Analysis of hepatitis C virus core encoding sequences in chronically infected patients reveals mutability, predominance, genetic history and potential impact on therapy of Cuban genotype 1b isolates


Abstract. – Background and Objectives: Hepatitis C virus (HCV) genotypes are relevant to epidemiological questions, vaccine development, and clinical management of chronic HCV infection. In the present work, we aimed at investigating HCV genotype, variability and genetic history of HCV isolates in Cuba from a sample of chronically infected patients.

Material and Methods: A prospective study, involving 73 Cuban anti-HCV positive patients, was carried out. RT-PCR and phylogenetic analysis was employed to determine HCV genotypes. Divergence dates and demographic parameters in a Bayesian coalescent framework were estimated, as implemented in BEAST v1.4.8.

Results: HCV RNA was undetectable in 15 patients that received antiviral therapy. All HCV RNA positive patients, 58, were infected with genotype 1, three of them with subtype 1a and 55 with subtype 1b. The analysis of the DNA sequence coding for a core fragment, spanning nt positions 435-816 (relative to strain H77), revealed high percent (96.7% ± 0.8%) nucleotide identity within Cuban HCV subtype 1b sequences. However, 56.7% and 20% of 30 analyzed individuals had changes in the core region in a six-month interval, at the nucleotide and amino acid level, respectively. Mutations involving aa changes were mainly found in the region encompassed between aa 70 and 106 of the core protein, with only one isolate showing a point mutation at position 43. Interestingly, some of the observed changes seem to be reversions and might in fact contribute to reducing the variability of this region. The estimated date for the most recent common ancestor of HCV genotype 1b Cuban isolates is 1969 (CI, 1953 to 1977).

Discussion: Analysis of HCV core encoding sequences from chronic patients reveals mutability of genotype 1b isolates in Cuba, which seem to be predominant and rapidly multiplied during the eighty decade of last century, and might limit the benefits obtained from current antiviral therapy.

Key Words: HCV, Genotype, Core, Genetic diversity, Epidemic, Transmission.

Introduction

Hepatitis C virus (HCV) is an enveloped, single strand RNA virus, belonging to the family Flaviviridae, genus Hepacivirus. More than 180 million people are infected with HCV worldwide, with risk of developing cirrhosis and hepatocellular carcinoma. There is no available vaccine against HCV and current treatments, based on interferon plus ribavirin, are expensive, frequently caused undesired adverse events and are effective in only approximately half of patients treated.

The viral genome, approximately 9.8 Kb long, encodes for a polyprotein, co- and posttranslationally processed in at least 10 viral proteins. Sequence comparisons of HCV variants recovered from different individuals and geographical regions have led to the identification and classification of at least six major genotypes, differing in up to 30-35 % of nucleotide sites over the complete genome. Each of them contains a series of more closely related subtypes that typically differ from each other by 20-25 % in nucleotide
sequences. Individual isolates of a given subtype typically differ by about 8%-10%, and, as HCV replicates as quasispecies, multiple variants exist even within individual patients.

HCV subtypes present distinct geographical distributions. Genotypes 1, 2 and 3 have a worldwide distribution. HCV genotype 4 is found in the Middle East and North Africa; genotype 5 has been found in South Africa, France and Belgium and genotype 6 is found mainly in Southeast Asia. Genotype 1 predominates in America, though there are differences in genotypes distribution through the continent. A predominance of HCV subtype 1a has been described in the Dominican Republic and Puerto Rico. Otherwise, a study carried out in Martinique, showed a prevalence of subtype 1b. In addition, previous studies reported a frequent detection of genotype 1b in Cuba. Nevertheless, data on HCV genotypes distribution in the Caribbean region are still limited and little is known about HCV origin and spread in this area.

Genotype of HCV is one critical factor for predicting response to standard antiviral therapy since genotype 1 and 4 respond worst, even after longer administration regimens, than genotypes 2 and 3. Therefore, different methods have been developed for determination of HCV genotypes and several HCV regions have been targeted with this purpose. However, the “gold standard” for HCV genotyping is direct sequencing of amplified Polymerase chain reaction (PCR) products followed by phylogenetic analysis from clinical material. Other methods include amplification of HCV RNA followed by hybridization with type-specific probes or digestion withendonucleases and restriction Fragment Length Polymorphism (RFLP) analysis or re-amplification with type-specific primers.

HCV core region have been shown useful for reliable identification of viral genotyping and subtyping since the sequence divergence is intermediate between 5’non-coding region and envelope proteins. The study of sequence diversity in HCV core region has additional value since this viral protein has been related with several elements potentially involved in HCV pathogenesis and persistence.

Here, we describe a molecular epidemiological analysis of HCV strains from a cohort of anti-HCV positive Cuban patients, based on the study of the DNA sequence coding for a HCV core fragment. We use coalescence-based population genetic methods in order to estimate the origin and dynamics of HCV genotype 1b strains. This is the first characterization of the genetic diversity and evolution of HCV in Cuba. The results shed new light on origin and history transmission of the HCV in the Caribbean region, as well as relative mutability of HCV core encoding region, with potential impact on standard antiviral treatment’s efficacy.

Material and Methods

Demographic Data of Patients
We included in this study 73 anti-HCV positive Cuban individuals that received therapy with interferon (IFN) plus ribavirin or were naïve to treatment. Informed consent was obtained from each patient, and ethical approval was granted by the Ethics Committee of the National Institute of Gastroenterology (La Habana, Cuba). Table I summarizes demographic data of individuals involved in the study.

RNA Extraction and RT-PCR
Plasma or serum specimens were obtained from each patient and stored at -80°C until its use. Table I. Demographic data of HCV infected individuals included in the study.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Patients (N = 73)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>39 (54.9)</td>
</tr>
<tr>
<td>Male</td>
<td>32 (45.1)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>47 ± 11</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>48 (39-55)</td>
</tr>
<tr>
<td>Skin color (%)</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>58 (82.9)</td>
</tr>
<tr>
<td>Black</td>
<td>5 (7.1)</td>
</tr>
<tr>
<td>Mixed</td>
<td>7 (10.0)</td>
</tr>
<tr>
<td>Potential route of infection (%)</td>
<td></td>
</tr>
<tr>
<td>Transfusion</td>
<td>37 (52.1)</td>
</tr>
<tr>
<td>Medical procedures</td>
<td>20 (28.2)</td>
</tr>
<tr>
<td>Frequent needle use</td>
<td>4 (5.6)</td>
</tr>
<tr>
<td>Unknown</td>
<td>10 (14.1)</td>
</tr>
<tr>
<td>Antiviral treatment (%)</td>
<td></td>
</tr>
<tr>
<td>Interferon/ribavirin</td>
<td>40 (54.8)</td>
</tr>
<tr>
<td>Naive</td>
<td>33 (45.2)</td>
</tr>
<tr>
<td>Genotype (%)</td>
<td></td>
</tr>
<tr>
<td>1a</td>
<td>3</td>
</tr>
<tr>
<td>1b</td>
<td>55</td>
</tr>
</tbody>
</table>

IQR: Interquartile range. *Information available for 71 patients. †Information available for 70 patients.
use. Paired samples were available from 30 patients within a period of six month. Viral RNA was extracted using QIAamp Viral RNA Mini Kit (QIAGEN, Valencia, CA, USA). The amplification was carried out using the One-step RT-PCR kit (QIAGEN). Primers 242 (5′-GTTCCCCGGCGGTGTCAGATCGTT-3′) and 243 (5′-CCTGTTGCTATAATTCACGGCGTCTC-3′) were designed to amplify a 432 bp DNA fragment of the core gene, spanning nt positions 435-816 (relative to strain H77). RT-PCR was carried out on 10 µl of extracted RNA (reaction mixture volume, 50 µl) and 0.6 µM of each primer.

Termocycling conditions consisted of 1 cycle-50ºC/30 min; followed by 1 cycle-95ºC/15 min and a touchdown PCR of 38 cycles -95ºC/30 s; 60ºC/30 s; 72ºC/30 s and a final extension cycle (72ºC/10 min). The annealing temperature was decreased during the first 5 cycles from 60ºC to 55ºC with a ramping rate of 1ºC per cycle. Saline solutions and sera negative for anti HCV-antibody served as negative controls. PCR products were sequenced bidirectionally in a specialized commercial laboratory (Macrogen Inc, Seoul, South Korea).

List of Reference Sequences

Reference sequences used for comparison with those obtained in this study were imputed from GenBank and presented here along with their accession number:

**Type 1:** AF009606, AF011751, AF271632, AB049088, AF139594, AF356827, D90208, M74815, U10204, U10205, AY051292, D14853;

**Type 2:** AB047639, AB047641, AF169004, D00944, AB030907, AF238486, AY232730, D10988, D50409, D14853, AB031663;

**Type 3:** AF046866, D17763, D28917, D49374, D63821;

**Type 4:** Y11604;

**Type 5:** AF064490, Y13184;

**Type 6:** Y12083, D84262, D84263, D63822, D84265, D84266.

Nucleotide Sequences Accession Numbers

New sequences obtained in the study have been submitted to GenBank and have been assigned accession number: AM494544, AM494545, AM494546, AM494547, AM494548, AM494549, AM494550, AM494551, AM494552, AM494553, AM494554, AM494555, AM494556, AM494557, AM494558, AM494559, AM494560, AM494561, AM494562, AM494563, AM494564, AM494565, AM494566, AM494567, AM494568, AM494569, AM494570, AM494571, AM494572, AM494573, AM494574, AM494575, AM494576, AM494577, AM494578, AM494579, AM494580, AM494581, AM494583, AM500641, AM500642, AM500643, GQ244326, GQ244327, GQ244328, GQ244329, GQ244330, GQ244331, GQ292447, GQ292448, GQ292450, GQ292451, CG247874, CG247875, CG247876, CG247877.

Phylogenetic Analysis

Nucleotide sequences were aligned using ClustalW, v1.8122. ModelGenerator, v0.8523 was used to identify the optimal evolutionary model that described our sequence dataset (Table II). The model with the smallest Akaike information criterion (AIC) score was used in subsequent analysis. Phylogenies were heuristically searched using NNI and SPR perturbation algorithms with the PhyML program24 available at http://mobyle.pasteur.fr/cgi-bin/portal.py. The robustness of the tree topology was assessed by bootstrap analysis with 1000 replicates. Trees were visualized with Megan program, v3.1.

Table II. Akaike information criterion (AIC) according to the substitution model applied on 58 partial core sequences of HCV genotype 1b from Cuba.

<table>
<thead>
<tr>
<th>Substitution model</th>
<th>Basal</th>
<th>Invariant site (+I)</th>
<th>Gamma-distributed rate variation (+G)</th>
<th>+I +G</th>
</tr>
</thead>
<tbody>
<tr>
<td>JC</td>
<td>15425</td>
<td>14334</td>
<td>13773</td>
<td>13745</td>
</tr>
<tr>
<td>K80</td>
<td>14815</td>
<td>13706</td>
<td>13093</td>
<td>13063</td>
</tr>
<tr>
<td>F81</td>
<td>15382</td>
<td>14282</td>
<td>13700</td>
<td>13674</td>
</tr>
<tr>
<td>HKY</td>
<td>14730</td>
<td>13615</td>
<td>13056</td>
<td>13028</td>
</tr>
<tr>
<td>TN</td>
<td>14715</td>
<td>13602</td>
<td>13058</td>
<td>13030</td>
</tr>
<tr>
<td>GTR</td>
<td>14647</td>
<td>13541</td>
<td>13024</td>
<td>12999</td>
</tr>
</tbody>
</table>
Coalescent-based Inference of HCV Population Dynamics

The past population dynamic of 1b HCV strains was estimated from the HCV core sequences using the BSP as implemented in program BEAST, v1.4.8. We obtained an external estimate of the evolutionary rate of the core fragment from independent, previously published sequences. A relaxed molecular clock approach was used (uncorrelated lognormal). The Markov Chain Monte Carlo analyses were run for \(70 \times 10^6\) states and sampled every 1000 states. BEAST output files were analyzed using Tracer 1.4.1 (http://tree.bio.ed.ac.uk/software/tracer/).

Results

HCV Genotyping

HCV RNA was detected in 58 out of 73 anti-HCV positive samples (79.4%). HCV RNA was undetectable in 15 patients that received antiviral therapy. Samples obtained from the 58 patients positive for HCV RNA were used for determination of viral genotype. No deletion or insertion was found in the amplified core region of the isolates tested. Figure 1 shows the distribution of the Cuban HCV isolates along with reference sequences. All Cuban HCV isolates clustered with genotype 1 strains. Three isolates clustered with sequences of subtype 1a (5.2%) and fifty-five isolates with subtype 1b (94.8%) (Figure 1). HCV 1b sequences were submitted to Blast program in order to find the most related sequences available in GenBank. For each sequence, the 200 closest sequences were downloaded. Only one sequence from each infected individual was retained, obtaining a subset of 131 sequences corresponding to the most related ones to the Cuban HCV isolates. These sequences were mainly from United States (U.S.) and Europe, although a few sequences were from Japan, China and Korea.

Analysis of Cuban HCV Subtype 1b Core Sequences

The nucleotide identity within Cuban HCV subtype 1b core sequences was 96.7% ± 0.8% (range 94.5 and 99.7%). Comparison with consensus sequence revealed that the change most frequently found at the aa level (in 17.2% of isolates), was V147A. In addition, two isolates had a change (R101H) in the DNA-binding motif.

---

Figure 1. Distribution of the Cuban isolates within the HCV prototype sequences from all six genotypes. Cuban HCV isolates are represented in white diamonds; whereas the reference strains are depicted with their accession number. The bootstrap value is indicated for each genetic group. Notice that all Cuban sequences were allocated to genotype 1.
SPRG (pos 99-102); three isolates had amino acid changes at the nuclear localization signal (pos 38-43), two of them with a non-conservative change R43G and one isolate with a conservative change R43K. The nuclear localization signal encompassed in the region aa 58-64, five tryptophan residues (in the region encompassing aa 76-107) and two phosphorylation sites (S99 and S116), were perfectly identical in all isolates.

On the other hand, thirty patients with detectable HCV RNA could be contacted for a second blood extraction 6 months later, being all samples also positive for this marker of infection. Comparison of sequences obtained six months apart from each patient revealed that there were nucleotide changes in 17 out of 30 individuals during this time (56.7%). Analysis of the deduced amino acid sequences revealed mutations in the analyzed region in six patients (20%) within the period of 6 months. Table III shows the position and mutations detected in each isolate. Mutations involving aa changes were mainly found in the region encompassed between aa 70 and 106 of the core protein, with only one isolate showing a point mutation at position 43.

**Estimation of Epidemic History**

In order to investigate the origin and spread of HCV subtype 1b in Cuba, we estimated divergence dates and demographic parameters in a Bayesian coalescent framework, as implemented in BEAST v1.4.8. The Bayesian skyline plots depict the estimated change in the effective number of infected individuals through time and are shown in Figure 2. The skyline plot shows that in Cuba there was a rapid, exponential spread of HCV during the last 5 years of the eighty-decade. The rate of growth appears to slow at the nineties to the present day. The estimated date for the most recent common ancestor (MRCA) is 1969 (CI, 1953 to 1977).

**Discussion**

In the present study, a cohort of 73 anti-HCV positive Cuban patients was recruited over 3 years. The patients investigated were epidemiologically diverse, showing a wide range of ages and routes of HCV infection. Blood transfusions and medical invasive procedures were the main potential modes of HCV infection in the analyzed population.

Several studies have already demonstrated that core gene shows sufficient variability to distinguish between HCV isolates and to establish phylogenetic relationships. Here, the phylogenetic analysis of the core region was able to classify the genotype in the 100% of the HCV isolates. All Cuban isolates tested from patients having detectable HCV RNA were classified as genotype 1.

HCV genotype 1 predominates in America and specifically the subtype 1a prevails in some Caribbean islands. Nevertheless, the genetic diversity of HCV in Cuba seems to be different to the diversity previously found in the Caribbean region. In this study, HCV subtype 1b was predominant and subtype 1a showed a low prevalence. A similar distribution of subtypes 1a and

<table>
<thead>
<tr>
<th>Cuban HCV isolates with mutations</th>
<th>Mutations</th>
<th>Amino acid position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu17</td>
<td>R instead of G</td>
<td>43</td>
</tr>
<tr>
<td>Cu22</td>
<td>Q instead of R</td>
<td>70</td>
</tr>
<tr>
<td>Cu34</td>
<td>A instead of T</td>
<td>75</td>
</tr>
<tr>
<td>Cu35</td>
<td>H instead of R</td>
<td>70</td>
</tr>
<tr>
<td>Cu38</td>
<td>H instead of P</td>
<td>71</td>
</tr>
<tr>
<td>Cu42</td>
<td>M instead of L</td>
<td>91</td>
</tr>
<tr>
<td>Total</td>
<td>S instead of T</td>
<td>75</td>
</tr>
<tr>
<td>Total</td>
<td>M instead of L</td>
<td>91</td>
</tr>
<tr>
<td>Total</td>
<td>N instead of S</td>
<td>106</td>
</tr>
</tbody>
</table>

Table III. Mutations detected in the core protein of Cuban HCV isolates in a six months time period.
1b were found in two previous studies with Cuban patients\textsuperscript{12,13}. However, in those studies mixed infections were detected, as well as genotypes 2 and 3, although at low frequency. We found no evidence of mixed infection, possibly as a consequence of the method used for genotyping, given that the sequences of the predominant variant are more likely to be obtained by direct sequencing of the PCR products. In our study, genotype 1 was the only genotype found. It should be taken into account that more than half patients involved in the study were non-responders to antiviral treatment and it is established that genotype 1 is considered a difficult HCV genotype to treat with IFN-based therapies\textsuperscript{2,31}. Moreover, the majority of those patients had received blood transfusions or blood-derived products, a route of infection sometimes associated with the transmission of this HCV genotype\textsuperscript{32,33}.

When compared with other published isolates, Cuban HCV sequences showed highest similarity with isolates from U.S. and Western Europe. The resemblance of Cuban HCV isolates to the North-American and European strains is not surprising given the close historical/geographical relationship between Cuba and these countries. Surprisingly, sequences from Latin America were not found to be similar to the Cuban ones, probably because there were few sequences published from this geographical area spanning the region of the core gene used for the analysis. This result suggests that HCV could have been introduced into Cuba from Europe or U.S.

The percent nucleotide identity within Cuban HCV subtype 1b isolates studied was high and similar to previous data obtained for HCV-chronic patients from other countries\textsuperscript{19}. This is in agreement with the relatively high level of sequence homology described for the core antigen between HCV isolates\textsuperscript{34}. However, the analysis of the HCV core sequences in the same individual in two times, six months apart, suggested that HCV core might not be as constant as previously thought. Point changes were frequently found in the patients during this period of time. At the aa level, mutations were mainly found in the region encompassed between aa 70 and 106 of the core protein. According to previous data, the identity of amino acid 70 of the HCV core region affected the sensitivity to IFN since Japanese HCV genotype 1b patients with glutamine at amino acid 70 of HCV showed resistance to IFN\textsuperscript{35}. Substitution of aa 91 has been also described as a useful pretreatment predictor of response to IFN plus ribavirin therapy\textsuperscript{36}. Some epitopes specific for T lymphocytes of certain HLA have been described in the regions with point mutations\textsuperscript{37,38}, which might also suggest that some kind of immune pressure could be responsible for these changes. Interestingly, some of the observed changes seem to return the sequence to the consensus at the particular aa position, which are in fact probably reversions and contribute to reducing the variability. In this scenario, despite the change in some individuals, our results support the conservation of HCV core among viral isolates as a region under immune pressure but probably critical for persistence in the host.

Finally, the MRCA was shown to date from 1953 to 1977 for genotype 1b isolates. This is similar to those found by Nakano in China and by Di Lello in Argentina\textsuperscript{39,40}. This result indicate a relatively short circulating time for HCV genotype 1b in Cuba, even after Brazil and the United States, which comprise common ancestors that are considerably older, dating from 1880 to 1920\textsuperscript{41}. Nevertheless, this difference would result from the use of different genomic regions.

In conclusion, analysis of HCV core encoding sequences from chronic patients reveals mutability of genotype 1b isolates in Cuba, which seem to be predominant and rapidly multiplied during the eighty decade of last century, with a potential negative impact on overall standard antiviral therapy efficacy in the country.

Acknowledgements

We are grateful to Dr. Hugo Nodarse, Dr. Enrique Arús; Dr. Marlén Castellanos and Dr. Juan Morales for constructive discussions and identification of patients. This work was partially supported with a Grant from Pan-American Health Organization.

References


20) WAGGONER SN, HALL CH, HAYN HS. HCV core protein interaction with gC1q receptor inhibits Th1 differentiation of CD4+ T cells via suppression of dendritic cell IL-12 production. J Leukoc Biol 2007; 82: 1407-1419.


26) TANAKA Y, HANADA K, MIKOMA M, YEO AE, SSHA JW, GOOBOPI T, ALTER HO. A comparison of the molecular clock of hepatitis C virus in the United States and Japan predicts that hepatocellular carcinoma


