Received: 10 November 2021 | Accepted: 4 February 2022

DOI: 10.1002/hep4.1929

ORIGINAL ARTICLE

Analysis of direct-acting antiviral-resistant hepatitis C virus haplotype diversity by single-molecule and long-read sequencing

Kozue Yamauchi | Mitsuaki Sato | Leona Osawa | Shuya Matsuda | Yasuyuki Komiyama | Natsuko Nakakuki | Hitomi Takada | Ryo Katoh | Masaru Muraoka | Yuichiro Suzuki | Akihisa Tatsumi | Mika Miura | Shinichi Takano | Fumitake Amemiya | Mitsuharu Fukasawa | Yasuhiro Nakayama | Tatsuya Yamaguchi | Taisuke Inoue | Shinya Maekawa | Nobuyuki Enomoto

Department of Gastroenterology and Hepatology, Faculty of Medicine, University of Yamanashi, Yamanashi, Japan

Correspondence
Shinya Maekawa, Department of Gastroenterology and Hepatology, Faculty of Medicine, University of Yamanashi, 1110 Shimokato, Chuo, Yamanashi 409-3898, Japan.
Email: maekawa@yamanashi.ac.jp

Funding information
This study was partly supported by the Japan Agency for Medical Research and Development (grant numbers 21fk0210047h0003, 21fk0210058s0703, 21fk0210065s0602, 21fk0210072s0602, and 21fk0210067s0902) and Grants-in-Aid for Scientific Research (grant numbers 19H03638 and 20K08281)

Abstract

The method of analyzing individual resistant hepatitis C virus (HCV) by a combination of haplotyping and resistance-associated substitution (RAS) has not been fully elucidated because conventional sequencing has only yielded short and fragmented viral genomes. We performed haplotype analysis of HCV mutations in 12 asunaprevir/daclatasvir treatment-failure cases using the Oxford Nanopore sequencer. This enabled single-molecule long-read sequencing using rolling circle amplification (RCA) for correction of the sequencing error. RCA of the circularized reverse-transcription polymerase chain reaction products successfully produced DNA longer than 30 kilobase pairs (kb) containing multiple tandem repeats of a target 3 kb HCV genome. The long-read sequencing of these RCA products could determine the original sequence of the target single molecule as the consensus nucleotide sequence of the tandem repeats and revealed the presence of multiple viral haplotypes with the combination of various mutations in each host. In addition to already known signature RASs, such as NS3-D168 and NS5A-L31/Y93, there were various RASs specific to a different haplotype after treatment failure. The distribution of viral haplotype changed over time; some haplotypes disappeared without acquiring resistant mutations, and other haplotypes, which were not observed before treatment, appeared after treatment. *Conclusion*: The combination of various mutations other than the known signature RAS was suggested to influence the kinetics of individual HCV quasispecies in the direct-acting antiviral treatment. HCV haplotype dynamic analysis will provide

novel information on the role of HCV diversity within the host, which will be useful for elucidating the pathological mechanism of HCV-related diseases.

INTRODUCTION

Persistent infection with hepatitis C virus (HCV) causes cirrhosis and hepatocellular carcinoma. It is estimated that 1 million people in Japan[1] and 58 million people worldwide have chronic HCV infection, and 0.3 million people die annually from hepatitis C globally, mostly from cirrhosis and hepatocellular carcinoma (https://www.who.int/news-room/fact-sheets/detail/hepatitis-c). Introduction of direct-acting antivirals (DAAs) began in 2011, and now the sustained virological response (SVR) rate of treated patients is as high as over 95%.[2] In 2016, the World Health Organization proposed a global health sector strategy with the goal of eliminating HCV infection by 2030 and developed an action plan to facilitate this goal.[3] However, problems remain that need to be addressed to eliminate HCV, such as virus resistance to DAAs, progression of fibrosis and development of hepatocellular carcinoma even after SVR, and social barriers for detecting and treating all infected patients.[3–5]

Mutations in the viral gene are involved in the cause of DAA-resistant virus,[6] disease progression,[7] and hepatocarcinogenesis[8] as well as host factors, and analysis of the viral gene is essential for its elucidation.[9] Until now, the HCV gene has been analyzed by traditional short-read Sanger or deep sequencing.[10] The presence of signature mutations was clinically found regarding resistance mutations to DAA,[11] and these mutations were shown to cause a decrease in drug sensitivity the cell-culture system and the replicon system.[12,13] However, these signature mutations have so far been analyzed separately only in part of the HCV genome, and it is not clear how the haplotype, which is a combination of the entire structure or various mutations, affects the disease state and therapeutic effect.[14] This is because these sequencing technologies have been able to sequence only short molecules up to several hundred base pairs (bp) and cannot discriminate whether detected distant mutations are on a single viral genome molecule or on different ones.[15]

Recently, a long-read single-molecule sequence, which is a third-generation sequencing technology, has been developed and can be applied to the analysis of viral genes.[16] These techniques can analyze the nucleotide sequence of a single molecule in several tens of kilobases (kb) and can analyze the large-scale structure of a gene. In this study, we revealed the change in combinations of HCV mutations (haplotype) on single HCV molecules in DAA treatment failure, using the Oxford Nanopore sequencer, which is a third-generation sequencing technology.[17,18] Using third-generation sequencing, long read is possible, but the read accuracy is low at about 90%. To overcome this problem, we used rolling circle amplification (RCA)-Nanopore sequencing in which the circularized polymerase chain reaction (PCR) amplification product of the HCV genome was amplified by RCA to produce long tandem repeats (concatemer) of the original sequence. These long tandem repeats of single molecules were sequenced by a Nanopore sequencer to correct sequencing errors and to obtain consensus, enabling single-molecule long-read sequencing.[19]

PATIENTS AND METHODS

Patients

Twelve patients with chronic HCV genotype 1b infection and who were unsuccessfully treated with asunaprevir (ASV)/daclatasvir (DCV) treatment (60 mg DCV once/day and 200 mg ASV twice/day for 24 weeks of the standard protocol) for chronic hepatitis C in 2015 at the University of Yamanashi Hospital were enrolled in this study. Serum samples were collected from all patients during treatment and store at −20°C before analysis. HCV genome sequences in the serum of each patient were analyzed at the start of ASV/DCV treatment, at treatment failure, and at the end of follow-up. The clinical course of each patient, including the time points of sample collection and viral titers, are shown in Table 1. The study was reviewed and approved by the Institutional Review Board and performed with the necessary patient consent. This study protocol complied with all provisions of the Declaration of Helsinki.

RNA extraction and reverse transcription

Total RNA was extracted in 30 μL of RNA solution from 200 μL of serum by using the QIAmp Mini Elute Virus Spin Kit (Qiagen, Germany). To synthesize complementary DNA (cDNA) as described,[20] 7 μL of RNA solution was reacted at 49°C for 65 min (followed by 85°C for 5 min) in 20 μL of reverse-transcription (RT) solution of SuperScript III reverse transcriptase (Thermo Fisher Scientific, Waltham, MA), RV1 primer[21] or a pangenomic primer Vir7 (oligo-2′-deoxyadenosine [oligo dA_{20}]),[20] deoxyribonucleotide triphosphates (dNTPs), and betain (Sigma Aldrich). Two units of ribonuclease H were then added before a final incubation at 37°C for 20 min.

TABLE 1 Course of ASV/DCV therapy and sampling time points for sequencing analyses

Case No.	Cause of Treatment Failure	Duration of Treatment (Week)	Time Period (Week) From the Start of ASV/DCV Administration to Sampling Time Points[a]						HCV-RNA Titer (LogIU/mL) of Samples for Nanopore Sequence		
			Start of Treatment		Treatment Failure		End of Follow-Up		Start of Treatment	Treatment Failure	End of Follow-Up
			Nanopore	Direct	Nanopore	Direct	Nanopore	Direct			
1	NVR[b]	7	−9	−9	11	11	247	135	6.8	7.0	6.5
2	Breakthrough	5	0	−45	5	8	55	73	5.5	4.7	5.9
3	Breakthrough	13	−9	−58	15	13	161	53	7.0	5.6	7.1
4	Breakthrough	24	0	−103	28	28	133	53	6.2	6.4	5.3
5	Breakthrough	10	−2	−15	45	45	98	52	5.5	5.3	5.8
6	Breakthrough	24	−8	−18	28	28	151	151	6.2	6.6	6.2
7	Breakthrough	22	0	−54	27	22	79	79	4.8	6.8	7.0
8	Relapse	24	−8	−25	35	35	146	52	6.1	5.7	6.0
9	Relapse	24	−12	−41	46	41	156	67	6.6	6.6	7.0
10	Relapse	24	0	−13	38	38	177	177	5.8	5.2	4.9
11	Tx withdrawn by AE	2	−4	−24	34	30	160	160	6.6	6.1	6.7
12	Tx withdrawn by AE	2	−14	−14	13	7	125	36	6.1	6.0	6.4

Abbreviations: AE, adverse event; NVR, null viral response; Tx, treatment.

[a]Sampling time points were different between direct and Nanopore sequencing in some cases.

[b]NVR (HCV RNA not becoming undetectable during the treatment course).

PCR reaction

Approximately 3 kb of the sequence spanning the NS3 and NS5A regions of the HCV genome were PCR amplified using PrimeSTAR GXL (Takara, Japan). PCR primers designed based on a published report[21] are listed in Table 2. The first PCR was performed with cDNA in 2.5 μL of RT solution (equivalent to 5.83 μL of serum) using the first-round sense and antisense primers, dNTPs, and betaine. The cycling condition was preincubated at 98°C for 2 min, 40 cycles at 98°C for 10 seconds, 60°C for 15 seconds, and 68°C for 3 min, with a final extension for 3 min. The nested PCR was performed with the first PCR product as described for the first PCR, using the second-round sense and antisense primers. If needed, the second-round sense primers with barcode sequences of 24 nucleotides (following the Nanopore PCR Barcoding Kit; Oxford Nanopore Technologies, Oxford, United Kingdom) attached at the 5′ end were used so that the PCR products from each sample were identifiable after pooled Nanopore sequencing. After extraction of the target PCR products from agarose gel electrophoresis using the Monarch DNA Gel Extraction Kit (New England Biolabs, Ipswich, MA), PCR products were quantified using a pico green double-stranded DNA assay kit (Invitrogen, Japan), the concentration of each sample was adjusted for standardization, and pooled samples were prepared for further analyses.

Ligation

Circular ligation of pooled PCR products was performed with the Quick Ligation Kit (New England Biolabs). The mixture was incubated at room temperature for 5 min. Ligated products were purified and concentrated using the Monarch PCR and DNA Cleanup Kit (New England Biolabs). Plasmid-safe adenosine triphosphate-dependent deoxyribonuclease (DNase) (Epicentre Biotechnologies, Madison, WI) digest was performed at 37°C for 1 hour to remove noncircularized products. The DNase was then heat inactivated at 70°C for 30 min.

RCA reaction

The circularized DNA template molecules were purified and concentrated by the Monarch PCR and DNA Cleanup Kit (New England Biolabs). The RCA reaction was performed using the TruePrime RCA Kit (Expedeon). Conditions for RCA were 30°C for 24 hours before heat inactivation at 65°C for 10 min. RCA products were purified with Agencourt AMPure XP beads (Beckman Coulter) and eluted in distilled water. Then NEB buffer 2 and T7 endonuclease (New England Biolabs) were added and incubated at 37°C for 90 min to resolve potential branching generated by the RCA. Debranched products were purified using AMPure XP beads and eluted. To eliminate low-molecular DNA, the Short-Read Eliminator Kit (Circulomics, Baltimore, MD) was used according to the manufacturer's protocol.

Nanopore sequencing reaction

Libraries were prepared with 1D amplicon/cDNA by ligation protocol (SQK-LSK109; Oxford Nanopore Technologies) and sequenced using R9.4.1 or R10.3 flowcells on a MinION Mk1B sequencer using MinKnow software on a personal computer or MinIT system (Oxford Nanopore Technologies) for 24 to 48 hours. Barcode sequences were appended on the templates using the Nanopore Barcode Kit (Oxford Nanopore Technologies), or the identical sequences were attached to the 5′ of the second-round sense primers for later multiplex analysis.

Bioinformatic analyses

The raw Nanopore reads of the RCA concatemers were base called using MinKnow and Guppy software (Oxford Nanopore Technologies), generating fastq files from fast5 files. We filtered the fastq reads of each sequencing run (those with a quality score <7–10 according to total read counts to avoid processing an excessive amount of sequence data, and length <30 kb) by using seqkit (https://bioinf.shenwei.me/seqkit/), Nanofilt (https://github.com/wdecoster/nanofilt), and Filtlong (https://github.com/rrwick/Filtlong) to remove short low-quality reads. Distribution of read quality and length was confirmed using NanoPlot (https://github.com/wdecoster/NanoPlot). We then used RCAcorrect (https://github.com/hr283/RCAcorrect) to demultiplex and to obtain the consensus sequence from concatemers of the longest 100 reads amplified from each sample by RCA from a single circularized molecule for error correction and variant calling. Ambiguous sites were called N for nucleic acids and X for amino acids. The RCAcorrect were modified for HCV sequencing, demultiplexing, and minimap2 (https://github.com/lh3/minimap2) instead of the BWA-MEM algorithm (https://github.com/lh3/bwa) so that sequence reads with gaps of 1% or more length of the target sequence and those with insertions were not used for alignment. Multiple alignments of quasispecies generated by RCAcorrect were visualized using Aliview (https://github.com/AliView/AliView), and we generated neighbor-joining phylogenetic trees by using MEGA X (https://www.megasoftware.net/), without excluding signature resistance-associated substitution (RAS). Several haplotypes were detected by visual inspection of the phylogenetic trees, and sequence alignments according to the haplotypes were generated using ete3 (https://pypi.org/project/ete3/). Statistical analysis

TABLE 2 Primers for HCV genome PCR amplification

Primer Name	Sequence (5′-3′)[a]	Binding Position[b]	Case No.					
			All (Except Right)	1	2	4	5	8
Sense Primer								
FW3[21]	ACAGGTCGGGACAAGAACCAG	3483–3503	S-1, F-1, E-1	F-1	E-1, R-1	S-1, E-1	S-1, F-1, E-1	S-1, E-1
FW4[21]	ACAAGAACCAGGTCGATGGGGAG	3493–3515	S-2, F-2, E-2		E-2	E-2	S-2, E-2	S-2
FW3new	CACAGGCCGGGACAAGAACCA	3482–3502		F-1	F-1	S-1	F-1	E-1
FW4new	GTCGAGGGGGAGGTTCAAGTGGT	3504–3526		F-2	F-2	S-2	F-2	E-2
FW5	GTYCARGTGGTTTCYACCGCWACRCA	3516–3541		S-1, E-1	S-1			F-1
FW6	TCYTTCYTGGCGACCTGYRTCAAYGG	3543–3568		S-2, E-2	S-2			F-2
FWout	CTTGGTTGCATCATCACTAGCCTCACAGG	3459–3487				F-1		
FWin	ATCCCCCACCCGCGTAACCTCCACGTACTC	3480–3508				F-2		
Antisense primer								
RV1[21]	CCCGTCACGTAGTGGAAATC	6633–6652	S-1, F-1, E-1	F-1	E-1, R-1	S-1, E-1	S-1, F-1, E-1	S-1, E-1
RV2[21]	GCGTAACCTCCACGTACTCC	6605–6624	S-2, F-2, E-2		E-2	E-2	S-2, E-2	S-2
RV2new	CGCGTAACCTCCACGTACTC	6606–6625		F-2	F-2	S-2	F-2	E-2
RV5	CTCAGCRGCYACCCGCCACAGCGCC	6581–6605		S-1, E-1	S-1			F-1
RV6	TGGARTARTTTGGMGCYGGGGAGGG	6555–6579		S-2, E-2	S-2			F-2
RVout	GTCAGTGGTCATGCCCGTCACGTAGTGGAA	6636–6665				F-1		
RVin	ATCCCCCACCCGCGTAACCTCCACGTACTC	6606–6635				F-2		

Abbreviations: 1, first-round PCR; 2, second-round PCR; E, end of follow-up; F, treatment failure; S, start of treatment.

[a]For degenerate primers, B = C or G or T; H = A or C or T; M = A or C; N = A or C or G or T; R = A or G; S = C or G; W = A or T; V = A or C or G; D = A or G or T; Y = C or T.

[b]Genome binding position with reference to HCV GT1b strain, HC-J4L6S; GenBank accession AF054247.1.

was performed using metadata-driven comparative analysis tool for sequences (meta-CATS) (https://www.viprbrc.org/brc/home.spg?decorator=flavi_hcv) to identify haplotype-associated substitution (HAS) and RAS among start of treatment, treatment failure, and end of follow-up. To quantify the change of viral haplotype diversity, the normalized Shannon entropy was calculated as described.[22]

Control experiments of Nanopore sequencing

In addition to sequencing HCV from clinical samples, we also applied our sequencing methods to plasmids to investigate the specificity and sensitivity for mutation detection of our approach using a template from plasmids containing different HCV sequences. We used the HCV plasmid full-genome construct of pCV-J4L6S[23] (kindly provided by Dr. Yanagi, National Institutes of Health, Bethesda, MD), and pORN/C-5B[24] (kindly provided by Dr. Ikeda, Okayama University, Okayama, Japan) as the control template for PCR, RCA, and Nanopore sequencing, along with the representative clinical samples obtained from patients with HCV infection and liver cirrhosis (LC). To evaluate the threshold values at which fewer sequences can be detected, PCR products from the plasmids having pORN/C-5B were mixed at 2%, 9%, 17%, and 50% ratios with those from pCV-J4L6S and were Nanopore sequenced.

Sanger sequencing and deep sequencing of signature RAS

Total RNA was extracted from 150 μL of serum, using the QIAquick Viral Mini Kit (Qiagen), and cDNA was synthesized with random hexamer and SuperScript II (Thermo Fisher) and amplified by PCR separately for the NS3 and NS5A regions containing target RASs.[22,25] Direct (populational) sequencing of the PCR product for targeted signature RASs was carried out by Sanger sequencing and deep sequencing with the next-generation sequencer GS Junior System (Roche, Branford, CT) using the described protocol.[22,25]

RESULTS

Outline of Nanopore sequencing

The graphical outline of the long-read single-molecule sequencing using the Nanopore sequencer is shown in Figure 1. We successfully obtained long reads more than 30 kb consisting of the concatemer of the single-molecule HCV sequence and determined their consensus sequences to correct Nanopore sequencing errors by aligning the tandem-repeated sequences amplified by RCA. The approximate 3.2-kb product amplified by RT-PCR was circularized and amplified by RCA, and the synthesis of high-molecular weight DNA was confirmed by agarose gel electrophoresis. When the RCA product was treated with a restriction enzyme (*Sac I*), which generates 2.8 kb and 0.4 kb fragmentes, or with *Ale I*, which generates 1.9-kb and 1.3-kb fragments from each repeat sequence, DNA fragments with expected sizes were detected on agarose gel electrophoresis, confirming that the circularized molecules were amplified by RCA as a concatemer (Figure 1).

By processing the PCR-RCA product derived from the HCV plasmid with the short-read eliminator before Nanopore sequencing, it was possible to efficiently remove reads of 10 kb or less. The proportion of reads over 30 kb was approximately 0.016% but increased to 1.5% after the short-read eliminator treatment (filtered reads/total reads, 126/780,000 versus 1945/130,804; filtering condition, quality score <8, length <30 kb). A snapshot of an example of a 70-kb long read that was searched, using the Basic Local Alignment Search Tool (BLAST), with the HCV sequence (https://blast.ncbi.nlm.nih.gov/Blast.cgi) is shown in Figure 1. HCV sequences formed a concatemer as tandem repeats, and consensus sequencing of these repeats reconstructed the original sequence of the single HCV molecule. When these reads were subjected to multiplex sequencing using barcodes, it was possible to analyze and separate multiple samples from different sera in a single-sequence run.

Control experiments of Nanopore sequencing

Nanopore sequencing using plasmid containing pCV-J4L6S as templates showed under a 0.1% substitution rate, indicating Nanopore sequencing with RCA achieved high accuracy (lower right graphs in Figure 1). In contrast, about a 1% substitution rate was observed in clinical samples from patients with LC, suggesting the reflection of quasispecies. It was possible to detect as low as 2% minor sequences in analysis of 100 reads in a semiquantitative manner when PCR products from different plasmids (pCV-J4L6S and pORN/C-5B) were mixed in various proportions, indicating high specificity of the present method for detecting quasispecies.

Nanopore sequencing results from clinical samples

Details of the results of nine independent Nanopore sequencing runs (run numbers 1–9), including a total of 36 samples consisting of three time points (start of treatment, treatment failure, and end of treatment)

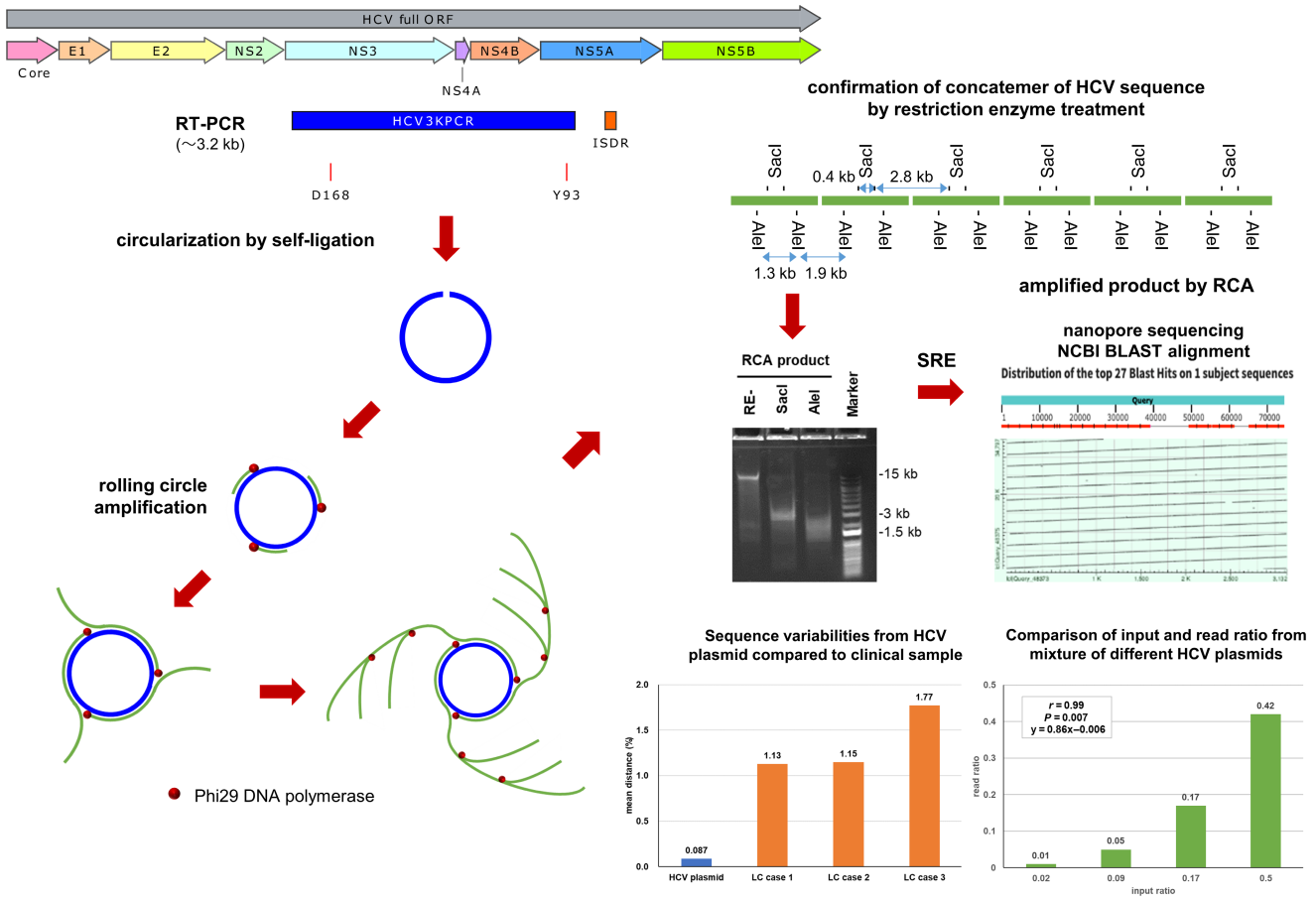


FIGURE 1 Method of RCA for Nanopore sequencing. Single-molecule long-read sequencing analysis of HCV subgenome was carried out with RCA and Nanopore sequencing. Approximately 3 kb of HCV fragment was generated by RT-PCR and circularized by self-ligation. RCA was performed, and generation of the concatemer of HCV sequence was confirmed by restriction enzyme treatment. Following the treatment with SRE to remove short concatemer, Nanopore sequencing was performed. The representative single-molecule sequencing result using BLAST alignment showed the repeat of the amplified sequence. Nanopore sequencing using plasmid containing pCV-J4L6S as templates had under 0.1% substitution rate, showing Nanopore sequencing with RCA achieved high accuracy; about 1% substitution rate was detected in LC cases, suggesting the presence of quasispecies. It was possible to detect as low as 2% of minor sequences in 100-reads analysis in a proportion-dependent semiquantitative manner when PCR products from different plasmids (pCV-J4L6S and pORN/C-5B) were mixed in various proportions, indicating the high specificity of the present method for detection of quasispecies. ISDR, interferon sensitivity-determining region; NCBI, National Center for Biotechnology Information; ORF, open reading frame; RE-, without restriction enzyme; SRE, short-read eliminator

from 12 patients (case numbers 1–12) are shown in Table 3. Each sequencing run included different combinations of patients and time points because of an arbitrary order of sample preparation. Read counts, length, and quality scores are shown before and after filtering with length >30 kb and quality >7–10 and after demultiplexing. Total counts of reads were 22,164–384,000 (depending on the applied amounts of samples and run time) and were sufficient for downstream analyses. After filtering the short or low-quality reads, 735–3802 reads were usable for demultiplexing, and the read sets from a total of 36 different samples were obtained with read counts of 121–310, length of 30.0–107.8 kb, and quality score 7.0–12.2. The sum of read counts after demultiplexing does not necessarily match with read counts after filtering because each sequencing run included samples for other experiments. From each read set after filtering and demultiplexing, the longest 100 reads were selected to determine the consensus of the multiple repeats in the concatemer generated by RCA from the single molecules.

Signature RAS detected by populational and Nanopore sequencing

A summary of the signature RAS detected by Nanopore and direct (populational) sequencing at NS3-D168, NS5A-P29, NS5A-P32, NS5A-L31, and NS5A-Y93 in 12 patients is shown in Table 4.[12] All patients had RASs in these positions during the course with changes over time. Nanopore sequencing detected most of the signature mutations in a quantitative manner. Although

TABLE 3 Details of Nanopore sequencing results from clinical samples

		Before Filtering		After Filtering			After Demultiplexing				
Run No.	Total Read Counts	Read Length, kb Median (Range)	Quality Score Median (Range)	Usable Read Counts	Read Length, kb Median (Range)	Quality Score Median (Range)	Case No.	Time Point	Usable Read Counts	Read Length, kb Median (Range)	Quality Score Median (Range)
1	32,062	5.5 (0–87.8)	9.2 (7.0–14.5)	511	37.8 (30.0–87.6)	10.0 (9.0–11.9)	5	start	194	35.3 (30.0–87.6)	10.0 (9.0–11.8)
							5	end	107	35.6 (30.0–83.3)	10.0 (9.0– 11.9)
							3	start	191	35.6 (30.0–66.6)	9.9 (9.0–11.6)
2	22,164	7.1 (0–89.9)	8.2 (7.0–12.3)	1189	36.2 (30.0–96.4)	7.9 (7.0–10.7)	3	end	190	38.0 (30.0–89.9)	8.0 (7.0–10.7)
							9	start	147	39.8 (30.0–82.6)	7.9 (7.0–10.5)
							9	failure	310	36.0 (30.0–83.4)	8.6 (7.0–10.6)
							9	end	304	36.0 (30.0–82.2)	8.1 (7.0–10.4)
3	21,054	2.6 (0–99.6)	7.6 (7.0–10.5)	735	35.8 (30.0–99.6)	7.6 (7.0–9.2)	6	start	279	35.4 (30.0–88.8)	7.5 (7.0–8.8)
							6	failure	125	36.5 (30.0–99.6)	7.6 (7.0–9.1)
							6	end	302	35.8 (30.0–88,8)	7.7 (7.0–9.2)
4	384,000	4.3 (0–186.7)	8.2 (7.0–14.4)	3147	35.7 (30.0–87.9)	10.2 (8.0–12.2)	3	failure	152	35.6 (30.0–65.6)	10.4 (9.0–11.7)
							7	start	122	35.6 (30.0–73.9)	10.6 (10.0–12.2)
							7	failure	124	35.1 (30.0–61.1)	10.5 (10.0–12.2)
							7	end	120	39.7 (30.0–85.3)	11.0 (10.0–12.2)
							8	start	275	38.7 (30.0–85,3)	10.2 (9.0–11.9)
							8	failure	124	37.7 (30.0–64.0)	10.2 (9.0 11.8)
							8	end	225	35.9 (30.0–72.3)	10.2 (9.0–11.9)
							12	start	133	35.8 (30.0–62.9)	10.2 (9.0–12.0)
							12	end	199	35.9 (30.0–69.6)	10.7 (10.0–12.2)
5	222,000	4.9 (0–159.0)	8.6 (7.0–13.4)	2923	37.2 (30.0–138.7)	9.3 (9.0–10.1)	2	failure	121	35.2 (30.0–95.8)	9.3 (9.0–10.0)
							2	end	121	35.8 (30.0–81.7)	9.5 (9.0–9.9)
							4	end	121	36.5 (30.0–107.8)	9.4 (9.0–9.8)
							5	failure	121	36.1 (30.0–63.1)	9.3 (9.0–9.7)
							10	start	125	36.4 (30.0–84.8)	9.5 (9.0–10.0)
							11	start	121	38.0 (30.0–97.2)	9.4 (9.0–9.9)
							11	failure	119	38.6 (30.0–103.1)	9.6 (9.0–10.0)
							11	end	121	38.9 (30.0–92.4)	9.4 (9.0–9.9)
							12	failure	125	38.9 (30.0–81.7)	9.4 (9.0–9.9)

(Continues)

TABLE 3 (Continued)

Run No.	Total Read Counts	Before Filtering		After Filtering			After Demultiplexing				
		Read Length, kb Median (Range)	Quality Score Median (Range)	Usable Read Counts	Read Length, kb Median (Range)	Quality Score Median (Range)	Case No.	Time Point	Usable Read Counts	Read Length, kb Median (Range)	Quality Score Median (Range)
6	212,000	5.8 (0–208.1)	10.1 (7.0–14.6)	3802	38.2 (30.0–151.5)	11.1 (10.0–12.6)	4	start	129	38.2 (30.0–73.5)	11.5 (11.0–12.3)
							10	end	121	36.4 (30.0–89.6)	10.4 (9.0–12.2)
7	104,998	8.3 (0–199.2)	9.5 (7.0–13.6)	2437	37.1 (30.0–156,3)	10.5 (9.0–12.2)	2	start	127	36.9 (30.0–66.5)	10.9 (10.0–11.7)
							4	failure	126	36.5 (30.0–71.3)	11.2 (11.0–11.7)
							10	failure	128	37.6 (30.0–90.5)	10.8 (10.0–12.2)
8	164,000	6.4 (0–152.4)	9.3 (7.0–14.6)	1973	35.9 (30.0–105.7)	10.6 (10.0–12.7)	1	start	182	35.2 (30.0–76.6)	10.6 (10.0–11.8)
							1	failure	184	36.5 (30.0–83.4)	11.0 (10.0–12.0)
							1	end	187	35.6 (30.0–80.5)	11.0 (10.0–12.1)

Abbreviations: end, end of follow–up; failure, treatment failure; start, start of treatment.

sampling time points of Nanopore sequencing were not necessarily the same for all cases with direct sequencing (Table 1), information of populational sequencing results from the same treatment phases (start of treatment, treatment failure, and end of follow-up) were useful for evaluating the validity of Nanopore sequencing. Analyses of 100 reads at each time point could find the minor RAS that was not detected by direct sequencing, finding most patients had a minor viral population with various types of substitutions at these signature RAS positions. The results of target deep sequencing of case 2 and case 10 showed a similar proportion of major substitution with Nanopore sequencing (Table 5). In case 5, NS5A-P29 deletion mutant, which is highly resistant to pibrentasvir, (https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/209394Orig1s000MicroR.pdf) was successfully detected with Nanopore sequencing. Nanopore sequencing revealed that the proportion of signature RAS found at treatment failure decreased at the end of follow-up, indicating that selective pressures of the DAA drove the changes of the viral population during the period 50–236 weeks after treatment failure (Table 1). Moreover, single-molecule long-read sequencing by the Nanopore sequencer enabled us to investigate the relation between these RASs in each virus as well as other substitutions consisting of viral haplotype, as described below.

Single-molecule, long-read sequencing of HCV haplotype during DAA treatment

The alignment (left panel) and phylogenetic (right panel) analysis of 100 long reads at three time points (start of treatment, treatment failure, end of follow-up) in 12 patients is shown in Figure 2. Amino acid substitutions from the consensus for each patient are depicted in alignment panels. Phylogenetic trees with a scale bar for genetic distance for each patient showed that several haplotypes can be detected in each patient. In the alignments of individual HCV genomes from single patients, numerous substitutions composing quasispecies found over the entire length of the PCR-amplified NS3 to NS5A region revealed the large-scale structures of quasispecies that have not been elucidated by short-read deep sequencing to date. Multiple nontypical RASs could be detected in a patient-specific manner. There were multiple HASs other than signature RASs in each patient. Haplotypes were not always consistent with the three time points, indicating that each haplotype reacted differently to DAA. In each haplotype, there were haplotype-specific RASs that changed over time during treatment in a haplotype-specific manner. Genetic distance of each phylogenetic tree can be estimated with scale bars that differ among cases; for example, cases 1 and 7 showed much lower genetic distances compared to other cases.

TABLE 4 Signature RAS found in NS3 and NS5A during ASV/DCV therapy

Case No.	NS3-D168						NS5A-L31						Treatment Failure to End of Follow-Up (Week)[c]
	Start of Treatment		Treatment Failure		End of Follow-Up		Start of Treatment		Treatment Failure		End of Follow-Up		
	Nanopore[a]	Direct[b]	Nanopore	Direct	Nanopore	Direct	Nanopore	Direct	Nanopore	Direct	Nanopore	Direct	
1	D(97) T(2) N(1)	D	T(77) D(20) N(3)	T	D(100)	D	L(64) V(35) I(1)	L	V(94) L(5) I(1)	V	V(98) L(2)	V	236
2	D(98) N,V(1)	D	T(85) D(7) N(6) A,E(1)	T	D(95) E(2) N(2) T(1)	D	I(98) L(2)	I	I(96) V(2) M,L(1)	I	I(92) L(4) V(3) F(1)	I	50
3	D(96) E(3) N(1)	D	Y(96) D(4)	Y	D(99) N(1)	D	L(99) V(1)	L	V(95) L(2) F,I,L(1)	V	V(89) L(11)	V	146
4	D(100)	D	V(99) D(1)	E/V	D(93) E(3) T(3) H(1)	D	L(97) I(3)	L	L(100)	L/V	V(76) L(13) I(8) M(2) F(1)	V	105
5	E(96) D(4)	E	E(78) D(11) T(7) K(2) R,N(1)	E	E(96) D(4)	E	L(100)	L	L(87) I(11) M,V(1)	L	L(100)	L	53
6	E(100)	E	E(100)	E	E(100)	E	L(85) M(9) V(6)	L	V(84) L(10) M(6)	V/M	M(99) V(1)	M	123
7	D(100)	D	E(94) D(6)	E	E(93) D(7)	E	L(98) M(2)	L	M(96) L,V(2)	M	M(96) L(2) I,V(1)	M	52
8	D(94) T(3) E(2) N(1)	D	A(96) T(2) V(2)	A	D(97) E(2) N(1)	D	L(89) I(7) V(3) F(1)	L	M(96) L,V(2)	M	M(84) L(12) I(4)	M	111
9	D(91) E(5) N(2) H,K(1)	D	E(92) D(7) N(1)	E	E(94) D(4) H,K(1)	E	L(93) F(3) V(3) M(1)	L	F(81) V(12) M,L(3) I(1)	F/V	F(93) L(5) M(2)	F	110
10	D(96) E(3) N(1)	D	H(53) D(44) E(3)	H	D(98) E(2)	D	L(95) M(5)	L	M(91) F(5) I(2) L,V(1)	M	M(83) V(11) L(4) F,I(1)	M/V	139
11	D(95) Y(3) E(2)	D	D(91) E(7) N,Y(1)	D	D(94) E(4) N,Y(1)	D	L(91) M(6) V(3)	L	V(75) L(12) M(11) F,I(1)	V	L(57) V(31) M(11) I(1)	L/V	126
12	D(100)	D	D(100)	D	D(100)	D	L(98) I(2)	L	L(99) V(1)	L	L(100)	L	112

Case No.	NS5A-Y93						Other Signature RAS						Treatment Failure to End of Follow-Up (week)
	Start of Treatment		Treatment Failure		End of Follow-Up		Start of Treatment		Treatment Failure		End of Follow-Up		
	Nanopore	Direct	Nanopore	Direct	Nanopore	Direct	Nanopore	Direct	Nanopore	Direct	Nanopore	Direct	
1	H(100)	H	H(100)	H	H(100)	H							236
2	Y(99) H(1)	Y	H(95) Y(5)	H	Y(98) H(2)	Y							50
3	Y(99) H(1)	Y	H(100)	H	H(89) Y(11)	H							146
4	Y(100)	Y	Y(100)	Y	Y(96) H(4)	Y	P32P(99) V(1)	P32P	P32L(93) P(7)	P32L/P	P32P(100)	P32P	105
5	Y(99) H(1)	Y	Y(91) H(9)	Y	Y(100)	Y	P29P(100)	P29P	P29del(100)	P29del	P29del(100)	P29del	53
6	Y(88) H(9) C(2) R(1)	Y	H(76) R(12) Y(8) C(4)	H	Y(59) H(34) R(7)	Y							123
7	Y(86) H(14)	Y	H(96) Y(4)	H	H(99) Y(1)	H							52
8	Y(93) H(7)	Y	H(97) Y(3)	H	H(86) Y(14)	H							111
9	Y(94) H(6)	Y	H(98) Y(2)	H	H(97) Y(3)	H							110
10	Y(86) H(14)	Y	H(98) Y(2)	H	H(97) Y(3)	H							139
11	Y(90) H(9) D(1)	Y	Y(80) H(20)	Y	H(63) Y(37)	H/Y							126
12	Y(100)	Y	Y(100)	Y	Y(100)	Y							112

Abbreviations: A, Alanine; C, Cysteine; D, Aspartate; E, Glutamate; F, Phenylalanine; G, Glycine; H, Histidine; I, Isoleucine; K, Lysine; L, Leucine; M, Methionine; N, Asparagine; P, Proline; Q, Glutamine; R, Arginine; S, Serine; T, Threonine; V, Valine; W, Tryptophan; Y, Tyrosine.

[a]Numbers in parentheses in the Nanopore column are the percentage of reads with corresponding amino acids by Nanopore sequencing.

[b]Two amino acids divided by the slash in the Direct column mean the mixture of amino acids; the first letter indicates the higher peak in the electropherogram.

[c]Time period for Nanopore sequencing.

TABLE 5 Comparison of Nanopore sequencing and target deep sequencing for signature RAS

Case No.	NS5A-L31				NS5A-Y93			
	Start of Treatment		Treatment Failure		Start of Treatment		Treatment Failure	
	Nanopore	Target Deep	Nanopore	Target Deep	Nanopore	Target Deep	Nanopore	Target Deep
2	I(98) L(2)	I(99.91)	I(96) V(2) M(1) L(1)	I(99.57) M(0.26) V(0.13)	Y(99) H(1)	Y(99.97)	H(95) Y(5)	H(99.57) R(0.32) Y(0.10)
10	L(95) M(5)	L(99.92)	M(91) F(5) I(2) L(1) V(1)	M(95.10) I(2.60) V(1.96) L(0.19)	Y(86) H(14)	Y(84.13) H(15.41) C(0.31)	H(98) Y(2)	H(99.58) R(0.28) Y (0.11)

Note: Numbers in parentheses are the percentage of reads with corresponding amino acids.

Abbreviations: C, Cysteine; F, Phenylalanine; H, Histidine; I, Isoleucine; L, Leucine; M, Methionine; R, Arginine; V, Valine; Y, Tyrosine.

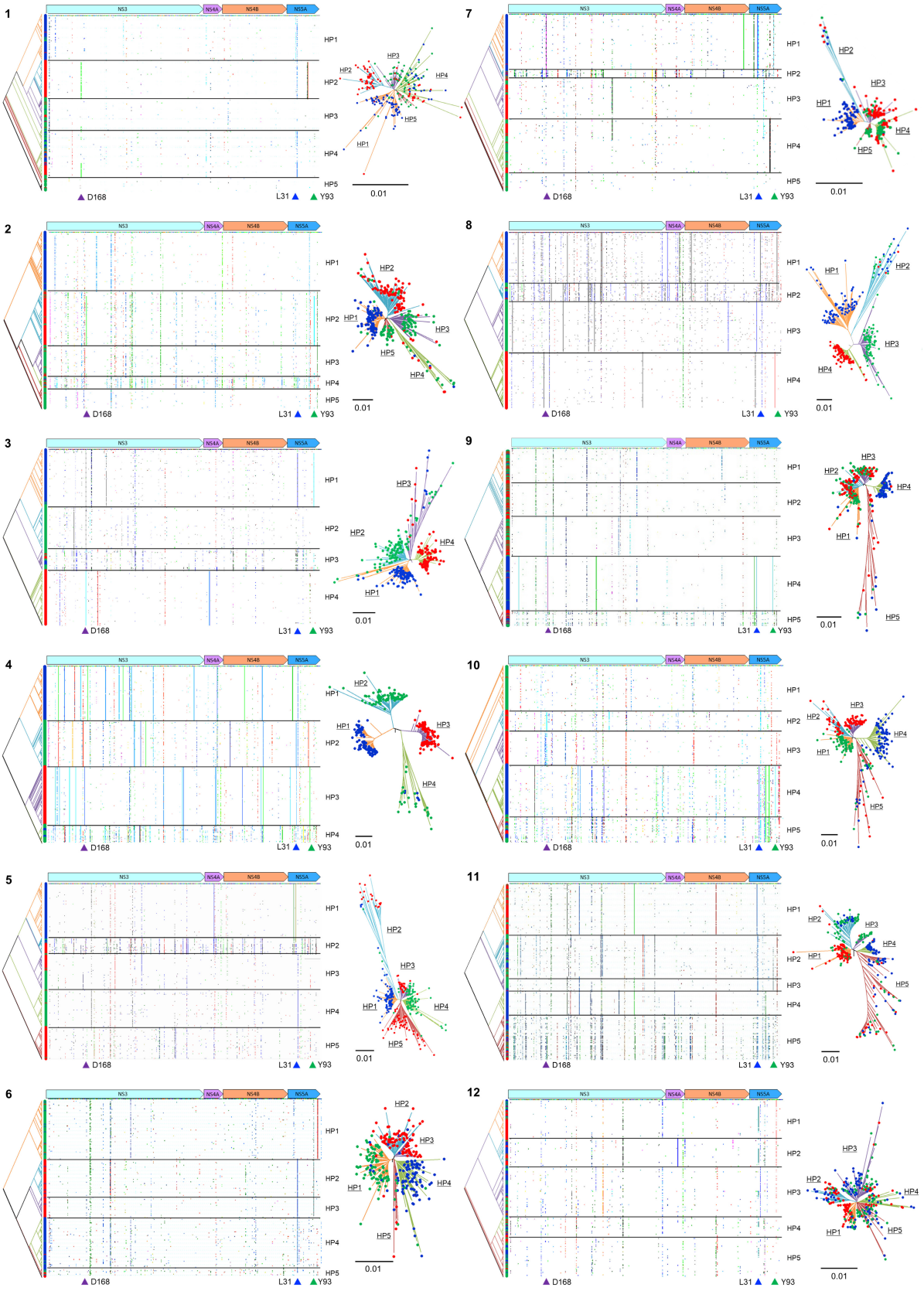
Analysis of HCV haplotype diversity during treatment

To elucidate the relations between HAS and RAS, serial changes of HAS during the DAA treatment were analyzed. The dynamic changes of HAS in each patient during the course with different amino acid substitutions among time points or haplotypes for each case are shown in Figure 3. Colored barcode patterns depicting the different amino acid residues and heights reflect the proportions of different residues at each position. These RASs and HASs extracted statistically by meta-CATS analysis with significant p values (5×10^{-5}) demonstrated right-most columns for each HAS or RAS. The composition of haplotypes changed dynamically during the course in each patient, as shown in the blue, red, and green bars. In addition, haplotype type-specific RASs, which are only found in specific haplotypes, were found in 10 of 12 patients. This analysis clearly showed the dynamic changes of haplotype composition that take place during DAA failure, which have been considered to be caused by a simple RAS-driven decrease of DAA susceptibility of the HCV mutant. Instead, DAA failure is related to the selections of HCV haplotypes with changes of haplotype-specific RAS. The normalized Shannon entropies of the 12 patients for each time point were (mean [SD]) 0.470 [0.213], 0.549 [0.260], and 0.701 [0.195] for start of treatment, treatment failure, and end of follow-up, respectively. These were significantly different among time points (p = 0.0338 with one-way repeated-measures analysis of variance; p = 0.0335 for start of treatment versus end of follow-up with Bonferroni correction).

DISCUSSION

In this study, we analyzed the HCV genomes related to DAA treatment failure by using Nanopore, a third-generation sequencer that enables long-read single-molecule sequencing of a target with more than 10 kb.[17] As a result, we clarified for the first time that DAA-resistant HCV consisted of resistant virus haplotypes, quasispecies with various HASs, and changed over time during the treatment with different RASs. This indicates that resistance to DAA,[6] growth fitness in the host,[26] and the ability to escape from the immune reaction are different depending on the haplotype[27] and not on the simple individual signature mutations of HCV genomes as previously thought. The HCV haplotype may be involved in a pathological process, such as fibrosis or carcinogenesis, caused by persistent infection of HCV. It is expected that the newly developed haplotyping by Nanopore sequencing will contribute to the elucidation of pathogenesis of viral diseases not limited to HCV.[16, 21, 28]

HCV is a single-stranded RNA virus; its RNA polymerase does not have proof-reading activity and forms



a quasispecies population with various mutations in the host.[10] Traditional Sanger direct sequencing of PCR products cannot reveal this quasispecies,[29] and sequencing of cloned PCR products can analyze only a small number of quasispecies clones spanning only a partial part of HCV genomes up to several hundred

FIGURE 2 Alignment and phylogenetic tree in a single-molecular HCV haplotype during DAA treatment. Blue-filled circles indicate the start of DAA therapy; red-filled circles indicate treatment failure after DAA therapy; green-filled circles indicate the end of follow-up. Phylogenetic trees were generated using the neighbor-joining method implemented in MEGA X, and the alignment was made by ete3. The alignment (right panel) and phylogenetic (left panel) analysis of 100 long reads at three time points (start of treatment, treatment failure, end of follow-up) from case 1 to case 12 are shown. Phylogenetic trees of each patient show that several haplotypes can be detected in each patient. In the alignment of individual HCV genomes from single patients, numerous substitutions composing quasispecies were found over the entire length of the PCR-amplified NS3 to NS5A region, and there are multiple HASs other than a signature RAS in each patient. In each haplotype, there were haplotype-specific RASs that changed during the treatment in a haplotype-specific manner. HP, haplotype

base pairs.[30] In recent years, massive parallel sequencing (100–300 bp), so-called short-read deep sequencing, has become possible by next-generation sequencing, and analysis of HCV quasispecies involved in DAA resistance has been performed for a small part of the HCV genome around signature RASs, showing that individual resistance mutations are associated with treatment failure[22,25,31] and the presence of mutations associated with liver carcinogenesis.[8] However, it has been difficult to analyze the entire structure of HCV with the combination of these individual characteristic mutations (haplotype) from a single viral genome in clinical samples with mixed quasispecies.[32]

The development of a sequencer that enables the long read of a single molecule, such as Nanopore[17] and PacBio,[33] is expected to facilitate analyzing how the haplotype of the above-mentioned viral genome affects the clinical picture. However, these third-generation sequencers have a high sequence error rate of around 10% as a trade-off of long reads.[34] In recent years, a technique has been developed that makes a target molecule circular to amplify and to read a single sequence multiple times as a concatemer to reveal the target sequence with error correction. PacBio with single-molecule real-time (SMRT) technology[35] and Nanopore with RCA technology[36] were developed to amplify and to sequence a concatemer of a target sequence of several tens of kilobases from a circular molecule. Applications of HCV sequencing of PacBio SMRT[37] and hepatitis B virus (HBV) sequencing of Nanopore RCA[38] have been reported.

In this study, Nanopore RCA was applied for the first time to the analysis of the DAA-resistant haplotype of HCV. When a PCR product derived from an HCV plasmid consisting of a single sequence was analyzed by this method, the accuracy was over 99.9% for a long read of an approximate 3000-bp single molecule, and it was shown to have sufficient specificity. On the other hand, when PCR products derived from HCV plasmids with different sequences were spiked, this analysis showed that mutations of at least 2% proportion could be detected. We also correlated the results of Nanopore and target deep sequencing of signature RASs in selected clinical cases and confirmed that proportions of major mutations were similar at each site (Table 5, case 2 and case 10), but those of minor mutations were not exactly the same. This might be caused by PCR selection bias for minor variants because the primer sets used for deep and Nanopore sequencing were different. In addition, if the number of Nanopore reads is further increased, it may be possible to detect variants with a lower frequency in the future.

In the present study, 100 reads were finally analyzed from the PCR-amplified HCV target sequence, and possible selection biases should be considered for evaluation of diversity. The first is PCR duplication bias due to the small number of templates. The number of HCV genomes subjected to PCR was equivalent to those in approximately 5.83 μL of serum, which corresponds to 1/170 of serum HCV titer (IU/mL). The HCV titer was 4.7 to 7.1 (median, 6.15) logIU/mL, which means approximately 50,000 to 12,500,000 (median, 1,400,000) copies of HCV genome per milliliter (Table 1). Therefore, 300 to 70,000 (median, 8000) copies of the HCV genome were subjected to PCR and amplified for downstream analysis. Therefore, analysis of 100 reads can reflect the diversity of the original sample without severe, if any, bias. However, if a larger number of reads are analyzed, it is important to secure enough HCV genome for PCR.

The second is PCR selection bias originating from unequal primer binding or amplification efficiency due to sequence variations among templates. FW3, FW4, RV1, and RV2 primers were effective for 76% (109/144) of PCR reactions, suggesting that binding sites of these primers are well preserved (Table 1). However, other samples still needed different primers, suggesting that PCR selection bias cannot be eliminated whenever PCR is used. Actually, in treatment failure of case 4, mixed RASs detected by direct sequence (D168E, L31V) were not apparent in Nanopore sequencing from the same sample (Table 4), possibly because primers used for direct and Nanopore sequencing were not identical. PCR-free, single-molecule, long-read Nanopore sequencing may be one of the solutions, and it has been successfully applied for HBV, which has a double-stranded circular DNA genome ready for RCA.[38] This technology is still challenging for HCV, which has a single-stranded linear RNA genome requiring additional RT and circularization steps, which result in much lower efficiencies. Further technological breakthrough is needed for complete removal of these PCR biases. We selected the longest

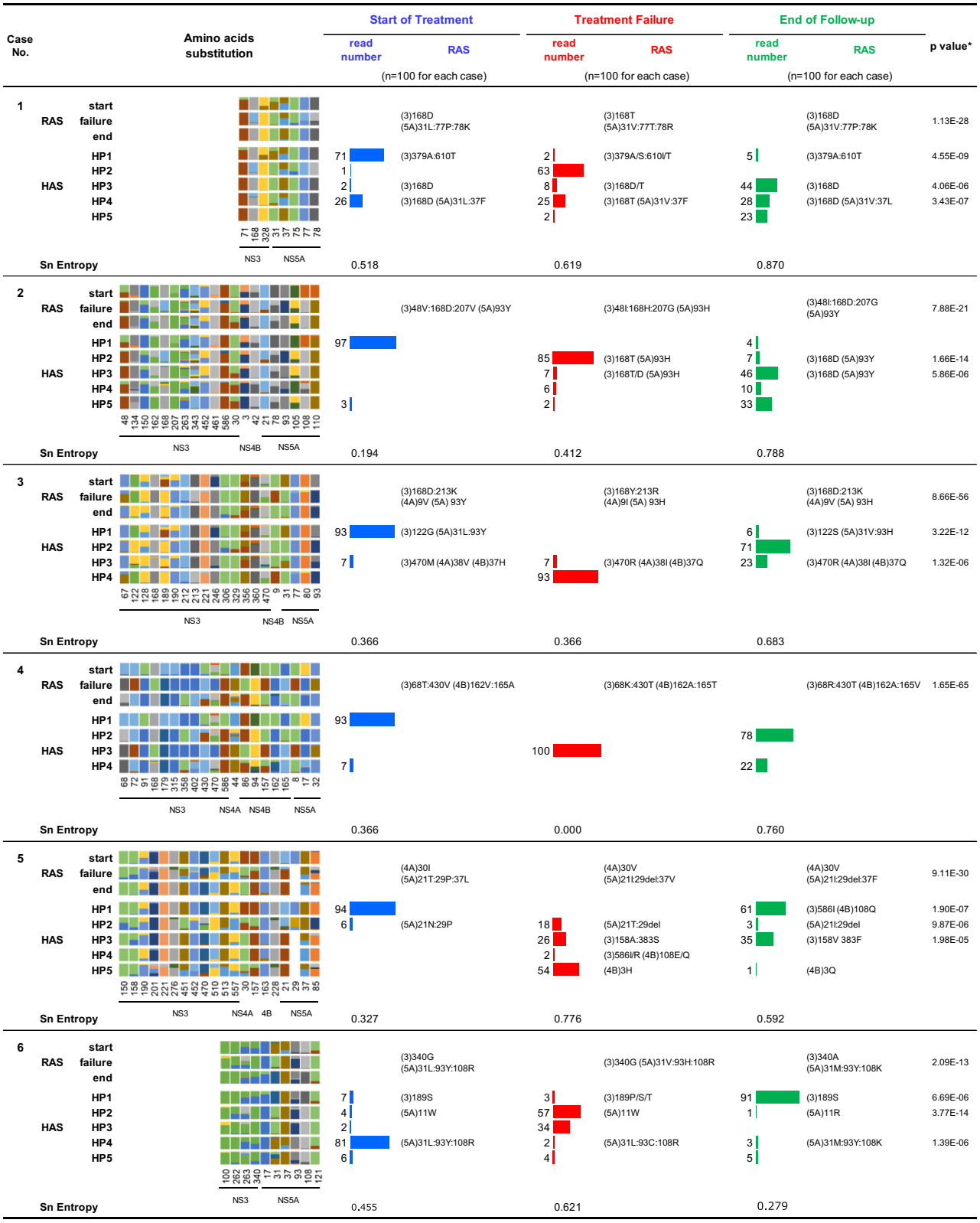


FIGURE 3 RAS and HAS in each patient and their changes during DAA treatment. *p value is the upper limit of each time point. (3), NS3; (4A), NS4A; (4B), NS4B; (5A), NS5A; end, end of follow-up; failure, treatment failure; Sn entropy, Shannon entropy; start, start of treatment. A, Alanine; C, Cysteine; D, Aspartate; E, Glutamate; F, Phenylalanine; G, Glycine; H, Histidine; I, Isoleucine; K, Lysine; L, Leucine; M, Methionine; N, Asparagine; P, Proline; Q, Glutamine; R, Arginine; S, Serine; T, Threonine; V, Valine; W, Tryptophan; Y, Tyrosine; X, any amino acid

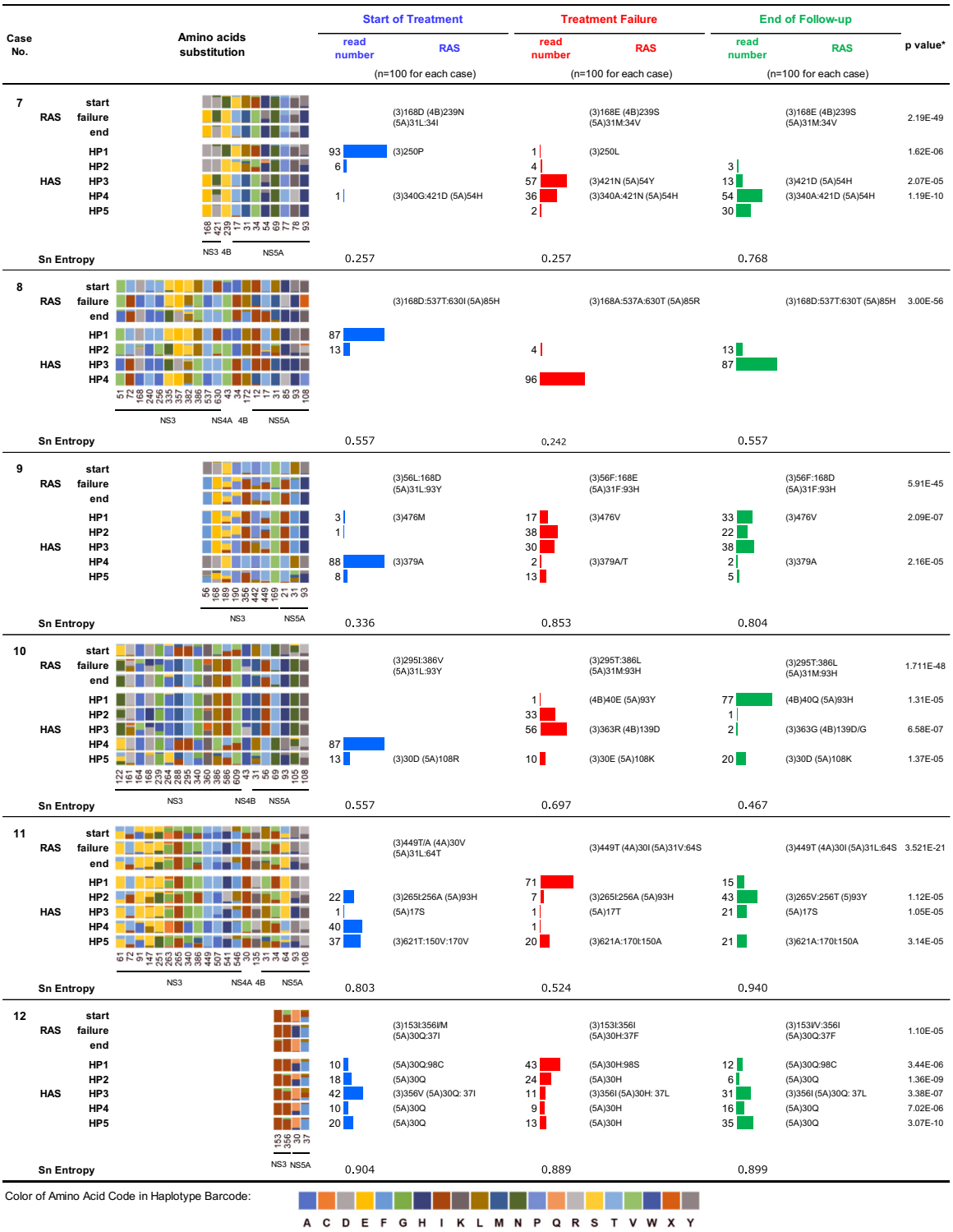


FIGURE 3 (Continued)

100 reads for phylogenetic analysis after RCA with the TruePrime Kit, which enables uniform amplification of mixed templates because it does not use random primers.[39] Therefore, the final 100 reads were randomly selected from PCR-amplified molecules.

Currently, the standard treatment for HCV is the combination of three classes of DAA, NS3 protease inhibitor, NS5A inhibitor, and NS5B polymerase inhibitor.[3] The NS3 protease inhibitor was the first class of DAAs developed, and its use started in 2010 in combination with interferon. NS5A inhibitor is the second class of DAAs and is a key drug contained in most regimens. Various amino acid substitutions, such as D168 mutation for NS3 resistance and L31/Y93 resistance

for NS5A resistance, were shown *in vitro* to alter DAA susceptibility of HCV and may be related to a therapeutic effect *in vivo*.[40] The emergence of HCV-resistant virus against multiple drugs became a serious problem after treatment failure of combination DAA treatments with multiple RASs related to different classes of DAAs. Proper and careful drug selection with genetic analyses is recommended for treatment of such multidrug-resistant HCV.[41]

However, in conventional techniques, it is difficult to confirm that these mutations for different drug resistance exist on a single genome.[11,12,40] This study showed that such multi-resistant haplotypes appeared after treatment failure of DAA treatment with different RASs in a single viral genome. From the phylogenetic tree analysis, we successfully visualized resistant haplotypes contributing to treatment failure because some quasispecies before treatment with multiple mutations were selected at treatment failure. Furthermore, we also clarified that resistant haplotypes are replaced with haplotypes having lower resistance and possibly higher proliferative fitness when the selective pressure disappears after drug administration. We did not exclude signature RAS for phylogenetic analysis to detect haplotype-specific RAS. With a combination of RASs in a single viral genome, the existence of the DAA-resistant haplotype was revealed for the first time by using Nanopore sequencing. Similar results have been recently shown with PacBio SMRT, and further accumulation of knowledge for the multidrug-resistant HCV haplotype is expected.[21,28]

In addition, the present study successfully estimated viral diversities with genetic distance of the phylogenetic tree and found their differences among patients. Some cases (case 1 and case 7) had lower diversity compared to the others, and its clinical significance should be further studied. Shannon entropies increased over time, suggesting divergence of haplotype complexity is related to treatment failure to DAA with the emergence of different haplotypes with different RASs. Haplotypes that became dominant later were not always detected at earlier time points; for example, in case 3, 4, and 8, major haplotypes at treatment failure were not detected at the start of treatment (Figure 3). These newly appearing haplotypes were not necessarily *de novo* but might be due to the small sampling size of 100 reads that can hardly detect haplotypes with proportions lower than 1%. Analysis with much larger number of reads will be needed to better track the minor haplotypes.

In this study, we performed a single-molecule sequence of the 3.2-kb region, which is about one third of the 9.6-kb genome of HCV. In the future, it will be necessary to elucidate the relationship between the haplotype of full-length HCV genome mutations and pathophysiology.[20] Because the bioinformatics method we used was designed to detect only small deletions up to 1% of the target sequence, we could not evaluate the large structural variants (deletion, insertion, and recombination) of the HCV genome at a single-molecule level. Structural variation of the HCV genome has been reported, such as large deletions related to more advanced liver disease, suggesting greater pathogenicity of the defective virus[42–46] or recombination,[28,47] but the details are still unclear. Recently, we applied the present method modified for the whole HCV genome mutation and deletion analysis and detected large deletions of the structural protein gene previously reported at the single-molecule levels (data not shown). It is expected that single-molecule long-read sequencing will facilitate understanding the entire HCV genome structure and its pathogenicity.

The Nanopore sequencer has characteristics that are different from conventional sequencers, such as small size, low cost of a USB connection, and real-time sequence information with few pretreatments, and it can be used not only in the laboratory but also in field research and clinical practice.[34,48] The problem of a high error rate has been a common property of the long-read sequencer, but this problem may be overcome by introducing a technique to sequence the target repeatedly as a concatemer, such as RCA applied in this study. In 2021, Oxford Nanopore Technologies announced the development of Q20+ Chemistry, which enables single-molecule long-read sequencing with an accuracy of 99% or higher (https://nanoporetech.com/q20plus-chemistry). The combination of this new technology with RCA will make it easier to perform single-molecule sequencing of a larger target, such as the whole HCV genome. In the future, it is expected that it will be applied not only to the analysis of hepatitis virus genomes but also to various pathogens, such as emerging infectious diseases like severe acute respiratory syndrome coronavirus 2.[49,50]

In the present study, we successfully clarified the kinetics of DAA-resistant haplotypes of HCV by combining Nanopore and RCA for single-molecule long-read sequencing. It is expected that this method can be applied not only to elucidate the relation between HCV genome structure and pathologies but also to analyze various viral diseases.

ACKNOWLEDGEMENT

We thank Dr. Koji Moriishi and Dr. Atsuya Yamashita for their scientific advice and Ms. Takako Ohmori and Ms. Tomoko Nakajima for their technical assistance.

CONFLICT OF INTEREST

Dr. Enomoto consults for, is on the speakers' bureau for, and received grants from Gilead, AbbVie, Merck Sharp and Dohme, and Bristol-Myers Squibb. The other authors have nothing to report.

AUTHOR CONTRIBUTIONS

Kozue Yamauchi performed most of the experiments. Mitsuaki Sato performed the populational and deep

sequencing. Leona Osawa, Shuya Matsuda, Yasuyuki Komiyama, Natsuko Nakakuki, Hitomi Takada, Ryo Katoh, Masaru Muraoka, Yuichiro Suzuki, Akihisa Tatsumi, Mika Miura, Shinichi Takano, Fumitake Amemiya, and Mitsuharu Fukasawa helped in clinical analyses. Yasuhiro Nakayama, Tatsuya Yamaguchi, and Taisuke Inoue assisted discussions. Shinya Maekawa and Nobuyuki Enomoto designed the study.

REFERENCES

1. Tanaka J, Akita T, Ko K, Miura Y, Satake M; Epidemiological Research Group on Viral Hepatitis and its Long -term Course, Ministry of Health, Labour and Welfare of Japan. Countermeasures against viral hepatitis B and C in Japan: an epidemiological point of view. Hepatol Res. 2019;49:990–1002.
2. Maekawa S, Enomoto N. The, "real-world" efficacy and safety of DAAs for the treatment of HCV patients throughout Japan. J Gastroenterol. 2018;53:1168–9.
3. Ghany MG, Morgan TR; AASLD-IDSA Hepatitis C Guidance Panel. Hepatitis C guidance update: American Association for the Study of Liver Diseases-Infectious Diseases Society of America Recommendations for Testing, Managing, and Treating Hepatitis C Virus Infection. Hepatology. 2020:686–721.
4. Mendizabal M, Alonso C, Silva MO. Overcoming barriers to hepatitis C elimination. Frontline Gastroenterol. 2019;10:207–9.
5. Heffernan A, Cooke GS, Nayagam S, Thursz M, Hallett TB. Scaling up prevention and treatment towards the elimination of hepatitis C: a global mathematical model. Lancet. 2019;393:1319–29.
6. Kai Y, Hikita H, Morishita N, Murai K, Nakabori T, Iio S, et al. Baseline quasispecies selection and novel mutations contribute to emerging resistance-associated substitutions in hepatitis C virus after direct-acting antiviral treatment. Sci Rep. 2017;7:41660.
7. Miura M, Maekawa S, Takano S, Komatsu N, Tatsumi A, Asakawa Y, et al. Deep-sequencing analysis of the association between the quasispecies nature of the hepatitis C virus core region and disease progression. J Virol. 2013;87:12541–51.
8. El-Shamy A, Pendleton M, Eng FJ, Doyle EH, Bashir A, Branch AD. Impact of HCV core gene quasispecies on hepatocellular carcinoma risk among HALT-C trial patients. Sci Rep. 2016;6:27025.
9. Tsukiyama-Kohara K, Kohara M. Hepatitis C virus: viral quasispecies and genotypes. Int J Mol Sci. 2017;19:23.
10. Bukh J. The history of hepatitis C virus (HCV): Basic research reveals unique features in phylogeny, evolution and the viral life cycle with new perspectives for epidemic control. J Hepatol. 2016;65(Suppl.1):S2–S21.
11. Dietz J, Susser S, Vermehren J, Peiffer K-H, Grammatikos G, Berger A, et al.; European HCV Resistance Study Group. Patterns of resistance-associated substitutions in patients with chronic HCV infection following treatment with direct-acting antivirals. Gastroenterology. 2018;154:976–88.
12. Li DK, Chung RT. Overview of direct-acting antiviral drugs and drug resistance of hepatitis C virus. Methods Mol Biol. 2019;1911:3–32.
13. Pethe MA, Rubenstein AB, Khare SD. Data-driven supervised learning of a viral protease specificity landscape from deep sequencing and molecular simulations. Proc Natl Acad Sci U S A. 2019;116:168–76.
14. Abdelrahman T, Hughes J, Main J, McLauchlan J, Thursz M, Thomson E. Next-generation sequencing sheds light on the natural history of hepatitis C infection in patients who fail treatment. Hepatology. 2015;61:88–97.
15. Guinoiseau T, Moreau A, Hohnadel G, Ngo-Giang-Huong N, Brulard C, Vourc'h P, et al. Deep sequencing is an appropriate tool for the selection of unique hepatitis C virus (HCV) variants after single genomic amplification. PLoS One. 2017;12:e0174852.
16. Takeda H, Yamashita T, Ueda Y, Sekine A. Exploring the hepatitis C virus genome using single molecule real-time sequencing. World J Gastroenterol. 2019;25:4661–72.
17. Uchida Y, Kouyama J-I, Naiki K, Uemura H, Tsuji S, Sugawara K, et al. A case of genotype-3b hepatitis C virus in which the whole genome was successfully analyzed using third-generation nanopore sequencing. Hepatol Res. 2019;49:1083–7.
18. Riaz N, Leung P, Barton K, Smith MA, Carswell S, Bull R, et al. Adaptation of Oxford nanopore technology for hepatitis C whole genome sequencing and identification of within-host viral variants. BMC Genom. 2021;22:148.
19. Wilson BD, Eisenstein M, Soh HT. High-fidelity nanopore sequencing of ultra-short DNA targets. Anal Chem. 2019;91:6783–9.
20. Bull RA, Eltahla AA, Rodrigo C, Koekkoek SM, Walker M, Pirozyan MR, et al. A method for near full-length amplification and sequencing for six hepatitis C virus genotypes. BMC Genom. 2016;17:247.
21. Takeda H, Ueda Y, Inuzuka T, Yamashita Y, Osaki Y, Nasu A, et al. Evolution of multi-drug resistant HCV clones from pre-existing resistant-associated variants during direct-acting antiviral therapy determined by third-generation sequencing. Sci Rep. 2017;7:45605.
22. Sato M, Maekawa S, Komatsu N, Tatsumi A, Miura M, Muraoka M, et al. Deep sequencing and phylogenetic analysis of variants resistant to interferon-based protease inhibitor therapy in chronic hepatitis induced by genotype 1b hepatitis C virus. J Virol. 2015;89:6105–16.
23. Yanagi M, St Claire M, Shapiro M, Emerson SU, Purcell RH, Bukh J. Transcripts of a chimeric cDNA clone of hepatitis C virus genotype 1b are infectious in vivo. Virology. 1998;244:161–72.
24. Ikeda M, Abe K, Dansako H, Nakamura T, Naka K, Kato N. Efficient replication of a full-length hepatitis C virus genome, strain O, in cell culture, and development of a luciferase reporter system. Biochem Biophys Res Commun. 2005;329:1350–9.
25. Miura M, Maekawa S, Sato M, Komatsu N, Tatsumi A, Takano S, et al. Deep sequencing analysis of variants resistant to the non-structural 5A inhibitor daclatasvir in patients with genotype 1b hepatitis C virus infection. Hepatol Res. 2014;44:E360–7.
26. Jensen SB, Fahnøe U, Pham LV, Serre SBN, Tang QI, Ghanem L, et al. Evolutionary pathways to persistence of highly fit and resistant hepatitis C virus protease inhibitor escape variants. Hepatology. 2019;70:771–87.
27. Di Lello FA, Culasso AC, Campos RH. Inter and intrapatient evolution of hepatitis C virus. Ann Hepatol. 2015;14:442–9.
28. Yamashita T, Takeda H, Takai A, Arasawa S, Nakamura F, Mashimo Y, et al. Single-molecular real-time deep sequencing reveals the dynamics of multi-drug resistant haplotypes and structural variations in the hepatitis C virus genome. Sci Rep. 2020;10:2651.
29. Maekawa S, Sakamoto M, Miura M, Kadokura M, Sueki R, Komase K, et al. Comprehensive analysis for viral elements and interleukin-28B polymorphisms in response to pegylated interferon plus ribavirin therapy in hepatitis C virus 1B infection. Hepatology. 2012;56:1611–21.
30. Montoya V, Olmstead A, Tang P, Cook D, Janjua N, Grebely J, et al. Deep sequencing increases hepatitis C virus phylogenetic cluster detection compared to Sanger sequencing. Infect Genet Evol. 2016;43:329–37.
31. Chen J, Zhao Y, Sun Y. De novo haplotype reconstruction in viral quasispecies using paired-end read guided path finding. Bioinformatics. 2018;34:2927–35.

32. Singer J, Thomson E, Hughes J, Aranday-Cortes E, McLauchlan J, da Silva Filipe A, et al. Interpreting viral deep sequencing data with GLUE. Viruses. 2019;11:323.
33. Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics. 2015;13:278–89.
34. Boldogkői Z, Moldován N, Balázs Z, Snyder M, Tombácz D. Long-read sequencing – a powerful tool in viral transcriptome research. Trends Microbiol. 2019;27:578–92.
35. Nakano K, Shiroma A, Shimoji M, Tamotsu H, Ashimine N, Ohki S, et al. Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. Hum Cell. 2017;30:149–61.
36. Li C, Chng KR, Boey EJH, Ng AHQ, Wilm A, Nagarajan N. INC-Seq: accurate single molecule reads using nanopore sequencing. GigaScience. 2016;5:34.
37. Bergfors A, Leenheer D, Bergqvist A, Ameur A, Lennerstrand J. Analysis of hepatitis C NS5A resistance associated polymorphisms using ultra deep single molecule real time (SMRT) sequencing. Antiviral Res. 2016;126:81–9.
38. McNaughton AL, Roberts HE, Bonsall D, de Cesare M, Mokaya J, Lumley SF, et al. Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). Sci Rep. 2019;9:7081.
39. Picher ÁJ, Budeus B, Wafzig O, Krüger C, García-Gómez S, Martínez-Jiménez MI, et al. TruePrime is a novel method for whole-genome amplification from single cells based on TthPrimPol. Nat Commun. 2016;7:13296.
40. Baumert TF, Berg T, Lim JK, Nelson DR. Status of direct-acting antiviral therapy for hepatitis C virus infection and remaining challenges. Gastroenterology. 2019;156:431–45.
41. European Association for the Study of the Liver. EASL recommendations on treatment of hepatitis C 2018. J Hepatol. 2018;69:461–511.
42. Cheroni C, Donnici L, Aghemo A, Balistreri F, Bianco A, Zanoni V, et al. Hepatitis C virus deletion mutants are found in individuals chronically infected with genotype 1 hepatitis C virus in association with age, high viral load and liver inflammatory activity. PLoS One. 2015;10:e0138546.
43. Iwai A, Marusawa H, Takada Y, Egawa H, Ikeda K, Nabeshima M, et al. Identification of novel defective HCV clones in liver transplant recipients with recurrent HCV infection. J Viral Hepat. 2006;13:523–31.
44. Ohtsuru S, Ueda Y, Marusawa H, Inuzuka T, Nishijima N, Nasu A, et al. Dynamics of defective hepatitis C virus clones in reinfected liver grafts in liver transplant recipients: ultradeep sequencing analysis. J Clin Microbiol. 2013;51:3645–52.
45. Karamichali E, Chihab H, Kakkanas A, Marchio A, Karamitros T, Pogka V, et al. HCV defective genomes promote persistent infection by modulating the viral life cycle. Front Microbiol. 2018;9:2942.
46. Li Q, Tong Y, Xu Y, Niu J, Zhong J. Genetic analysis of serum-derived defective hepatitis C virus genomes revealed novel viral cis elements for virus replication and assembly. J Virol. 2018;92:e02182-17.
47. Galli A, Bukh J. Comparative analysis of the molecular mechanisms of recombination in hepatitis C virus. Trends Microbiol. 2014;22:354–64.
48. Petersen LM, Martin IW, Moschetti WE, Kershaw CM, Tsongalis GJ. Third-generation sequencing in the clinical laboratory: exploring the advantages and challenges of nanopore sequencing. J Clin Microbiol. 2019;58:e01315-19.
49. Bull RA, Adikari TN, Ferguson JM, Hammond JM, Stevanovski I, Beukers AG, et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. Nat Commun. 2020;11:6272.
50. Cole C, Byrne A, Adams M, Volden R, Vollmers C. Complete characterization of the human immune cell transcriptome using accurate full-length cDNA sequencing. Genome Res. 2020;30:589–601.

How to cite this article: Yamauchi K, Sato M, Osawa L, Matsuda S, Komiyama Y, Nakakuki N, et al. Analysis of direct-acting antiviral-resistant hepatitis C virus haplotype diversity by single-molecule and long-read sequencing. Hepatol Commun. 2022;6:1634–1651. https://doi.org/10.1002/hep4.1929