Phylogenetic Characterization of Genotype 4 Hepatitis C Virus Isolates from Argentina

VICTORIA ALFONSO,1 DIEGO FLICHMAN,1 SILVIA SOOKOIAN,2 VIVIANA ANDREA MBAYED,1,* AND RODOLFO HECTOR CAMPOS1

Cátedra de Virología, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires,1 and Unidad de Hepatología, Hospital Cosme Argerich,2 Buenos Aires, Argentina

Received 16 October 2000/Returned for modification 27 December 2000/Accepted 23 February 2001

Among 114 patients infected with hepatitis C virus, three genotype 4 isolates, unusual in Argentina, were detected by phylogenetic analysis over different genomic regions. The patients were not related. One sample was associated with Egyptian sequences, and the others were associated with a Zairean isolate, a fact which reinforces the idea that they are from independent sources.

Hepatitis C virus (HCV) is the major causative agent of chronic liver disease in developed countries. It belongs to the Flaviviridae family and has a positive-sense RNA genome and high mutation rate of its own. Different isolates show substantial nucleotide sequence variability distributed throughout the viral genome. This variability gives us the opportunity to classify the virus in six phylogenetically distinct groups, which have been previously denominated genotypes but are now described as clades 1 to 6 for standarization (19). Clades 1, 2, 4, and 5 may correspond to genotypes 1, 2, 4, and 5, while clades 3 and 6 may comprise the seven remaining genotypes.

It has been observed that the genotypes have a particular geographic distribution. Some seem to have spread worldwide (genotypes 1a, 1b, 2a, and 2c), while others have been found in very specific regions only (genotypes 5a, 6a, and 4).

Genotyping is a marker frequently required for the study and follow-up of patients chronically infected with HCV. Therefore, for regular monitoring, the HCV genotypes of isolates from 114 infected patients were determined by using the conventional methods of the line probe assay (24) and/or restriction fragment length polymorphism analysis (25). These procedures were carried out when the patients started their antiviral treatments. The ensuing general epidemiological pattern (genotype 1, 61.4% of isolates; genotype 2, 12.3%; genotype 3, 23.7%; genotype 4, 2.6%) squared with the prevalence previously reported in Argentina (7, 13, 16). Yet, it should be noted that three samples, which were denominated AR45, AR46, and AR47, turned out to be genotype 4. Genotype 4 has been reported to be the most prevalent genotype in Central Africa, the Middle East, and Egypt but not in Western countries. Surprisingly, there is an unexpectedly high prevalence of genotype 4 in southern Spain (20) among intravenous drug users (IDUs).

In Argentina, several studies have ruled out the presence of this genotype (7, 14, 17). Only one study, which enrolled hemophilic patients and used restriction fragment length polymorphism analysis for genotyping, has shown a small proportion of patients coinfected with multiple HCVs of different genotypes, including genotypes 4 and 5 (16).

There has been some argumentation over the most useful way to identify and assign genotypes of HCV, but now there is a consensus that the most suitable procedure involves comparisons based on phylogenetic analyses of nucleotide sequences of at least two coding regions (19).

The present phylogenetic study was performed for clear classification and also to allow a direct comparison of the evolutionary relationships of the three isolates that correspond to an unusual genotype in Argentina. Different genomic regions of these isolates were amplified and sequenced: NS5b, the core, and also the 5′ untranslated region (5′UTR). Viral RNA was reverse transcribed and amplified with the primers and parameters indicated in Table 1, according to Kwok and Higuchi’s recommendations (10). Amplified DNA was purified and directly sequenced with a Thermo Sequenase radiolabeled terminator cycle sequencing kit (USB Corporation, Cleveland, Ohio) along both chains. Confirmatory sequencing was performed on reamplified DNA. The sequences obtained and sequences of representatives of the six HCV clades which were previously employed for standardization of genotyping (19) were used to reconstruct the phylogenetic trees with programs from the PHYLIP package (Neighbor and Seqboot) (6) and fastDNAml (5, 12).

The different methods, which were applied to the different genomic regions, yielded congruent topologies (all of them assigned the three isolates to clade 4). The best clustering and robustness of the trees were achieved when we used the complete core region (positions 1 to 585) and a fragment of 304 nucleotides (nt) of NS5b (positions 7955 to 8259) (Fig. 1). Nucleotide positions were numbered according to the work of Choo et al. (3).

Although the 5′UTR is the most conserved region in the HCV genome, it was possible to identify many of the different genotypes from phylogenetic analysis with this genomic fragment (positions −251 to −72). However, clade 6 was not clearly defined, and the general robustness of the groups was
lower than that obtained with the NS5b or core fragment (not shown).

Besides the variety of HCV genotypes, there is considerable diversity among isolates with the same genotype. In genotype 4 several subtypes have already been identified (1, 11, 13), even in the same geographic region where this virus type predominates (8, 18, 23).

Further analysis was done to delve into the evolutionary relationships among the Argentinian isolates themselves and with other previously described genotype 4 viruses from different geographical areas. It has been taken into account that not all the sequences of genotype 4 available at present comprise both the core and NS5b regions. Therefore, phylogenetic comparison included, on the one hand, some sequences which

![FIG. 1. Maximum-likelihood analysis (fastDNAml) of representative HCV sequences (designation [GenBank accession numbers]: ED43 [Y11604], BEBE1 [D50469], EUH1480 [Y13184], EUHK2 [Y12083], HC-G9 [D14853], HC-J6 [D00944], HC-J8 [D10988], HCV-1 [M62321], HCV-J [D90208], Jk046 [D63822], HCV-Tr [E10839], JK049 [D63821], NZL1 [D17763]) and the three isolates reported in this work (AR45, AR46, and AR47). The bootstrap value of neighbor-joining congruent trees is indicated for each genetic group. The branch lengths of the phylogenetic tree are proportional to the genetic distances. (A) Phylogenetic tree obtained by using the complete core region (positions 1 to 585); (B) phylogenetic tree obtained by using a fragment of 304 nt of NS5b (positions 7955 to 8259).]

### TABLE 1. Reverse transcription and PCR amplification parameters and primers

<table>
<thead>
<tr>
<th>Region</th>
<th>Reactiona</th>
<th>Primer</th>
<th>Primer reference</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS5b</td>
<td>RT</td>
<td>Random primers</td>
<td>Modified from reference 4</td>
<td>90 min at 37°C</td>
</tr>
<tr>
<td>PCR</td>
<td>5’ CTGTCATAGCCTCCTGGA 3’, positions 8294 to 8274; 5’ ATGATAAGGGTCTTTGAC 3’, positions 7916 to 7935</td>
<td>5 min at 94°C; 30 cycles of 1 min at 94°C, 1 min at 50°C, and 2 min at 72°C; and a final extension of 10 min at 72°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td>RT</td>
<td>Random primers</td>
<td>This paper</td>
<td>90 min at 37°C</td>
</tr>
<tr>
<td></td>
<td>First-round PCR</td>
<td>5’ CTCCCGAAGCGCGGACG 3’, positions 696 to 679; 5’ CGAAAGGCCTGTTGACTG 3’, positions –70 to –51</td>
<td>The same as for NS5b PCR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Second-round PCR</td>
<td>5’ ATGCTTAGTGGACAGCTCG 3’, positions 635 to 615; 5’ TTGTGCTACTGCTGATAGGT 3’, positions –61 to –40</td>
<td>The same as for NS5b PCR</td>
<td></td>
</tr>
<tr>
<td>5’ UT</td>
<td>RT</td>
<td>Random primers</td>
<td>Modified from reference 24</td>
<td>90 min at 37°C</td>
</tr>
<tr>
<td>PCR</td>
<td>5’ GGTGACGCTTACGAGACCT 3’, positions –1 to –21; 5’ CTGTGAGAACACTGCTTCAAGC 3’, positions –298 to –273</td>
<td>5 min at 94°C; 40 cycles of 30 s at 94°C, 30 s at 55°C, and 45 s at 72°C; and a final extension of 7 min at 72°C</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a RT, reverse transcription.
comprised a common fragment of 222 nt of the NS5b region (positions 7975 to 8196) and, on the other hand, those isolates that shared a 210-nt sequenced fragment of the core gene (positions 39 to 248) (Fig. 2). The analysis of the NS5b fragment indicated that isolate AR45 was related to the Egyptian sequences EG13, EG19 (22), and ED43 (2). The association with Egyptian samples was also observed when the core fragments were compared, but in this case the available Egyptian sequences were EG21, EG29, EG33 (21) (collected in the same survey in which EG13 and EG19 were collected), and ED43.

As for isolate AR46, it was closely related to sample DK13 from Zaire (1) when the analysis was performed with the core fragment. This published sequence, described as subtype 4b, did not comprise an NS5b region and no related isolate was found previously (23). So sample AR46 appeared in the NS5b phylogenetic tree associated only with another Argentinian sample, AR47. The core fragment of this last sample could not be amplified.

These results suggest that at least two sources of infection were detected in the three Argentinian samples. One of them involved sample AR45. Samples AR46 and AR47 were genetically related, so a common source could not be ruled out in this case.

The patients, who attended the same hospital, were not related. Even though the majority of HCV genotype 4-harboring individuals in Spain were IDUs, our three patients did not display a common route of infection. Only one of them was an IDU, while a second patient was a health care worker. The third patient had only a tattoo as a risk factor. Moreover, the three patients denied having traveled to areas of endemicity. These facts reinforce the hypothesis that not all of the isolates had a common source.

It should be observed that the virus did not appear to be restricted to any specific population (such as the IDUs in Spain). In addition to this, Sánchez Quijano et al. (20) reported that most of the genotype 4 infections were detected in patients with troublesome genotyping and were exposed when a special procedure was applied. Therefore, the prevalence of genotype 4 in Argentina might be higher than expected.

In conclusion, three independent cases of HCV clade 4 infections were reported in Argentina, a country where this genetic group is infrequent. A phylogenetic characterization of these isolates was carried out based on the core, NS5b, and 5'UTR subgenomic regions. Genetic characteristics, reinforced by epidemiological aspects, indicated the existence of at least two sources of infection in the three cases. The coexistence of diverse subtypes of this genotype is described for regions which display a high prevalence of genotype 4 infections. Argentina has had a low prevalence of this HCV genotype up to now. Future detection of genotype 4 in this area may deserve special attention.

**Nucleotide sequence accession numbers.** The GenBank accession numbers for the sequences reported in this work are AF308573 to AF308580.

This study was supported by grants from the Universidad de Buenos Aires (SECyT-UBA, TB14), the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, PIP723/98), the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT, PICT 97 01610), and the Ministerio de Salud Pública de la Nación (Beca Carrillo-Oñativia 2000).

**REFERENCES**


4. Enomoto, N., A. Takada, T. Nakao, and T. Date. 1990. There are two major
1021–1025.
Department of Genetics, University of Washington, Seattle.
Avagnina, D. Benbassat, J. Rey, and M. Nakatsuno. 1999. Distribution of
the genotypes of hepatitis C virus in intravenous drug addicts in Argentina.
Medicina 59:49–54.
8. Fretz, C., D. Jeannel, L. Stuyver, V. Herve, F. Lunel, A. Boudifa, C. Mathiot,
G. de The, and J. J. Fournel. 1995. HCV infection in a rural population of
the Central Africa Republic (CAR); evidence for three additional subtypes of
Investigation of the pattern of hepatitis C virus sequence diversity in differ-
ent geographical regions: implications for virus classification. The Interna-
a tool for construction of phylogenetic trees of DNA sequences using max-
Duval, and D. Dhumeaux. 1995. Relationship between hepatitis C virus
genotypes and sources of infection in patients with chronic hepatitis C. J.
1995. Hepatitis C (HCV) genotype and viral titer distribution among Ar-
gentinian hemophiliac patients in the presence or absence of human immu-
17. Querleri, J. F., B. H. Robertson, V. Mathet, S. D. Sinha, I. Badia, B. Frider,
Genomic and phylogenetic analysis of hepatitis C virus strains from Argen-
epidemiology of hepatitis C virus throughout Egypt. J. Infect. Dis. 182:698–
707.
Gojobori, G. Maertens, M. Mikozami, O. Nainan, S. Netesov, K. Nishioka,
T. Shin-i, P. Simmonds, D. Smith, L. Stuyver, and A. Weiner. 1998. Classi-
fication, nomenclature, and database development for hepatitis C virus
(HCV) and related viruses: proposals for standardization. Arch. Virol. 143:
2493–2503.
Leal, J. Macias, and E. Lissen. 1997. Unexpected high prevalence of hep-
21. Simmonds, P., F. McOmish, P. L. Yap, S. W. Chan, C. K. Lin, G. Dusheiko,
A. A. Saeed, and E. C. Holmes. 1993. Sequence variability in the 5' non-
coding region of hepatitis C virus: identification of a new virus type and
22. Simmonds, P., E. C. Holmes, T. A. Cha, S. W. Chan, F. McOmish, B. Irvine,
E. Beall, P. L. Yap, J. Kolberg, and M. S. Urdea. 1993. Classification of
hepatitis C virus into six major genotypes and a series of subtypes by phy-
Maertens. 1994. Classification of hepatitis C viruses based on phylogenetic
analysis of the envelope 1 and nonstructural 5B regions and identification of
ment of a simple restriction fragment length polymorphism (RFLP) based
assay for HCV genotyping and comparative analysis with genotyping and