This is a pre- or post-print of an article published in
Dynamic changes in viral population structure and compartmentalization during chronic hepatitis C virus
infection in children
Dynamic changes in viral population structure and compartmentalization during chronic hepatitis C virus infection in children

María Inés Gismondi\textsuperscript{a}\textsuperscript{1,*}, Juan María Díaz Carrasco\textsuperscript{a}\textsuperscript{*,} Pamela Valva\textsuperscript{a}, Pablo Daniel Becker\textsuperscript{c}\textsuperscript{2}, Carlos Alberto Guzmán\textsuperscript{c}, Rodolfo Héctor Campos\textsuperscript{b}, María Victoria Preciado\textsuperscript{a}\textsuperscript{#}

\textsuperscript{a}Laboratorio de Biología Molecular, División Patología, Hospital de Niños Ricardo Gutiérrez, Buenos Aires, Argentina; \textsuperscript{b}Cátedra de Virología, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Ciudad de Buenos Aires, Argentina; and \textsuperscript{c}Department of Vaccinology and Applied Microbiology, Helmholtz Centre for Infection Research, Braunschweig, Germany.

\textsuperscript{1} Present address: Instituto de Biotecnología, CICVyA, INTA, Hurlingham, Argentina.

\textsuperscript{2} Present address: Department of Immunobiology, King’s College London, London, UK.

* These authors contributed equally to this work.

# Corresponding author. Lic. Juan M. Díaz Carrasco. Laboratorio de Biología Molecular, División Patología, Hospital de Niños Ricardo Gutiérrez, Gallo 1330, C1425EFD Ciudad de Buenos Aires, Argentina. Phone: +5411-4962-9138. Fax: +5411-4963-4122. E-mail: diazcarrascojuan@yahoo.com.ar

Author’s e-mail addresses:

M.I. Gismondi mgismondi@cnia.inta.gov.ar

J.M. Díaz Carrasco diazcarrascojuan@yahoo.com.ar

P. Valva valvapamela@yahoo.com
P.D. Becker pablo.becker@kcl.ac.uk

C.A. Guzmán carlos.guzman@helmholtz-hzi.de

R.H. Campos rcampos@ffyb.uba.ar

M.V. Preciado preciado@conicet.gov.ar
Abstract

Classic phylogenetic and modern population-based clustering methods were used to analyze hepatitis C virus (HCV) evolution in plasma and to assess viral compartmentalization within peripheral blood mononuclear cells (PBMCs) in 6 children during 3.2 to 9.6 years of follow-up. Population structure analysis of cloned amplicons encompassing hypervariable region 1 led to the distinction of two evolutionary patterns, one highly divergent and another one genetically homogeneous. Viral adaptability was reflected by co-evolution of viral communities switching rapidly from one to another in the context of divergence and stability associated with highly homogeneous communities which were replaced by new ones after long periods. Additionally, viral compartmentalization of HCV in PBMCs was statistically demonstrated, suggesting their role as a pool of genetic variability.

Our results support the idea of a community-based structure of HCV viral populations during chronic infection and highlight a role of the PBMC compartment in the persistence of such structure.

Keywords: compartmentalization; evolution; hepatitis C virus; PBMC; pediatric infection; population structure
Introduction

Hepatitis C virus (HCV) produces a chronic liver infection that may lead to cirrhosis and hepatocellular carcinoma, which are the main indications for liver transplantation among adult patients in developed countries (1-2). However, the outcome of HCV infection acquired in childhood is uncertain, and HCV-associated liver disease shows a milder clinical course with a slower progression to fibrosis (3).

HCV is a small enveloped ~9.6 kb positive stranded RNA virus of the Flaviviridae family. The viral envelope consists of a lipid bilayer derived from the host cell in which the two glycoproteins E1 and E2 are anchored as a heterodimer (4). Both glycoproteins are essential for HCV entry into hepatocytes (5). Hypervariable region 1 (HVR1), encompassing the first 27 amino acids at the N-terminal region of E2, is the most variable region of the HCV genome (6) and it has been widely used in studies examining intrahost evolutionary dynamics (33, 46-53, 63). As previously demonstrated both in vitro and in vivo, HVR1 contains a linear neutralizing epitope (7-8) and thus it is subjected to immune selective pressure of the host. Despite its variability, this region constitutes a basic stretch, in which the positively charged residues occupy specific positions (9). Accordingly, HVR1 would be implicated in interactions with negatively charged molecules present on the cellular surface or on plasma lipoproteins to which HCV associates (9-10). Indeed, HVR1 has been related to the binding with the CD81 receptor on the cellular surface (11), and its relevance for functional interaction of E2 with the lipoprotein receptor SR-B1 has been demonstrated (10, 12-14).

HCV replicates primarily in the liver, although extrahepatic localizations including circulating lymphoid cells which form part of the peripheral blood mononuclear cells (PBMCs) have been described (15-21). Only selected HCV variants from the plethora
of sequences present in plasma would be able to associate with these cells, a phenomenon called viral compartmentalization. Compartmentalized HCV sequences have been demonstrated both in immunocompromised (18, 22-23) and immunocompetent (24-25) individuals, and in the setting of liver transplantation (18, 26-27). Additionally, PBMC-associated HCV sequences have also been detected after resolution of HCV infection thus indicating occult infection (28-29). However, no reports exist about compartmentalization in pediatric patients with HCV infection.

In the infected patient, HCV exists as a mixture of different but phylogenetically related sequences referred to as quasispecies (30). During the course of infection, dynamic changes in viral population size and composition occur, thus facilitating viral escape to the host’s immune response and establishment of persistence (31). Consequently, HCV evolution reflects the continuous adaptation of the virus to the constantly changing selection pressures in the host’s environment. In that sense, compartmentalization could be a mechanism contributing to viral variability, persistence and pathogenesis in the infected host. In this work, classic phylogenetic and modern population-based sequence analysis methods were applied to evaluate if previous observations obtained using samples from adult patients suggesting a community-based structure of viral populations and HCV compartmentalization within PBMC during chronic infection also hold true for pediatric patients.

**Materials and methods**

**Ethics Statement**

A written informed consent was obtained from all participants’ parents. The study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki as
reflected in a priori approval by the Ricardo Gutierrez Children's Hospital Bioethics and Research Committees.

**Subjects and samples**

This retrospective study was conducted on 6 pediatric patients with chronic HCV infection. No patient was co-infected with HBV or HIV. Some patients were enrolled before HCV interferon treatment was widely applied to all infected children, so only 3/6 patients received alpha interferon (α-IFN, 3MU subcutaneously, 3 times a week), which was withdrawn after at least 6 months due to lack of response. Serum alanine aminotransferase (ALT) levels and HCV viral load in plasma were taken from clinical records. Formalin-fixed and paraffin-embedded (FFPE) liver sections were obtained from the files of our hospital’s Pathology Division.

Multiple consecutive blood samples (3-5 ml) were taken from each patient by vein puncture between years 1999 and 2010. Plasma was immediately separated and stored at −80°C. PBMCs (mean number $5.30 \times 10^6 \pm 2.69 \times 10^6$ cells) were isolated by centrifugation on Histopaque® (SIGMA), followed by two washes with sterile phosphate-buffered saline (PBS) 1X. Cells were finally resuspended in sterile PBS 1X and stored at −80°C. The supernatants of the washing steps were aliquoted and frozen at −80°C.

**Liver samples histology.** Inflammatory activity and fibrosis were assessed in FFPE liver sections using the modified Knodell scoring system (Histological Activity Index, HAI) and METAVIR (32). According to HAI, each biopsy specimen was categorized as minimal (≤3), mild (4-6), moderate (7-12) or severe hepatitis (>12).

**Reverse transcription (RT) and PCR amplification.** RNA was extracted from 200 µl plasma and PBMCs with Trizol reagent (Invitrogen). RNA was denaturated at 70°C for
5 min, and a 351 bp fragment encompassing HVR1 was amplified by RT-nested PCR as previously described (33). To rule out contamination of PBMCs with plasma-derived sequences, supernatants from both washing steps during PBMC isolation were tested for HCV RNA by means of a more sensitive amplification reaction of HCV 5’UTR (34).

Cloning of PCR products and sequencing. Purified PCR products were subjected to the addition of 3’A-overhangs using Taq DNA polymerase (Invitrogen), cloned in a pCR2.1 vector (Invitrogen) and transformed into Escherichia coli DH5α-competent cells. Plasmid DNA from transformants producing white colonies was extracted from minipreps cultured overnight by alkaline lysis (35). Purified recombinant plasmid DNA was sequenced bidirectionally in an Automated 3700 DNA Analyzer (Applied Biosystems, Foster City/CA/USA) using the ABI PRISM® BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems).

Sequence analysis. Nucleotide sequences corresponding to all samples from each patient were aligned using CLUSTALX program (36). All analysis listed below were performed from nucleotide alignments (300 bp), which correspond to clonal sequences encompassing HVR1 and excluding oligonucleotide sequences. A total number of 597 clones were analyzed, of which 370 were derived from plasma and 227 from PBMCs.

a. Quasispecies complexity. Global sequence variability at the nucleotide level was estimated from multiple alignments of all plasma or PBMCs-derived sequences using Shannon’s entropy $S = -\sum_i (p_i \ln p_i)$, where $p_i$ is the frequency at which a given residue is found at position i. Genetic complexity (number of different variants within each sample) in plasma and PBMCs was calculated with HVR1 amino acid sequences and expressed as normalized $S (S_N)$, being $S_N = S/\ln N$, where N is the total number of clones in each sample. In this case, $p_i$ is the frequency of each clone in the quasispecies.
$S_N$ theoretically varies from 0 (only one variant detected) to 1 (maximum quasispecies complexity).

b. **Phylogenetic analysis.** (i) **Maximum Likelihood (ML) Trees.** Phylogenetic analysis was performed with PhyML program version 3.0 (37). The most appropriate model of evolution for each multiple alignment was inferred using Akaike’s information criterion, as implemented in jModeltest v.0.1.1 (38) (Supplementary Table 1). ML trees were constructed by a heuristic search with the previously selected model of substitution. Bootstrap analysis (1000 pseudoreplicates) was performed on each ML tree. (ii) **Networks.** Phylogenetic reconstruction was assessed through a maximum-parsimony based heuristic algorithm implemented in Network 1.6.1.0 program using the median-joining method (39). The resulting multitude of plausible trees is expressed as a network which displays alternative potential evolutionary paths in the form of cycles.

c. **Population studies.** The model-based Bayesian clustering approach implemented in BAPS program version 5.4 (40) was used to estimate the number of genetically homogeneous clusters or populations. The prior for the maximum number of clusters (parameter $K$) was set between 1 and 20. Five independent runs for each dataset were performed, and all the partitions obtained had a posterior probability greater than 0.9.

d. **Compartmentalization of viral sequences.** The association between phylogeny and compartmentalization was statistically assessed in a Bayesian framework implemented in BaTS program (41), which takes into account the phylogenetic uncertainty in the analysis. The input trees were obtained using BEAST software version 1.6.2 (42). All Monte Carlo Markov Chains (MCMC) were run for $5 \times 10^7$ generations in order to achieve an Effective Sample Size (ESS) >200. The substitution rate parameter was estimated by MCMC analysis, which was run under two molecular
clock models: Strict and Relaxed Uncorrelated Lognormal. The best clock model was
selected by Bayes Factor. For BaTS analysis, two states were defined according to the
compartment from which the sequences were obtained (plasma or PBMCs) in order to
calculate three compartmentalization indexes: Parsimony Score (PS), Association Index
(AI) and Monophyletic Clade (MC). The expected value of the indexes under the non-
association hypothesis was estimated by 500 randomized sets.

e. **Selection analysis.** Individual codons under positive or negative selection were
detected through an integrative analysis taking into account the results of four different
selection methods (SLAC, FEL, REL and MEME) implemented in the DataMonkey
webserver (43-45). The best-fitting model of nucleotide substitution was used. An alpha
value of 0.05 was used to assess significance, and selection was considered when at
least two methods matched a significant result.

**Nucleotide sequence accession numbers.** The GenBank/EMBL/DDBJ accession
numbers of the sequences determined in this work are EU0498394 to EU498673,
KC175915 to KC176072, and EU045934 to EU046001.

**Results**

**Clinical characterization of chronic infection during follow-up**

Six children with chronic HCV infection were included in this study. Clinical,
biochemical and virological features of the patients studied are shown in Table 1 and
Figure 1. Three patients received α-IFN treatment during follow-up. Serum ALT levels
were normal or slightly elevated in patients not undergoing treatment (cases 1-3),
whereas they proved moderately to highly elevated in cases 4-6, rising approximately
three times during α-IFN therapy. Viral load in plasma was continuously elevated in
samples from all patients, whether treated or not.
Regarding pathogenesis, liver biopsies obtained during follow-up showed predominantly minor liver damage in all patients (Supplementary Table 2). Hepatitis ranged from mild to moderate and fibrosis was mostly low. Steatosis was higher in treated patients. Besides, intralobular lymphoid follicles were observed only in three samples from different patients. Lastly, all liver samples showed a similar proportion of infected hepatocytes (11-22%, data not shown). Taken together, these observations indicate that liver damage in this cohort is predominantly mild, in accordance to previous reports on pediatric cohorts (3).

**Evolutionary dynamics in plasma**

The existence of multiple viral “communities” co-evolving in the infected host has been suggested recently (46). This idea prompted us to investigate the genetic structure of HCV populations within infected hosts using a Bayesian analysis of population structure (BAPS). To this end, an HVR1-encompassing fragment was amplified from consecutive plasma samples and molecular clones were obtained and sequenced (Supplementary Table 3). The phylogenetic clusters (communities) inferred by BAPS were superimposed to the corresponding maximum likelihood (ML) trees describing the phylogenetic relationships between clonal sequences from each patient (Figure 2). Both approaches showed consistent results since most of the communities obtained overlapped well supported clades reconstructed by ML. However, BAPS analysis allowed identification of independent populations otherwise not evident.

Taking the results of both types of analysis together, two levels of viral evolution were observed within infected hosts. On the one hand, a small number of viral communities co-evolved throughout follow-up in all patients, switching intermittently from one to another (cases 1 to 4) or being replaced by a new community after a long period of
genetic stability in some cases (cases 5 and 6, see the communities’ frequency plots in Figure 2). On the other hand, a more detailed observation of the genetic structure of HCV populations showed the continuous evolution of viral sequences belonging to each community, probably reflecting viral adaptation to particular environmental conditions within the infected host at each sampling time point. In this sense, α-IFN treatment did not abrogate the viral communities present in treated patients neither during therapy nor after its withdrawal (Figure 2, cases 4 to 6), reflecting the existence of “naturally resistant” HCV variants in them. Most interestingly, the highly homogeneous clusters that emerged in patients 5 and 6 towards the end of the studied period formed new viral communities that replaced the pre-existing one; which until month 26 and 61, respectively, had been the only community present in each infected child.

Compartmentalization of viral sequences. Viral sequences associated to PBMCs were analyzed in available samples to study compartmentalization during follow-up. PBMC-derived HCV RNA was detected in all patients, whereas it was undetectable in the supernatants of two washing steps performed during PBMC isolation. Moreover, amplification and molecular cloning of the HVR1-encompassing fragment from plasma and PBMC samples allowed the identification of unique sequences in each compartment in certain samples, ruling out contamination of PBMC with plasma-derived sequences. Both plasma and PBMC samples showed a high number of variable sites (176/300 and 107/300 sites, respectively) which were mostly concentrated within the HVR1 region (Figure 3A and Supplementary Figure 2), thus demonstrating the heterogeneous character of the viral quasispecies present in both compartments. Intrasample HVR1 complexity was lower in PBMCs than in plasma at most samples (Figure 3B), denoting the existence of more homogeneous variants in PBMCs. Nevertheless, the amino acid
profile of this stretch was similar in both types of samples (Figure 3C), ruling out the existence of an amino acid signature related to HCV compartmentalization at least within the HVR1 region. Lastly, a deep analysis of sites subject to selection both in plasma and PBMCs demonstrated a concentration of positively selected sites within HVR1, particularly in cases 1 to 4. Regarding purifying selection, negatively selected sites predominated outside HVR1 (Supplementary Figure 2).

As previously reported (33), almost no HVR1 variability was detected in cases 5 and 6 both in plasma and PBMCs (data not shown). Noteworthy, in case 6 there was an increment in complexity around 115 months of follow-up, which was concurrent with a worsening of the clinical status of this patient, as perceived in the liver biopsy (Supplementary Table 2).

The detection of different viral sequences in plasma and PBMCs does not unequivocally demonstrate compartmentalization. Thus, the overall association between PBMC compartmentalization and phylogeny was statistically assessed in a Bayesian framework (BaTS). Both the Association Index (AI) and the Parsimony Score (PS) showed statistically significant differences (p<0.001) with the null distribution (no association), indicating compartmentalization of PBMC-associated sequences in all cases (Table 2).

Of note, the compartment-specific Monophyletic Clade (MC) index showed a significant association with PBMCs in all cases, which denotes compartmentalization of phylogenetically related sequences. This result was further confirmed by the Slatkin-Maddison test (47) (data not shown).

To better characterize PBMC-associated viral variants, their relation with plasma-derived sequences and their dynamic behavior throughout follow-up in our patients, two types of phylogenetic analyses were carried out. The first consisted in a maximum parsimony-based network approach to study the phylogenetic link between plasma and
PBMC-derived sequences. As previously mentioned by other authors (46), this kind of analysis accurately represents viral evolution, particularly when using closely related sequences, since this method takes the phylogenetic uncertainty into account. Nevertheless, phylogeny was also reconstructed by ML to check whether there is an inconsistency in the tree topology between both methods. After the comparative analysis minor differences were found (data not shown). The second analysis was the above mentioned BAPS, which results were superimposed to the corresponding phylogenetic networks (Figure 4). In line with the compartmentalization tests performed in BaTS, in all patients’ networks large clusters of homogeneous sequences derived from PBMCs were observed, which sometimes were very distinct respect to concomitant plasma sequences, thus explaining the statistically significant results obtained by BaTS. However, in some samples, sequences from plasma and PBMCs were closely related or even identical during follow-up. Compartmentalized HCV variants exhibited a dynamic behavior in consecutive paired plasma-PBMC samples of each patient, although without showing a clear phylogenetic association between serial PBMC samples. Together, these results reveal a continuous genetic flow between both compartments during chronic infection and suggest some kind of constraint for lymphotropism imposed to particular viral variants.

As mentioned above, viral sequences present in plasma were grouped into viral communities that evolved simultaneously in the infected patient over time. We hypothesized that a) PBMC-derived sequences could be partitioned into the previously defined viral communities observed in plasma samples (Figure 2) or b) they could constitute independent communities associated with the cellular compartment. The BAPS analysis performed with all sequences from both compartments showed that in the majority of cases PBMC-derived sequences did not form plasma-independent
communities (Figure 4); instead, they were distributed among the previously identified
viral communities along with sequences present in plasma concomitantly or not.

Besides, one community described in Figure 2 (Case 2, Com 3) split into two sub-
communities containing more closely related plasma and/or PBMC derived sequences
(Com 3A and Com 3B in Figure 4), likely due to a more accurate phylogenetic signal
contained in these sequences. Somehow, these homogeneous clusters from PBMCs
allowed further partition of this community. Interestingly, although PBMC-derived
variants were generally associated to the same viral communities previously described
in plasma, they did not exactly mimic their dynamics over time (compare graphics in
Figures 2 and 4). This observation became particularly striking among treated patients 4
and 6, where viral sequences present concomitantly in plasma and PBMCs were
partitioned into different communities, reflecting the pool character of the cellular
compartment and its contribution to viral diversity.

Discussion

Intrahost evolution of the HVR1 region has been extensively studied in adult patients
with chronic HCV infection (46, 48-50), and to a lesser extent in pediatric individuals
(33, 51-54). However, long-term follow-up studies in children in the absence of HIV
co-infection are scarce. Moreover, population structure and compartmentalization of
HCV quasispecies, treated here, had never been thoroughly evaluated in children before.

As already mentioned, this novel kind of analysis of viral sequences more accurately
reflects viral evolution during chronic infection, since it takes into account the whole
population circulating in the infected host at any time point and does not found on an
evolutionary model as classic phylogenetic methods assuming bifurcating trees do [53].
In this study, different evolutionary patterns not clearly associated neither with α-IFN treatment nor with the route of transmission were distinguished. Instead, our results illustrate viral adaptation to a continuously changing environment within the infected host, probably as a result of the host’s immune selective pressure, as previously seen by other authors in both pediatric and adult patients (46, 50, 53, 55). Importantly, we have demonstrated that PBMC compartmentalization of HCV commonly exists in children and that it plays a significant role as a pool of genetic variability, thus contributing to viral persistence.

As expected, sequence variability was mostly concentrated within HVR1 (56). Interestingly, intrasample variability was generally lower in PBMCs than in plasma, and phylogenetic networks showed highly homogeneous clusters of PBMC-derived sequences, which reflect some kind of constraint for the association of HCV particles with PBMCs. Conflicting results have been reported about HCV infecting PBMCs (29, 57) or being attached on the surface of these cells without infection (58-59). The close phylogenetic relationship observed between plasma and PBMC variants indicates that a continuous genetic flow between these two compartments commonly exists and therefore suggests that PBMC-derived sequences do not follow a plasma-independent evolutionary pathway. Instead, the model of a temporal association of PBMCs with certain lymphotropic variants better fits our results, since many communities contained at least one homogeneous cluster of PBMC clones. Furthermore, codon selection analyses showed a very similar profile between both compartments in all patients, again suggesting that forces driving evolution are not compartment-dependent.

Chronically infected children are generally asymptomatic and usually have mild abnormal ALT levels, with a small proportion of them showing an aggressive course leading to end-stage hepatic disease (60-62). In accordance, our series displayed
predominantly mild liver damage, with the exception of case 6 who showed a rapid
progression of injury, requiring multiple biopsies during the follow-up and resulting in
severe hepatitis and incomplete cirrhosis. The relationship between the clinical course
of chronic hepatitis C and viral evolution has been addressed before in both adult (63-
65) and pediatric patients (53), where a correlation between hepatic injury and the
development of a mono- or oligoclonal viral population during long-term follow-up has
been described, as well as mild or no liver damage was correlated with a heterogeneous
viral quasispecies. In accordance, in cases 1 to 4 the slow progression of hepatitis
correlated with a complex population structure in which continuous emergence and/or
disappearance of viral communities was observed over time. Additionally, in these
cases a significantly higher positive selection on individual codons within HVR1 was
evident, indicative of strong immunologic pressure on this region, which is a major
target of neutralizing antibodies. Thus, in these cases the BAPS clusters mostly reflect
the population structure imposed by selection. By contrast, in cases 5 and 6, HVR1
invariability was noted during long periods of time as extensively described elsewhere
(33). Particularly in case 6, the slow evolutionary rate was associated with a rapid liver
disease progression. Moreover, the abrupt worsening of the patient’s clinical status at
115 months correlated with a switch in the quasispecies community after 5 years of an
almost complete invariability in the HVR1 amino acid sequence. Strikingly, at the same
time point, a pronounced compartmentalization of the viral populations was observed
which led to the segregation of a new PBMC-exclusive community not previously
detected during follow-up.

Additional evidence about the role of PBMC compartmentalization in viral
persistence can be inferred from case 4, in which an effect of α-IFN treatment on viral
quasispecies was observed. Quasispecies complexity in plasma decreased during
treatment and simultaneously the main viral community in plasma (Figure 2, Com1) declined its dominance while other minor communities grew. After treatment withdrawal, Com1 predominated again and finally, at the late chronic phase, a diversification pattern was observed. Interestingly, complexity in PBMCs increased during treatment (Figure 3, 5 months), and that increment corresponded to diversifying PBMC variants in Com3 that were closely related to previous PBMC-specific clusters at 0 and 3 months, which were poorly represented in plasma before that moment. This dynamic pattern of compartmentalization suggests that those lineages in Com3 detected in PBMCs at the beginning of follow-up were somehow “naturally resistant” to α-IFN treatment and served as a viral pool, filling the plasma compartment temporarily during treatment.

Finally, taken together our results support the notion of a community-based structure of the viral populations circulating within an infected host during chronic infection and point out a role of the PBMC compartment in the persistence of such structure. The relevance of the use of new evolutionary models to evaluate HCV chronic infection in the context of population structure has been recently highlighted (66). This kind of analyses permits the identification of independent lineages co-evolving in an infected host for long periods, which enable to clarify why lineages that remain unobserved for several years do not go extinct before later reappearing. By means of these approaches, our results highlight the significant role of compartmentalization in the maintenance of chronicity during pediatric HCV infection.

Acknowledgments

M.V.P., M.I.G. and R.H.C. are members of the National Research Council (CONICET) Research Career Program. J.M.D.C is a doctoral fellow of CONICET. P.V. is a post-
doctrinal fellow of CONICET. This study was funded in part by CONICET PIP 2005 n°5359 and PIP 2010 n°51.

References


**Figure captions**

**Figure 1. Biochemical and virological features of the patients studied.** Serum alanine aminotransferase (ALT) levels during follow-up are shown as filled areas on the left Y axis and HCV viral load in plasma is shown as black boxes on the right Y axis. Vertical discontinuous lines indicate samples used for phylogenetic analysis in this study. Duration of α-IFN treatment (cases 4 to 6) is shown as a black box on top of each plot. Serum ALT levels <32 IU/l were considered as normal.

**Figure 2. Evolutionary dynamics of HCV quasispecies in plasma.** Maximum-likelihood phylogenetic trees were reconstructed with all plasma-derived nucleotide sequences from each patient. Colored dots represent clones from different time points. Numbers at tree nodes indicate bootstrap values. The scale bar indicates number of substitutions (nucleotides/site). Viral communities identified by BAPS analysis are represented as colored areas and superimposed with the phylogenetic trees for each case. Temporal variations in the composition of viral communities detected in each patient during follow-up are depicted in graphics under each phylogenetic tree. Colored areas indicate independent communities. Duration of α-IFN treatment (cases 4 to 6) is shown as a black box on top of each plot.

**Figure 3. Complexity of plasma and PBMC-associated viral quasispecies.** (A) Variability of the amplified region at the nucleotide level, as represented by the entropy at each position in a sequence alignment including all plasma (left) or PBMC (right)
sequences from cases 1 to 6. (B) Intra-sample complexity within HVR1 as calculated by the normalized Shannon entropy of amino acid sequences. Values obtained for plasma and PBMC samples are represented as black and white symbols, respectively. (C) Sequence logo representations based on HVR1 amino acid alignments of plasma (top) and PBMC (bottom) derived sequences from cases 1 to 6. The frequency of each residue at each position is indicated by the relative height of the corresponding letter, and the overall height of the stack denotes the degree of sequence conservation. Logos were obtained using WebLogo program (67).

Figure 4. Phylogenetic analysis of plasma and PBMC-compartmentalized viral sequences. Maximum parsimony-based networks were constructed with all plasma and PBMC derived nucleotide sequences from each patient. Each node corresponds to a unique sequence, and the diameter of each circle refers to the frequency of that sequence. Colors within the circles represent clones from different time points as indicated in the panels above each network. Plasma-derived sequences have the same color-code as in Figure 2. Viral communities obtained by BAPS present in both compartments are represented as white areas and superimposed to each network. Viral communities are numbered as in Figure 2, and have an almost identical composition regarding plasma-derived clones. In case 2, Com 3 has been split in two sub-communities 3A and 3B. Temporal variations in the composition of viral communities detected in PBMC compartment are depicted in graphics under each network. Colored areas indicate independent communities, following the same color-code as in Figure 2. Duration of α-IFN treatment (cases 4 to 6) is shown as a black box on top of each plot.
Supplementary Figure 1. Intra-patient analysis of nucleotide variability of the amplified region. Multiple nucleotide sequence alignments including all plasma or PBMC sequences from each patient were used to calculate Shannon’s entropy as described in Material and Methods. The shaded area denotes the HVR1 region.

Supplementary Figure 2. Positively and negatively selected sites within the analyzed region. Horizontal bars represent evidence of selection at individual codons inferred from plasma (P) and PBMCs (C) of each patient. Bars are numbered at the left according to case number. Positively and negatively selected sites are denoted as white and black squares, respectively, and grey squares indicate no evidence of selection. Codons are numbered at the top according to HCV isolate H77 (genotype 1a). The HVR1 region is delimited by vertical lines and highlighted in gray.
Table 1  Clinical description of the patients included in this study

<table>
<thead>
<tr>
<th>Patient No.</th>
<th>Age&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Gender&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Route of infection</th>
<th>HCV genotype</th>
<th>Time of follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
<td>F</td>
<td>Transfusion</td>
<td>1a</td>
<td>50 mo</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>F</td>
<td>HCV+/HIV+ mother</td>
<td>1a</td>
<td>38 mo</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>M</td>
<td>Unknown</td>
<td>1a</td>
<td>51 mo</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>M</td>
<td>Unknown</td>
<td>1a</td>
<td>41 mo</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>F</td>
<td>HCV+ mother</td>
<td>1a</td>
<td>63 mo</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>M</td>
<td>HCV+ mother</td>
<td>1a</td>
<td>115 mo</td>
</tr>
</tbody>
</table>

<sup>a</sup> Age at the beginning of follow-up  
<sup>b</sup> F stands for female, M for male
Table 2. Results of Bayesian Tips-Significance Tests (BaTS).

<table>
<thead>
<tr>
<th>Case</th>
<th>Statistic</th>
<th>Observed mean (95% CI)</th>
<th>Null mean (95% CI)</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AI</td>
<td>1.75 (1.12-2.38)</td>
<td>3.85 (3.28-4.44)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>PS</td>
<td>12.64 (11.00-14.00)</td>
<td>20.68 (18.60-22.37)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>MC (Plasma)</td>
<td>25.03 (25.00-26.00)</td>
<td>5.47 (4.01-7.37)</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>MC (PBMC)</td>
<td>5.12 (5.00-6.00)</td>
<td>2.35 (1.86-3.10)</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>AI</td>
<td>3.45 (2.45-4.44)</td>
<td>6.32 (5.29-7.19)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>PS</td>
<td>27.13 (24.00-30.00)</td>
<td>37.51 (33.41-40.82)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>MC (Plasma)</td>
<td>6.11 (6.00-7.00)</td>
<td>4.17 (3.29-5.10)</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>MC (PBMC)</td>
<td>5.26 (4.00-8.00)</td>
<td>3.65 (2.80-4.97)</td>
<td>0.05</td>
</tr>
<tr>
<td>3</td>
<td>AI</td>
<td>2.94 (1.98-3.90)</td>
<td>5.61 (4.66-6.30)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>PS</td>
<td>20.56 (17.00-24.00)</td>
<td>31.87 (29.21-33.59)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>MC (Plasma)</td>
<td>20.49 (20.00-24.00)</td>
<td>5.57 (4.44-7.16)</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>MC (PBMC)</td>
<td>6.12 (5.00-8.00)</td>
<td>2.74 (2.26-3.25)</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>AI</td>
<td>1.42 (0.88-1.95)</td>
<td>5.87 (4.97-6.71)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>PS</td>
<td>11.25 (10.00-12.00)</td>
<td>35.35 (31.68-38.68)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>MC (Plasma)</td>
<td>17.51 (12.00-20.00)</td>
<td>3.69 (2.93-5.20)</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>MC (PBMC)</td>
<td>12.48 (12.00-15.00)</td>
<td>3.91 (3.11-5.22)</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>AI</td>
<td>0.73 (0.20-1.30)</td>
<td>3.10 (2.64-3.56)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>PS</td>
<td>6.80 (5.00-9.00)</td>
<td>16.34 (15.05-17.19)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>MC (Plasma)</td>
<td>15.38 (11.00-24.00)</td>
<td>6.84 (5.52-9.22)</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>MC (PBMC)</td>
<td>6.46 (4.00-11.00)</td>
<td>1.94 (1.57-2.24)</td>
<td>0.01</td>
</tr>
<tr>
<td>6</td>
<td>AI</td>
<td>2.12 (1.30-2.93)</td>
<td>5.94 (5.30-6.47)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>PS</td>
<td>14.45 (12.00-16.00)</td>
<td>33.81 (31.27-35.49)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>MC (Plasma)</td>
<td>13.99 (11.00-23.00)</td>
<td>5.66 (4.52-8.09)</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>MC (PBMC)</td>
<td>19.83 (18.00-20.00)</td>
<td>2.81 (2.42-3.42)</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Association Index (AI), Parsimony Score (PS) and Monophyletic Clade (MC) test the global association between a trait and tree topology. The observed mean and its associated 95% confidence intervals were obtained by analyzing 10,000 trees sampled during the Bayesian phylogenetic reconstruction. The null mean and its associated confidence intervals were obtained after randomly distributing the traits in the phylogeny (100 replicates). Significance level is the p-value for the statistical hypothesis test for equality between the observed index and that expected under no-association.