Molecular epidemiology of hepatitis A virus infections in Catalonia, Spain, 2005–2009: Circulation of newly emerging strains

Unai Pérez-Sautu a, M. Isabel Costafreda a, Josep Lite b, Rosa Sala c, Irene Barrabeig c, Albert Bosch a,*, Rosa M. Pintó a,**

a Enteric Virus Laboratory, Department of Microbiology and Institute of Nutrition and Food Safety, University of Barcelona, Barcelona, Spain
b CatLab, Viladecavalls, Spain
c Department of Health, Autonomous Government of Catalonia, Barcelona, Spain

ARTICLE INFO

Received 13 April 2011
Received in revised form 10 June 2011
Accepted 24 June 2011

Keywords:
Hepatitis A virus
Genotype
Food imports
MSM

ABSTRACT

Background: In spite of annual vaccination campaigns, hepatitis A cases increased in Catalonia (North-East Spain) in the period 2002–2005 calling for the elucidation of the underlying mechanisms associated to the epidemiological shifts.

Objective: The molecular characterization of the circulating strains to trace their origin and the study of the effects of vaccination on the incidence of sporadic and outbreak-associated cases.

Study design: Forty-eight different hepatitis A virus (HAV) strains isolated from sporadic and outbreaks cases during 2005–2009 in Catalonia were molecularly characterized.

Results: Seventeen out of 48 strains were imported from endemic areas through traveling, immigration and food trade, 12 were endemic strains circulating in the men having sex with men (MSM) group and 1 was from a Roman child. The remaining 18 could not be associated to any specific origin and thus were considered autochthonous. Forty-eight percent of the strains belonged to subgenotype IA, 40% to subgenotype IB and 2% to subgenotype IIA. The remaining 10% belonged to an undetermined subgenotype equidistant from IA and IB.

Conclusions: During the period 2005–2009, the annual attack rates remained around 3.5 and even increased up to 6.5 in the first half of 2009. This increase with respect to the period 1999–2001, in which vaccination campaigns started to be implemented, is explained by an increase in the number of outbreaks. The predominant subgenotypes were IA and IB. However a considerable amount of strains imported from Peru through consumption of contaminated shellfish belonged to an undetermined subgenotype that may constitute a new candidate subgenotype IC.

© 2011 Elsevier B.V. All rights reserved.

1. Background

The incidence of hepatitis A in Catalonia (North-East Spain) significantly decreased during the period 2000–2001 due to the implementation since 1999 of a vaccination program among preadolescents.1 The effectiveness of the vaccination campaign is clearly reflected in the attack rate reduction from 6.2 cases per 100,000 inhabitants in the period 1996–98 to 2.6 in the period 1999–2001. However since 2001, the number of cases raised, in spite of annual vaccination campaigns, with an increase of the attack rate up to 3.5 in the period 2002–2005.

In low-moderate endemicity areas, hepatitis A infection mainly develops as outbreaks among the adolescent and adult populations, as some sporadic cases and to a much lesser extent as asymptomatic cases in children. Virus introduction from endemic areas is mainly done through consumption of imported foods and traveling. In recent years, several outbreaks related with consumption of imported contaminated food have been reported in Europe.3–6 Several outbreaks related to travels have also been reported.7,8 Additionally, and particularly in the case of Catalonia, the high immigration flows from North Africa, South America and other areas of high endemicity might represent the introduction of new viral strains. A high risk group for hepatitis A infection is men having sex with men (MSM) and several outbreaks affecting the MSM group have been recently reported in different countries across Europe.9,10

Comparative studies of HAV strains have suggested that sequence relatedness can be correlated with the geographical origin of the virus.11 Nucleotide sequences may be used as molecular
markers for the elucidation of origins and modes of transmission of the virus. Different genomic regions have been used to differentiate HAV genotypes. However, partial genomic sequences will never achieve the reliability of the complete genome. This is the rationale of the choice of long genomic regions covering at least the entire VP1 including its 2A junction (European HAV Network) for a more broad molecular typing of HAV. Using this latter region, HAV strains are classified into six genotypes: I, II and III of human origin and IV, V and VI of simian origin. Genotypes I, II and III are further subdivided into subgenotypes, being their nucleotide distance cut-off of 7.5% and 5.1% in the VP1X2A and complete VP1 regions, respectively.

2. Objective

The molecular characterization of strains circulating in Catalonia in the period 2005–2009 and the study of the effects of vaccination on the incidence of sporadic and outbreak-associated cases.

3. Study design

3.1. Serum samples

Serum specimens (n = 157) were collected from anti-HAV IgM-positive patients in Catalonia from April 2005 to July 2009. Samples were from both sporadic (n = 54) and outbreak (n = 103) related cases. Epidemiological data including age, sex, vaccination status, suspected source of infection, recent travels, mode of transmission (person to person, sexual contact, intravenous drug administration) and sexual orientation were recorded.

3.2. RNA extraction

RNA was extracted from 150 μl of serum samples with the Nucleospin RNA virus kit (Macherey Nagel).

3.3. Detection and typing

HAV RNA was amplified by RT-PCR. Two independent two-step-RT-PCR assays were conducted: one corresponding to the N-terminal fragment of the VP1 coding region and another to the C-terminal fragment. Altogether both fragments included the VP3/VP1 junction, the entire VP1 and the VP1/2A junction. The previously described primer pairs NH2-VP1 and COOH-VP1 were used in both reactions, respectively. The Expand Reverse Transcriptase and the Expand HiFi PCR System enzyme kits (Roche) were used.

3.4. Sequencing

DNA was purified using the High Pure PCR Product Purification kit (Roche) and sequenced using the ABI Prism® Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit v3.1 (Applied Biosystems) and the ABI Prism 3700 automatic sequencer (Applied Biosystems). The reverse strand was sequenced and all mutations confirmed in the forward strand.

3.5. Phylogenetic analysis

Whenever possible, sequences of both amplicon fragments (1057 bp) were used for the phylogenetic analysis. Otherwise, sequences of the amplicon corresponding to the C-terminal part of the VP1 coding region (521 bp) were used.

Nucleotide sequences were compared with those deposited in GenBank and EMBL using the BLAST N and CLUSTAL W softwares. Boot-Strap phylogenetic trees were constructed using the MEGA software package, version 4.0.2 with the following settings: neighbor-joining trees using the kimura 2-parameter model and a 1000 pseudoreplicas bootstrap resampling. Sequences have been deposited in GenBank with accession numbers HQ401214 to HQ401267.

4. Results

4.1. Study population

Analysed samples (n = 157) represented 14% of total declared cases (Table 1). Fifty-four samples (34.4%) were from sporadic cases and 103 (65.6%) from outbreaks. This represent a certain bias towards the analysis of outbreaks since sporadic cases accounted for 45.7% of total hepatitis cases while outbreak-associated cases accounted for 54.3%. As can be observed in Table 1 the total annual attack rate per 100,000 inhabitants clearly increased in 2009 due to a huge outbreak. However, the attack rate per 100,000 inhabitants not taking into consideration the outbreak-associated cases remained low along the studied period.

Among sporadic cases, 52% of patients were male and 48% female and the average age was 24.8 years (21 cases in the 0–12 years group; 0 cases in the 12–21 years group corresponding to the vaccinated group; 33 cases in the >21 years group). Travel information could be obtained for 23 out of 54 sporadic cases (42.6%) and of these 14 (61%) were confirmed to be travel-related. Travel-related cases corresponded to local people visiting endemic countries or immigrant people visiting their countries of birth. The most visited country was Morocco followed by Egypt and, to a lesser extent, Peru and India. The origin of the remaining 31 sporadic cases (57.4%) could not be traced.

Seven outbreaks affecting 225 patients were studied and of these 103 samples were analyzed (Table 2). The total number of outbreaks of the period was of 135 and the total number of patients was 609.

4.2. Phylogenetic analysis

HAV-typing was achieved in 129 out of 157 analyzed samples. Most of the non-typable samples corresponded to sera obtained long after the onset of symptoms. Forty-eight different strains were distinguished among these 129 isolates. Twenty-three strains (102 isolates) belonged to subgenotype IA, 19 strains belonged to subgenotype IB (21 isolates) and 1 strain belonged to subgenotype IIa (1 isolate). The remaining 5 strains (5 isolates) belonged to an undetermined subgenotype equidistant from IA and IB.
Table 2

<table>
<thead>
<tr>
<th>Date</th>
<th>Setting</th>
<th>Index case</th>
<th>Average age (gender)(^a)</th>
<th>Source of infection</th>
</tr>
</thead>
<tbody>
<tr>
<td>May06</td>
<td>Family</td>
<td>Adopted Ethiopian child</td>
<td>36.0 m (2M)</td>
<td>Immigration</td>
</tr>
<tr>
<td>Jun06</td>
<td>Nursery</td>
<td>Boy returning from Colombia</td>
<td>33.0 y (1F)</td>
<td>Immigration</td>
</tr>
<tr>
<td>May07</td>
<td>Family</td>
<td>Husband of a couple</td>
<td>2.3 y (5M, 3F)</td>
<td>Unknown</td>
</tr>
<tr>
<td>Jul08</td>
<td>Nursery</td>
<td>Father of a child</td>
<td>29.6 y (5M, 7F)</td>
<td>Unknown</td>
</tr>
<tr>
<td>Oct08</td>
<td>General</td>
<td>Unknown</td>
<td>29.3 y (1M, 1F)</td>
<td>Imported clams</td>
</tr>
<tr>
<td>Sep08–Mar09</td>
<td>MSM</td>
<td>Unknown</td>
<td>33.0 y (185M)</td>
<td>Risky sexual practices</td>
</tr>
<tr>
<td>Oct08</td>
<td>School</td>
<td>Boy returning from Egypt</td>
<td>6.0 y (3M, 1F)</td>
<td>Travel-related</td>
</tr>
</tbody>
</table>

In outbreaks declared in nurseries or schools, as well as that affecting the MSM community, vaccination to contacts was implemented as a measure to prevent spread of the infection. The foodborne outbreak, that in Catalonia affected very few people but that was responsible for over 100 cases in the rest of Spain, caused the immobilization of around 1000 tons of shellfish and the ban of all Peruvian shellfish imports in the European Union (14).

\(^a\) Average age is expressed in years (y) or months (m). Children and adults data are shown separately. The number of males (M) and females (F) is depicted in brackets.

Phylogenetic trees of sequences from the C-terminal fragment of VP1 (Fig. 1) or the complete VP1 gene (Fig. 2) were generated. Sequences from samples isolated off the period under study were also included to see a broader overview.

Among subgenotype IA strains two clusters were detected in association with the MSM group. The MSM04-06 cluster contained strains from MSM patients isolated during the period 2004–2006 with one strain previously isolated in 2004 (BCN05-
04 with 1 single isolate) and two strains in the present period of study, one in 2005 (BCN04-05 with 1 single isolate) and another in 2006 (2 identical isolates BCN52, BCN68). However, a more significant number of strains were isolated during the MSM-outbreak 2008–2009 (MSM08-09). Nine different strains (with an average of 99% homology among them) were isolated in this outbreak. Additionally, another strain, isolated before the beginning of the outbreak in 2 samples, belonged to this same cluster. The strain was isolated from male patients although information regarding their sexual behavior was not available (BCN87 and BCN92; labeled with an asterisk in the trees), and thus although phylogenetically-related their epidemiological relationship remains unclear. Simultaneously to the MSM08-09 outbreak, 2 strains also belonging to subgenotype IA were detected in several sporadic cases. No epidemiological information could be obtained for these samples, although the phylogenetic analysis revealed that these cases were caused by 2 strains which did not belong to the MSM08-09 cluster.

Among subgenotype IB strains, two clusters of closely related strains could be clearly differentiated. The so-called Egyptian cluster related with travels to Egypt and including sporadic and outbreak cases. The strain isolated from the outbreak May-2006 (Suria06) was closely related with this Egyptian cluster and in fact the origin was Ethiopia. The so-called Moroccan cluster consisted of strains detected in sporadic cases related with travels to this country. This cluster also included strains from sporadic cases not related with recent travels (BCN60, BCN74, BCN75, BCN78, BCN80 and BCN227; labeled with an asterisk in the trees), whose close relationship with the aforementioned Moroccan strains, suggested a common geographical origin.

Apart from subgenotype IA and IB strains, one subgenotype IIIA strain was detected in one sample corresponding to a child of Roman origin.

### 4.2. A new subgenotype I (IC) endemic in Peru is circulating in Spain

Two strains were isolated from patients who had consumed coquinas clams imported from Peru (Shellfish08-106, Shellfish08-110). These strains form a closely related cluster with other strains isolated during another shellfish-borne outbreak in 1999 (Valencia strains), being the source of infection the same type of coquinas imported from Peru. The relatedness of these strains in spite of the time lapse between their isolations indicates that this cluster is highly endemic in Peru. Another three strains (BCN30, BCN70 and BCN72) isolated before the foodborne outbreak of 2008 belonged to the same cluster and probably represent strains circulating in our area since the first importation from Peru in 1999. The strains forming this cluster belong to an undetermined subgenotype midway from IA and IB following the phylogenetic information (Figs. 1 and 2). The maximum nucleotide homology of this cluster with reference strains belonging to subgenotype IA (CR326 and GBM, GenBank accession numbers M10033 and X75215, respectively) or IB (HM175 and MBB, GenBank accession numbers M14707 and M20273, respectively) is of 94.6% with both subgenotypes in the VP1X2A region and 93.5% (IA) and 91.0% (IB) in the VP3XVP1X2A region (Table 3). Given these values, and the independent branch in the phylogenetic tree of the VP3XVP1X2A region (Fig. 2), it is not possible to ascertain to which subgenotype these samples belong. Consequently a new subgenotype, IC, is proposed.

#### Table 3

<table>
<thead>
<tr>
<th>Strains</th>
<th>CR326 (IA)</th>
<th>GBM (IA)</th>
<th>HM175 (IB)</th>
<th>MBB (IC)</th>
<th>30 (IC)</th>
<th>70/72 (IC)</th>
<th>106 (IC)</th>
<th>110 (IC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR326</td>
<td>98.2</td>
<td>96.3</td>
<td>91.4</td>
<td>91.4</td>
<td>94.4</td>
<td>94.6</td>
<td>92.8</td>
<td>92.8</td>
</tr>
<tr>
<td>(IA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBM</td>
<td>100</td>
<td>91.7</td>
<td>94.0</td>
<td>94.6</td>
<td>94.6</td>
<td>92.8</td>
<td>93.4</td>
<td></td>
</tr>
<tr>
<td>(IA)</td>
<td>99.4</td>
<td>91.7</td>
<td>93.8</td>
<td>94.4</td>
<td>93.5</td>
<td>93.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98.6</td>
<td>90.8</td>
<td>91.6</td>
<td>–</td>
<td>93.3</td>
<td>93.1</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HM175</td>
<td>98.2</td>
<td>98.2</td>
<td>96.4</td>
<td>92.3</td>
<td>91.7</td>
<td>92.3</td>
<td>93.4</td>
<td></td>
</tr>
<tr>
<td>(IB)</td>
<td>98.8</td>
<td>99.4</td>
<td>95.4</td>
<td>91.4</td>
<td>91.4</td>
<td>91.5</td>
<td>91.5</td>
<td></td>
</tr>
<tr>
<td>99.1</td>
<td>98.9</td>
<td>94.1</td>
<td>–</td>
<td>91.0</td>
<td>90.8</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBB</td>
<td>98.2</td>
<td>98.2</td>
<td>100</td>
<td>94.0</td>
<td>94.6</td>
<td>92.8</td>
<td>93.4</td>
<td></td>
</tr>
<tr>
<td>(IB)</td>
<td>98.8</td>
<td>99.4</td>
<td>100</td>
<td>91.0</td>
<td>91.7</td>
<td>91.4</td>
<td>91.2</td>
<td></td>
</tr>
<tr>
<td>98.9</td>
<td>98.6</td>
<td>99.7</td>
<td>–</td>
<td>90.8</td>
<td>90.4</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>98.2</td>
<td>98.2</td>
<td>100</td>
<td>99.4</td>
<td>97.6</td>
<td>98.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(IC)</td>
<td>98.8</td>
<td>99.4</td>
<td>100</td>
<td>98.6</td>
<td>98.5</td>
<td>96.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70/72</td>
<td>98.2</td>
<td>98.2</td>
<td>100</td>
<td>98.2</td>
<td>98.8</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(IC)</td>
<td>98.8</td>
<td>99.4</td>
<td>100</td>
<td>98.6</td>
<td>98.8</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99.1</td>
<td>98.9</td>
<td>100</td>
<td>99.7</td>
<td>–</td>
<td>97.6</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>98.2</td>
<td>98.2</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(IC)</td>
<td>98.8</td>
<td>99.4</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99.1</td>
<td>98.9</td>
<td>100</td>
<td>99.7</td>
<td>–</td>
<td>100</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>98.2</td>
<td>98.2</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(IC)</td>
<td>98.8</td>
<td>99.4</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figures at the top right indicate results of pairwise comparison of the nucleotide sequences and figures at the bottom left indicate results of pairwise comparison of the deduced amino acid sequences. First raw corresponds to comparisons made with the VP1X2A region (168 nt). Second raw corresponds to comparisons made with the carboxy-terminal VP1 region (521 nt). Third raw, when available, corresponds to comparisons made with the VP3XVP1X2A region (1057 nt). Additional information: maximum homology of the whole capsid region (2457 nt) of strains belonging to the proposed IC subgenotype (isolated in a previous period to that of the present study) with the herein included prototype IA and IB strains is of 93%, below the 95% and 94% homology observed among IA strains or IB strains, respectively.
5. Discussion

Molecular characterization of HAV strains isolated from sporadic and outbreak-related cases in Catalonia during the period 2005–2009 was undertaken. Around 17% of the outbreak-related cases (103 out of 609) and 11% of the sporadic cases (54 out of 512) were analyzed. Overall 11.5% of the circulating isolates (129 isolates) were molecularly characterized.

Sixteen different strains were detected in the outbreaks, 2 in a shellfishborne outbreak, 9 in the MSM08–09 outbreak and 1 in each of the remaining outbreaks. While isolation of different strains in shellfishborne outbreaks may be expected, the isolation of so many strains in a MSM outbreak whose origin is likely to be from a single individual is less expected. Nevertheless it may be explained by the quasispecies distribution of HAV and the selective pressure imposed by a partial immune response in vaccinated HAV+ patients. Thirty-one different strains were detected from 54 sporadic cases. Three strains were detected in both sporadic and outbreaks cases. In total 48 different strains were isolated. Since only 11.5% of the circulating strains were molecularly characterized an actual higher diversity is expected to occur. Circulation of mixtures of strains is very common in non-endemic areas, while circulation of very few or even single strains is the rule in endemic areas.

Four out of the seven outbreaks were caused by travel-imported strains. Importation of HAV strains has also been very recently described in Germany. Most travel-related strains belong to subgenotype IB and originated in Africa, with the exception of one subgenotype IA, whose origin was Colombia. On the contrary, autochthonous strains belong to subgenotypes IA, IB and IIA. No subgenotype IIA strains were detected in contrast to the recently described isolation in neighboring France.

The critical role of shellfish global trade in the introduction of HAV in Spain has very recently been addressed. The shellfishborne strains could not be clearly classified as subgenotype IA or IB and clustered together with previously isolated ones also imported through shellfish from Peru. This cluster was previously identified as belonging to subgenotype IB, yet a more exhaustive analysis suggests the occurrence of a third subgenotype I, i.e. IC, probably derived, by the quasispecies dynamics, from subgenotype IA, the most common in South America. At this moment this candidate new subgenotype IC is already circulating both in Peru and Spain.

While massive vaccination campaigns are compulsory implemented in Catalonia, at the age of 12 since 1999, and most inhabitants older than 50 were naturally immunized in their childhood, most of the age group 24–50 is naive to the virus. Thus vaccination of MSM individuals and travelers to endemic countries among this group must be recommended. However, hepatitis outbreaks due to the consumption of contaminated food are harder to prevent without adoption of regulatory measures of virus detection in food or vaccination of this age group.

Competing interests

None.

Ethical approval

Not required.

Acknowledgments

This work was supported in part by grant BIO2008–01312 (Ministry of Science and Innovation, Spain; http://web.micinn.es/), 2009 SGR 24 and Xarxa de Referència de Biotecnologia de Catalunya (Generalitat de Catalunya; www.gencat.cat/agaur). The Enteric Virus Laboratory participates in the European Network on hepatitis A.

References