Infection, Genetics and Evolution 9 (2009) 860–866

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

The role of positive selection in hepatitis C virus

José M. Cuevas [a,b,*], Michael Gonzalez [a], Manuela Torres-Puente [a], Nuria Jiménez-Hernández [a,b,c], María A. Bracho [a,b,c], Inmaculada García-Robles [a], Fernando González-Candelas [a,b,c], Andrés Moya [a,b,c]

[a] Instituto Cavanilles de Biodiversidad y Biología Evolutiva and Departamento de Genética, Universidad de Valencia, València, Spain
[b] CIBER en Epidemiología y Salud Pública (CIBERESP), Spain
[c] Área Genómica y Salud, CSISP, Valencia, Spain

ARTICLE INFO

Article history:
Received 30 December 2008
Received in revised form 6 March 2009
Accepted 11 May 2009
Available online 20 May 2009

Keywords:
Hepatitis C virus
HVR1
HVR2
HVR3
ISDR
PKR-BD
Positive selection
V3 domain

ABSTRACT

Hepatitis C virus (HCV) is a major health problem worldwide, infecting an estimated 170 million people. In this study, we have employed a large data set of sequences (14,654 sequences from between 25 and 100 clone sequences per analyzed region and per patient) from 67 patients infected with HCV genotype 1 (23 subtype 1a and 44 subtype 1b). For all patients, a sample prior to combined therapy with alpha interferon plus ribavirin was available, whereas for some patients additional samples after 6 or 12 months of treatment were also available. Twenty-seven patients responded to treatment (12 subtype 1a and 15 subtype 1b) and forty patients did not respond to treatment (11 subtype 1a vs. 29 subtype 1b). Two regions of the HCV genome were analyzed, one compressing the hypervariable regions (HVR1, HVR2 and HVR3) of the envelope 2 glycoprotein and another one including the interferon sensitive determining region (ISDR) and the V3 domain of the NS5A protein. Previously (Cuevas, J.M., Torres-Puente, M., Jiménez-Hernández, N., Bracho, M.A., García-Robles, I., Wrobel, B., Carnicer, F., del Olmo, J., Ortega, E., Moya, A., González-Candelas, F., 2008b. Genetic variability of hepatitis C virus before and after combined therapy of interferon plus ribavirin. Plos One 3 (8), e3058), several amino acid positions in both regions analyzed were detected to be under positive selection. Here, we have compared the amino acid composition of each positively selected position between responder and non-responder patients for both subtypes. If we exclude some non-conclusive cases, no clear differences were detected in any case. In conclusion, identifying specific positions as completely discriminatory of treatment response seems to be a difficult task. Our results, in concordance with previous studies, suggest that HCV evasion strategies are more likely based on a global increased variability, which would yield combinations of mutations with an increased resistance, than on the fixation of specific amino acids conferring resistance to antiviral treatment or immune response. In this sense, the particular systemic response from each patient could play an essential role in determining the outcome of the antiviral treatment.

1. Introduction

Hepatitis C virus (HCV) is a newly emerged human pathogen that has become a major health problem worldwide, infecting an estimated 170 million people (Wasley and Alter, 2000; Shepard et al., 2005). 20% of those infected are at high risk of developing severe liver disease, including cirrhosis and hepatocellular carcinoma (reviewed in Afdhal, 2004). The current antiviral therapy combines the use of pegylated interferon-α with ribavirin (Fried et al., 2002), but it is only effective in about 50% of cases and involves a significant toxicity (Heathcote and Main, 2005; Dienstag and McHutchison, 2006).

HCV is an enveloped positive-sense RNA virus of the genus *Hepacivirus* in the *Flaviviridae* family. Its 9.6-kb genome encodes for a polyprotein of about 3000 amino acids. This polyprotein is cleaved by both host and viral proteases to generate three structural (core, E1, E2) and seven non-structural (p7, NS2-NS5B) proteins (reviewed in Lindenbach and Rice, 2005). As other RNA viruses, HCV is characterized by a high genetic diversity, which likely plays an important role in viral persistence. More specifically, its mutation rate has been estimated to be in the order of 1.5×10^{-3} to 2.0×10^{-3} nucleotide substitutions per site per genome per year (Bukh et al., 1995).

The study of adaptive strategies in HCV has been focused on those regions seemingly involved in immune system evasion. This is the case of the three hypervariable regions, HVR1 (Okamoto et al., 1992), HVR2 (Kato et al., 1992) and HVR3 (Troesch et al.,

* Corresponding author at: Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universitat de Valencia, P.O. Box 22085, 46071 València, Spain.
Tel.: +34 963 543 686; fax: +34 963 543 670.
E-mail address: cuevast@uv.es (J.M. Cuevas).

1567-1348/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2009.05.007

2006). In particular, HVR1 has been extensively studied, and the rate and nature of nucleotide substitutions within this region during the early stages of infection appear to be correlated with treatment outcome (Farci et al., 2000). In addition, the genetic variability in other regions such as the protein kinase-R binding domain (PKR-BD), which includes the putative interferon sensitivity determining region (ISDR), or the V3 domain in the NS5A gene has also been studied. A potential role in responsiveness to interferon has been hypothesized for these domains (Gale et al., 1998; Durante et al., 2003).

Several studies have attempted to associate different parameters of HCV genetic heterogeneity, such as genetic diversity or number of mutations (or amino acid substitutions), with treatment outcome (reviewed in Hofmann et al., 2005; Goyal et al., 2007). In this sense, the information provided by analyses of variability is useful but limited, because no specific amino acids can be assigned as clear candidates for participating in immune system evasion. To shed more light on this issue, positive selection analyses are currently being used to determine the particular sites involved in immune escape (Sheridan et al., 2004; Chen and Wang, 2007; Cuevas et al., 2008b; Torres-Puente et al., 2008c).

In a previous study (Cuevas et al., 2008b), we performed positive selection analyses for all the biologically meaningful regions mentioned above, as well as for the surrounding and intervening regions, in a set of 22 HCV-infected patients (7 subtype 1a and 15 subtype 1b) which did not respond to antiviral therapy with interferon alpha-2a plus ribavirin. In the present study, we have expanded these analyses to a cohort of 67 HCV-infected patients (27 responders and 40 non-responders), which included the 22 ones analyzed previously. In order to find specific patterns of adaptation, amino acid composition comparisons at each positively selected position were performed between the groups of responder and non-responder patients, separately for both 1a and 1b subytpes. However, no clear patterns of adaptation were found in any case.

2. Materials and methods

2.1. Samples

The set of samples used in this study has been described previously (Torres-Puente et al., 2007; Torres-Puente et al., 2008b). Serum samples from 67 HCV infected patients were chosen for this study; 23 patients were infected with HCV subtype 1a and 44 with subtype 1b. HCV genotyping was performed according to Ohno et al. (1997).

As previously described, serum samples were obtained before the patients were subjected to treatment with alpha interferon plus ribavirin (Supplementary Tables S1 and S2). For 22 patients

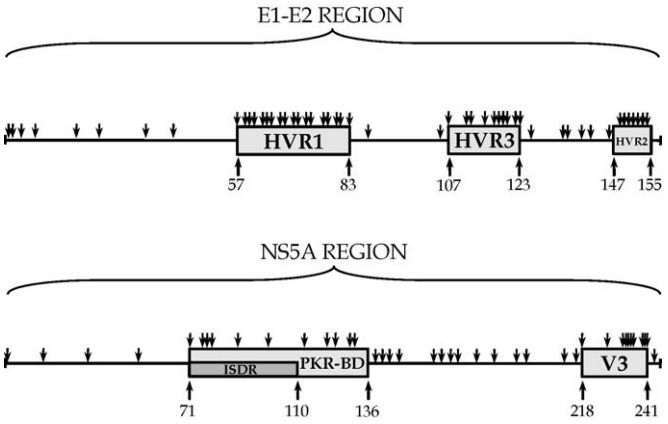

Fig. 1. Schematic representation of the E1–E2 and NS5A regions analyzed in our study, including the hypervariable regions, the PKR-BD, the V3 domain and their surrounding and intervening regions. For each region, arrows in the upper half indicate positions previously detected to be under positive selection (Cuevas et al., 2008b).

(7 with subtype 1a and 15 with subtype 1b), additional serum samples were also obtained after 6 or/and 12 months of antiviral treatment (Cuevas et al., 2008b). Twenty-seven patients responded to treatment (12 with HCV subtype 1a and 15 with subtype 1b) and forty patients did not (11 with subtype 1a and 29 with subtype 1b). Samples were obtained in different hospitals from the Comunidad Valenciana, Spain (Tables S1 and S2).

Two HCV genome regions were studied: one corresponded to a 472 nt fragment encompassing genes encoding proteins E1 and E2 (from nucleotide 1310 to 1781 in the HCV-J reference genome sequence, GenBank accession number AF009606) including the three hypervariable regions (HVR1, HVR2 and HVR3) plus three adjacent and intervening subregions (Table 1), and referred to as E1–E2 region. The other region corresponded to a 743 nt fragment from gene NS5A (nucleotides 6742–7484) including the PKR-BD, which contained the ISDR and the V3 domain, besides three additional intervening and adjacent regions (Table 1) and referred to as NS5A region (Fig. 1).

2.2. Experimental procedures

RNA extraction, reverse transcription, amplification, cloning and sequencing, were explained in detail elsewhere (Jiménez-Hernández et al., 2007). For the E1–E2 region, we obtained about 100 clones from each serum sample, yielding a total of 9228 sequences (Table S1). For the NS5A region, we obtained between 25 and 96 clones per serum sample and 5426 sequences in total were determined (Table S2). HCV sequences obtained in this study

Table 1
Summary of positions under positive selection in at least one patient at the different subregions defined for both E1–E2 and NS5A regions (Cuevas et al., 2008b) (Na: number of amino acid positions in the corresponding subregion).

Region	Subregion		Positions	Na	Number of selected positions	Positively selected positions
E1–E2	E1		1–56	56	8	2, 3, 6, 8, 17, 22, 32, 38
	HVR1		57–83	27	18	57, 59, 60, 61, 63, 64, 65, 67, 68, 70, 71, 73, 74, 77, 78, 80, 81, 83
	E2-1		84–106	23	2	87, 105
	HVR3		107–123	17	10	107, 111, 112, 115, 117, 118, 119, 120, 122, 123
	E2-2		124–146	23	6	126, 133, 134, 137, 139, 146
	HVR2		147–155	9	7	148, 149, 150, 151, 152, 153, 154
NS5A	NS5A-1		1–70	70	4	1, 15, 31, 49
	PKR-BD	ISDR	71–110	40	6	71, 76, 78, 80, 90, 101
		Rest	111–136	26	5	114, 122, 125, 130, 132
	NS5A-2		137–217	81	14	139, 142, 145, 149, 162, 165, 167, 169, 174, 181, 191, 195, 211, 216
	V3 domain		218–241	24	10	218, 226, 232, 233, 234, 235, 236, 239, 240, 241
	NS5A-3		242–247	6	1	244