────────────────────────────────────── ────────────────────────────
              710       720       730       740       750       760       770       780       790       800
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CON-GII   AVSGALDGLGDKTDSTFGLVSIQIANYNHSDEYLSFSCYLSVTEQSEFYFPRAPLNSNAMLSTETMMSRIAAGDLESSVDDPRSEEDRRFESHIESRKPY
2008-6    ....................................................................................................
2007-2    ....................................................................................................
2007-1    ....................................................................................................
2008-2    ....................................................................................................
2007-3    ....................................................................................................
2008-4    ....................................................................................................
2009-1    ....................................................................................................
2004-2    ....................................................................................................
2004-1    ....................................................................................................
2008-1    ....................................................................................................
2008-3    ....................................................................................................
2008-5    ....................................................................................................
2009-2    ....................................................................................................
CF53-BERNE ...................................................................................................
SLF88-IIB ....................................................................................................
CON-IA    ...........................................................S............................C....
CON-IB    ...........................................................S........................K......C....
CON-III   ......................................................T...M.S...D..L................K.....K....

                         2A
          ───────────────────────────
              810       820       830
          ....|....|....|....|....|....|.
CON-GII   KELRLEVGKQRLKYAQEELSNEVLPPPRKMKGLFSQ
2008-6    ...................................
2007-2    ...................................
2007-1    ...................................
2008-2    ...................................
2007-3    ...................................
2008-4    ...................................
2009-1    ...................................
2004-2    .......R...........................
2004-1    ...................................
2008-1    ...................................
2008-3    ...................................
2008-5    ...................................
2009-2    .........................V.....
CF53-BERNE .................................
SLF88-IIB ...................................
CON-IA    ...................................
CON-IB    ...................................
CON-III   ........................I..V...
```

FIG. 2—*Continued.*

Phylogenetic analysis of IRES sequences identified a single cluster for clinical genotype II sequences, which included the IIB strain SLF88. Interestingly, the CF53/Berne sequence clustered apart from clinical sequences, probably reflecting its adaptation to cell culture (data not shown).

Phylogenetic analysis of P1 sequences showed two main clusters of subgenotype IIA strains, supported by a high bootstrap value (Fig. 3a). The first cluster included P1 sequences from 10 patients: 8 patients who had traveled to West Africa and 2 with no travel risk factors. These two strains (2007-2 and 2007-3) were recovered from a mother and her child. The second cluster included 3 sequences from autochthonous cases. As with the IRES analysis, the P1 sequence of the CF53/Berne isolate clustered apart.

Strain 9F94 (7) was included in the VP1 phylogenetic tree (Fig. 3b). VP1 analysis evidenced the same two main clusters as P1 analysis, but the VP1 clusters were not supported by bootstrap values greater than 70%. Strain 9F94 appeared closer to West African isolates than to autochthonous isolates, perhaps due to its Moroccan origin. The CF53/Berne strain clustered apart.

Analysis of the N terminus of VP1 comprised all 14 clinical IIA isolates and strain 9F94 and yielded the same two main clusters as P1 analysis, with high bootstrap values (data not shown).

The VP1/2A analysis included all 14 clinical isolates and strain BRAB13, recovered from a Dutch patient who had traveled in Spain (Fig. 3c). The phylogenetic tree showed the same two main clusters, but with bootstrap values lower than 70%. Again, the CF53/Berne isolate belonged to a different lineage. Interestingly, strain BRAB13 (23) clustered with West African sequences.

DISCUSSION

To date, several full-length sequences of genotype I and genotype III have been published (5, 9–11, 17, 22). In contrast, genotype II has rarely been reported worldwide, so genomic information is sparse. The availability of full-length sequences for each IIB and IIA isolate (4, 14) excluded the possibility that this genotype was an artifact of classification derived from insufficient sequence analysis of recombinant HAV strains, as suggested previously (6, 7). However, other than the CF-53/Berne reference isolate, whose full-length sequence has been determined, a single IIA strain has been reported and was characterized only by its VP1/2A sequence (23). In the present work, we report 13 unique IIA strains isolated from 14 patients and suggest their likely African origin. Our results evidence a low genetic variability of the IIA subgenotype, regardless of the genomic region used for analysis.

Comparison of the sequences of our clinical strains to those of the CF53/Berne isolate revealed three amino acid differences localized in VP3 (residues 162, 176, and 179). These variations may be considered a result of the cell adaptation of the CF-53 isolate, from which the sequence was obtained after 12 cell culture passages (14). This isolate clustered apart, no matter which region was used for analysis.

As reported for the other genotypes (3, 12, 22), the 5′ UTR was the most conserved region of genotype II strains, and the phylogenetic analysis performed with this region was unable to distinguish between the IIA and IIB subgenotypes. In the P1 region, very few nonsynonymous mutations were observed, and the amino acid variability of capsid proteins was less than 1% despite nearly 5% nucleotide variability. Studies using full-length VP1 and the whole P1 sequences have also shown very low rates of nonsynonymous substitutions compared to synonymous substitutions for genotypes I and III (11, 22). This pattern of divergence is probably due to negative selection forces that do not allow amino acid replacements (1). The low capsid variability correlates with a very low antigenic variability of HAV, since only a single serotype has been described, and this is a striking difference from the other members of the family *Picornaviridae* (13, 20). To date, several antigenic epitopes of the capsid protein have been reported, comprising amino acids from both the VP1 (residues 11 to 25, 102, 171, 176, and 221) and VP3 (residues 70, 74, and 110 to 121) regions (2, 15, 16, 18). Among the 13 clinical isolates reported here, only 2004-1 harbored one amino acid substitution in the antigenic epitope of VP1 comprising residues 11 to 25, suggesting a possible altered antigenicity of this particular strain. However, no clinical impact was observed; the clinical features of the IIA-infected patients were not different from those of patients infected with other genotypes.

HAV viremia measured a mean of 4.7 days after the onset of symptoms ranged from 2.64 to 5.31 $\log_{10}$ copies ml^{-1}, with a mean value of 4.0 $\log_{10}$ copies ml^{-1}. Surprisingly, the mean HAV viral load was nearly 2 $\log_{10}$ units higher for subgenotype IA than for the other subgenotypes, despite a similar delay from disease onset to sampling. This difference may indeed reflect a higher replication capacity of IA isolates, or it could, more likely, be related to less-efficient quantification of other genotypes/subtypes by our in-house assay, as also reported for commercial assays (21). Less-efficient quantification may be due to the design of the primers and probe. The set of primers and the probe of our quantification assay target the 5′ UTR upstream of the IRES region; thus, we could not ascertain the validity of this hypothesis. By measuring viremia on the first day of clinical diagnosis, Costa-Mattioli et al. obtained values ranging from 3.2 to 6.9 $\log_{10}$ copies ml^{-1} (mean, 5.8 $\log_{10}$ copies ml^{-1}) (8). These slightly higher viremia results may be explained by overrepresentation of IA subtypes in the previous series.

From 2002 to 2007, our survey identified a few IIA strains acquired in mainland France with no specific risk factors reported. The first imported IIA strain was isolated from an anthropologist returning from West Africa. This prompted us to investigate whether IIA strains could originate from Africa, as subgenotype IIB strains do. The 2008 mandatory notification records enabled us to identify five additional IIA strains from patients who had contracted HAV infections in West Africa. Finally, among our 14 IIA-infected patients, nearly two-thirds had acquired HAV in West Africa, especially in Cameroon. Besides, the data provided by the continuous surveillance of circulating HAV strains in France confirmed the very low incidence of genotype II, representing less than 1% of the isolates collected within 7 years. The limited detection of genotype II may suggest either a unique source of infection, a specific route of transmission, or an inability to maintain infections in humans. The present study proves that the geographical origin of subgenotype IIA was not unique. The phy-
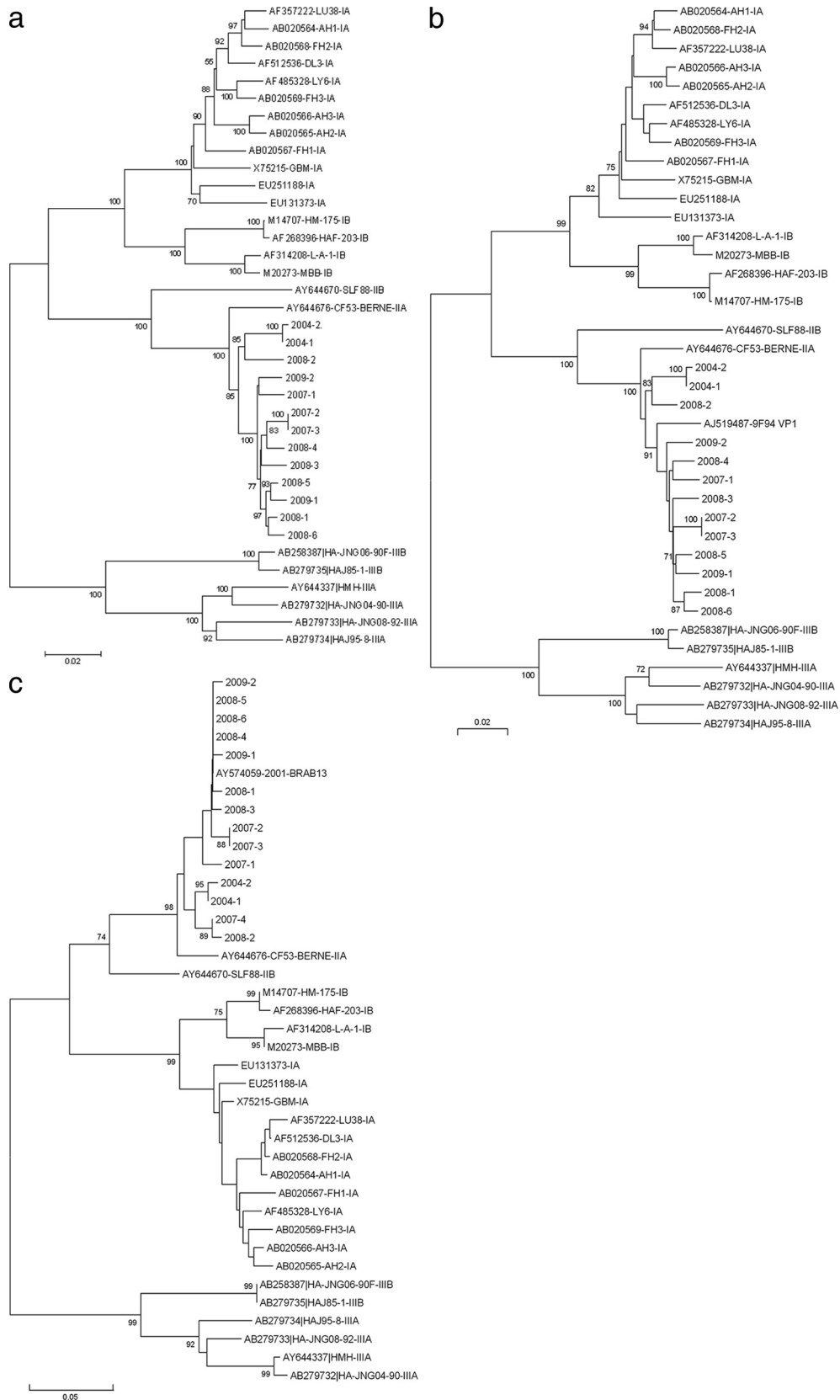
FIG. 3. Phylogenetic analysis of clinical and reference isolates. Reference sequences are labeled with their GenBank accession numbers, strain codes, and defined subgenotypes according to the most recent classification of HAV isolates. Clinical isolates are labeled with their year of isolation and a number referring to the chronological order of isolation. Trees were constructed from a Kimura two-parameter genetic-distance matrix using the neighbor-joining method. (a) P1 sequences (2,295 nucleotides); (b) VP1 sequences (900 nucleotides); (c) VP1/2A sequences (486 nucleotides). Bootstrap values, obtained after 1,000 replications of bootstrap sampling, are given at the branches. Bars at the bottom of each phylogenetic tree denote distance.

logenetic analysis of P1 and the N terminus of VP1 allowed us to distinguish two clusters reliably, with high bootstrap values: a cluster including African and some autochthonous strains and a distinct genetic lineage with autochthonous strains only. The two patients with autochthonous cases whose strains clustered with West African isolates lived in poor sanitary conditions. They were probably infected through close contact with an infected patient returning from West Africa. For the other four autochthonous cases, the origin of contamination remains elusive. These results suggest either another, still to be discovered geographical origin or a French reservoir to be explored. The two clusters we identified for IIA strains might have resulted from a genetic divergence of HAV strains over time, and not from a geographical difference. However, strains 9F94 and BRAB13, isolated in 1994 and 2001, clustered with West African strains isolated from 2007 to 2009. Very recently, another IIA isolate was identified in a patient returning from Chad (data not shown). Nevertheless, the phylogenetic analysis using VP1/2A sequences revealed that this new IIA strain clustered with West African isolates and not with autochthonous isolates.

In conclusion, this study provides new information on the epidemiological and genetic characteristics of HAV subgenotype IIA strains. The genetic variability within this genotype is extremely low both in the 5′ UTR and in the P1 and P2A regions. Most IIA strains isolated in France are imported by travelers returning from West Africa. According to phylogenetic analysis, some patients with autochthonous cases may have been infected through contact with infected individuals returning from West Africa. However, the source of infection for about one-third of the cases remains elusive. The hypothesis of another geographical origin or of a French reservoir should be investigated.

REFERENCES

1. **Aragonès, L., A. Bosch, and R. M. Pinto.** 2008. Hepatitis A virus mutant spectra under the selective pressure of monoclonal antibodies: codon usage constraints limit capsid variability. J. Virol. **82:**1688–1700.
2. **Bosch, A., J. F. Gonzalez-Dankaart, I. Haro, R. Gajardo, J. A. Perez, and R. M. Pinto.** 1998. A new continuous epitope of hepatitis A virus. J. Med. Virol. **54:**95–102.
2a. **Bower, W. A., O. V. Nainan, X. Han, and H. S. Margolis.** 2000. Duration of viremia in hepatitis A virus infection. J. Infect. Dis. **182:**12–17.
3. **Brown, E. A., S. P. Day, R. W. Jansen, and S. M. Lemon.** 1991. The 5′ nontranslated region of hepatitis A virus RNA: secondary structure and elements required for translation in vitro. J. Virol. **65:**5828–5838.
4. **Ching, K. Z., T. Nakano, L. E. Chapman, A. Demby, and B. H. Robertson.** 2002. Genetic characterization of wild-type genotype VII hepatitis A virus. J. Gen. Virol. **83:**53–60.
5. **Cohen, J. I., J. R. Ticehurst, R. H. Purcell, A. Buckler-White, and B. M. Baroudy.** 1987. Complete nucleotide sequence of wild-type hepatitis A virus: comparison with different strains of hepatitis A virus and other picornaviruses. J. Virol. **61:**50–59.
6. **Costa-Mattioli, M., A. Di Napoli, V. Ferre, S. Billaudel, R. Perez-Bercoff, and J. Cristina.** 2003. Genetic variability of hepatitis A virus. J. Gen. Virol. **84:**3191–3201.
7. **Costa-Mattioli, M., V. Ferre, D. Casane, R. Perez-Bercoff, M. Coste-Burel, B. M. Imbert-Marcille, E. C. Andre, C. Bressollette-Bodin, S. Billaudel, and J. Cristina.** 2003. Evidence of recombination in natural populations of hepatitis A virus. Virology **311:**51–59.
8. **Costa-Mattioli, M., S. Monpoeho, E. Nicand, M. H. Aleman, S. Billaudel, and V. Ferre.** 2002. Quantification and duration of viraemia during hepatitis A infection as determined by real-time RT-PCR. J. Viral Hepat. **9:**101–106.
9. **Endo, K., M. Takahashi, K. Masuko, K. Inoue, Y. Akahane, and H. Okamoto.** 2007. Full-length sequences of subgenotype IIIA and IIIB hepatitis A virus isolates: characterization of genotype III HAV genomes. Virus Res. **126:**116–127.
10. **Fujiwara, K., O. Yokosuka, K. Fukai, F. Imazeki, H. Saisho, and M. Omata.** 2001. Analysis of full-length hepatitis A virus genome in sera from patients with fulminant and self-limited acute type A hepatitis. J. Hepatol. **35:**112–119.
11. **García-Aguirre, L., and J. Cristina.** 2008. Analysis of the full-length genome of hepatitis A virus isolated in South America: heterogeneity and evolutionary constraints. Arch. Virol. **153:**1473–1481.
12. **Joshi, M. S., A. M. Walimbe, and S. D. Chitambar.** 2008. Evaluation of genomic regions of hepatitis A virus for phylogenetic analysis: suitability of the 2C region for genotyping. J. Virol. Methods **153:**36–42.
13. **Lemon, S. M., R. W. Jansen, and E. A. Brown.** 1992. Genetic, antigenic and biological differences between strains of hepatitis A virus. Vaccine **10**(Suppl. 1):S40–S44.
14. **Lu, L., K. Z. Ching, V. S. de Paula, T. Nakano, G. Siegl, M. Weitz, and B. H. Robertson.** 2004. Characterization of the complete genomic sequence of genotype II hepatitis A virus (CF53/Berne isolate). J. Gen. Virol. **85:**2943–2952.
15. **Nainan, O. V., M. A. Brinton, and H. S. Margolis.** 1992. Identification of amino acids located in the antibody binding sites of human hepatitis A virus. Virology **191:**984–987.
16. **Nainan, O. V., G. Xia, G. Vaughan, and H. S. Margolis.** 2006. Diagnosis of hepatitis A virus infection: a molecular approach. Clin. Microbiol. Rev. **19:**63–79.
17. **Paul, A. V., H. Tada, K. von der Helm, T. Wissel, R. Kiehn, E. Wimmer, and F. Deinhardt.** 1987. The entire nucleotide sequence of the genome of human hepatitis A virus (isolate MBB). Virus Res. **8:**153–171.
18. **Ping, L. H., and S. M. Lemon.** 1992. Antigenic structure of human hepatitis A virus defined by analysis of escape mutants selected against murine monoclonal antibodies. J. Virol. **66:**2208–2216.
19. **Robertson, B. H., R. W. Jansen, B. Khanna, A. Totsuka, O. V. Nainan, G. Siegl, A. Widell, H. S. Margolis, S. Isomura, K. Ito, et al.** 1992. Genetic relatedness of hepatitis A virus strains recovered from different geographical regions. J. Gen. Virol. **73:**1365–1377.
20. **Sánchez, G., A. Bosch, and R. M. Pinto.** 2003. Genome variability and capsid structural constraints of hepatitis A virus. J. Virol. **77:**452–459.
21. **Sánchez, G., S. Populaire, S. Butot, T. Putallaz, and H. Joosten.** 2006. Detection and differentiation of human hepatitis A strains by commercial quantitative real-time RT-PCR tests. J. Virol. Methods **132:**160–165.
22. **Stene-Johansen, K., T. O. Jonassen, and K. Skaug.** 2005. Characterization and genetic variability of hepatitis A virus genotype IIIA. J. Gen. Virol. **86:**2739–2745.
22a. **Stene-Johansen, K., G. Tjon, E. Schreier, V. Bremer, S. Bruisten, S. L. Ngui, M. King, R. M. Pinto, L. Aragonès, Z. Mazick, S. Corbet, L. Sundqvist, H. Blystad, H. Norder, and K. Skaug.** 2007. Molecular epidemiological studies show that hepatitis A virus is endemic among active homosexual men in Europe. J. Med. Virol. **79:**356–365.
23. **Tjon, G. M., C. J. Wijkmans, R. A. Coutinho, A. G. Koek, J. A. van den Hoek, A. C. Leenders, P. M. Schneeberger, and S. M. Bruisten.** 2005. Molecular epidemiology of hepatitis A in Noord-Brabant, The Netherlands. J. Clin. Virol. **32:**128–136.