CD4+ T-cell recovery with suppressive ART induced rapid sequence evolution in HCV envelope but not NS3

Lin LIU,
Division of Infectious Diseases and Global Medicine, Department of Medicine, University of Florida College of Medicine, Gainesville, FL 32610, USA

David NARDO,
Department of Pharmacology, University of Florida College of Medicine, Gainesville, FL 32610, USA

Eric LI, and
Division of Infectious Diseases and Global Medicine, Department of Medicine, University of Florida College of Medicine, Gainesville, FL 32610, USA

Gary P. WANG, M.D., Ph.D.
Division of Infectious Diseases and Global Medicine, 1600 SW Archer Road, ARB R2-110, Gainesville, FL 32610, USANorth Florida/South Georgia Veterans Health System, Gainesville, FL 32608, USA. Telephone: 352-294-5481; Fax: 352-392-6481
Gary P. WANG: gary.wang@medicine.ufl.edu

Abstract

Objectives—CD4+ T-cell depletion from human immunodeficiency virus (HIV) infection leads to a global decline in anti-hepatitis C virus (HCV) envelope neutralizing antibody (nAb) response, which may play a role in accelerating liver fibrosis. An increase in anti-HCV nAb titers has been reported during antiretroviral therapy (ART) but its impact on HCV remains poorly understood. The objective of this study is to determine the effects of ART on long-term HCV evolution.

Design/methods—We examined HCV quasispecies structure and long-term evolution in HIV/HCV co-infected subjects with ART-induced CD4+ T-cell recovery, and compared to subjects with CD4+ T-cell depletion from delayed ART. We applied a single-variant sequencing (SVS) method to construct authentic viral quasispecies and compared sequence evolution in HCV envelope, the primary target for humoral immune responses, and NS3, a target for cellular immunity, between the two cohorts.

Results—The SVS method corrected biases known to skew the proportions of viral variants, revealing authentic HCV quasispecies structures. We observed higher rates of HCV envelope sequence evolution in subjects with ART-induced CD4+ T-cell recovery, compared to subjects with CD4+ T-cell depletion from delayed ART (P=0.03). Evolutionary rates for NS3 were considerably lower than the rates for envelope (P<0.01), with no significant difference observed between the two groups.

Correspondence to: Gary P. WANG, gary.wang@medicine.ufl.edu.
Conclusions—ART-induced CD4+ T-cell recovery results in rapid sequence evolution in HCV envelope, but not in NS3. These results suggest that suppressive ART disproportionally enhances HCV-specific humoral responses more than cellular responses, resulting in rapid sequence evolution in HCV envelope but not NS3.

Keywords
hepatitis C virus; human immunodeficiency virus; evolution; antiretroviral therapy; neutralizing antibody response; single variant sequencing

Introduction

Over 35 million people globally are living with human immunodeficiency virus (HIV),[1] and a quarter of these individuals are co-infected with hepatitis C virus (HCV).[2] Compared to HCV mono-infection, HIV co-infection accelerates HCV-related liver fibrosis, cirrhosis, and hepatocellular carcinoma, increasing the overall mortality.[3, 4] With effective antiretroviral therapy (ART), liver disease has surpassed acquired immunodeficiency syndrome (AIDS) to become a leading cause of hospitalization and death in this co-infected population.[5]

The mechanisms underlying accelerated liver disease in HIV-HCV co-infection remain poorly understood. During the transition from acute to chronic HCV infection, humoral immune response initiates late but then broadens and persists, whereas cellular immune response wanes over time.[6–10] In chronic infection, cellular immunity is largely dysfunctional due to HCV-induced mechanisms involving PD-1, RIG-1 and others.[11–14] Without reversal of T-cell exhaustion, humoral immunity alone, specifically neutralizing antibodies (nAbs), remains ineffective in clearing chronic HCV infection.[15] However, nAbs can block HCV propagation in vitro and in vivo, potentially modulating disease progression in chronic HCV.[7, 16–21] HIV infection causes CD4+ T-cell depletion and progressive T-cell dysfunction, which may influence nAbs functions.[22, 23] Bailey et al. recently reported a significant reduction of anti-HCV envelope binding antibody titers, neutralizing antibody (nAb) titers, and nAb breadth in patients who had low CD4+ T-cell counts after incident HIV infection.[24] Lee et al. demonstrated that HIV/HCV co-infected subjects who received ART developed increased anti-HCV nAb titers.[25] These findings indicate that HIV co-infection decreases HCV-specific nAb response, raising the possibility that preserving or restoring anti-HCV humoral immunity with ART may play a role in decreasing the rate of liver fibrosis progression.

Studies of viral dynamics and evolution can provide important insights into host-virus interactions. To date, few studies have investigated the effects of HIV co-infection and ART on HCV quasispecies evolution. An early study showed some positive selections against the hypervariable region (HVR)-1 of HCV envelope in patients receiving ART.[26] Another study showed a decreased HVR1 complexity and diversity in HIV/HCV co-infected patients who progressed to end-stage liver disease compared to those with compensated chronic hepatitis.[27] HCV genome is under constant selective pressure from host immunity. Envelope (E1E2) gene sequences are selected primarily by nAbs, whereas nonstructural
genes (p7, NS2 – NS5B) are selected by cellular immunity.[7, 8, 20, 28–30] Thus, a comparison of HCV evolutionary rates between structural (i.e. envelope) and nonstructural (e.g. NS3) gene sequences may provide information about the relative importance of humoral and cellular immunity.

One major obstacle in defining viral quasispecies is quantifying the relative proportions of different sequence variants in the overall viral population. Both clonal sequencing and endpoint limiting-dilution polymerase chain reaction (PCR) are labor-intensive, and their sequencing depth is limited by the numbers of clones that could be analyzed. High throughput next-generation sequencing (NGS) technologies yield hundreds of thousands of sequence reads, but technical artifacts during sample preparation, including PCR and sequencing errors, template resampling, and PCR amplification bias, are known to skew the proportion and accuracy of viral variants in a quasispecies population. To overcome these artifacts, Jabara and colleagues developed a Primer-ID approach,[31, 32] which incorporates a random sequence tag during reverse transcription to the cDNA copy of each RNA template. Because each initial template is tagged with a unique Primer-ID barcode sequence, technical artifacts from allelic skewing, template resampling, and nucleotide misincorporation are corrected by generating a consensus sequence for each unique Primer-ID. However, deep sequencing using the Roche/454 pyrosequencing platform introduces a large number of homopolymeric errors that complicates automated sequence alignments and bioinformatics analysis.[33–36] Furthermore, because the Primer-ID strategy requires a large number of reads for consensus calls, the lower sequencing throughput of Roche/454 limits the number of samples that could be analyzed simultaneously. Thus, overcoming these technical challenges could allow for more accurate and rapid determination of authentic quasispecies population structure.

Here, we examined the evolution of HCV E1E2 envelope (structural) and NS3 (nonstructural) gene segments in HIV/HCV co-infected subjects receiving either early or delayed ART. We found significantly higher rates of sequence evolution for HCV envelope in subjects who had early ART and CD4+ T-cell recovery compared to those with progressive CD4+ T-cell depletion due to delayed ART. This difference in evolutionary rates between the two groups was not observed for NS3. Using a single-variant sequencing (SVS) approach that modifies the primer-ID strategy for sequencing on the Illumina platform, we not only constructed authentic HCV quasispecies but also identified naturally occurring resistance-associated variants (RAVs) circulating at less than 0.1% of the viral population. To our knowledge, this is the first study that demonstrates the effects of CD4+ T-cell recovery from suppressive ART on long-term evolution of HCV quasispecies.

**Materials and Methods**

**Samples and methods**

Serum samples from subjects with HIV/HCV co-infection were identified and selected from the University of Pennsylvania Center for AIDS Research (CFAR) repository. HCV RNA was extracted from serum samples using QIAmp Viral RNA kit (Qiagen). CDNA was synthesized using SuperScript III Reverse Transcriptase (Invitrogen) and RT primers that added random sequence tags to cDNA copies of viral RNA templates (Figure S1). E1E2 and
NS3 gene segments were amplified by nested-PCR followed by gel extraction. Pooled DNA library was sequenced on an Illumina MiSeq sequencer. Following a Q30 filter that removed low quality reads, a total of 25.7 Gigabases of nucleotides were generated. For each unique sequence tag, a consensus sequence was determined using MAFFT (http://mafft.cbrc.jp/alignment/software/), based on an alignment of at least three reads that share the same sequence tag. Nucleotide sequences have been deposited in NCBI SRA under BioProject accession number PRJNA290532. Additional details on molecular biology and bioinformatics methods can be found in SI Materials and Methods.

Statistical analysis

Mann-Whitney rank sum test was used to calculate significance of nonsynonymous substitution rates between early versus delayed ART groups or between different gene segments; Student’s $t$ test was used when normality was satisfied. A paired $t$ test was used to compare the number of drug resistant mutations called by consensus and raw sequences from the same samples. A $P$ value of $<0.05$ was considered statistical significant.

Results

Subjects and samples

We queried the specimen repository at University of Pennsylvania Center for AIDS Research (CFAR) and selected six subjects that had detailed clinical and laboratory data and a sufficient number of samples available, and also met the following criteria: (1) HIV and HCV antibody positivity, (2) followed for at least 5 years, (3) received either early ART with CD4+ T-cell recovery of $>200$ cells/uL between the first and the last samples (referred to as the “early ART” group below), or delayed ART with a decline of CD4+ T-cell counts of $>200$ cells/uL between the first and the last samples (the “delayed ART” group). A total of six subjects were included with three subjects in each group (Figure 1). The average duration of follow-up was 10.8 years.

SVS identified authentic HCV quasispecies at both individual and population levels

Over 16 million paired-end Illumina reads (over 9 billion bases) were filtered using a set of stringent criteria to remove low-quality reads (Figure 2). A total of 91,326 (51,099 for E2 and 40,227 for NS3) consensus sequences were generated, averaging 1,054 (range 233 to 5032) consensus sequences per sample from an average of 147,387 (range 26,169 to 328,582) raw reads. Approximately 75% of consensus sequences were formed using 3 to 100 reads, whereas the remaining 25% were built from over 100 reads (maximum 98,124).

To calculate the background error rate, we synthesized HCV RNA in vitro using a plasmid containing subtype 1a HCV sequence (H77c).[37] We amplified the E1E2 gene segments containing the hypervariable region 1 (HVR1), and an NS3 gene segment of the in vitro transcript using the SVS method. We then compared HCV quasispecies before and after SVS correction (Figure S2, H77c and H77c-dup for duplicate). Before SVS correction, only 55% of raw E1E2 and NS3 sequences were identical to the H77c sequence. After SVS correction, only 55% of raw E1E2 and NS3 sequences were identical to the H77c sequence. After SVS correction, 99.38% of E1E2 and NS3 consensus sequences were identical to H77c (99.44% and 98.93% for E1E2 and NS3, respectively). The overall error rate prior to SVS correction...
was $1.53 \times 10^{-3}$ errors per nucleotide, which decreased to $1.41 \times 10^{-5}$ errors per nucleotide (>100 fold reduction) after SVS correction.

Analysis of clinical samples from HIV/HCV co-infected subjects showed that prior to SVS correction (Figure S2, Raw column), minority variants (defined as <1% of viral population, indicated by color gray) dominated the overall E1E2 and NS3 sequences. Following SVS correction (Figure S2, SVS column), the proportion of minority variants (color gray) was significantly reduced. Thus, SVS correction identified the authentic sequences of dominant variants (color red) and unmasked the structure of HCV quasispecies population. Furthermore, in more than half of E1E2 and more than one-third of NS3 amplicons, sequences of dominant variants were different before and after SVS correction, indicating that the SVS procedure not only restored the accurate composition of quasispecies but also corrected the technical artifacts that altered the identity of viral sequences.

**Distinct evolutionary patterns of HCV quasispecies between structural and non-structural genes**

To investigate long-term evolution of HCV quasispecies, we analyzed the phylogenetic relationship of major sequence variants (defined as >=1% of viral population), which together represents ~60% of all consensus sequences from patient samples (>40,000 sequences). Phylogenies for the structural gene E1E2 envelope (Figure 3) and the non-structural gene NS3 (Figure S3) showed different patterns. For E1E2, separate clusters of sequence variants were observed from different time points (Figure 3, Subject 2 and 5), showing a predominantly temporal order of evolution, i.e. shifting to new clades from one visit to the next. In contrast, many sequence overlaps between visits were observed for NS3 (Figure S3), consistent with a transitional order of evolution. These differences in evolutionary patterns were more pronounced in subjects 2 and 5, in which more samples were available for analysis. In addition, the envelope sequences showed more diverse major variants (i.e. shorter horizontal bars in E1E2 trees in Figure 3 and more grey areas in Figure S2) compared to NS3. Quasispecies complexity was also significantly higher for E1E2 compared to NS3 (0.44 vs. 0.24 for E1E2 and NS3, respectively, P = 4 x 10^{-7}).

**ART-induced CD4+ T-cell recovery is associated with rapid HCV envelope sequence evolution**

To evaluate the impact of suppressive ART on HCV quasispecies evolution, we compared the rates of nonsynonymous substitutions between the “early ART” and the “delayed ART” groups (Figure 4). Overall, envelope sequences showed higher rates of nonsynonymous substitutions than NS3 (average rate of $4.27 \times 10^{-3}$ and $1.80 \times 10^{-4}$ nonsynonymous substitutions per nonsynonymous site per year for E1E2 and NS3 segments, respectively; P<0.01). The “early ART” group had a higher rate of nonsynonymous evolution in HCV envelope compared to the “delayed ART” group (average rate of $6.07 \times 10^{-3}$ and $2.47 \times 10^{-3}$ for early ART and delayed ART, respectively; P=0.03) (Figure 4B), driven primarily by the more rapid sequence turnover in HVR1 in the early ART group (Figure 4A, sliding window analysis). In contrast, no significant difference in evolutionary rates for NS3 was found between the two groups (average rate of $2 \times 10^{-5}$ and $3.5 \times 10^{-4}$ for the early and the delayed ART groups, respectively; P=0.16) (Figure 4B).
We next examined temporal changes in HVR1 amino acid (AA) sequences using high-resolution sequence logo analysis (Figure 5). In the early ART group, HVR1 sequences were replaced sequentially by new sequences. In contrast, HVR1 sequences remained relatively unchanged for 2 of 3 subjects in the delayed ART group. Interestingly, for subject 2, several AA changes in HVR1 sequence occurred at year 5.9, which coincided with an increase in CD4+ T-cell count (from 114 to 553 cells/μL). The AA sequence reverted at year 8 following the decline of CD4+ T-cell from 553 to 278 cells/μL. Taken together, these data suggest that ART-induced CD4+ T-cell recovery was associated with rapid amino acid sequence evolution in HCV envelope but not NS3.

**SVS identified authentic NS3 resistance-associated variants in HIV/HCV co-infection**

Naturally occurring NS3 resistance associated variants (RAVs) have been reported for both HCV mono-infected and HIV/HCV co-infected patients,[38, 39] but their temporal dynamics is unknown. Using SVS, we quantified the temporal evolution of authentic NS3 RAVs. Naturally occurring RAVs were identified in 5 of the 6 subjects (mean detection limit of 0.1%) (Figure S4). RAV carrying the Q80K mutation was a dominant quasispecies in 3 subjects, and the V55A RAV was dominant in 1 subject. A total of 13 minor RAVs were identified in 5 subjects, all circulating at very low frequencies (0.04–0.86%). Among these, VI170T was found in 7 longitudinal samples from 4 subjects, while each of the remaining minor RAVs was observed in only one subject. Subject 4 had no detectable NS3 RAVs. Overall, 71% (24 of 34) of the samples harbored at least one RAV, of which 53% (18 of 34) had a dominant RAV and 35% (12 of 34) had minor RAVs. SVS correction removed over 95% of amino acid substitutions erroneously called as RAVs by conventional deep sequencing (Figure S4, S5). Additional details can be found in SI Results.

**Discussion**

We present an in-depth analysis of HCV quasispecies evolution in HIV/HCV co-infected subjects followed up to 11 years. Using single-variant sequencing (SVS), we showed that ART-induced CD4+ T-cell recovery was associated with rapid non-synonymous sequence evolution in HCV envelope, and that subjects with CD4+ T-cell depletion and delayed ART had significantly lower rates of HCV envelope evolution. In contrast, the rates of non-synonymous evolution in NS3 were considerably lower compared to HCV envelope with no significant difference between the two groups. These data are consistent with our hypothesis that ART-induced CD4+ T-cell recovery enhances anti-HCV envelope antibody response, but not anti-HCV cellular response.

Evolution of envelope sequences (especially HVR1) is driven largely by neutralizing antibodies (nAbs).[7, 20, 28, 40, 41] During acute HCV infection, an early robust nAbs response is associated with rapid envelope sequence evolution and spontaneous clearance of HCV.[6, 7, 40] In patients who progress to chronic infection, nAb responses are delayed but remain persistent, consistent with increasing rates of evolution in the envelope region.[6–10] Thus, rapid HCV envelope evolution in the early ART group suggests enhanced nAb response upon CD4+ T-cell recovery (Figure 4 & 5). In contrast, lower rates of HCV envelope evolution, particularly with a lack of amino acid changes in the HVR1 for a
decade, in the delayed ART group suggest a decline in nAb response from CD4+ T-cell depletion (Figure 4 & 5). Our results are consistent with the recent findings that CD4+ T-cell depletion by HIV infection results in a global decline in the anti-HCV nAb response, and that increased nAb titers were observed in patients on suppressive ART.[24, 25]

We observed no significant difference in rates of NS3 evolution between the early and the delayed ART groups. Nonsynonymous evolutionary rate in subjects with rising CD4+ T-cell counts approached zero, indicating minimum selective pressure on NS3 over a decade despite CD4+ T-cell recovery. Nonstructural genes such as NS3 are presumably under selective pressure from HCV-specific cellular immune responses that are likely impaired by HCV-induced mechanisms during chronic infection.[8, 28, 29, 42] Given the largely comparable low evolutionary rates among nonstructural genes (p7, NS2-NS5B) during early chronic HCV infection,[28] our data suggest that HIV infection and ART disproportionally affect HCV-specific humoral responses more than cellular responses, resulting in rapid sequence evolution in the envelope but not NS3. We also note that ART had little effects on HCV RNA levels or viral clearance, consistent with the previous findings that rising nAb response alone is insufficient for HCV clearance, unless accompanied by a reversal of T-cell exhaustion.[7, 15, 20, 25, 28, 42] Nonetheless, since HIV co-infection leads to accelerated progression and ART slows the progression of HCV-related liver diseases, this disproportional effect on nAb responses and evolution of HCV envelope suggests a potential protective role of nAbs in HCV liver disease progression.

Combined with previous data,[24, 25], our results support the recent finding that co-infected patients who achieved HIV viral suppression with ART had lower rates of hepatic decompensation compared to those who did not maintain HIV suppression.[43, 44] However, despite suppressive ART, co-infected patients continue to have an increased risk of hepatic decompensation compared to HCV-monoinfected individuals,[45] possibly related to ongoing immune dysregulation not reversed by ART. Taken together, these results are consistent with the current DHHS guideline that recommends initiation of ART among HIV/HCV co-infected patients regardless of CD4+ T cell count [46], and support the most recent AASLD/IDSA recommendations that prioritize HIV/HCV co-infected individuals for consideration of early HCV therapy to reduce the risk of liver disease progression.[47]

SVS represents a substantial technical advance in quantifying accurate proportions of dominant and minority variants compared to conventional PCR-based deep sequencing. First, the number of reads that shared the same random sequence tags varied widely (range: 1 to 98,124). This suggests that conventional PCR introduces substantial template re-sampling and PCR amplification bias, and therefore can severely skew the proportions of the initial viral populations if uncorrected. SVS correction markedly reduced the effects of PCR bias and template resampling. Second, SVS corrects a large number of nucleotide misincorporation errors, as indicated by the analysis of minority variants (Figure S2) and resistance-associated variants (Figures S4 & S5). Strikingly, these artifacts led to incorrect identification of more than half of E1E2 and more than one-third of NS3 dominant variants, highlighting the limitations and the potential risks of conventional clonal or deep sequencing without correction for technical artifacts. Third, the Illumina workflow obviates the need for homopolymer error correction that frequently complicates the analysis of pyrosequencing.
data.[34–36] Furthermore, the low error rate for Illumina minimizes base call errors in the primer tags that could lead to the creation of artificial variants and distort quasispecies distributions.[33, 48] Thus, SVS allowed us to determine viral variants and quasispecies population with high accuracy and sensitivity. We note that errors from first-strand DNA synthesis during reverse transcription could not be corrected using SVS. However, these errors are common to all amplification strategies that require first-strand DNA synthesis from viral RNA. Finally, although errors may also be introduced in sequence tags during PCR or nucleotide sequencing, they likely contribute little compared to the effects of template resampling and PCR/sequencing errors.

Emergence of resistance-associated variants (RAV) is a potential concern for patients receiving direct-acting antiviral (DAA) therapy. Drug resistance develops rapidly with protease inhibitors monotherapy.[38, 39] This is not surprising because polymorphic RAVs (e.g. V55A, Q80K) and minor RAVs pre-exist in treatment-naïve individuals. In 4 of 6 subjects, HCV populations were dominated by Q80K or V55A over all time points analyzed. In addition, minority RAVs were identified in at least one locus in all but one subject. Among the 10 amino acid sites associated with NS3 resistance,[49–51] no RAVs were observed in 4 AA sites in any of the samples analyzed (Figure S4), suggesting that substitutions at these sites may incur a high fitness cost in vivo. Each minority RAV was observed in only one subject, except for IV170T, which was found in 4 of 6 subjects at multiple time points. These results are consistent with the previous data indicating that IV170T has minimal fitness cost compared to wild type HCV.[52] Although SVR rates are high and are comparable between HCV and HIV/HCV co-infection with new DAAs, treatment failure still occurs in a small number of individuals.[53, 54] Going forward, understanding the significance of pre-existing RAVs and their selection during treatment failure would likely require sensitive and accurate methods such as SVS, which can detect authentic RAVs with minimal bioinformatics or statistical manipulations to correct for technical artifacts.

In summary, we report that early ART-induced CD4+ T-cell recovery is associated with increased rates of HCV evolution in the envelope, but not NS3. Since HCV envelope is the primary target of nAbs, these data suggest HCV-specific nAb response modulated by ART as a potential mechanism for slowing the progression of liver fibrosis in HIV/HCV co-infection. From the technical perspective, the SVS approach should be broadly applicable to evolutionary studies of other hypervariable viruses such as HIV and sensitive detection of drug resistant mutations during antiviral therapy.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgments**

We thank Dr. Ian Frank and University of Pennsylvania Center for AIDS Research (CFAR) for providing the samples and laboratory data for this study. We thank Lisa Zhao, Zhubene Mesbah, Lorenzo Bean and Mariana Kirst for helpful discussions. This study was supported by NIH/NIAID K08 AI077713, Merck Investigator Studies Program IISP 39046, and the Gatorade Trust through funds distributed by the University of Florida, Department of Medicine.
Funding information: This work was supported by the National Institutes of Health and the National Institute of Allergy and Infectious Diseases [K08 AI077713]; the Merck Investigator Studies Program [IISP 39046]; and the Gatorade Trust through funds distributed by the University of Florida, Department of Medicine.

References


Figure 1. Clinical laboratory parameters and duration of antiretroviral therapy for HIV/HCV co-infected subjects analyzed in this study

For each subject, CD4+ T-cell counts are shown on the left vertical axis. HIV RNA (copies/mL) and HCV RNA (IU/mL) levels are shown on the right vertical axis. The gray arrow bar above each subject indicates the duration of antiretroviral therapy (ART). Serum samples analyzed for HCV gene sequencing are circled on the HCV RNA plot. Subjects 1–3 ("delayed ART" group) showed a gradual decline in CD4+ T-cell counts prior to initiation of ART. Subjects 4–6 ("early ART" group) received early ART and showed an increase in CD4+ T-cell counts over time.
Starting reads with matching sequencing barcodes (16.2 M, 100%)

Do the paired-end reads have a combined length of no less than 540 bp?
  yes (14.5 M, 89.1%)
  no  → Discard

Does the average quality score of each read higher than Q30 (error rate <0.1%)?
  yes (12.0 M, 74.2%)
  no  → Discard

Does read have a correct sequencing tag (location, size, control bases*)?
  yes (11.7 M, 72.0%)
  no  → Discard

Can the paired-end reads be stitched together based on at least 10 bp of matching overlap?
  yes (11.7 M, 71.9%)
  no  → Discard

Are there at least 3 reads per sequencing tag?
  yes (10.7 M, 65.6%)

A total of 91.3 k consensus sequences built (Average 117 reads per consensus sequence)

Figure 2. Workflow for Illumina MiSeq paired-end sequence processing and construction of consensus sequences

Raw Illumina MiSeq paired-end reads were filtered and stitched using the criteria shown to construct consensus sequences. At each step, reads that did not satisfy the criteria were discarded. Consensus sequences were built based on a minimum of 3 reads that had identical sequence tags. Control bases (asterisk) indicate the 5 non-N, non-binding, degenerate bases, and were used to remove sequences that had sequencing errors. M = million reads; k = thousand reads.
Figure 3. Phylogenetic analysis revealed temporal evolution of HCV E1E2 quasispecies over 10 years
A master maximum likelihood (ML) tree (center) was built using dominant sequences from each sample. Sequences from the same subject are shown using the same color in the master tree (center). For each subject, detailed maximum likelihood phylogeny of HCV population is shown using representative E1E2 sequences (≥1% in HCV populations). Minority variants (<1% in HCV populations) were excluded for clarity. Longitudinal sequences are indicated by rainbow colors. The length of each horizontal bar indicates the proportion of each variant within the viral population. Highly abundant variants (>80% of the population)
are indicated by double bars (e.g. Subject 4, blue bar). When two samples from the same year were available, variants from the later samples are marked with an asterisk (e.g. Subject 2 and 5). Bootstrap values are shown at major nodes in the master tree (center) and are omitted in the individual trees for clarity. Reference sequences are Bole1a (subtype 1a, solid circle, GenBank accession number: JQ791196.1[55]), H77 (subtype 1a, empty circle, GenBank accession number: AF009606) and Con1 (subtype 1b, solid triangle, GenBank accession number: AJ238799). This tree contains sequence information from >16,000 SVS consensus sequences (>50% of all E1E2 consensus sequences) constructed from over 2 million paired-end reads.
Figure 4. Increased rate of sequence evolution in HCV envelope but not NS3 in subjects with ART-induced CD4+ T-cell recovery

(A) Sliding window analysis showing rates of nonsynonymous evolution across E1E2 envelope (upper panel) and NS3 (lower panel) gene segments. Rates of nonsynonymous substitutions were calculated by comparing the first and the last sample of each subject and adjusted by time (interval between visits) using Varplot (Version 1.7, see SI methods). The x-axis indicates the centered codon positions of the respective gene segments. Horizontal bars depict the relative positions in each region in envelope gene segment, i.e. E1, hypervariable region (HVR)-1, and E2. A 10-codon window with 1-codon increment was
used. (B) Comparison of rates of nonsynonymous evolution between envelope and NS3
gene segments and between early versus delayed ART groups. Rates were calculated by
comparing the first and the last sample for each subject using MEGA (Version 6, see SI
methods). Mean values are indicated with a short horizontal line in each group. Single-
asterisk (*) indicates a P value of <0.05 whereas double-asterisks (**) indicates a P value of
<0.01. For both A and B, an average of 1,099 consensus sequences per sample were used for
the calculation of evolutionary rates.
Figure 5. Rapid HCV envelope evolution is driven largely by rapid HVR1 amino acid sequence turnover in subjects with ART-induced CD4+ T-cell recovery. Changes in amino acid (AA) sequences in the hypervariable region (HVR)-1 were demonstrated using sequence logos generated by VisSPA (Version 1.6.6, see SI methods). For initial samples, type 1 logos were generated from consensus sequences to demonstrate quasispecies composition. For subsequent visits, type 2 logos were generated where only AA changes relative to the initial viral sequences were shown. The height of the type 2 logos indicates relative abundance of AA changes.[28, 40] Time intervals (years) from the initial visit to subsequent visits are shown on the left. CD4+ T-cell count and HCV RNA levels are
shown on the right. An average of 986 consensus sequences per sample were used to generate the sequence logos.