Infection, Genetics and Evolution 43 (2016) 186–196

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

Evolution of hepatitis C virus in HIV coinfected patients under antiretroviral therapy

Mariano Sede [a,b], Micaela Parra [a,b], Julieta M. Manrique [a,c], Natalia Laufer [a,b], Leandro R. Jones [a,c,*,1], Jorge Quarleri [a,b,**,1]

[a] Consejo Nacional de Investigaciones Científicas y Técnicas, Av. Rivadavia 1917, C1083ACA Buenos Aires, Argentina
[b] Instituto de Investigaciones Biomédicas en Retrovirus y Sida (INBIRS), Facultad de Medicina, Universidad de Buenos Aires, Paraguay 2155-Piso 11, C1121ABG Buenos Aires, Argentina
[c] Laboratorio de Virología y Genética Molecular, Facultad de Ciencias Naturales sede Trelew, Universidad Nacional de la Patagonia San Juan Bosco, 9 de Julio y Belgrano S/N, 9100 Trelew, Chubut, Argentina

ARTICLE INFO

Article history:
Received 19 February 2016
Received in revised form 13 May 2016
Accepted 23 May 2016
Available online 24 May 2016

Keywords:
HCV
HIV
E1/E2 gene
Haplotypes
Phylogeny
Temporal structuring

ABSTRACT

Five patients (P) were followed-up for an average of 7.73 years after highly active antiretroviral therapy (HAART) initiation. Patients' immune and virological status were determined by periodical CD4 + T-cell counts and HIV and HCV viral load. HCV populations were studied using longitudinal high throughput sequence data obtained in parallel by virological and immunological parameters. Two patients (P7, P28) with sub-optimal responses to HAART presented HCV viral loads significantly higher than those recorded for two patients (P1, P18) that achieved good responses to HAART. Interestingly, HCV populations from P7 and P28 displayed a stable phylogenetic structure, whereas HCV populations from P1 and P18showeda significant increase in their phylogenetic structure, followed by a decrease after achieving acceptable CD4 + T-cell counts (>500 cell/μl). The fifth patient (P25) presented high HCV viral loads, preserved CD4 + T-cell counts from baseline and all along the follow-up, and displayed a constant viral phylogenetic structure. These results strongly suggest that HAART-induced immune recovery induces a decrease in HCV viral load and an increase in the HCV population phylogenetic structure likely reflecting the virus diversification in response to the afresh immune response. The relatively low HCV viral load observed in the HAART responder patients suggests that once HCV is adapted it reaches a maximum number of haplotypes higher than that achieved during the initial stages of the immune response as inferred from the two recovering patients. Future studies using larger number of patients are needed to corroborate these hypotheses.

1. Introduction

Hepatitis C virus (HCV) is an enveloped RNA virus belonging to the *Flaviviridae* family genus *Hepacivirus*. More than 80% of acutely infected people evolve to viral persistence and the global prevalence of HCV infection among human immunodeficiency virus type 1 (HIV-1)-infected patients may be as high as 90% as a result of common transmission routes (Clausen et al. 2014). HIV-related immunosuppression or immune deregulation is associated with accelerated liver disease in individuals with HIV–HCV coinfection (Miro et al. 2015). Accordingly, patients on highly active antiretroviral therapy (HAART) undergo a slow HCV disease progression and reduced HCV-related mortality rates (Pascual-Pareja et al. 2009). Therefore, host immunity can affect both viral replication and mutation, possibly influencing the dynamics of HCV evolution. In this regard, the low fidelity RNA-dependent RNA-polymerase contributes to a high mutational rate incorporating about one mutation per genome replication round, thus generating closely-related yet genetically distinct variants commonly called *quasispecies* (Bittar et al. 2013).

HCV contains a small RNA genome (~9.6 kb) that encodes several structural and non-structural proteins (Webster et al. 2015). Among the former, two envelope glycoproteins, E1 and E2, form non-covalent heterodimers that are involved in the adsorption process and the receptor-mediated endocytosis (Kong et al. 2013). The E2 glycoprotein plays a critical role in host cell recognition by humoral (Wahid and Dubuisson 2013) and cell-mediated immune responses (Shirai et al. 1999). Different domains are distinguished along the gene, including the hypervariable region 1 (HVR1), which exhibits the highest degree of variability along the HCV genome (Kato et al. 1994). Thus, this region is frequently the preferred marker to analyze the HCV adaptation to host-selective pressures and intra-host diversification (Sala and Wain-Hobson 2000). The second hypervariable region of E2, HVR2, is located downstream of HVR1 and is potentially involved in cell surface receptor binding

* Correspondence to: L.R. Jones, Consejo Nacional de Investigaciones Científicas y Técnicas, Av. Rivadavia 1917, C1083ACA Buenos Aires, Argentina.
** Correspondence to: J. Quarleri, Instituto de Investigaciones Biomédicas en Retrovirus y Sida (INBIRS), Facultad de Medicina, Universidad de Buenos Aires, Paraguay 2155-Piso 11, C1121ABG Buenos Aires, Argentina.
E-mail addresses: lrj000@gmail.com (L.R. Jones), quarleri@fmed.uba.ar (J. Quarleri).
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.meegid.2016.05.032
1567-1348/

(Yagnik et al. 2000). An additional variable domain, termed HVR3, is positioned between HVR1and HVR2 (Troesch et al. 2006). The mutational change at the neutralizing epitopes on HVR1 avoids immune recognition and contributes to immune evasion (Bankwitz et al. 2014; Ray et al. 2005). Likewise, other strategies such as the number of glycosylation of acceptor sites (Helle et al. 2011; Pantua et al. 2013) and the HCV ability to infect neighboring cells by direct cell-to-cell transmission (Brimacombe et al. 2011) are contributing mechanisms through which highly sophisticated, host-adapted sequences can evolve. As with other RNA viruses, genetic recombination is a mechanism related to HCV molecular evolution. Intra-subtype recombinants can accommodate recombination breakpoints at multiple sites along the genome, including the E1/E2 junction (Gao et al. 2007; Palmer et al. 2014; Scheel et al. 2013; Sentandreu et al. 2008).

The aim of the present study is to analyze the intra-host HCV molecular evolution *in vivo* in HIV-coinfected patients. For this goal data were obtained by next-generation sequencing to gain a greater level of sequence depth and expanded timeframes over which the analysis were implemented (a mean of 7.73 years). Such data allowed inferring the phylogenetic diversity and temporal structuring of HCV lineages in different patients, the impact of intra-subtype recombination, and the relevance of positive (immune) selective pressures exerted on domains of E2 glycoprotein.

2. Materials and methods

2.1. Patients and clinical parameters

Five subjects were selected from a previous study of HCV quasispecies evolution among HIV-coinfected patients on uninterrupted HAART (Sede et al. 2014a). Patients included in the present study were those with a richer sampling which allowed obtaining a more extensive viral genomic database by next-generation sequencing to support a deeper study of the HCV molecular evolution including the unexplored temporal structuring phenomenon. Twenty-nine samples encompassing an average of 7.7 years from the five chronic HCV-1a interferon-based treatment-naïve patients were collected and analyzed by ultra-deep pyrosequencing. As shown in Table 1, these consecutive samples from each patient were numbered with a capital letter in alphabetical order (A, B, C, *etc.*) followed by a number between parentheses that represents the months after the HAART initiation (0, 24, 35, *etc.*). At each sampling time, CD4 + T cell counts were performed (flow cytometry double platform, BD FACS Canto; BD Biosciences, San Diego/California, USA). This study was conducted in line with the Declaration of Helsinki, and the research protocol was approved by the Ethics Review Board of Huesped Foundation, Buenos Aires, Argentina. All patients were provided with medical care and counseling required by their condition. Serum samples were collected from whole blood by centrifugation and kept at −80 °C until use.

2.2. RNA extraction and HCV-E1/E2 RT-nested PCR

Viral RNA was isolated from plasma samples with Trizol LS following the manufacturer instructions (GIBCO, LifeTechnologies, USA). The nucleotide sequence of HCV-H77 was used as reference (GenBank accession no. AF009606). The HCV-E1/E2 encoding sequence was amplified using primers EF1 (5′ GGATATGATGATGAACTGGTC3′, position1307–1327) and E2R1 (5′ RAARCARTCCGTGGGGCA, position 2069–2086), and then primers E2F2 (5′GGACATGATCGCTGGTGCTCA3′, position 1376–1396) and E2R2 (5′ TCCGCACACTTTGGTGAATCC 3′, position 2016–2036) in a nested reverse transcriptase polymerase chain reaction (RT-PCR). Primers used in the second round were modified by 5′Tag extensions, which provided binding sites for the multiplex identifiers (MIDs). Tags were as follows: sense-Tag 5′-CACGACGTTGTAAAACGA-3′; antisense-Tag 5′-CAGGAAACAGCTATGACC-3′. The MIDs allowed the identification of the samples after completing the pyrosequencing procedure (see below).

The RT program included 60 min at 42 °C followed by enzyme deactivation at 85 °C for 5 min. In order to minimize the probability to generate PCR-mediated artificial chimeras, several measures were performed (Di Giallonardo et al. 2013; Lahr and Katz 2009). Briefly, PCR cycling conditions were optimized by increasing the elongation time to minimize the occurrence of prematurely terminated extension

Table 1
Laboratory results for each patient during the follow-up.

Patient	Sampling time (months)	CD4 T cell (cells/mm^3)	HIV VL (log copies/mL)	HCV VL (log copies/mL)
1	A (0)	307	4.41	5.64
	B (12)	17	<1.70	5.32
	C (23)	86	<1.70	5.45
	D (35)	161	<1.70	5.80
	E (44)	202	<1.70	5.80
	F (50)	332	<1.70	5.85
	G (65)	538	<1.70	5.85
	H (77)	400	<1.70	5.85
18	A (24)	306	<1.70	4.18
	B (48)	463	<1.70	4.40
	C (63)	740	<1.70	3.85
	D (73)	619	<1.70	5.23
	E (87)	573	3.55	5.31
	F (108)	651	<1.70	6.04
25	A (31)	524	<1.70	6.78
	B (41)	891	<1.70	5.99
	C (71)	485	<1.70	6.57
	D (88)	743	<1.70	>7.60
	E (100)	858	<1.70	7.23
28	A (0)	305	5.37	5.84
	B (24)	256	<1.70	>7.60
	C (41)	458	<1.70	>7.60
	D (65)	479	<1.70	6.28
7	A (0)	253	3.59	>7.60
	B (36)	236	3.96	6.36
	C (57)	240	<1.70	>7.60
	D (83)	388	4.72	5.85
	E (93)	379	3.73	6.30
	F (114)	337	<1.70	6.50

events, increasing dNTP (0.4 mM) and oligonucleotide (1 μM each) concentrations, and omitting the final extension step. The first PCR program included 25 cycles each of 30s at 94 °C, 60s at 54 °C and 2 min at 72 °C. The PCR product was used in the nested PCR and the program included 20 cycles of 30s at 94 °C, 60s at 58 °C, 2 min at 72 °C. For both rounds of amplification a highly processive and high fidelity proof reading polymerase (Advantage High Fidelity 2, HF-2) was used (Clontech, Mountain View, CA). PCR products were resolved by electrophoresis in1% agarose gels.

In order to ensure a good sampling as well as to optimize the study of the genetic heterogeneity of viral population present in 0.75 ml of plasma by UDS analysis, PCR amplicons from 3 reactions were pooled for each sample. Strict preventative measures were taken to avoid inter-sample PCR cross-contamination (Kwok and Higuchi 1989).

2.3. Ultra-deep sequencing (UDS)

Ultra deep sequencing (UDS) was carried out on the Roche/454 Life Sciences "Genome Sequencer-FLX" (GS-FLX) pyrosequencer. For this purpose, PCR-purified amplicons were quantified using Quant-iT PicoGreen (Invitrogen, Life Technologies, MI, USA) after PCR amplification. In addition, an Agilent 2100 bioanalyzer (Agilent Life Science, Santa Clara, California, USA) was used to verify the quality and length of amplicons.

After quality controls, PCR amplicons were pooled in equimolar concentrations; subsequently PCR amplicons were combined at an appropriate ratio with DNA capture microbeads. Emulsion PCR was performed, and DNA and beads were washed, purified and prepared for pyrosequencing according to the manufacturer's instructions. HCV-E1/E2 sequences amplified on each bead were determined by pyrosequencing on the GS-FLX (Bushman et al. 2008). The raw sequence output ('reads') generated by Roche/454 GS-FLX platform was processed through the native amplicon pipeline using default settings.

In order to minimize the potential bias that sequencing errors can introduce in next generation sequencing data, we filtered the data obtained based on sequence length and quality of base calls. Also, reads containing at least one base displaying a base call value ≤10 were discarded, and no sequence with Ns was kept in the processed data set. In addition, an average (mean) quality value ≥30 was required for each read in order to pass the quality control. Finally, low-quality read ends were trimmed out. All of these pre-processing steps were done using PRINSEQ v.0.15 (Schmieder and Edwards 2011). Then, data were further filtered out by the presence of frame shifts, and low frequency k-mers were corrected by the Kec program (Skums et al. 2012) as described previously (Sede et al. 2014b).

2.4. Evolutionary and sequence analyses

Maximum likelihood phylogenetic trees were performed from datasets obtained by UDS from each patient throughout the longitudinal study. HCV E1/E2 nucleotide sequences were aligned with MAFFT using iterative refinement methods (Katoh and Standley 2014). Maximum likelihood trees were obtained with FastTree-2 (Price et al. 2010), using a GTR + CAT model (General Time Reversible plus CAT approximation, which was obtained by MrAIC.pl program (Nylander 2004)). Trees were visualized and annotated in Dendroscope 3 (Huson and Scornavacca 2012). During the tree inference, FastTree scales searched for intensity by increasing automatically the number of Minimum Evolution and Maximum Likelihood Nearest Neighbor Interchange cycles using a factor directly proportional to the number of sequences in the dataset (Price et al. 2010). In addition, the program provides likelihood-based support values based on the Shimodaira-Hasegawa (SH) test (Shimodaira 1999), allowing for estimating accurate branch supports in reasonable times for very large phylogenetic trees.

Viral diversity was evaluated from phylogenetic trees, *i.e.* phylogenetic diversity (*PD*), and by the species richness metric, where haplotypes were analogous to species. These analyses were performed using the PD function in the Picante package (Kembel et al. 2010). The total phylogenetic diversity is equivalent to the sum of the branch lengths of a tree. In addition, Picante allows partitioning of the data, which permit to dissect the total diversity with respect to any factor of interest. In this set up, the phylogenetic diversity of a given data partition is equivalent to the sum of the lengths of the branches interconnecting the taxa corresponding to that partition. Here, the data were partitioned by patient and sampling times. Given that diversity indices can be biased by differences in sequence effort (Fig. S2) (Gregori et al. 2014), a weighted index (*w-PD*) was devised for the partitioned analyses. This metric is obtained simply by dividing *PD* by the total viral diversity of the corresponding patient. Thus, *w-PDs* are equivalent to the fraction of viral diversity contributed by each individual sampling time.

In order to quantitatively evaluate the structuring phenomenon, the Mean Phylogenetic Distance (MPD) and the Mean Nearest Taxon Distance (MNTD) statistics were used with the R packages *Picante* as described previously (Jones et al. 2016), except that here patients were equated to Webb et al.'s species pools and sampling times to communities (Webb et al. 2002). Statistically standardized effects of phylogenetic structure were obtained as $(D_{OBS} - D_R)/sdD_R$, where D_{OBS} represent the distances (*i.e.* either MPD or MNTD) obtained from the actual data, D_R corresponds to the distances obtained from taxa distributed randomly on the phylogeny and sdD_R is the standard deviation of the distances in the randomized data. Null distributions were generated by randomizing the phylogenies' tip labels. Under this setup, time-structured viral haplotypes' distributions result in negative standardized effects. The bootstrap trees used for these analyses were generated with FastTree after resampling the corresponding datasets by the Seqboot program of the Phylip package (Felsenstein 1989).

The intra-host selective pressure was analyzed. In order to achieve this goal, E1/E2 variability at each amino acid position was calculated as the entropy, defined as $-\Sigma P(si)\log P(si)$, where P(si) is the probability of a given amino acid (s) appearing at a given position (i). Similarly, the selective pressure was calculated separately for the three E2 HVR domains. Entropy was computed and consensus amino acid sequence was determined using the Entropy-ONE Web tool (http://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy_one.html) (Korber et al. 1994).

Mean genetic distance defined as the mean of all pairwise *p* distance comparisons (number of amino acid differences divided by the total number of amino acid sites compared), and its standard deviation were calculated using MEGA version 6.0 (Tamura et al. 2013).

The number of nonsynonymous (amino acid replacement) nucleotide substitutions per nonsynonymous site (dN), the number of synonymous (silent) nucleotide substitution per synonymous site (dS), and the dN/dS ratio were calculated by two methods. First, by Maximum Likelihood analysis of natural selection codon-by-codon using the joint Maximum Likelihood reconstructions of ancestral states under a Muse-Gaut model (Muse and Gaut 1994) of codon substitution and General Time Reversible model of nucleotide substitution. To estimate ML values, a tree topology was automatically computed. The test statistic dN - dS is used for the detection of codons that have undergone positive selection. The probability of rejecting the null hypothesis of neutral evolution (p-value) is calculated (Kosakovsky Pond and Frost 2005; Suzuki and Gojobori 1999). The p-values less than 0.05 are considered significant at a 5% level. Normalized dN - dS for the statistical test is obtained using the total number of substitutions in the tree (measured in expected substitutions per site). It is useful when making comparisons across data sets. Maximum Likelihood estimation of dN and dS were conducted using HyPhy software package (Pond et al. 2005). Codon positions included 1st + 2nd + 3rd + Noncoding. Secondly, a codon-based test of positive selection for analysis averaging over all sequence pairs was carried out according to the Nei–Gojobori method with the Jukes–Cantor

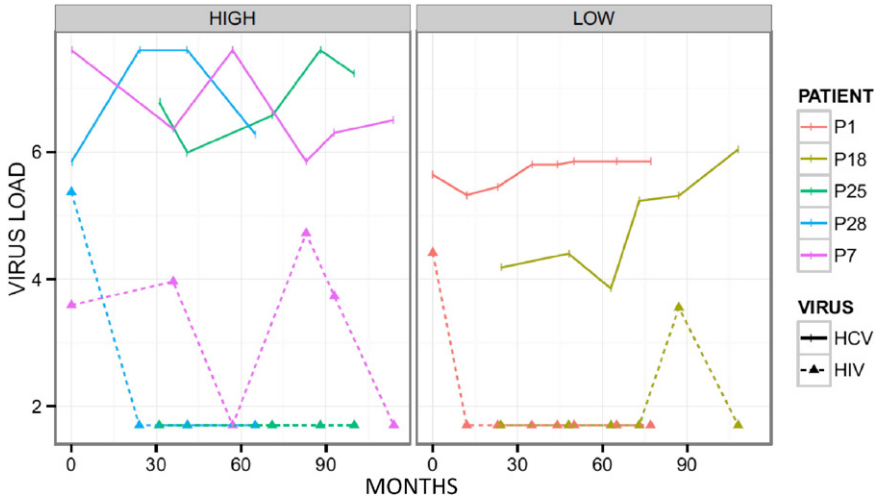


Fig. 1. Longitudinal data for hepatitis C virus (HCV) and human immunodeficiency virus (HIV) viral loads (expressed as log copies/mL) for the five patients studied here (P1, P28, P25, P28, and P7). The data were classified in two blocks (high and low HCV viral load) based on the observed HCV loads. Viral loads were significantly different among the two groups of patients (Kolmogorov-Smirnov test, $p < 0.01$). Time units are "months after HAART initiation".

correction (Zhang et al. 1998) with a dN/dS ratio > 1 taken as an index of positive selection for specific amino acid residues. The significance of positive selection was determined using analysis of variance (Z test). Both analyses were carried out as implemented in MEGA 6.0 (Tamura et al. 2013).

2.5. Recombination inference

Potential recombination events between divergent E1/E2 nucleotide sequences were assessed by two methods: Recco (Maydt and Lengauer 2006) and SplitsTree4 (Huson and Bryant 2006). These techniques were selected considering the large size of our datasets, the mean diversity among nucleotide sequences, and the nucleotide extension of the amplicon. First, we carried out the pairwise homoplasy index (PHI) test as implemented in SplitsTree4 (http://splitstree.org) to assign a probability that the aligned sequences within a sample set contained recombinants (Huson and Bryant 2006). P-values <0.05 were indicative of recombination (Bruen et al. 2006). Second, the Recco algorithm developed as being a fast, simple and sensitive method for detecting recombination in a set of sequences and locating putative recombination breakpoints is based on cost minimization.

2.6. Sequence information

Amino acid residues were numbered according to the polyprotein of the H77 reference sequence (AF009606). The GenBank accession

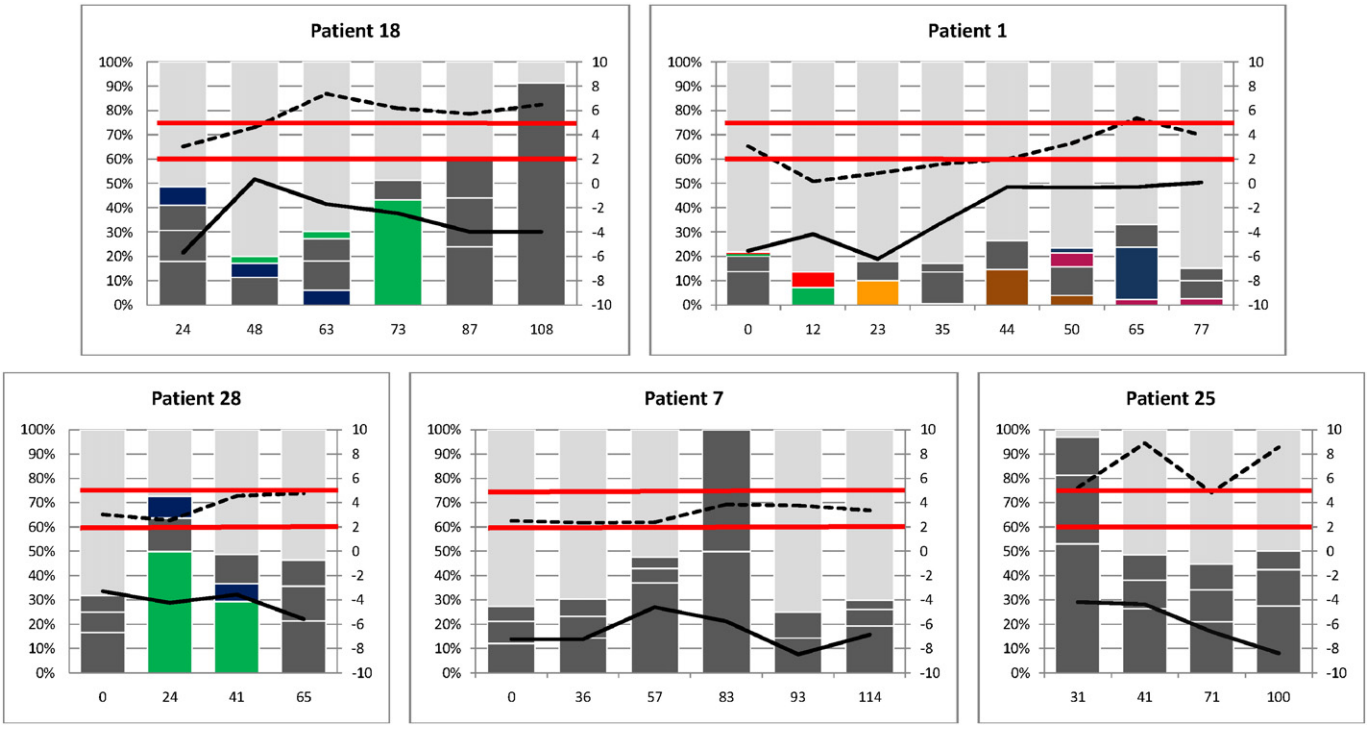
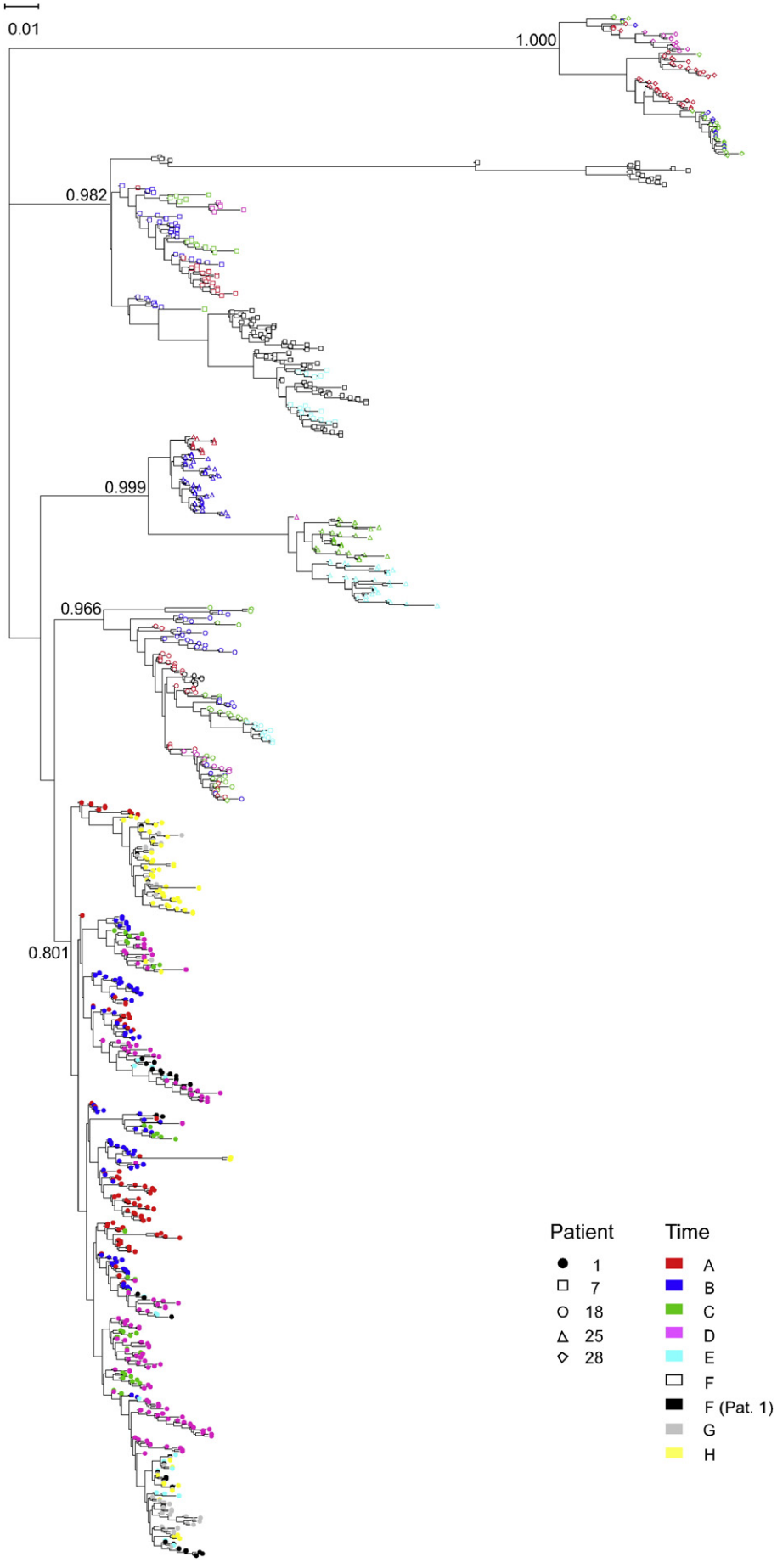


Fig. 2. Analysis of the HCV haplotypes frequency dynamic for each patient along the follow up. Dark gray represent the most frequent haplotypes detected in a single sampling time; light gray include haplotypes present in very low frequencies; colors represent viral haplotypes that were present in more than one sampling time. The *x* axis depicts the sampling times in months. The right *y* axis depicts the longitudinal CD4+ T-cell counts (×100 cells/mm^3) and Mean Phylogenetic Distances' standardized effects for the five studied patients shown by dashed and solid lines, respectively. The horizontal red lines represent the interval between 200 and 500 cell/mm^3 CD4+ T-cells.


0.01
1.000
0.982
0.999
0.966
0.801
Patient
1
7
18
25
28
Time
A
B
C
D
E
F
F (Pat. 1)
G
H

numbers for the haplotype sequences reported here are KR993061 to KR994681.

3. Results

HCV E1 and E2 envelope gene sequences were amplified from serial serum samples taken before and during HAART therapy from 5 patients coinfected with HCV and HIV-1. Laboratory results (HIV and HCV plasma viral load – VL – as well as the CD4 T-cell count) for each patient along the follow-up were obtained (Table 1 and Figs. 1, 2, and S1). Patients 1 and 18 displayed significantly lower HCV viral loads (Fig. 1; $p < 0.01$, Kolmogorov-Smirnov test) and hence, they were included into the "low" HCV viral load group with levels lower than log 6.3 copies/mL (Martinot-Peignoux et al. 2000). These patients presented a sustained increment of CD4 + T cell counts along time. In contrast, patients 7, 25 and 28 presented relatively steady CD4 + T cell counts (Figs. 2, and S1) and high HCV viral load levels (Supplementary Table 1).

3.1. Handling and error correction of UDS data

A total of 14,728 partial HCV E1/E2 gene sequence reads were analyzed. The number of reads per patient ranged between 1665 and 5650. Taking into account the potential errors generated by next generation sequencing techniques, we filtered the data obtained based on sequence length, quality of base calls, frame shifts and k-mer frequencies as described previously (Sede et al. 2014b). After such analyses, nucleotide sequences were edited and translated, spanning at least amino acid positions 353–484 of the HCV-1a reference sequence (GenBank accession No. AF009606), and corresponding to the 31 C-terminal amino acid residues of E1 and the 100 N-terminal amino acid residues of E2.

3.2. Phylogenetic analyses

Maximum Likelihood analyses recovered moderately (SH > 70) and strongly (SH > 90) supported monophyletic groups of viral haplotypes corresponding to each patient (Fig. 3). The haplotypes from each sampling time per patient were highly related to each other, as is revealed by the clustering of terminals with the same color code in Fig. 3. The degree to which viral haplotypes from different sampling times are distributed along the tree (which we refer to as a phylogenetically or temporally *dispersed* pattern) was more pronounced in some patients (P1 and P18) whereas for the other (P7, P25, and P28) viral haplotypes are clumped together for each given sampling time (which we refer to as a temporally *structured* pattern) (Fig. 2).

In order to evaluate the significance of the structuring phenomenon, we used the Mean Phylogenetic Distance (MPD) and the Mean Nearest Taxon Distance (MNTD) statistics (Webb et al. 2002). Despite the good support observed for the patients clades (Fig. 3), intra-patient supports were quite low (data not shown). Thus, in order to incorporate the potential effect of phylogenetic uncertainty, each analysis was repeated with 100 trees obtained from resampled datasets (*i.e.* bootstrap trees). Temporal structuring was corroborated for all the patients (Fig. 4a; $p < 0.01$), confirming the structured nature of HCV evolution suggested by the Maximum Likelihood tree. Notwithstanding, patients 25 and 7 presented the greater (*i.e.* "more negative") MPD standardized effects (mean values −5.38 and −6.68, respectively, compared to mean values of −2.90 and −2.48, respectively for P18 and P1). Likewise, Patient 28 displayed an intermediate degree of structure (−4.23). Despite the vast majority of MNTD effect sizes were negative, that is supported by the structured haplotypes' distributions (see Materials and methods for details), they were distributed more homogeneously in comparison with MPD values (Fig. 4a). This phenomenon is quite expectable because MNTD depends on how close is each sequence and its closest relative from other sampling times (blocks).

Interestingly, it was found a relationship between the temporal variations of phylogenetic structure and the HCV viral loads. Patients P7, P25 and P28 shown a steady increment of phylogenetic structure (reflected by the decreasing effect size values), whereas P1 and P18 displayed a trend towards dispersed patterns (reflected by increasing effect size values, followed by an inflexion point (just before month 40 in Fig. 4b), from which structuring began to increase again. Remarkably, this inflexion point occurred approximately by the time these patients reached the 500 CD4 + T-cells per microliter threshold (Figs. 2 and 4b).

Phylogenetic diversity (PD) and haplotype richness (HR) were homogeneous among the patients regardless of treatment response and sampling time (Table 2 and Fig. S1). The small differences observed among PD and HR may be due to differences in the number of sequence reads available for each patient, in agreement with observations made for other diversity metrics (Gregori et al. 2014). There was no evident trend for the phylogenetic diversity along the follow-up. However, the tree branch separating sampling times B and C of P25 clade was much larger than the rest of branches inside this patient's clade (Fig. 3). A similar situation was observed for Patient 7, for whom some of the reads from sampling time F clustered into a very divergent clade. These topological patterns suggest the occurrence of rampant divergence events in the corresponding HCV populations. Coincidentally, these divergent events were accompanied by the occurrence of dissimilar intra-host positive selection (see below).

3.3. E1/E2 amino acid sequence variability and diversity

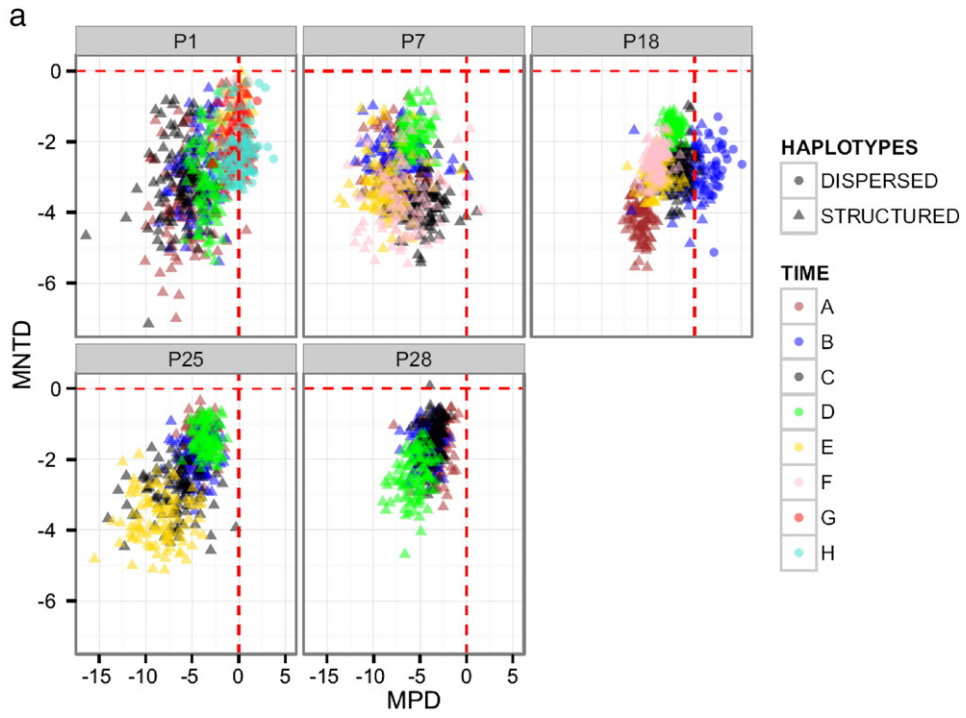
Amino acid sequence variability at the E1/E2 as well as at the corresponding hypervariable domains were estimated by the mean, uncorrected pairwise distances (*p* distances). As expected, the variability at HVR1 was higher than at HVR2 and HVR3, contributing to the total variability three times more than the rest of the E1/E2 sequence. These mean *p* distance values (and standard deviation) for all patients at each sampling time are given in Table 2.

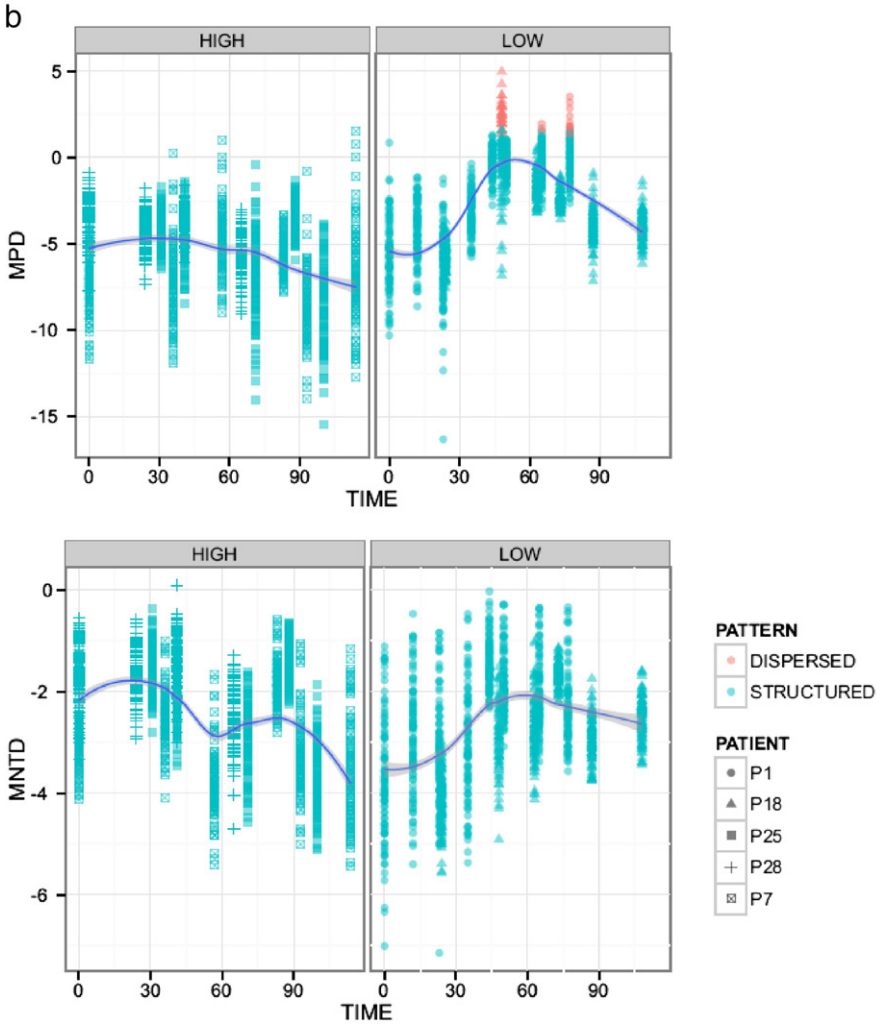
Accordingly to the *p* distance analyses, sequence positional entropy was also distributed quite unevenly along the E1/E2 sequences. The 27 HVR1 amino acids cluster (amino acid positions 384–410) displayed a higher score than HVR2 (amino acid positions 474–482) and HVR3 (amino acid positions 431–466) scores. Furthermore, the entropy along HVR1 positions was patient-dependent (Fig. 5). For P1, P7 and P25 very low entropy values were observed for residues 397, 398, and 408–410, which are critical for cell entry mediated by binding of the E2 protein to the scavenger receptor class B (Guan et al. 2012). The amino acid diversity in the region spanning positions 399–407 also depicted a patient-dependent pattern. P1, P18, and P28 showed higher values than P7 and P25. This subdomain is dispensable for HCV entry and contains the sole neutralizing epitope. Residues 1–13 can affect HCV infectivity and may influence the binding of HVR1specific antibodies, but are also dispensable for HCV entry (Guan et al. 2012). Similar levels of entropy were observed at this subdomain, except for P25.

3.4. Longitudinal evaluation of the intra-host selective pressure

The presence and extent of selective pressures exerted on HCV E1/E2 sequences were longitudinally examined by using codon-by-codon based test. All the patients studied exhibited positive selection at

Fig. 3. Maximum likelihood tree for the HCV haplotypes obtained from all patients along the follow-up. Patient sources are coded by symbols (Patient 1: closed circles; Patient 7: squares; Patient 18: open circles; Patient 25: triangles; Patient 28: diamonds). Sampling times are depicted by the symbols' colors as follows: A: red, B: blue, C: green, D: fuchsia, E: turquoise, F: open circles (except for Patient 1. for who time F was coded by closed. Black circles), G: gray, and H: yellow. Branch lengths are proportional to the number of substitutions per aligned position. Numbers close to branch correspond to SH support values. The tree was midpoint rooted. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)


a
P1
P7
P18
P25
P28
MNTD
MPD
HAPLOTYPES
DISPERSED
STRUCTURED
TIME
A
B
C
D
E
F
G
H


b
HIGH
LOW
MPD
TIME
HIGH
LOW
MNTD
TIME
PATTERN
DISPERSED
STRUCTURED
PATIENT
P1
P18
P25
P28
P7

selected amino acid residues of HVR1 and, to a lesser extent of HVR3. Remarkably, two patients (P1, P18) displayed more prominent (regarding the number of amino acid residues affected) positive selection exerted at the HVR1 and HVR3 domains in concomitance with their more profound CD4 + T cell restoration ($p = 0.001$; Supplementary Table 1) and reduction of HCV viral load (Tables 2 and Supplementary Table 1).

3.5. Recombination analyses

Two methods, the parsimony (or PHI) test and RECCO, were used to evaluate the occurrence of recombination along the genomic region studied. The PHI test detected evidence of recombination for the five patients ($p < 0.002$), which was further supported by the RECCO analyses. Overall, our analyses revealed a quite low intra-patient recombination rate (Table 3) in comparison to previous studies (Palmer et al. 2014; Sentandreu et al. 2008). However, such frequencies could have been underestimated due to the shortness of the sequences studied.

Based on the inferred recombinant breakpoints positions among the patients studied, they would have been mainly located between HVR1 and HVR3 in 11 out of 20 events; the remaining ones were positioned within the HVR2 and HVR3 coding regions. In no case the HVR1 was involved. Sequences generated by recombination preserve the polyprotein reading frame, which is not expected in PCR-mediated recombination where selection cannot purge deleterious mutants. Only in a single case there was a coincident start and end breakpoints in two recombinants in a host (p18_B_125 and p18_C_ 445). In addition, for each subject the intra-subtype recombination appears to occur at different times during infection. A plausible recombinant (p28_A_190) was detected only in P28 prior to HAART initiation while in the remaining ones, the recombinants emerged once (P1), twice (P18, P25), or thrice (P7) during antiretroviral therapy (Table 3). Such results support that they are not experimental artifacts.

4. Discussion

We analyzed longitudinally the intra-host molecular evolution of HCV in HIV-coinfected patients starting previous to HAART initiation. The evolution of HCV quasispecies in HIV/HCV-coinfected patients according to the pattern of response to HAART was previously studied (Babik and Holodniy 2003; Bernini et al. 2011; Sede et al. 2014a; Solmone et al. 2006). However in these reports the HCV characterization was performed by molecular cloning of PCR products followed by Sanger sequencing. So, at the present work, in order to gain a greater level of sequence depth by using next-generation sequencing, variants of the E1/E2 HCV genomic region were characterized from five chronically infected patients who were followed for 5.6 to 9.5 years.

Phylogenetic analyses conducted in this study showed that HCV quasispecies exhibited dissimilar temporal structuring among the five patients studied. Among all ones such structuring was statistically significant after its measurement by applying community phylogenetic metrics (Webb et al. 2002). Interestingly, MPD and MNTD effects for the patients that presented the higher HCV loads (P7, P25 and P28) displayed a steady tendency toward smaller values along time. Conversely, for patients that presented the lower HCV viral loads and a sustained recovery of their CD4 + T-cell populations (P1 and P18), structuring decreased up to an inflexion point that occurred approximately by the time these patients reached the 500 cells per microliter threshold (Figs. 2, 4a, and S1). This last observation strongly suggested that the recovering CD4 + T-cell populations in P1 and P18 induced afresh immune responses against the higher HCV replicators after starting HAART, resulting in the detection of HCV variants that otherwise would have remained hidden –including those generated by genetic recombination- due to the overabundance of higher replicators, and thus in increasingly dispersed phylogenetic pattern. Once the host immune response is recovered (≥500 CD4 + T-cells/mm^3), the virus adapts to the afresh immune response and the viral population is shaped by competition. Hence, successive generations of high replicators dominate the population, resulting in a sustained increase in temporal structuring from this point (Figs. 2 and 4b). As the immune response is expected to preferentially target the more abundant HCV lineages, this process would result in a flatten haplotype distribution, characterized by the presence of several lineages intermingled all along the intra-patient phylogenies and along several sampling times. Conversely, for those patients with relatively high HCV viral loads and more stable CD4 + T-cell populations (P7, P25 and P28), the virus possibly have completely adapted to the immune pressures reaching thus maximum carrying capacities. We believe that a similar scenario might have taken place after P1 and P18 reached a CD4 + T-cell count threshold, possibly 500 cells per microliter, hence structuring increased since this point of maximal dispersal. Interestingly, selection seemed to be intermittent along time (Table 2), which apparently was reflected by the topological anomalies observed for P25 and P7. This strongly suggests that HCV populations at carrying capacity might evolve by natural selection (positive selection detected), random drift (positive selection not detected) or both. Furthermore, this also suggests that, despite the high HCV viral loads observed for P7, P25 and P28, the corresponding effective population sizes might be small enough to result in pronounced genetic drifting. The E2 domains were subjected to differential levels of intra-host selective pressure. Positive selection and amino acid sequence variability were the highest in HVR1. The intra-host positive selection pressure was detected in all patients at almost all sampling times during the follow-up mainly involving HVR1 and HVR3 amino acid residues. However, it was more prominent for those patients showing a sustained recovery of their CD4 + T-cell populations (P1 and P18). Intra-host selective pressure can be exerted at least in part by host HCV-specific immune responses, where both cellular and humoral immunity are involved. Moreover, positive selection associated with cytotoxic T lymphocyte escape was observed in chronic HCV infection (Uebelhoer et al. 2008) that could explain the intra-host amino acid sequence variability. Besides immune reconstitution other factors such as stage of HCV infection, history of innate and adaptive responses to HCV infection, duration of HIV coinfection, and stage of HIV infection before therapy, as well as host genetic factors, should be considered. The steady high level of HCV viral load and the predominance of the negatively selected E1/E2 variants in all subjects suggest a deep adaptation of the viral populations to their host (Ramachandran et al. 2011). The minor modification in HCV viral load levels observed during the study period likely reflects the absence of variations in the effectiveness of neutralizing immune responses, and therefore, could be associated with inefficient HCV clearance rather than with optimal replicative fitness. Nevertheless, neither CD4 + T cell counts nor HCV viral loads were significantly different when positive selection pressure was detected, as reported previously by our group (Sede et al. 2014a).

In vivo, we have observed a low frequency of HCV intra-subtype recombinant haplotypes that supports the hypothesis of detrimental

Fig. 4. (a) Standardized effects of phylogenetic structure on Mean Pairwise Distances (MPD) and Mean Nearest Taxon Distances (MNTD) among sampling times for the five patients studied here (P1, P7, P18, P25 and P28). Significantly structured patterns ($p < 0.01$) are depicted as indicated in the figure legend. The analyses were multiplied along 100 bootstrap trees with the aim of capturing the potential effect of phylogenetic uncertainty. Samples A-H were taken at baseline (A) and at different, successive moments after starting HAART: *P1* months 12, 23, 35, 44, 50, 65 and 77; *P7* months 36, 57, 83, 93 and 114; *P18* months 24, 48, 63, 73, 87 and 108; *P25* months 8, 31, 40, 71, 88 and 100; *P28* months 24, 41 and 65. **(b)** Longitudinal trends of standardized effect sizes of phylogenetic structure on Mean Pairwise Distances (MPD) and Mean Nearest Taxon Distances (MNTD) among sampling times (TIME, in months) for the five patients studied here (P1, P7, P18, P25 and P28). Patients were classified in two blocks (High/Low) based on the observed HCV loads (please see Fig. 1. Significantly structured patterns ($p < 0.01$) are depicted as indicated in the legend. The analyses were multiplied along 100 bootstrap trees with the aim of capturing the potential effect of phylogenetic uncertainty (each point plotted corresponds to a single bootstrap tree). The fitted lines (blue lines) and confidence bounds (gray shaded bands) were obtained by local regression. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Estimate of selective pressure (expressed as dN/dS), amino acid sequence variability (as mean of pairwise *p* distances), and weighted phylogenetic diversity and haplotype richness on HCV E1/E2 region from each patient during the follow-up.

Patient	dN/dS		P-distance[c]	w-PD[d]	HR[e]
	Codon-by-codon[a] Total	Prob (Stat)[b] Total HVR1			
1A	419;451;477	1 (−2.56)	0.12 (1.18)	0.025 (0.006)	0.23009	124
1B	**399**;**404**;477	1 (−1.68)	0.13 (1.13)	0.018 (0.006)	0.20011	140
1C	**394**;470	1 (−2.55)	0.49 (0.02)	0.023 (0.006)	0.14791	50
1D	371;434;468;469	1 (−1.73)	0.34 (0.40)	0.036 (0.009)	0.36820	169
1E	**393**;**396**;**399**;**402**	1 (−1.24)	0.01 (2.30)	0.044 (0.01)	0.12719	34
1F	**401**;463	1 (−1.99)	0.04 (1.81)	0.042 (0.01)	0.18927	51
1G	382;436;464;479	1 (−2.10)	0.25 (0.67)	0.041 (0.01)	0.17580	130
1H	438;462	1 (−2.64)	0.43 (0.17)	0.036 (0.009)	0.23660	79
7A	**386**	1 (−2.75)	0.29 (0.55)	0.015 (0.003)	0.14332	31
7B	**385**;**407**;**410**	1 (−1.63)	0.11 (1.25)	0.033 (0.008)	0.19307	56
7C	**407**	1 (−1.12)	1 (−0.12)	0.037 (0.008)	0.17274	54
7D	430	0.29 (0.56)	1 (−0.13)	0.012 (0.004)	0.09624	18
7G	**385**;**388**;443;444;464	1 (−2.87)	0.13 (1.12)	0.027 (0.005)	0.16363	25
7H	382;439;440;460	1 (−0.86)	0.12 (1.19)	0.055 (0.012)	0.61431	125
18A	**402**;**408**	1 (−1.11)	0.009 (2.37)	0.021 (0.005)	0.25640	39
18B	**392**;**403**;**408**;**409**;422;432	1 (−1.44)	0.19 (0.86)	0.021 (0.005)	0.52779	35
18C	**392**;**393**;**404**;**410**;434;474	1 (−1.74)	0.27 (0.61)	0.057 (0.01)	0.51917	33
18D	**389**;443	1 (−2.52)	0.25 (0.67)	0.020 (0.003)	0.18103	37
18E	**407**	1 (−0.54)	0.03 (1.84)	0.022 (0.005)	0.23798	25
18F	–	1 (−4.69)	1 (−1.17)	0.020 (0.004)	0.155078	23
25A	**400**	1 (−0.99)	0.27 (0.61)	0.011 (0.001)	0.19019	32
25B	438;453	0.41 (0.22)	0.07 (1.41)	0.023 (0.005)	0.33640	68
25C	–	1 (−1.84)	0.350 (0.38)	0.025 (0.007)	0.42860	38
25D	–	1 (−0.53)	1 (−0.51)	0.012 (0.002)	0.21641	2
25E	**391**;450;463	1 (−0.63)	0.165 (0.97)	0.031 (0.006)	0.50690	40
28A	374;**385**;411;443	0.20 (0.81)	0.0004 (3.47)	0.041 (0.009)	0.68285	72
28B	372	1 (1.95)	0.04 (1.70)	0.027 (0.004)	0.56821	22
28C	–	1 (1.42)	0.02 (2.07)	0.030 (0.005)	0.63369	41
28D	435	1 (−1.93)	0.03 (1.78)	0.034 (0.007)	0.50622	28

[a] Maximum Likelihood analysis of natural selection codon-by-codon in the total E1/E2 amplicon (named as "Total"). Only amino acid positions that have undergone positive selection are shown (those bolded are into HVR1). Such condons exhibit significant values of P (the probability of rejecting the null hypothesis of neutral evolution). Maximum Likelihood computations of dN and dS were conducted using HyPhy software package (Pond et al. 2005).

[b] Codon-based test of positive selection for analysis averaging over all sequence pairs at intra-host level for both the "Total" and the "HVR1" domain. The probability of rejecting the null hypothesis of strict-neutrality (dN = dS) in favor of the alternative hypothesis (dN > dS) (Prob values) is shown. The test statistic (dN - dS) is shown in the Stat values between parentheses. The variance of the difference was computed using the bootstrap method (100 replicates) and those values statistically significant (a significant probability of positive selection) are underlined ($p < 0.05$, Z test).

[c] Estimates of average uncorrected genetic distances (p) over all E1/E2 sequence pairs. The number of base differences per site from averaging over all sequence pairs is shown. Standard error estimate(s) are shown between parentheses and were obtained by a bootstrap procedure (100 replicates).

[d] Weighted phylogenetic diversity (w-PD).

[e] Haplotype richness (HR) along the follow up as obtained from the maximum likelihood evidence from E1/E2 nucleotide sequences.

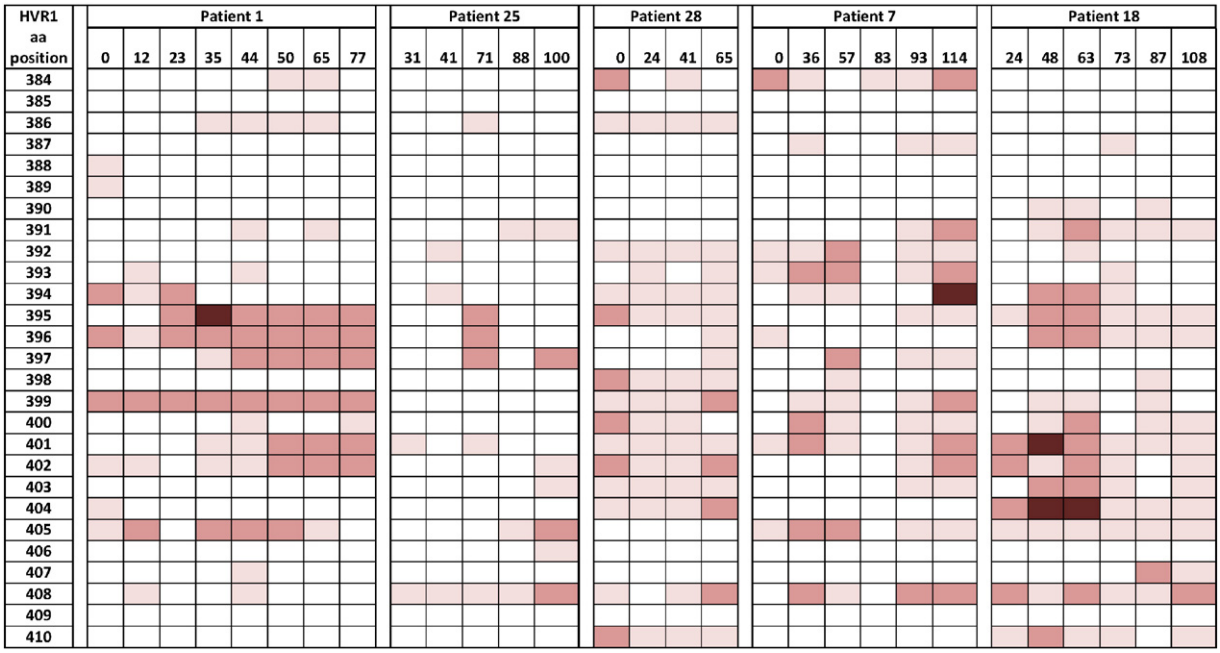


Fig. 5. Heat maps showing amino acid sequence diversity in HVR1 domain. Variability at each amino acid position was computed as described in Material & methods. Amino acid positions correspond to the AF009606 reference sequence (H77 isolate). The relative frequency of individual entropy values per amino acid is shown in a color scale as follows: □ □ ■ ■. Values on top of the columns represent the sampling times (months).

effect on viral fitness making these recombinant variants in disadvantage to compete against dominant parental strains for replication space within the quasispecies composition. To date, there are three reports on putative HCV intra-subtype recombinant strains, detected by analyzing the NS5A or E2/E1 sequences (Gao et al. 2007; Palmer et al. 2014; Scheel et al. 2013; Sentandreu et al. 2008). The nearby breakpoints identified appeared to be distributed randomly in HCV samples from different patients at different sampling times, supporting the presence of potential recombination hotspots. These were mainly located in the region between HVR1 and HVR3 that exhibit a high sequence identity that may facilitate the template switching of the polymerase complex ("copy choice") during the recombination event.

In conclusion, this longitudinal study shows the impact of HAART-induced CD4 + T-cell recovery in shaping intra-host HCV evolution. A host-dependent HCV temporal structuring was evidenced, where the contribution of the intra-subtype recombination was scarce. Such scenarios also depicted dissimilar magnitudes of positive selection pressure mainly focused on the HVR1 domain. A limitation of our study is the low number of patients analyzed. However, this study had the advantages of implementing high throughput sequencing technologies and potent sequence analysis tools. Notwithstanding that, further studies using a larger number of patients are deserved.

Acknowledgements

This work was supported by the Argentinean National Agency of Scientific and Technological Promotion (grant number PICT 2012-0422 to JQ and PICT PRH-14 120 to LRJ); the University of Buenos Aires

Table 3
Recombination inference analysis by Recco software showing breakpoints positions (Start and End). Sequence p-values <0.05 were considered as significant.

Sequence ID	Start[a]	End[a]	Savings[b]	Seq pv[c]
p1_D_554	172	236	5	0.023
p1_D_173	247	257	3	0.033
p7_C_232	358	416	5	0.002
p7_F_181	291	356	3.2	0.002
p7_F_235	191	245	2.8	0.059
p7_F_559	268	296	2.2	0.003
p7_F_367	171	267	2.2	0.010
p7_B_207	209	296	2	0.027
p7_F_157	181	265	2	0.037
p18_C_179	178	239	9	0.001
p18_B_38	182	217	5	0.002
p18_B_102	182	310	3	0.004
p18_B_437	218	310	3	0.006
p18_B_125	178	209	3	0.010
p18_B_237	301	324	3	0.021
p18_C_441	178	209	3	0.040
p25_E_89	201	422	3	0.029
p25_C_267	267	422	3	0.034
p28_A_190	191	267	4.2	0.001
p28_C_309	172	279	4	0.030

[a] Nucleotide position according to reference sequence (HCV isolate H77; GenBank accession no. AF009606).
[b] Number of mismatches saved when explaining the read by a recombination event between two viral strains and mutations, rather than by a single strain and mutations only.
[c] P-value of each recombination event regarding the savings distribution.

(grant number 20020110100034); and the Argentinean National Scientific and Technical Research Council (grant number PIP-11220110101089).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.meegid.2016.05.032.

References

Babik, J.M., Holodniy, M., 2003. Impact of highly active antiretroviral therapy and immunologic status on hepatitis C virus quasispecies diversity in human immunodeficiency virus/hepatitis C virus-coinfected patients. J. Virol. 77, 1940–1950.

Bankwitz, D., Vieyres, G., Hueging, K., Bitzegeio, J., Doepke, M., Chhatwal, P., Haid, S., Catanese, M.T., Zeisel, M.B., Nicosia, A., Baumert, T.F., Kaderali, L., Pietschmann, T., 2014. Role of hypervariable region 1 for the interplay of hepatitis C virus with entry factors and lipoproteins. J. Virol. 88, 12644–12655.

Bernini, F., Ebranati, E., De Maddalena, C., Shkjezi, R., Milazzo, L., Lo Presti, A., Ciccozzi, M., Galli, M., Zehender, G., 2011. Within-host dynamics of the hepatitis C virus quasispecies population in HIV-1/HCV coinfected patients. PLoS One 6, e16551.

Bittar, C., Jardim, A.C., Yamasaki, L.H., Carareto, C.M., Pinho, J.R., Lemey, P., de Carvalho-Mello, I.M., Rahal, P., 2013. On hepatitis C virus evolution: the interaction between virus and host towards treatment outcome. PLoS One 8, e62393.

Brimacombe, C.L., Grove, J., Meredith, L.W., Hu, K., Syder, A.J., Flores, M.V., Timpe, J.M., Krieger, S.E., Baumert, T.F., Tellinghuisen, T.L., Wong-Staal, F., Balfe, P., McKeating, J.A., 2011. Neutralizing antibody-resistant hepatitis C virus cell-to-cell transmission. J. Virol. 85, 596–605.

Bruen, T.C., Philippe, H., Bryant, D., 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172, 2665–2681.

Bushman, F.D., Hoffmann, C., Ronen, K., Malani, N., Minkah, N., Rose, H.M., Tebas, P., Wang, G.P., 2008. Massively parallel pyrosequencing in HIV research. AIDS 22, 1411–1415.

Clausen, L.N., Lundbo, L.F., Benfield, T., 2014. Hepatitis C virus infection in the human immunodeficiency virus infected patient. World J. Gastroenterol. 20, 12132–12143.

Di Giallonardo, F., Zagordi, O., Duport, Y., Leemann, C., Joos, B., Kunzli-Gontarczyk, M., Bruggmann, R., Beerenwinkel, N., Gunthard, H.F., Metzner, K.J., 2013. Next-generation sequencing of HIV-1 RNA genomes: determination of error rates and minimizing artificial recombination. PLoS One 8, e74249.

Felsenstein, J., 1989. PHYLIP - phylogeny inference package (version 3.2). Cladistics 5, 164–166.

Gao, F., Nainan, O.V., Khudyakov, Y., Li, J., Hong, Y., Gonzales, A.C., Spelbring, J., Margolis, H.S., 2007. Recombinant hepatitis C virus in experimentally infected chimpanzees. J. Gen. Virol. 88, 143–147.

Gregori, J., Salicru, M., Domingo, E., Sanchez, A., Esteban, J.I., Rodriguez-Frias, F., Quer, J., 2014. Inference with Viral Quasispecies Diversity Indices: Clonal and NGS Approaches (Bioinformatics).

Guan, M., Wang, W., Liu, X., Tong, Y., Liu, Y., Ren, H., Zhu, S., Dubuisson, J., Baumert, T.F., Zhu, Y., Peng, H., Aurelian, L., Zhao, P., Qi, Z., 2012. Three different functional microdomains in the hepatitis C virus hypervariable region 1 (HVR1) mediate entry and immune evasion. J. Biol. Chem. 287, 35631–35645.

Helle, F., Duverlie, G., Dubuisson, J., 2011. The hepatitis C virus glycan shield and evasion of the humoral immune response. Viruses 3, 1909–1932.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.

Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst. Biol. 61, 1061–1067.

Jones, L.R., Sede, M., Manrique, J.M., Quarleri, J., 2016. Virus evolution during chronic hepatitis B virus infection as revealed by ultradeep sequencing data. J Gen Virol 97, 435–444.

Kato, N., Ootsuyama, Y., Sekiya, H., Ohkoshi, S., Nakazawa, T., Hijikata, M., Shimotohno, K., 1994. Genetic drift in hypervariable region 1 of the viral genome in persistent hepatitis C virus infection. J. Virol. 68, 4776–4784.

Katoh, K., Standley, D.M., 2014. MAFFT: iterative refinement and additional methods. Methods Mol. Biol. 1079, 131–146.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., Webb, C.O., 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26, 1463–1464.

Kong, L., Giang, E., Nieusma, T., Kadam, R.U., Cogburn, K.E., Hua, Y., Dai, X., Stanfield, R.L., Burton, D.R., Ward, A.B., Wilson, I.A., Law, M., 2013. Hepatitis C virus E2 envelope glycoprotein core structure. Science 342, 1090–1094.

Korber, B.T., MacInnes, K., Smith, R.F., Myers, G., 1994. Mutational trends in V3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type 1. J. Virol. 68, 6730–6744.

Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22, 1208–1222.

Kwok, S., Higuchi, R., 1989. Avoiding false positives with PCR. Nature 339, 237–238.

Lahr, D.J., Katz, L.A., 2009. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. Biotechniques 47, 857–866.

Martinot-Peignoux, M., Le Breton, V., Fritsch, S., Le Guludec, G., Labouret, N., Keller, F., Marcellin, P., 2000. Assessment of viral loads in patients with chronic hepatitis C with AMPLICOR HCV MONITOR version 1.0, COBAS HCV MONITOR version 2.0, and QUANTIPLEX HCV RNA version 2.0 assays. J. Clin. Microbiol. 38, 2722–2725.

Maydt, J., Lengauer, T., 2006. Recco: recombination analysis using cost optimization. Bioinformatics 22, 1064–1071.

Miro, J.M., Stock, P., Teicher, E., Duclos-Vallee, J.C., Terrault, N., Rimola, A., 2015. Outcome and management of HCV/HIV coinfection pre- and post-liver transplantation. A 2015 update. J. Hepatol. 62, 701–711.

Muse, S.V., Gaut, B.S., 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. 11, 715–724.

Nylander, J.A.A., 2004. MrAIC.pl. Program distributed by the editor. Evolutionary Biology Centre. Uppsala University (Available at) http://www.ebc.uu.se/systzoo/staff/nylander.html.

Palmer, B.A., Dimitrova, Z., Skums, P., Crosbie, O., Kenny-Walsh, E., Fanning, L.J., 2014. Analysis of the evolution and structure of a complex intrahost viral population in chronic hepatitis C virus mapped by ultradeep pyrosequencing. J. Virol. 88, 13709–13721.

Pantua, H., Diao, J., Ultsch, M., Hazen, M., Mathieu, M., McCutcheon, K., Takeda, K., Date, S., Cheung, T.K., Phung, Q., Hass, P., Arnott, D., Hongo, J.A., Matthews, D.J., Brown, A., Patel, A.H., Kelley, R.F., Eigenbrot, C., Kapadia, S.B., 2013. Glycan shifting on hepatitis C virus (HCV) E2 glycoprotein is a mechanism for escape from broadly neutralizing antibodies. J. Mol. Biol. 425, 1899–1914.

Pascual-Pareja, J.F., Caminoa, A., Larrauri, C., Gonzalez-Garcia, J., Montes, M.L., Diez, J., Grande, M., Arribas, J.R., 2009. HAART is associated with lower hepatic necroinflammatory activity in HIV-hepatitis C virus-coinfected patients with CD4 cell count of more than 350 cells/microl at the time of liver biopsy. AIDS 23, 971–975.

Pond, S.L., Frost, S.D., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676–679.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490.

Ramachandran, S., Campo, D.S., Dimitrova, Z.E., Xia, G.L., Purdy, M.A., Khudyakov, Y.E., 2011. Temporal variations in the hepatitis C virus intrahost population during chronic infection. J. Virol. 85, 6369–6380.

Ray, S.C., Fanning, L., Wang, X.H., Netski, D.M., Kenny-Walsh, E., Thomas, D.L., 2005. Divergent and convergent evolution after a common-source outbreak of hepatitis C virus. J. Exp. Med. 201, 1753–1759.

Sala, M., Wain-Hobson, S., 2000. Are RNA viruses adapting or merely changing? J. Mol. Evol. 51, 12–20.

Scheel, T.K., Galli, A., Li, Y.P., Mikkelsen, L.S., Gottwein, J.M., Bukh, J., 2013. Productive homologous and non-homologous recombination of hepatitis C virus in cell culture. PLoS Pathog. 9, e1003228.

Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864.

Sede, M., Jones, L.R., Moretti, F., Laufer, N., Quarleri, J., 2014a. Inter and intra-host variability of hepatitis C virus genotype 1a hypervariable envelope coding domains followed for a 4-11 year of human immunodeficiency virus coinfection and highly active antiretroviral therapy. Virology 471–473, 19–28.

Sede, M.M., Moretti, F.A., Laufer, N.L., Jones, L.R., Quarleri, J.F., 2014b. HIV-1 tropism dynamics and phylogenetic analysis from longitudinal ultra-deep sequencing data of CCR5- and CXCR4-using variants. PLoS One 9, e102857.

Sentandreu, V., Jimenez-Hernandez, N., Torres-Puente, M., Bracho, M.A., Valero, A., Gosalbes, M.J., Ortega, E., Moya, A., Gonzalez-Candelas, F., 2008. Evidence of recombination in intrapatient populations of hepatitis C virus. PLoS One 3, e3239.

Shimodaira, H.H., M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 3.

Shirai, M., Arichi, T., Chen, M., Masaki, T., Nishioka, M., Ikeda, K., Takahashi, H., Enomoto, N., Saito, T., Major, M.E., Nakazawa, T., Akatsuka, T., Feinstone, S.M., Berzofsky, J.A., 1999. T cell recognition of hypervariable region-1 from hepatitis C virus envelope protein with multiple class II MHC molecules in mice and humans: preferential help for induction of antibodies to the hypervariable region. J. Immunol. 162, 568–576.

Skums, P., Dimitrova, Z., Campo, D.S., Vaughan, G., Rossi, L., Forbi, J.C., Yokosawa, J., Zelikovsky, A., Khudyakov, Y., 2012. Efficient error correction for next-generation sequencing of viral amplicons. BMC Bioinf. 10 (13 Suppl.), S6.

Solmone, M., Girardi, E., Lalle, E., Abbate, I., D'Arminio Monforte, A., Cozzi-Lepri, A., Alessandrini, A., Piscopo, R., Ebo, F., Cosco, L., Antonucci, G., Ippolito, G., Capobianchi, M.R., 2006. Evolution of HVR-1 quasispecies after 1-year treatment in HIV/HCV-coinfected patients according to the pattern of response to highly active antiretroviral therapy. Antivir. Ther. 11, 87–94.

Suzuki, Y., Gojobori, T., 1999. A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol. 16, 1315–1328.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Troesch, M., Meunier, I., Lapierre, P., Lapointe, N., Alvarez, F., Boucher, M., Soudeyns, H., 2006. Study of a novel hypervariable region in hepatitis C virus (HCV) E2 envelope glycoprotein. Virology 352, 357–367.

Uebelhoer, L., Han, J.H., Callendret, B., Mateu, G., Shoukry, N.H., Hanson, H.L., Rice, C.M., Walker, C.M., Grakoui, A., 2008. Stable cytotoxic T cell escape mutation in hepatitis C virus is linked to maintenance of viral fitness. PLoS Pathog. 4, e1000143.

Wahid, A., Dubuisson, J., 2013. Virus-neutralizing antibodies to hepatitis C virus. J. Viral Hepat. 20, 369–376.

Webb, C.O., Ackerly, D.D., McPeek, M.A., Donoghue, M.J., 2002. Phylogenies and community ecology. Annu. Rev. Ecol. Syst. 33, 475–505.

Webster, D.P., Klenerman, P., Dusheiko, G.M., 2015. Hepatitis C. Lancet 385, 1124–1135.

Yagnik, A.T., Lahm, A., Meola, A., Roccasecca, R.M., Ercole, B.B., Nicosia, A., Tramontano, A., 2000. A model for the hepatitis C virus envelope glycoprotein E2. Proteins 40, 355–366.

Zhang, J., Rosenberg, H.F., Nei, M., 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. U. S. A. 95, 3708–3713.