Analysis of Hepatitis C Virus Intrahost Diversity across the Coding Region by Ultradeep Pyrosequencing

Michael Lauck,[a] Mónica V. Alvarado-Mora,[b] Ericka A. Becker,[a] Dipankar Bhattacharya,[d] Rob Striker,[c,d] Austin L. Hughes,[e] Flair J. Carrilho,[b] David H. O'Connor,[a] and João R. Rebello Pinho[b]

Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, Wisconsin, USA[a]; Laboratory of Tropical Gastroenterology and Hepatology, São Paulo Institute of Tropical Medicine, and Department of Gastroenterology, School of Medicine, University of São Paulo, São Paulo, Brazil[b]; W. S. Middleton Memorial Veteran's Association Hospital, Madison, Wisconsin, USA[c]; Departments of Medicine and Medical Microbiology & Immunology, University of Wisconsin-Madison, Madison, Wisconsin, USA[d]; and Department of Biological Sciences, University of South Carolina, Columbia, South Carolina, USA[e]

Hepatitis C virus (HCV) is the leading cause of liver disease worldwide. In this study, we analyzed four treatment-naïve patients infected with subtype 1a and performed Roche/454 pyrosequencing across the coding region. We report the presence of low-level drug resistance mutations that would most likely have been missed using conventional sequencing methods. The approach described here is broadly applicable to studies of viral diversity and could help to improve the efficacy of direct-acting antiviral agents (DAA) in the treatment of HCV-infected patients.

According to the World Health Organization, 130 to 170 million people are persistently infected with hepatitis C virus (HCV) and are at risk of developing severe liver disease and hepatocellular carcinoma (19). HCV is an enveloped RNA virus containing a single-stranded positive-strand RNA genome of approximately 9,600 nucleotides (3, 30). The virus encodes a single polyprotein precursor that is subsequently cleaved into at least 10 different proteins, including the structural core (C) and envelope (E1, E2, and p7) proteins as well as the nonstructural proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B) (8).

HCV replication is characterized by a high rate of virus production and a corresponding high degree of genetic diversity in circulating viruses. This is due to the lack of efficient proofreading by the HCV RNA-dependent RNA polymerase (3). As a result, the HCV population in each patient consists of closely related but nonidentical genomes, referred to as viral quasispecies (21, 22).

DNA sequencing has been altered radically with the development of second-generation pyrosequencing techniques. Recent work by our group and others has employed pyrosequencing both for whole-genome shotgun sequencing and for amplicon-based sequencing of short regions of human and simian immunodeficiency viruses (4, 5, 34). These methods demonstrate a new approach for studying the complexity of the viral population within a host and identifying minor genomic variants.

Direct-acting antivirals (DAAs), also known as "specifically targeted antiviral therapy for hepatitis C" (STAT-C), are the newest and most promising therapeutic option in HCV treatment (28). Several DAAs have been developed that inhibit different viral proteins, including the NS3 protease, the NS5b polymerase, and the NS5A replication complex (13, 25, 28). Two NS3 protease inhibitors were recently approved for the treatment of HCV-infected patients: telaprevir (Vertex, J&J) and boceprevir (Merck). These are the first new HCV-specific drugs in 20 years (6). A large number of new drugs are in development for the treatment of hepatitis C, including second-generation protease inhibitors such as ITMN-191 (R7227), BI 201335, NM283, R1626, MK-7009, BMS-650032, and PHX1766. While many of these compounds have more powerful antiviral activity than first-generation protease inhibitors, their utility is limited by the development of viral

mutations conferring cross-resistance (13). Resistance mutations differ depending on the specific drug used and the HCV subtype, though mutations conferring resistance to all currently approved drugs have already been described (10, 14, 28). Furthermore, uncommon variants of the viral quasispecies with reduced susceptibility to DAAs can occur naturally even before treatment begins.

While targeted sequencing has been used to analyze differences in HCV variability in HCV-monoinfected and HIV-HCV-coinfected subjects (32), as well as to determine antiviral resistance mutations against protease inhibitors (14, 26), no study has employed second-generation sequencing techniques to examine HCV subtype 1a heterogeneity across the entire coding region. Here we combined pyrosequencing with a transposon-based fragmentation method to perform genomewide ultradeep sequencing of four HCV-1a genomes, allowing analysis of viral sequence heterogeneity and identification of minor variants conferring preexisting HCV-specific drug resistance. Early identification of DAA resistance mutations in hepatitis C virus-infected patients (1, 2, 12) may support the use of drug resistance testing before DAA prescription (28). Our general approach could also facilitate longitudinal studies of HCV evolution.

MATERIALS AND METHODS

Patients and plasma specimens. Plasma samples were obtained from four treatment-naïve, anonymously selected patients after qualitative and genotypic testing at the University of Wisconsin Hospital and Clinics. All four patients were subjected to serial HCV PCR testing at the University of Wisconsin Hospital and Clinics prior to May 2011 (FDA approval of telaprevir and boceprevir). The initial HCV infection was determined at those facilities, and none of the patients were receiving therapy at the time of analysis. Furthermore, none of the patients received protease inhibitors or investigational drugs. All HCV samples were shown to be genotype 1a,
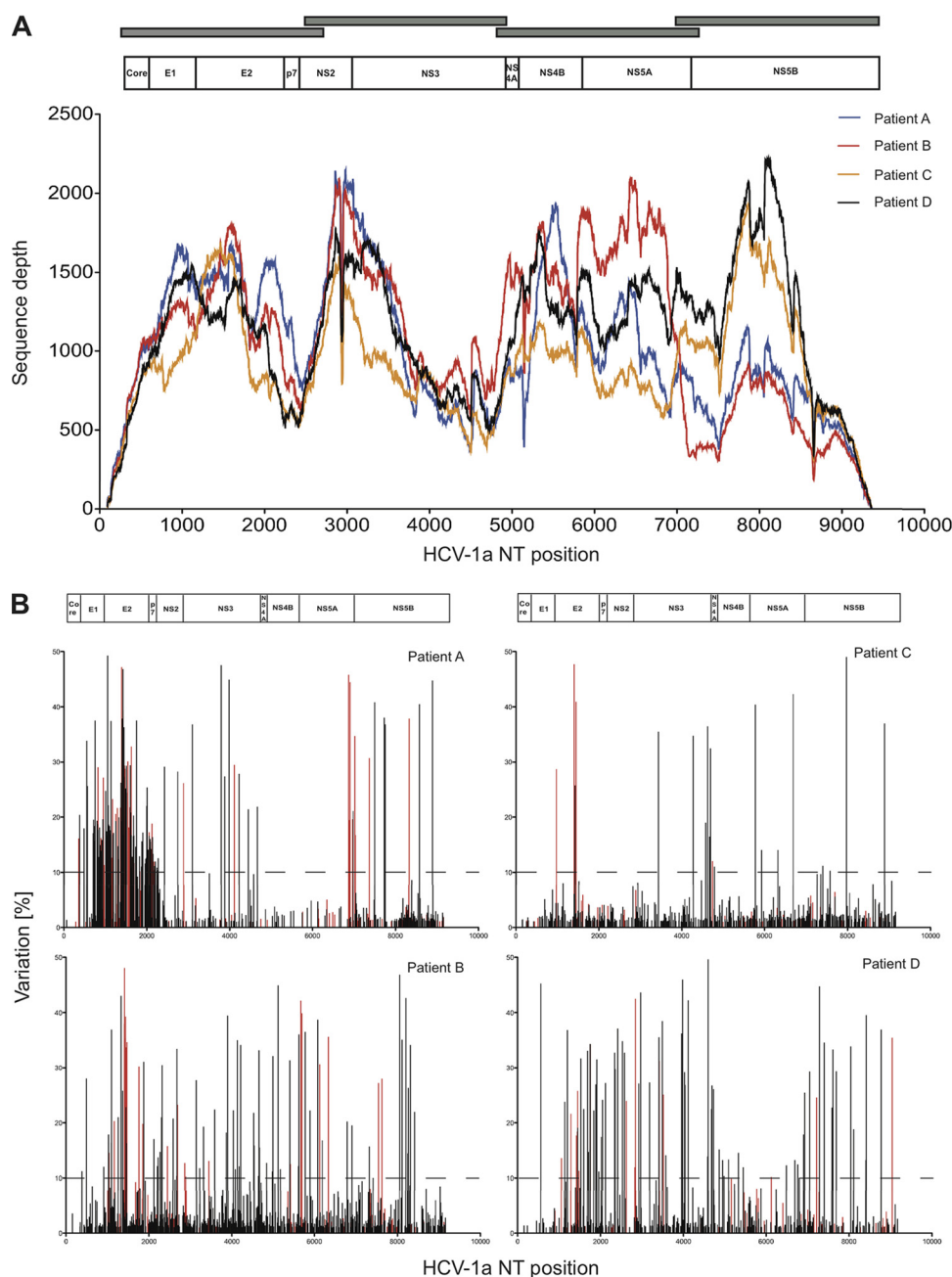
FIG 1 Genomewide pyrosequencing of HCV-1a, with analysis of intrahost variability. (A) Schematic representation of HCV-1a amplification strategy and sequence coverage obtained for each patient. Gray bars denote RT-PCR amplicons relative to the HCV-1a genome. The nucleotide numbering is based on the sequence of HCV strain H77 (GenBank accession no. NC_004102). (B) Graphs indicating the SNP frequency for each patient along the consensus genomic sequence. Synonymous mutations are indicated in black, and nonsynonymous mutations are shown in red. The HCV-1a genomic organization is depicted at the top of the figure. The nucleotide numbering is based on the sequence of HCV strain H77 (GenBank accession no. NC_004102).

with viral loads above 1×10^7 IU/ml. All protocols were evaluated and approved by the University of Wisconsin Institutional Review Board (IRB) prior to starting the experiments. Since there were no links to identifiable patient subject information, it was determined that this research was not human subject research.

RNA extraction and genomewide pyrosequencing. Cell-free plasma was obtained from EDTA-anticoagulated whole blood by use of Ficoll-Paque Plus (GE Healthcare, Piscataway, NJ) and density centrifugation. Viral RNA in the plasma was isolated using a Qiagen QIAamp MinElute virus spin kit (Qiagen, Valencia, CA) according to the manufacturer's

instructions. We designed four overlapping PCR amplicons of approximately 2.5 kb each to amplify nucleotide positions 16 through 9302 of the HCV genome, including all but 4 bp of the coding region (Fig. 1A). Viral RNA was reverse transcribed and amplified using a SuperScript III high-fidelity one-step reverse transcription-PCR (RT-PCR) kit (Invitrogen, Life Technologies, Carlsbad, CA). Specific primer sequences are available upon request. The reverse transcription-PCR conditions were as follows: 50°C for 30 min; 94°C for 2 min; 40 cycles of 94°C for 15 s, 55°C for 30 s, and 68°C for 3 min; and 68°C for 5 min. Following PCR amplification of the genomes, PCR amplicon bands were isolated using gel electrophoresis

(1% agarose) and were purified using a Qiagen MinElute gel extraction kit (Qiagen, Valencia, CA). Each amplicon was quantified using Quant-IT HS reagents (Invitrogen, Life Technologies, Carlsbad, CA), and all amplicons from a single viral genome were pooled together at equimolar ratios. Each pool was then quantitated, and approximately 50 ng of each was used in a fragmentation reaction mix, using a Nextera DNA sample prep kit (Roche Titanium compatible; Epicentre Biotechnologies, Madison, WI). Fragmentation was performed according to the manufacturer's protocol. Final libraries representing each genome were characterized for average size by use of a high-sensitivity DNA Bioanalyzer chip on a model 2100 Bioanalyzer (Agilent Technologies, Loveland, CO) and were quantitated with Quant-IT HS reagents (Invitrogen, Life Technologies, Carlsbad, CA). Libraries were then subjected to emulsion PCR, and enriched DNA beads were loaded onto a picotiter plate and pyrosequenced with a Roche/454 GS Junior sequencer using titanium chemistry (454 Life Sciences, Branford, CT). All four HCV-1a genomes were sequenced in a single GS Junior run.

Phylogenetic analysis. Maximum likelihood phylogenetic trees were inferred using GARLi, version 0.951, software (36), which employs an extensive branch-swapping protocol and optimizes the substitution model iteratively during the search. The parameters were set to run until topology improvement was not observed for 10,000 consecutive generations with the nucleotide substitution general time-reversible (GTR) model, using default parameters. A phylogenetic tree was constructed based on 382 complete HCV-1a genome sequences and was visualized using FigTree, version 1.2.2 (http://tree.bio.ed.ac.uk/software/figtree/).

SNP analysis. Pyrosequencing data were analyzed using CLC Genomics Workbench (CLC Bio, Aarhus, Denmark). Initially, reads were trimmed to remove Nextera-specific transposon sequences as well as short and low-quality reads and then were assembled using the CLC *de novo* assembler. The resulting consensus sequences were used in a reference-guided alignment. The single nucleotide polymorphism (SNP) analysis was performed using CLC's SNP analysis tool, applying the following parameters: window length = 7, maximum gap and mismatch count = 2, minimum central quality base = 35, minimum average quality for window bases = 30, minimum coverage = ×100, and minimum variant frequency = 1%. By choosing high-quality values for both the central base (the SNP base) and the window bases (bases surrounding the central SNP base, both upstream and downstream), as well as having a minimum coverage of ×100, only areas of high quality and high coverage were considered for SNP calling. Furthermore, all homopolymer regions, which we defined as regions with four or more consecutive bases of the same kind, were excluded from the SNP analysis due to the likelihood of insertion or deletion errors in these regions with Roche/454 pyrosequencing.

Calculation of PCR and Roche/454 pyrosequencing error rate. In order to ensure that errors introduced by PCR as well as errors inherent to the Roche/454 pyrosequencing technology were below our minimum variant frequency threshold of 1%, we sequenced an HCV-2b/1a DNA clone (GenBank accession no. JF779679) by the same methodology that was employed previously to sequence HCV-1a from the four treatment-naïve patients. We designed four overlapping PCR amplicons of approximately 2.5 kb each to amplify nucleotides 96 to 9172 of the HCV genome (Fig. 2A). Viral DNA was amplified using a Platinum *Taq* high-fidelity DNA polymerase kit (Invitrogen, Life Technologies, Carlsbad, CA). This kit contains the same DNA polymerase as the SuperScript III high-fidelity one-step RT-PCR kit that was used to amplify RNA from the four treatment-naïve patients. Primers specific to the HCV-2b/1a DNA clone were designed and are available upon request. We applied the same PCR conditions as those used in the one-step RT-PCR: 94°C for 30 s; 40 cycles of 94°C for 15 s, 55°C for 30 s, and 68°C for 3 min; and a final extension at 68°C for 5 min. After quantification and pooling, the DNA was sequenced by Roche/454 pyrosequencing as described above. In order to determine whether errors introduced by PCR and Roche/454 pyrosequencing were below the 1% minimum variant frequency threshold, we performed a SNP analysis using the same parameters as those described above.

Analysis of mutations conferring resistance against HCV protease (NS3/4a), NS5A, and polymerase (NS5b) inhibitors. Previously reported markers of resistance to hepatitis C virus inhibitors were analyzed in the four treatment-naïve patients, including mutations at 16 previously reported positions in the NS3/4a region (C16S, V36A/M/L/G, A39V, Q41R, F43S/I/V/C, T54S/A, V55A, Q80R/K/H/G/L, R109K, S138T, M175L, R155K/T/I/S/M/Q/L/G, V158I, A156V/T/S/I/G/F, D168A/V/E/G/N/T/Y/I/H, V170A/T, L170T, and I170V) (13, 17, 24, 27), 12 in the NS5A region (M28T, F28S, L23F, Q30E/H/R, L31M/F/V, Q54L, P32L, P58S, M21, Y93H/M/N, C92R, R318W, and D320E) (7, 10, 31), and 13 in the NS5B region (S282T/R, L314F, C316Y/F/S, M411S, M423T/I, P495S/L/A/T, V499A, M414I, C445F, Y448CH, Y452H, P496A/S, and L419M/V) (12, 17, 28). This analysis was based on the previously performed SNP analysis. Each mutation was located in the HCV genome, and its frequency within each patient was determined.

RESULTS

We obtained between 29,567 and 37,627 sequence reads for each of the four HCV genomes, resulting in an average coverage depth between 916× and 1,125× across the coding region (Fig. 1A). This deep coverage creates a high-resolution view of the viral population, revealing both the number and frequency of mutations within the quasispecies (34). Comparison of the SNP patterns between each of the four samples showed substantial differences in their distribution and frequency as well as in the proportion of synonymous versus nonsynonymous SNPs (Fig. 1B). The majority of mutations detected were present in less than 10% of the viral quasispecies, indicating that many of those mutations may have resulted from genetic drift (Table 1). Overall, the number of synonymous substitutions per synonymous site (dS) significantly exceeded the number of nonsynonymous substitutions per nonsynonymous site (dN) throughout the viral genome for all four patients ($P < 0.001$) (Table 2).

To demonstrate that errors introduced by RT-PCR as well as errors inherent to the Roche/454 pyrosequencing technology were below our minimum variant frequency threshold of 1%, an HCV-2b/1a DNA clone was sequenced using the same methodology as that previously employed to sequence HCV-1a from four treatment-naïve patients. This resulted in 97,367 reads aligning to the HCV-2b/1a reference sequence and an average coverage of 2,838× (Fig. 2A). Since the HCV-2b/1a sequence was clonal, all SNPs present were assumed to have been introduced through PCR and Roche/454 pyrosequencing. Using the same analysis parameters as those described above, SNPs were determined, and all were found to be present at a frequency below 1% (Fig. 2B). In order to account for the additional errors introduced during the reverse transcription step, the total error rate introduced by PCR and Roche/454 pyrosequencing of this clonal population was calculated. We determined the total number of SNPs (4,442) and divided this number by the total number of bases that were aligned to the HCV-2b/1a reference sequence (25,346,127 bp). This resulted in an error rate of 0.0175%. By adding the error rate of SuperScript III reverse transcriptase (0.0027%) (15), we obtained a total error rate of 0.0202%. Assuming a Poisson distribution, the probability that an error would be seen at a frequency of 1% or more at a site with coverage of just 100×, which was the lowest acceptable coverage for a SNP to be valid, was 0.0002. The probability that an error would be seen at a frequency of 1% or more at a site with coverage of 579× (the mean coverage for all four samples from treatment-naïve patients) was below 0.0000001.

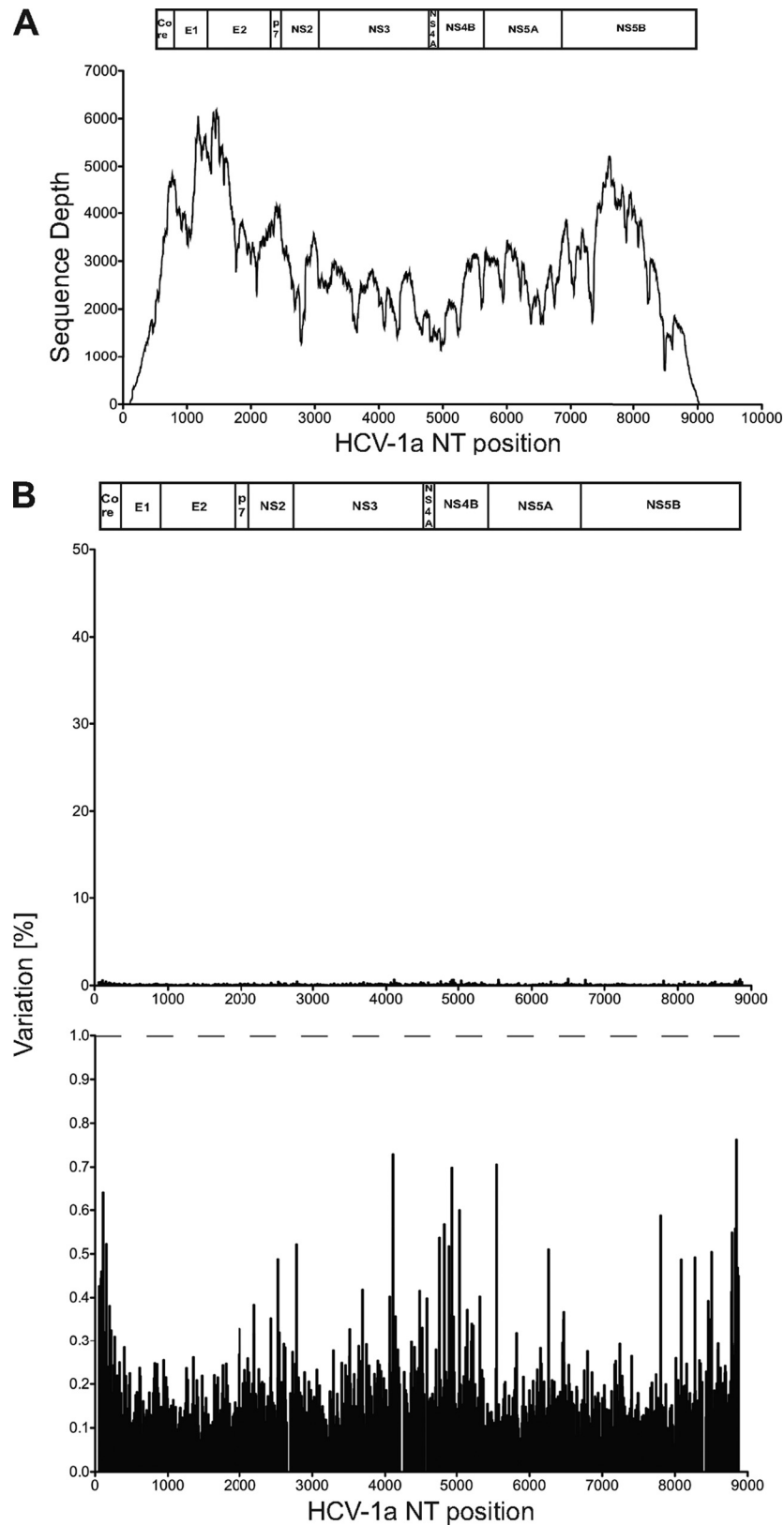In order to determine the phylogenetic relationships of the

FIG 2 Genomewide pyrosequencing of HCV-2b/1a DNA clone in order to determine error rates associated with PCR and Roche/454 pyrosequencing. (A) Sequence coverage obtained across the HCV-2b/1a genome. The HCV-2b/1a genomic organization is depicted at the top of the figure. The nucleotide numbering is based on the sequence of HCV strain H77 (GenBank accession no. NC_004102). (B) Graph indicating the SNP frequency across the HCV-2b/1a genome on a 50% y axis (top) and 1% y axis (bottom). The HCV-2b/1a genomic organization is depicted at the top of the figure. The nucleotide numbering is based on the sequence of HCV strain H77 (GenBank accession no. NC_004102).

TABLE 1 Summary of synonymous (S) and nonsynonymous (NS) mutation frequencies for each patient

| | No. of mutations | | | | | | | | | | | |
| | Patient A | | | Patient B | | | Patient C | | | Patient D | | |
Mutation frequency (%)	S	NS	Total (S + NS)	S	NS	Total (S + NS)	S	NS	Total (S + NS)	S	NS	Total (S + NS)
1–10	262	97	359	642	84	726	448	67	515	303	47	350
10–20	69	34	103	28	8	36	8	0	8	22	6	28
20–30	22	18	40	14	4	18	1	1	2	18	5	23
30–40	10	9	19	15	10	25	5	0	5	17	3	20
40–50	7	3	10	4	2	6	3	2	5	6	1	7
1–50	370	161	531	703	108	811	465	70	535	366	62	428

four nearly full-length HCV-1a sequences in relation to other known HCV-1a genomes, we performed a phylogenetic analysis based on consensus sequences generated from each of the four treatment-naïve patients. HCV subtype 1a isolates can be separated into two distinct subgenotypes (clade 1 and clade 2) (23). Analysis of the phylogenetic relationships confirmed the previous HCV-1a subtype classification, with viruses from all four patients grouping in clade 1 (black cluster) (Fig. 3).

To determine if any of the four treatment-naïve patients harbored preexisting mutations previously associated with resistance to the new HCV protease inhibitors, 16 sites in the NS3/4a region were examined. We also analyzed 12 sites in the NS5A region associated with resistance to NS5A inhibitors and 13 sites in the NS5B region associated with resistance to polymerase inhibitors (Table 3). Patient A showed a preexisting resistance mutation in the NS5B region (V499A), at a frequency of 98.7%. Patients B and D showed low-level drug resistance mutations in the NS5A (Q30R) and NS3/4A (I170V) regions that were present at frequencies of 2.5% and 1%, respectively (Table 3). No preexisting resistance mutations were detected for patient C.

DISCUSSION

Pyrosequencing provides a new opportunity to study the composition of the viral population in greater detail than was previously possible. Changes in the relative frequencies of variants may provide information about the fitness cost associated with

TABLE 2 Numbers of synonymous substitutions per synonymous site (dS) and nonsynonymous substitutions per nonsynonymous site (dN) in comparison to the consensus sequence for each gene

Protein	Median $dS \pm SE$[a]	Median $dN \pm SE$[a]	P value (paired t test)[b]
Core	0.0075 ± 0.0038	0.0002 ± 0.0001	NS
E1	0.0185 ± 0.0084	0.0030 ± 0.0020	NS
E2	0.0200 ± 0.0074	0.0041 ± 0.0016	NS
p7	0.0080 ± 0.0040	0.0008 ± 0.0004	NS
NS2	0.0072 ± 0.0013	0.0008 ± 0.0002	0.010
NS3	0.0111 ± 0.0024	0.0006 ± 0.0003	0.026
NS4a	0.0046 ± 0.0013	0.0002 ± 0.0002	0.045
NS4b	0.0067 ± 0.0031	0.0009 ± 0.0006	NS
NS5a	0.0076 ± 0.0014	0.0013 ± 0.0007	NS
NS5b	0.0080 ± 0.0009	0.0005 ± 0.0001	0.002
Mean ± SE for all proteins	0.0099 ± 0.0016	0.0012 ± 0.0004	<0.001

[a] Values represent the medians and standard errors for all four patients.
[b] NS, not significant.

a mutation or selective pressures exerted by the host's immune system. The fact that the majority of detected mutations were present in less than 10% of the viral population in our study may indicate that those mutations are selectively neutral, or nearly so, and thus subject to genetic drift. Still, their presence increases the overall genetic diversity and may enable the virus to respond more rapidly to changes in selection pressure. Interestingly, the prevalences of both synonymous (dS) and nonsynonymous (dN) substitutions differed for all four patients, and these substitutions were not particularly concentrated in the envelope region (Fig. 1B; Table 4). Although this region is used frequently in studies analyzing viral sequence heterogeneity (9, 18), the lack of focus on additional regions and our presented data indicate that the genetic variability of HCV might have been underestimated in previous studies.

Despite the high level of mutagenesis inherent to HCV replication, the number of synonymous substitutions per synonymous site (dS) significantly exceeded the number of nonsynonymous substitutions per nonsynonymous site (dN) throughout the viral genome for all four patients, providing evidence of purifying selection. This confirms similar results obtained by Sanger consensus sequencing (17, 20). Nonetheless, this does not exclude the possibility that in certain regions of the genome, dN could exceed dS due to locally mediated immune selection.

Due to HCV's high replication rate and the lack of an RNA polymerase proofreading system, variants potentially resistant to novel DAAs are constantly arising and may emerge quickly as the dominant species during antiviral treatment, even if initially present at a low frequency (3, 11). High-frequency mutations in the V499A codon of NS5B (>70%) have previously been reported for patients infected with HCV genotype 1 but were associated with only minor shifts in drug potency (16, 17, 31). While substitutions in the Q30R codon of NS5A are frequently selected in patients failing therapy with NS5A replication complex inhibitors (28), changes in the I170V codon in the NS3/4A region do not confer reduced susceptibility to boceprevir *in vitro* (33). These two minor variants would most likely be missed by traditional Sanger sequencing but were found using next-generation pyrosequencing. This demonstrates the advantage of generating a high-resolution view of the viral quasispecies with the presented method. Preliminary studies suggest that protease resistance mutations present in less than 1% of the viral quasispecies in a treatment-naïve individual can emerge as dominant mutations only days after treatment initiation (29), eventually leading to treatment failure as well as the development of cross-resistance to related compounds (10).

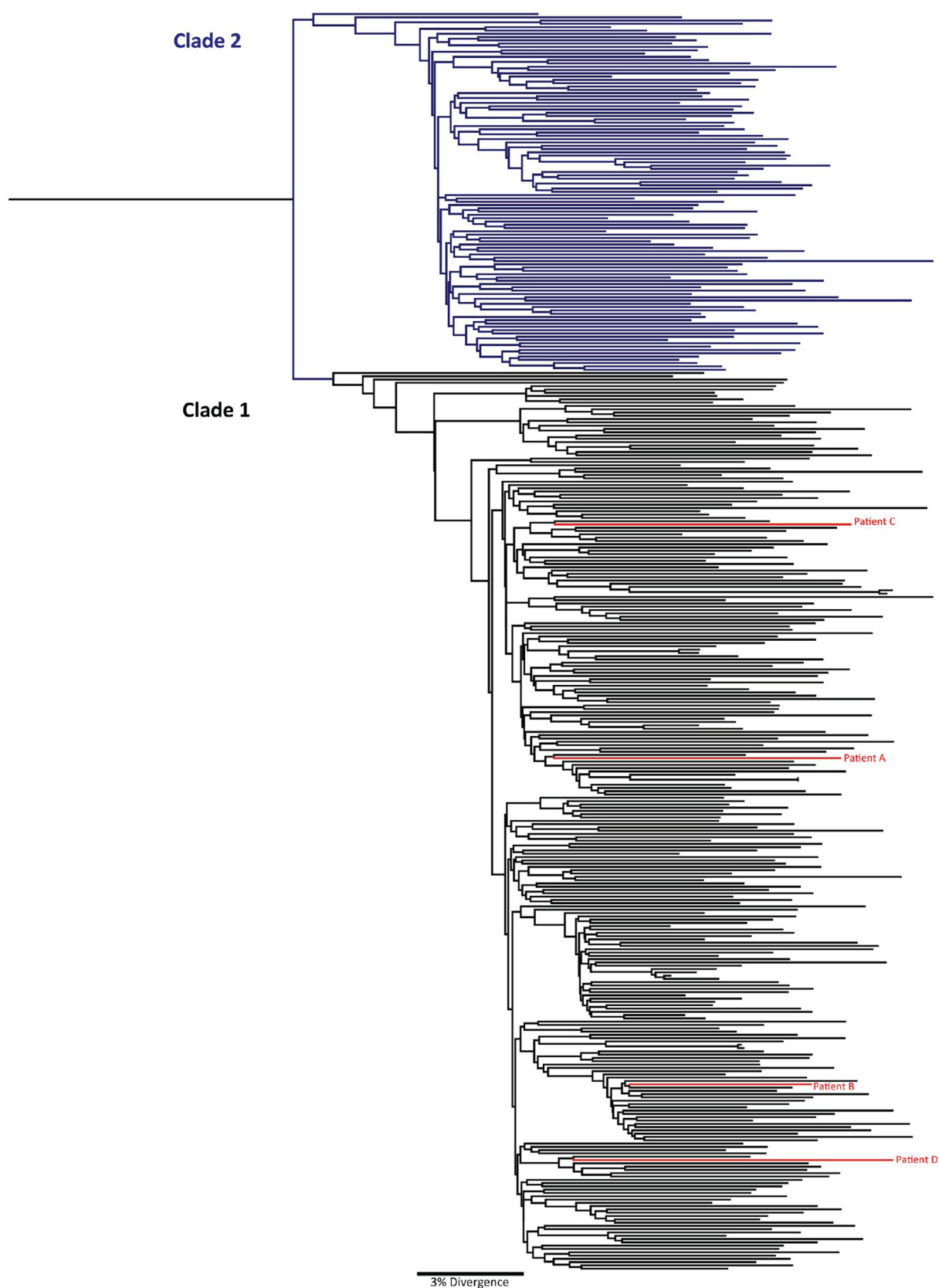**FIG 3** Phylogenetic tree showing evolutionary relationships between 382 complete HCV-1a genome sequences. A maximum likelihood phylogenetic tree is presented that demonstrates the separation of the HCV-1a isolates into two distinct clades: clade 1 (black) and clade 2 (blue). The four consensus sequences generated in this study (for patients A to D) are highlighted in red and grouped in clade 1.

TABLE 3 Characterization of preexisting drug resistance mutations identified in the NS3/4a, NS5A, and NS5B regions of viruses from patients A to D

HCV region	Mutation(s)[a]	Reference(s)	Prevalence (%) in viruses from patient			
			A	B	C	D
NS3/4a	C16S	17				
	V36A/M/L/G	17, 25, 28, 29				
	A39V	17				
	Q41R	17				
	F43S/I/V/C	13, 17				
	T54S/A	17, 28, 29				
	V55A	13, 28, 29				
	Q80R/K/H/G/L	13, 28				
	R109K	17				
	S138T	17				
	M175L	24				
	R155K/T/I/S/M/Q/L/G	13, 17, 25, 28, 29				
	V158I	24				
	A156V/T/S/I/G/F	13, 17, 28, 29				
	D168A/V/E/G/N/T/Y/I/H	13, 17, 28				
	I170V	33				99.0 (I), 1.0 (V)
NS5A	M28T or F28S	10				
	L23F	10				
	Q30R	10		97.5 (Q), 2.5 (R)		
	L31M/F/V	10, 28				
	Q54L	10				
	P32L	10				
	P58S	10				
	Y93H/M/N	10, 28				
	C92R	10				
	R262Q	7				
	R318W	7				
	D320E	7				
NS5b	S282T/R	17, 28				
	L314F	27				
	C316Y/F/S	17, 28				
	M411S	17, 28				
	M423T/I	17, 28				
	P495S/L/A/T	17, 28				
	V499A	12, 28	1.3 (V), 98.7 (A)			
	M414I	17				
	C445F	28				
	Y448CH	17, 28				
	Y452H	28				
	P496A/S	12, 28				
	L419M/V	28				

[a] Substitutions in bold were observed in this study.

Clearly, an understanding of low-frequency variants may be important in predicting DAA treatment outcomes.

In conclusion, we have reported the use of a transposon-based method that employs genomewide sequencing of HCV-1a with the ability to readily identify and quantify viral variants present at frequencies as low as 1%, thereby permitting an unprecedented assessment of HCV's diversity. The analysis of viral variability across the coding region allows identification of mutations previously associated with drug resistance and impaired or improved disease outcome. Additionally, these data may also accelerate the evaluation of mutational effects on other regions of the genome, as

well as the identification of new sites associated with resistance to drug treatment. While this method might not be applicable yet in a clinical setting, the basic methodology could be employed in the future, with greater multiplexing, in order to screen large populations of patients at a low cost per sample. This is particularly true with rapid advances in second- and third-generation sequencing technologies that are dramatically reducing the cost and time required to sequence whole genomes (35). Furthermore, the same approach could be utilized for longitudinal studies of HCV evolution in response to immune-mediated or drug-mediated pressure in a smaller number of subjects. While we have used this

TABLE 4 Mean numbers of synonymous substitutions per synonymous site (dS) and nonsynonymous substitutions per nonsynonymous site (dN) in comparison to the consensus sequence for each gene for patients A to D

Patient	Protein	Mean $dS \pm$ SE	Mean $dN \pm$ SE	P value (paired t test)[a]
A	Core	0.0185 ± 0.0108	0.0004 ± 0.0010	NS
	E1	0.0417 ± 0.0165	0.0089 ± 0.0045	NS
	E2	0.0407 ± 0.0121	0.0081 ± 0.0031	<0.01
	p7	0.0005 ± 0.0030	0.0001 ± 0.0008	NS
	NS2	0.0054 ± 0.0057	0.0007 ± 0.0012	NS
	NS3	0.0062 ± 0.0035	0.0003 ± 0.0005	NS
	NS4a	0.0010 ± 0.0049	0.0001 ± 0.0011	NS
	NS4b	0.0011 ± 0.0023	0.0001 ± 0.0005	NS
	NS5a	0.0037 ± 0.0033	0.0032 ± 0.0018	NS
	NS5b	0.0083 ± 0.0043	0.0008 ± 0.0008	NS
B	Core	0.0060 ± 0.0062	0.0000 ± 0.0000	NS
	E1	0.0200 ± 0.0116	0.0013 ± 0.0017	NS
	E2	0.0151 ± 0.0075	0.0048 ± 0.0024	NS
	p7	0.0153 ± 0.0170	0.0008 ± 0.0025	NS
	NS2	0.0100 ± 0.0075	0.0011 ± 0.0015	NS
	NS3	0.0164 ± 0.0057	0.0002 ± 0.0004	<0.01
	NS4a	0.0055 ± 0.0114	0.0006 ± 0.0023	NS
	NS4b	0.0154 ± 0.0086	0.0025 ± 0.0021	NS
	NS5a	0.0103 ± 0.0055	0.0010 ± 0.0010	NS
	NS5b	0.0104 ± 0.0048	0.0006 ± 0.0007	<0.05
C	Core	0.0019 ± 0.0034	0.0001 ± 0.0006	NS
	E1	0.0049 ± 0.0058	0.0008 ± 0.0013	NS
	E2	0.0061 ± 0.0048	0.0016 ± 0.0014	NS
	p7	0.0018 ± 0.0059	0.0004 ± 0.0017	NS
	NS2	0.0046 ± 0.0052	0.0002 ± 0.0006	NS
	NS3	0.0079 ± 0.0040	0.0014 ± 0.0031	NS
	NS4a	0.0044 ± 0.0101	0.0000 ± 0.0000	NS
	NS4b	0.0061 ± 0.0054	0.0002 ± 0.0005	NS
	NS5a	0.0072 ± 0.0046	0.0004 ± 0.0006	NS
	NS5b	0.0068 ± 0.0039	0.0002 ± 0.0003	NS
D	Core	0.0037 ± 0.0048	0.0000 ± 0.0000	NS
	E1	0.0076 ± 0.0072	0.0011 ± 0.0016	NS
	E2	0.0183 ± 0.0082	0.0016 ± 0.0014	<0.05
	p7	0.0145 ± 0.0166	0.0019 ± 0.0037	NS
	NS2	0.0087 ± 0.0072	0.0012 ± 0.0016	NS
	NS3	0.0139 ± 0.0052	0.0006 ± 0.0007	<0.05
	NS4a	0.0073 ± 0.0132	0.0000 ± 0.0000	NS
	NS4b	0.0044 ± 0.0046	0.0007 ± 0.0011	NS
	NS5a	0.0090 ± 0.0051	0.0005 ± 0.0007	NS
	NS5b	0.0066 ± 0.0039	0.0004 ± 0.0006	NS

[a] NS, not significant.

method to sequence human and simian immunodeficiency viruses as well as hepatitis C virus, this approach is applicable to virtually any heterogeneous viral pathogen.

ACKNOWLEDGMENTS

REFERENCES

1. **Bae A, et al.** 2010. Susceptibility of treatment-naive hepatitis C virus (HCV) clinical isolates to HCV protease inhibitors. Antimicrob. Agents Chemother. **54**:5288–5297.
2. **Bartels DJ, et al.** 2008. Natural prevalence of hepatitis C virus variants with decreased sensitivity to NS3.4A protease inhibitors in treatment-naive subjects. J. Infect. Dis. **198**:800–807.
3. **Bartenschlager R, Lohmann V.** 2000. Replication of hepatitis C virus. J. Gen. Virol. **81**:1631–1648.
4. **Bimber BN, et al.** 2009. Ultradeep pyrosequencing detects complex patterns of CD8$^+$ T-lymphocyte escape in simian immunodeficiency virus-infected macaques. J. Virol. **83**:8247–8253.
5. **Bimber BN, et al.** 2010. Whole-genome characterization of human and simian immunodeficiency virus intrahost diversity by ultradeep pyrosequencing. J. Virol. **84**:12087–12092.
6. **Butt AA, Kanwal F.** 2012. Boceprevir and telaprevir in the management of hepatitis C virus-infected patients. Clin. Infect. Dis. **54**:96–104.
7. **Coelmont L, et al.** 2010. DEB025 (alisporivir) inhibits hepatitis C virus replication by preventing a cyclophilin A induced cis-trans isomerisation in domain II of NS5A. PLoS One **5**:e13687.
8. **Farci P, et al.** 1991. A long-term study of hepatitis C virus replication in non-A, non-B hepatitis. N. Engl. J. Med. **325**:98–104.
9. **Figlerowicz M, et al.** 2010. Hepatitis C virus quasispecies in chronically infected children subjected to interferon-ribavirin therapy. Arch. Virol. **12**:1977–1987.
10. **Fridell RA, et al.** 2011. Distinct functions of NS5A in hepatitis C virus RNA replication uncovered by studies with the NS5A inhibitor BMS-790052. J. Virol. **85**:7312–7320.
11. **Garson JA, et al.** 1990. Detection of hepatitis C viral sequences in blood donations by "nested" polymerase chain reaction and prediction of infectivity. Lancet **335**:1419–1422.
12. **Gaudieri S, et al.** 2009. Hepatitis C virus drug resistance and immune-driven adaptations: relevance to new antiviral therapy. Hepatology **49**:1069–1082.
13. **Halfon P, Locarnini S.** 2011. Hepatitis C virus resistance to protease inhibitors. J. Hepatol. **55**:192–206.
14. **Hiraga N, et al.** 2011. Rapid emergence of telaprevir resistant hepatitis C virus strain from wildtype clone in vivo. Hepatology **54**:781–788.
15. **Ji JP, Loeb LA.** 1992. Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. Biochemistry **31**:954–958.
16. **Kukolj G, et al.** 2005. Binding site characterization and resistance to a class of non-nucleoside inhibitors of the hepatitis C virus NS5B polymerase. J. Biol. Chem. **280**:39260–39267.
17. **Kuntzen T, et al.** 2008. Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients. Hepatology **48**:1769–1778.
18. **Kurosaki M, Enomoto N, Marumo F, Sato C.** 1993. Rapid sequence variation of the hypervariable region of hepatitis C virus during the course of chronic infection. Hepatology **6**:1293–1299.
19. **Lavanchy D.** 2011. Evolving epidemiology of hepatitis C virus. Clin. Microbiol. Infect. **17**:107–115.
20. **Li H, et al.** 2011. Genetic diversity of near genome-wide hepatitis C virus sequences during chronic infection: evidence for protein structural conservation over time. PLoS One **6**:e19562.
21. **Martell M, et al.** 1992. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. J. Virol. **66**:3225–3229.
22. **Pawlotsky JM.** 2006. Hepatitis C virus population dynamics during infection. Curr. Top. Microbiol. Immunol. **299**:261–284.
23. **Pickett BE, Striker R, Lefkowitz EJ.** 2011. Evidence for separation of HCV subtype 1a into two distinct clades. J. Viral Hepat. **18**:608–618.
24. **Qiu P, et al.** 2009. Identification of HCV protease inhibitor resistance mutations by selection pressure-based method. Nucleic Acids Res. **37**:e74.
25. **Rong L, Dahari H, Ribeiro RM, Perelson AS.** 2010. Rapid emergence of

protease inhibitor resistance in hepatitis C virus. Sci. Transl. Med. **2**:30ra32.

26. **Schlutter J.** 2011. Therapeutics: new drugs hit the target. Nature **474**:S5–S7.

27. **Simen BB, et al.** 2009. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. J. Infect. Dis. **199**:693–701.

28. **Soriano V, et al.** 2011. Directly acting antivirals against hepatitis C virus. J. Antimicrob. Chemother. **66**:1673–1686.

29. **Susser S, et al.** 2009. Characterization of resistance to the protease inhibitor boceprevir in hepatitis C virus-infected patients. Hepatology **50**: 1709–1718.

30. **Suzuki T, Aizaki H, Murakami K, Shoji I, Wakita T.** 2007. Molecular biology of hepatitis C virus. J. Gastroenterol. **42**:411–423.

31. **Trevino A, et al.** 2011. Natural polymorphisms associated with resistance to new antivirals against HCV in newly diagnosed HIV-HCV-coinfected patients. Antivir. Ther. **16**:413–416.

32. **Verbinnen T, et al.** 2010. Tracking the evolution of multiple in vitro hepatitis C virus replicon variants under protease inhibitor selection pressure by 454 deep sequencing. J. Virol. **84**:11124–11133.

33. **Vierling JM, et al.** 2010. Frequencies of resistance-associated amino acid variants following combination treatment with boceprevir plus Pegintron (peginterferon alfa-2b)/ribavirin in patients with chronic hepatitis C, genotype 1 (G1). Hepatology **52**:702A.

34. **Wang GP, Sherrill-Mix SA, Chang KM, Quince C, Bushman FD.** 2010. Hepatitis C virus transmission bottlenecks analyzed by deep sequencing. J. Virol. **84**:6218–6228.

35. **Welch JS, Link DC.** 2011. Genomics of AML: clinical applications of next-generation sequencing. Hematology Am. Soc. Hematol. Educ. Program **2011**:30–35.

36. **Zwickl DJ.** 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. thesis. The University of Texas at Austin, Austin, TX.