New hepatitis C virus genotype 1 subtype naturally harbouring resistance-associated mutations to NS5A inhibitors

Laura Ordeig, Damir Garcia-Cehic, Josep Gregori, Maria Eugenia Soria, Leonardo Nieto-Aponte, Celia Perales, Meritxell Llorens, Qian Chen, Mar Riveiro-Barciela, Maria Buti, Rafael Esteban, Juan Ignacio Esteban, Francisco Rodriguez-Frias and Josep Quer

Abstract

Hepatitis C virus (HCV) is a highly divergent virus currently classified into seven major genotypes and 86 subtypes (ICTV, June 2017), which can have differing responses to therapy. Accurate genotyping/subtyping using high-resolution HCV subtyping enables confident subtype identification, identifies mixed infections and allows detection of new subtypes. During routine genotyping/subtyping, one sample from an Equatorial Guinea patient could not be classified into any of the subtypes. The complete genomic sequence was compared to reference sequences by phylogenetic and sliding window analysis. Resistance-associated substitutions (RASs) were assessed by deep sequencing. The unclassified HCV genome did not belong to any of the existing genotype 1 (G1) subtypes. Sliding window analysis along the complete genome ruled out recombination phenomena suggesting that it belongs to a new HCV G1 subtype. Two NS5A RASs (L31V+Y93H) were found to be naturally combined in the genome which could limit treatment possibilities in patients infected with this subtype.

BACKGROUND

Hepatitis C virus (HCV) is a single-stranded, positive-sense RNA virus of approximately 9600 nucleotides that belongs to the genus Hepacivirus (Flaviviridae family) [1]. Phylogenetic studies using full-length HCV genomes have confirmed the existence of seven major HCV genotypes, 67 subtypes [2] and 86 subtypes as of June 2017 according to the ICTV website (https://talk.ictvonline.org/ictv_wikis/flaviviridae/w/sf_flavi/56/hcv-classification), while additional not yet assigned isolates have been reported [3–5]. Based on genetic distances, the major genotypes differ from each other by >30 % of their nucleotide sequence, whereas subtypes typically differ by >15 %. Even different isolates from the same subtype can vary by up to 10 %. Another level of variation is found within infected individuals, in which HCV circulates as a complex mixture of closely related, but distinctly different genomes (1–3 % distance), known as a quasispecies [6]. Thus, HCV has a high level of genetic heterogeneity, and this has important implications for the diagnosis and treatment of this infection.

The use of direct-acting antivirals (DAAs) in combination, has been a major milestone in the treatment of chronic HCV infection, reaching viral elimination rates (sustained virological response, SVR) greater than 90 %. However, 2–10 % of these treatments fail and selection of drug resistance-associated substitutions (RASs) in the viral genome that confer decreased susceptibility to DAAs occur [7]. Of particular note, naturally harbouring RASs have also been detected in some HCV subtypes [8]. SVRs in HCV-infected patients can vary depending on the patient’s viral subtype [9–13]. Therefore, accurate HCV subtyping is emerging as a key factor when deciding the most suitable DAA combination therapy for affected patients [14–16].

High-resolution HCV subtyping based on next-generation sequencing has been implemented in our clinical laboratory [17, 18], and more than 2000 patients have been correctly
Fig. 1. Phylogenetic trees of the full genomes of the reference sequences for the complete set accepted and not yet classified G1 subtypes (green) compared with that of the new isolate (orange). Trees were constructed by neighbour-joining on the matrix of p-distances.
subtyped in routine practice. The high coverage of the method (2000–5000 reads) also enables detection of mixed infections involving more than one genotype/subtype with high sensitivity (1% prevalence in our geographic area). In addition, the technique, which uses phylogenetic analysis of a small fragment of the NS5B region, can confidently identify the 86 HCV subtypes and detect new subtypes.

**OBJECTIVES**

Characterization of a sample from an untreated patient from Equatorial Guinea that was identified as genotype 1 (G1) during routine clinical practice, but even though a 498-nucleotide fragment of NS5B was easily amplified, the sample could not be classified into any of the known G1 subtypes.

**RESULTS**

The complete 'new G1' sequence (9408 nt) has been deposited in the GenBank database with accession number KY348757. It includes the complete coding region and partial sequences from both 5'UTR and 3'UTR. Our sequence was aligned with the H77 reference sequence (accession number AF009606) [19] and representative complete genome sequences of the 13 G1 assigned subtypes by June 2017 [2]. Phylogenetic analysis showed that our sequence did not cluster with any of the confirmed G1 subtypes (Fig. 1). The genetic
p-distance between our sequence and the closest subtype (1b) was 0.211, whereas genetic distance between two sequences belonging to the same subtype was less than 0.10 (example: M58335-D90208 [1b], 0.089) (Table S1, available in the online version of this article). Therefore, our sequence could not be classified into any of the confirmed G1 subtypes and does not group with any of the not yet assigned isolates. The heatmap on the matrix of p-distances shows also that the new isolate classifies far from and well between, the known subtypes (Fig. 2).

To rule out potential viral recombination events between the various HCV subtype 1 genomes, nucleotide similarity curves were plotted along the HCV genome using the sliding window approach. The sliding window plots (Figs 3 and S1) showed that the genetic p-distance between our sequence and the nearest of the 13 G1 subtypes was higher than the minimum p-distance among subtypes along the complete genome, thus providing evidence that our isolate was not the result of recombination of different subtype 1 genomes (Fig. 3). The stochastic nature of the classification of Px at each window excludes a simple recombination of two pure subtypes (Fig. S1).

To further characterize the viral genome, we looked for RASs in the NS3, NS5A and NS5B protein sequences [12, 20]. One NS3, five NS5A (including L31V and Y93H) and one NS5B RASs were detected (Table 1). As the subtype closest to our virus was subtype 1b, we focused on mutations with a significant mean fold change in resistance for subtype 1b (Table 1). Interestingly, deep sequencing analysis demonstrated that the L31V and Y93H mutations were both present in the same genome.

**DISCUSSION**

The present report describes and characterizes a full-length genome of a new HCV isolate from a patient native to Equatorial Guinea. This novel isolate was identified during routine classification of the HCV virus in patient samples using a high-resolution HCV subtyping method [17]. Phylogenetic analysis, genetic distance analysis, and sliding window analysis of this genome (GenBank KY348757) clearly demonstrated that our isolate is not a recombinant product of different subtypes and that the genetic distance to the closest accepted reference sequence is higher than 15%. However, given the recent recommendations for a unified nomenclature system [2] and that no related viruses have
been detected in GenBank using complete and subgenomic sequences, it will remain as an unclassified lineage until a related sequence is discovered.

RASs in HCV can develop when DAA treatment fails. The level of resistance to a DAA inhibitor depends on the genetic barrier to resistance of the drug, but also on the type of RAS, the frequency at which a RAS is present in the viral isolate, and particularly, on the presence of RAS combinations in the same genome. In subtype 1b, single L31V or Y93H substitutions in the viral genome increase resistance to daclatasvir by 15-fold and 12-fold, respectively. However, when both mutations (L31V+Y93H) are present in the same genome, the virus becomes 5,425 times more resistant to daclatasvir than the wild-type virus [20]. The most interesting observation from our isolate is that the virus is a natural carrier of L31V+Y93H substitutions in the same genome, as well as H/P58S and E62Q, which have been associated with varying levels of resistance to DAA-based treatments in other subtypes. Hence, this subtype may be resistant to NS5A inhibitor drugs.

### Table 1. RASs in NS3, NS5A and NS5B

<table>
<thead>
<tr>
<th>Variant</th>
<th>HCV gene</th>
<th>Subtype</th>
<th>Telaprevir</th>
<th>Daclatasvir</th>
<th>Ledipasvir</th>
<th>Ombitasvir</th>
</tr>
</thead>
<tbody>
<tr>
<td>I132V</td>
<td>NS3</td>
<td>1a</td>
<td>2</td>
<td>ND</td>
<td>2–100</td>
<td>ND</td>
</tr>
<tr>
<td>Q30L</td>
<td>N55A</td>
<td>1a</td>
<td>ND</td>
<td>15</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>L31V</td>
<td>N55A</td>
<td>1a</td>
<td>1000</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>L31V</td>
<td>N55A</td>
<td>1b</td>
<td>15</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>H/P58S</td>
<td>N55A</td>
<td>1a/1b</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Y93H</td>
<td>N55A</td>
<td>1a</td>
<td>1600</td>
<td>1677</td>
<td>41 383</td>
<td>77</td>
</tr>
<tr>
<td>E446Q</td>
<td>N55B</td>
<td>1a</td>
<td>ND</td>
<td>12</td>
<td>ND</td>
<td>2–100</td>
</tr>
<tr>
<td>L31V+Y93h</td>
<td>N55A</td>
<td>1b</td>
<td>5425</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Amplified products were Sanger-sequenced using the same primers as those used in the nested PCR, and deep sequencing of the three DAA-HCV targeted regions (NS3, NS5A and NS5B) were performed according to a previously described method [17]. Chromatogram files were assembled, verified and edited using GeneDoc 2.7.000 (2006) software. An updated set of 22 complete reference genomes corresponding to all confirmed G1 subtypes (1a, 1b, 1c, 1d, 1e, 1f, 1g, 1h, 1i, 1j, 1k, 1l, 1m, 1n) along with seven not yet assigned isolates, by June 2017, was used to perform the classification study [2]. Phylogenetic trees and a heatmap of the matrix of p-distances were constructed (Figs 1 and 2). To study the eventuality of a recombination of different subtypes, an analysis by sliding windows was carried out (Figs 3 and S1). A classification to the nearest reference sequence, by p-distance, was performed by sliding windows of 500 bp in steps of 10 bp (Fig. 3). The scripts in R [21] used to obtain the figures are included as supplementary material. R libraries used were Biostrings, ape, stringr, RGBColorBrewer, psych. MUSCLE was used for the required multiple alignments [22].

### METHODS

The blood sample containing the unclassified HCV isolate was obtained from an HCV-HBV co-infected Equatorial Guinea woman, born in 1953 and diagnosed with HCV in 2008. A sample taken in September 2015 was delivered to our laboratory within 8 h after blood drawing, centrifuged and plasma was stored at −80 °C until analysis.

The whole genome was obtained by nested reverse transcription-PCR using three newly designed external G1 primer pairs and 11 new overlapping inner primer pairs (Table S2), numbered using isolate H77 (accession number AF009606) [19].

Funding information

This study was funded by Instituto de Salud Carlos III, PI15/00829, PI15/00856 and PI16/00337, co-financed by the European Regional Development Fund (ERDF). CIBERehd (Consorcio Centro de Investigación en Red de Enfermedades Hepáticas y Digestivas), which is funded by Instituto de Salud Carlos III, and by CDTI (Centro para el Desarrollo Tecnológico Industrial), Spanish Ministry of Economics and Competitiveness (MINECO), IDI-20151125. C.P. is supported by the Miguel Servet program of the Instituto de Salud Carlos III, grant CP14/00121, co-financed by the ERDF.

Acknowledgements

The authors thank Celine Cavallo for English language support and helpful editing suggestions.

Conflicts of interest

The authors declare that there are no conflicts of interest.
Ethical statement
We have not experimented with humans, only with serum excess that was left over after routine analysis. The experimental work with serum has been approved by an Ethical Committee and subjects gave informed consent for antiviral treatment.

References