Rapid hepatitis C virus divergence among chronically infected individuals

Mayra Cruz-Rivera,a† Juan Carlos Carpio-Pedroza,b† Alejandro Escobar-Gutiérrez,b Daniela Lozano,c Arely Vergara-Castaneda,d Pilar Rivera-Osorio,b Armando Martinez-Guarneros,b Carlos A. Vazquez Chacon,b Salvador Fonseca-Coronado,e and Gilberto Vaughanb*

Facultad de Medicina, Universidad Nacional Autónoma de México, Mexico City, Mexicoa; Instituto de Diagnóstico y Referencia Epidemiológicos, Mexico City, Mexico;b; Departamento de Biomedicina Molecular, Centro de Investigación y de Estudios Avanzados, Instituto Politécnico Nacional, Mexico City, Mexico;c; Facultad de Ciencias Químicas, Universidad La Salle, Mexico City, Mexico;d; and Laboratorio de Inmunobiología de Enfermedades Infecciosas, Unidad de Investigación Multidisciplinaria, Facultad de Estudios Superiores Cuautitlán, Universidad Autónoma de México, Estado de México, Mexicoe

Address correspondence to Gilberto Vaughan, gilvaughan@yahoo.com

Departamento de Investigaciones Inmunológicas, Instituto de Diagnóstico y Referencia Epidemiológicos, Carpio 470, México City 11340, Mexico. Phone: + (52) (55) 5342-7563; Fax: + (52) (55) 5341-3264.

† These authors contributed equally to this work.

* Current adscription: Division of Viral Hepatitis, Centers for Disease Control and Prevention.

1600 Clifton Rd, Atlanta, GA 30329.
Here, we analyze the viral divergence among HCV chronic cases infected with genotype 1. The intrahost viral evolution was assessed by deep sequencing using the 454 Genome Sequencer platform. The results showed a rapid nucleotide sequence divergence. This notorious short-term viral evolution is of the utmost importance for the study of HCV transmission because direct links between related samples were virtually lost. Thus, rapid divergence of HCV significantly affect genetic relatedness studies and outbreak investigations.

Keywords: HCV, divergence, evolution.
Globally, Hepatitis C virus (HCV) infects approximately 130 million people, with three million new infections occurring annually (1). Before the arrival of the new direct-acting antivirals (DAAs), the combined therapy based on long-acting pegylated alpha interferon (IFN) and ribavirin (RBV) was successful in only approximately 50% of cases infected with HCV genotype 1 (26). Diverse predictive markers including virus genotype, viral load, complexity of viral population and viral genome sequence have been reported to have some predictive value. Single nucleotide polymorphisms (SNPs) located near the interleukin-28B (IL-28B) gene were found to be associated with the outcome of antiviral treatment, spontaneous viral clearance, and progression to chronicity, which make them useful as predictive factors for personalizing antiviral therapy (10).

HCV is a positive-polarity, single-stranded RNA virus that belongs to the genus *Hepacivirus* in the family *Flaviviridae* (15). The HCV polymerase is an RNA-dependent RNA polymerase lacking proof-reading and error-correction mechanisms. As a result, HCV replication, like that of many other RNA viruses, is highly error prone (19). The HCV mutation rate has been estimated to be $2.5 \times 10^5$ mutations per nucleotide per genome replication (23). This characteristic high mutation rate is consistent with the high degree of intrahost genetic diversity commonly observed among infected individuals (5, 7). The high degree of genetic variability allows rapid evolution that ultimately favors escape from immune and antiviral drug pressures (9, 24). This genetic variation also allows the virus to evolve and optimize its fitness in each host (positive selection) under the functional constraints on replication (22). Furthermore, the high progeny production (approximately $10^{12}$ virions per day) facilitates rapid viral evolution (17).

Different HCV subgenomic regions exhibit different degrees of variability (28). As a consequence, several regions of the viral genome have been used for molecular studies (5, 13,
However, the HCV hypervariable region 1 (HVR1) of the envelope (E) gene has been the region of choice to conduct relatedness studies due to its greater variability. Additionally, the HVR1 exhibits more rapid sequence changes over time, a feature that confounds the identification of related HCV cases. Thus, HCV viral divergence further complicates the association of cases to their source of infection. Therefore, accurate and reliable methods for the identification of HCV transmission are of critical importance.

Transmission networks are difficult to identify for several reasons, including the lack of laboratory tests capable of distinguishing acute from chronic infections and the typical asymptomatic nature of acute HCV infections. Moreover, the characteristically long incubation periods among symptomatic cases make it difficult to link related cases to a common source of infection.

In addition, molecular approaches required to assess the intrahost viral genetic variation are quite sophisticated, time-consuming and expensive, further complicating the identification of transmission events. The main issues associated with the correct identification of viral variants present in any given clinical specimen are the complexity of such populations and the difficulty in detecting low-frequency variants. However, the development of next generation sequencing (NGS) platforms has significantly improved the analysis of the intrahost viral populations, providing a snapshot of the entire virus population, which makes it a viable alternative to other conventional methodologies.

The goal of this work was to analyze the viral population divergence among chronic cases infected with HCV genotype 1, reporting the use of intravenous drugs. All patients were anti-HCV treatment-naïve. Subjects were epidemiologically unrelated. The Ethical Review and Informed Consent approval was granted by the Ethical Committee of the National Reference.
Laboratory. Informed consent was obtained from all subjects. The patients’ blood samples were collected by venipuncture into PPT Vacutainer® tubes (Becton Dickinson, Franklin Lakes, NJ) and processed according to the manufacturer’s recommendations. IL-28B genotyping was performed as reported previously (8). HCV genotyping was performed as described elsewhere (5).

Initially, HCV chronic cases bearing the non-CC IL-28B genotype were given the option to start conventional dual therapy with IFN and RBV or to postpone treatment until the arrival of the new protease inhibitors. Four individuals reporting the use of intravenous drugs opted to wait for the new therapy regime. These patients (A-D) were sampled twice within 18 months before the start of anti-HCV therapy (18, 14, 12 and 7 months for patients A, B, C and D respectively).

Three patients were infected with subgenotype 1a (A, B and C) and one with subgenotype 1b (D). Nine unrelated HCV cases, seven individuals infected HCV genotype 1a (C1, C2, C3, C5, C6, C8 and C9) and two with genotypes 1b (C4 and C7), were also enrolled in the study and were used as controls. The characteristics of all individuals are listed in Table 1.

All serum samples from patients and controls were subjected to ultra-deep pyrosequencing to analyze the degree of divergence as previously described (5, 7). A comprehensive nucleotide alignment, containing 358 HCV HVR1 representative sequences from subgenotypes 1a and 1b (140 and 218, respectively), was obtained from the Los Alamos National Laboratory HCV database (http://hcv.lanl.gov/content/sequence/NEWALIGN/align.html) and used as unrelated sequences. The program ModelTest was used to establish the best model of RNA substitution for the data. The general time reversible model was chosen to create maximum-likelihood (ML) trees as implemented in MEGA5 (25). Polymorphism and divergence analyses were conducted using DnaSP software v5 (16). After denoising the data, phylogenetic
analysis (maximum-likelihood) was conducted to assess the changes in the architecture of the viral population overtime (Figure 1A). The results showed a high degree of divergence among paired specimens from all four patients (minimal nucleotides distances 22, 17, 32 and 26 for patients A, B, C and D, respectively). Minimal distances among unrelated cases ranged between 17-29 nucleotides. Interhost genetic distances between patients A and B were closer. In contrast, patient C displayed the most noticeable divergence among all four of these subjects. The genetic distances between the viral populations from patients and controls belonging to subgenotype 1a did not show statistically significant differences. Finally, the two viral populations sampled from patient D remained within the same sequence space despite the significant sequence divergence observed at the two time points. In summary, all four subjects displayed a remarkable degree of sequence divergence. The phylogenetic and minimal distance analyses showed no significant differences between the paired samples and the non-related controls regardless of the infecting viral genotype.

Here, we have shown the rapid, short-term, sequence divergence among chronic cases infected with HCV genotype 1 bearing the non-CC IL-28B genotype reporting the use of intravenous drugs. Additionally, the results demonstrated that the usage of next generation platforms provides a comprehensive snapshot of the viral population in clinical samples. HCV evolution is a critical factor that in many ways dictates the outcome of the disease. The mechanisms of persistence, a key feature of the biology of the virus, are not fully understood. The high mutation rate, characteristic of HCV infection, typically leads to a continuous diversification of the viral population (6). As a consequence, HCV circulates in vivo as a complex mixture of viral variants closely related to each other with a dynamic distribution subjected to a continuous process of genetic variation, competition, and selection. This genomic
heterogeneity allows rapid adaptation to changes in the microenvironment when the viral population is subject to selective forces (host immune response, antiviral therapy, etc.). Thus, the enormous variability of HCV represents a great challenge for disease control.

The evolution of HVR1 among these individuals highlights some of the most important limitations encountered when dealing with outbreak investigations. The results clearly showed that the high degree of divergence makes the identification of related cases extremely challenging. The remarkable changes in the composition of the viral populations at different times observed among these patients could be the result of potential reinfection or super infection. Indeed, both phenomena have been known to occur in individuals reporting the use of drugs (11). Thus, one could expect that the dramatic changes in the architecture of the viral population may be the result of viral replacement by related strains rather than by molecular evolution. The close relatedness among strains can be explained because networks of IDU are "tight-knit" groups that can maintain certain viral lineages that might have an advantage (fitness) in this particular setting.

The use of minimal genetic distances has been shown to be an appropriate approach to address relatedness among HCV cases (5). However, the high degree of divergence observed among these cases in such short period of time prevents the correct assignment of cases to clusters of transmission because the minimal distances overlap with those obtained from unrelated cases. Alternatively, and as shown in Figure 1, a phylogeny relationship can be used to establish transmission links. Additionally, the use of other informative regions (i.e., the NS5A gene) or larger segments of the genome can help overcome the limitations of current approaches used in molecular epidemiological studies. Nevertheless, overcoming HCV molecular evolution is not an easy task. As seen with patient D, in some cases phylogeny is not sufficient to correctly
link cases. As a consequence, the time of identification plays an important role in the accurate identification of cases and their source of infection. Therefore, the opportune identification of cases in a timely manner will still play a pivotal role in the recognition of transmission networks.

All four patients displayed the TT IL-28B genotype. Others have suggested the participation of IL-28B genotype in the evolution of HCV (30). Whether the host IL-28B genotype is responsible for the rapid sequence divergence observed among these patients requires further research. One of the limitations of our study was the lack of CC IL-28B genotype HCV infected subjects. However, the ethical restrictions surrounding this issue prevent the inclusion of individuals with that particular IL-28B genotype as these patients are likely to respond to the antiviral treatment. Therefore, it is highly improbable to carry out prospective studies looking at the molecular evolution of HCV among CC patients. The second critical shortcoming of this study was the small number of individuals. It is important to mention that with the arrival of the DAAs, the success rates have been improved significantly, and therefore, keeping patients from being treated is not justifiable. In this particular study, the individuals were originally enrolled before the licensing of the new antiviral drugs. Thus, the possibilities of enrolling large number of individuals suitable for long prospective follow-up studies are rather limited.

Recently, a study has shown the advantages of using next generation sequencing platforms over conventional methods for the analysis of the HCV intrahost variability (7). Ultra-deep sequencing is a powerful technology that allows for the analysis of the viral population in great detail, providing a comprehensive picture of the architectural arrangement of the viral population. Several reports have described the usage of this methodology in the study of HCV intrahost viral evolution (3, 12, 18, 27, 29). Thus, usage of more advanced technology such as
amplicon deep sequencing is more appropriate for a thorough evaluation of the viral population in clinical samples (5, 14).

In conclusion, HCV can exhibit a high degree of sequence divergence. These findings have important implications for outbreak investigations and genetic relatedness studies as links to related cases can be lost over a short period of time. Additionally, the use of NGS technologies will most likely help to better understand the molecular basis of HCV evolution.
ACKNOWLEDGMENTS

This work was partially supported by Proyecto PAPIIT TA200112, Dirección General de Asuntos del Personal Academico, Universidad Nacional Autónoma de México.
REFERENCES


Hepatitis C virus transmission bottlenecks analyzed by deep sequencing. J Virol **84**:6218-6228.

2012. A single nucleotide polymorphism in IL28B affects viral evolution of hepatitis C 
quasispecies after pegylated interferon and ribavirin therapy. J Med Virol:DOI:  
10.1002/jmv.23407.
Figure Captions

FIG 1 Phylogenetic analysis. Phylogenetic analysis (maximum likelihood) was conducted using the HCV HVR1 sequences from all cases and controls. Sequences from the first (closed circles) and second (open circles) time points are depicted in colors. Sequences from controls are also color coded. Additionally, sequences from the Los Alamos database are shown in black.
Table 1. Patient’s characteristics

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Gender</th>
<th>Risk factor</th>
<th>IL-28B genotype</th>
<th>HCV genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-A</td>
<td>28</td>
<td>Male</td>
<td>IDU</td>
<td>TT</td>
<td>1a</td>
</tr>
<tr>
<td>P-B</td>
<td>32</td>
<td>Male</td>
<td>IDU</td>
<td>TT</td>
<td>1a</td>
</tr>
<tr>
<td>P-C</td>
<td>25</td>
<td>Male</td>
<td>IDU</td>
<td>TT</td>
<td>1a</td>
</tr>
<tr>
<td>P-D</td>
<td>31</td>
<td>Male</td>
<td>IDU</td>
<td>TT</td>
<td>1b</td>
</tr>
<tr>
<td>C-1</td>
<td>36</td>
<td>Male</td>
<td>IDU</td>
<td>CC</td>
<td>1a</td>
</tr>
<tr>
<td>C-2</td>
<td>38</td>
<td>Male</td>
<td>IDU</td>
<td>CT</td>
<td>1a</td>
</tr>
<tr>
<td>C-3</td>
<td>63</td>
<td>Male</td>
<td>Transfusion</td>
<td>TT</td>
<td>1a</td>
</tr>
<tr>
<td>C-4</td>
<td>41</td>
<td>Female</td>
<td>IDU</td>
<td>TT</td>
<td>1b</td>
</tr>
<tr>
<td>C-5</td>
<td>40</td>
<td>Male</td>
<td>IDU</td>
<td>TT</td>
<td>1a</td>
</tr>
<tr>
<td>C-6</td>
<td>61</td>
<td>Male</td>
<td>Transfusion</td>
<td>CC</td>
<td>1a</td>
</tr>
<tr>
<td>C-7</td>
<td>65</td>
<td>Female</td>
<td>Transfusion</td>
<td>CC</td>
<td>1b</td>
</tr>
<tr>
<td>C-8</td>
<td>69</td>
<td>Female</td>
<td>Transfusion</td>
<td>CC</td>
<td>1a</td>
</tr>
<tr>
<td>C-9</td>
<td>59</td>
<td>Male</td>
<td>Transfusion</td>
<td>TT</td>
<td>1a</td>
</tr>
</tbody>
</table>