PO Box 2345, Beijing 100023, China
www.wjgnet.com
wjg@wjgnet.com

ELSEVIER

World J Gastroenterol 2005;11(19):2874-2884
World Journal of Gastroenterology ISSN 1007-9327

Multigene tracking of quasispecies in viral persistence and clearance of hepatitis C virus

Song Chen, Yu-Ming Wang

Song Chen, Yu-Ming Wang, Department of Infectious Diseases, Southwest Hospital, Third Military Medical University, Chongqing 400038, China
Supported by the National Natural Science Foundation of China, No. 39870694
Correspondence to: Song Chen, MD, Department of Infectious Diseases, Southwest Hospital, Third Military Medical University, 30 Gaotanyan Zhengjie, Shapingba District, Chongqing 400038, China. cs196@medmail.com.cn
Telephone: +86-23-68753196
Received: 2004-04-06 Accepted: 2004-05-24

Abstract

AIM: To investigate the evaluation of hepatitis C virus (HCV) quasispecies in the envelope region and its relationship with the outcome of acute hepatitis C.

METHODS: HCV quasispecies were characterized in specimens collected every 2-6 mo from a cohort of acutely HCV-infected subjects. We evaluated two individuals who spontaneously cleared viremia and three individuals with persistent viremia by cloning 33 1-kb amplicons that spanned E1 and the 5′ half of E2, including hypervariable region 1 (HVR1). To assess the quasispecies complexity and to detect variants for sequencing, 33 cloned cDNAs representing each specimen were assessed by a combined method of analysis of a single-stranded conformational polymorphism and heteroduplex analysis. The rates of both synonymous and nonsynonymous substitutions for the E1, HVR1 and E2 regions outside HVR1 were analyzed.

RESULTS: Serum samples collected from chronic phase of infection had higher quasispecies complexity than those collected from acute phase of infection in all individuals examined. The genetic diversity (genetic distance) within HVR1 was consistently higher than that in the complete E1 (0.0322±0.0068 *vs* -0.0020±0.0014, $P<0.05$) and E2 regions outside HVR1 (0.0322±0.0068 *vs* 0.0017±0.0011, $P<0.05$) in individuals with persistent viremia, but did not change markedly over time in those with clearance of viremia. For individuals with persistent viremia, the rate of nonsynonymous substitutions within the HVR1 region ($2.76\times10^{-3}\pm1.51\times10^{-3}$) predominated and gradually increased, as compared with that in the E1 and E2 regions outside HVR1 ($0.23\times10^{-3}\pm0.15\times10^{-3}$, $0.50\times10^{-3}\pm0.10\times10^{-3}$). By contrast, the rates of both nonsynonymous and synonymous substitutions for the E1 and E2 regions including HVR1 were consistently lower in individuals with clearance of viremia.

CONCLUSION: HCV persistence is associated with a complexity quasispecies and positive selection of HVR1 by the host immune system.

Key words: Hepatitis C virus; Immune system

Chen S, Wang YM. Multigene tracking of quasispecies in viral persistence and clearance of hepatitis C virus. *World J Gastroenterol* 2005; 11(19): 2874-2884
http://www.wjgnet.com/1007-9327/11/2874.asp

INTRODUCTION

Hepatitis C virus (HCV), a member of the *Flaviviridae* family, is the major cause of chronic liver disease worldwide[1]. HCV is a positive-sense single-strand RNA virus with a genome that encodes one large polyprotein in which putative structural proteins are located at the N-terminal end, and the putative nonstructural proteins are located at the C-terminal end[2]. One of the important characteristics of HCV is that its genome exhibits significant genetic heterogeneity as a result of the accumulation of mutations during viral replication. The genetic sequences of HCV variants are very heterogeneous, varying by more than 30% across the entire genome among the six major genotypes, 20% among subtypes, and up to 10% within a subtype[3]. Analogous to other RNA viruses, HCV circulates in an infected individual as a population of close-related, yet heterogeneous, sequences: the quasispecies[4-6]. The quasispecies distribution of HCV might have important biological consequences. It has been proposed that this genetic heterogeneity allows HCV to escape immune pressure and to establish chronic infection[7-9]. Furthermore, the existence of a heterogeneous population of HCV may influence the outcome of antiviral therapy; and resistance to treatment might result from selection of minor viral populations during this therapy[10]. Therefore, it is important to define accurately quasispecies populations of HCV.

Many analyses of viral quasispecies of HCV have been published. The majority of these studies have focused on the most variable part of the HCV genome, hypervariable region 1 (HVR1) of glycoprotein E2[11-13]. Mutation of this region of the genome is believed to be associated with viral persistence via immune escape mechanisms[14,15]. It is well known that genetic heterogeneity of HCV extends throughout the entire genome. However, it is still not known whether significant mutation occurred in other regions of

the HCV envelope genes during the chronic infection. In addition, most studies assessing the diversity of HCV quasispecies have been conducted by amplifying selected portions of the genome by PCR, isolating individual subgenomic fragments by a cloning product, and then characterizing the nucleotide sequence of each clone[16-18]. Evaluating the diversity of HCV quasispecies in clinical samples often requires the sequencing of a large number of clones, but because of the effort and expense, published studies have obtained sequence information from a small number of colonies per subject.

Two recent developments enabled us to investigate genetic variation of the HCV envelope genes and its relationship with the outcome of acute hepatitis C. First, we identified and characterized the long-term virologic outcomes for five individuals with acute HCV infection. Second, we developed a method for efficiently and accurately characterizing the HCV quasispecies[19-21]. In this study, these resources were used to examine viral complexity and distortion in amino acid sequences of subjects with persistent viremia *vs* those with clearance of viremia.

MATERIALS AND METHODS

Patients and samples

From November 1998 to January 2002, 284 current injection drug users (IDUs) totaling 125 HCV-infected individuals have been monitored in Chongqing. Five individuals were identified as HCV seroconverters when a sample tested positive for antibody to HCV following at least one negative result. After more than 3 years of semiannual follow-up subsequent to seroconversion, two distinct patterns of viremia were noted. For two subjects, HCV RNA was undetectable for a minimum of 2 years in at least two serum samples from each person. In contrast, for three subjects, HCV RNA remained detectable in the last sample tested. Clinical and virological backgrounds of the subjects studied are summarized in Table 1.

Detection of serum virological parameters

These samples were tested for antibodies to HCV (HCV EIA 2.0; Ortho Diagnostics Raritan, NJ) and, if these results were positive, by a strip immunoblot assay (RIBA HCV 2.0; Chiron Corporation, Emeryville, CA). HCV RNA was detected by a quantitative reverse transcriptase PCR (RT-PCR) assay (AMPLICOR HCV MONITO, Roche Diagnostic Systems, Branchburg, NJ), the linear range of which was determined to be 500-500 000 copies/mL of serum by our and other laboratories[22,23]. Liver tests, including alanine aminotransferase (ALT) levels in serum, were performed at the first clinical examination and repeated during follow-up. Hepatitis B surface antigen, anti-HBc, anti-HBe and anti-HIV IgM were negative in all subjects detected by ELISA. HCV subtype was determined by the RT-PCR-restriction fragment length polymorphism analysis targeted to the 5' non-coding region of the HCV[24].

Envelope region amplification

HCV RNA characterization was based on examination of 33 cloned cDNAs spanning the 1 025-nucleotide (nt) region thought to encode envelope protein E1 and a segment of E2, including HVR1. Total RNA was extracted from 100 μL serum using 500 μL of TRIzol LS Reagent (Life Technologies, Gaithersburg, MD) at room temperature, followed by chloroform extraction and isopropanol precipitation in the presence of 20 μg of glycogen (Boehringer Mannheim, Indianapolis, IN). The RNA pellet was washed with 75 mL/L ethanol and then air dried briefly and redissolved in 50 μL of diethyl pyrocarbonate-treated water with 10 mmol/L dithiothreitol (Promega, Madison, WI) and 5 U of RNasin ribonuclease inhibitor (Promega). After incubation at 65 ℃ for 5 min, 5 μL purified RNA was used to generate cDNA in a 20-μL reaction mixture at 37 ℃ for 1 h with 20 U of Moloney murine leukemia virus reverse transcriptase (Promega) and first-round PCR reverse primer.

The entire 20-μL cDNA synthesis reaction mixture was

Table 1 Molecular, biochemical, and serological characterization of five HCV primary infections

Subjects	Samples	Age (yr) /sex	Duration of infection (mo)	Genotype	Log_{10} [HCVRNA][1]	HCV RNA	ALT level (nkat/L)[2]	Result of ELISA
A	A1	30/M	0	3b	7.30	++++	1 833.70	-
	A2		6		6.20	+++	583.45	+
	A3		14		5.30	+++	616.79	+
B	B1	28/F	0	3b	7.50	++++	1 933.72	-
	B2		6		6.40	+++	1 150.23	+
	B3		12		6.90	+++	1 533.64	+
C	C1	33/M	0	1b	6.41	+++	683.47	+
	C2		24		5.25	+++	383.41	+
	C3		34		4.16	++	166.70	+
E	E1	20/M	0	1a	6.80	+++	4 017.47	-
	E2		3.5		4.20	++	766.82	+
	E3		9		0	–	583.45	+
	E4		24		0	–	400.08	-
F	F1	28/M	0	1b	6.00	++++	3 067.28	-
	F2		2		5.60	+++	1 066.88	+
	F3		7		0	–	666.80	+
	F4		18		0	–	483.43	-

[1]Number of HCV RNA molecules per milliliter plasma. Time zero is the time where the first sample was available; others indicate are times after time zero. [2]Normal value, <666.80 nkat/L.

used for the first-round PCR in a 25-μL reaction mixture containing 0.75 U Expand HF polymerase mixture (Boehringer Mannheim), 1.5 mmol/L $MgCl_2$, 0.2 mmol/L concentration of deoxynucleoside triphosphates, and 500 μmol/L concentrations of primers. The mixed oligonucleotides primers were used for RT-PCR (Table 1). Degenerate bases are indicated with standard codes of the International Union of Pure and Applied Chemistry. Nucleotide positions were numbered according to the HCV-J6 sequence. One microliter of the first-round reaction mixture was added to the second-round PCR, which had the same reagents as in the first round except for primers. Thermal-cycling conditions for the inner and outer reactions were pre-denaturation for 120 s at 94 ℃, followed by 35 amplification cycles of 45 s at 94 ℃, 45 s at 60 ℃, and 120 s at 72 ℃ (during the last 25 cycles, the elongation time was increased by 20 s per cycle).

Cloning of cDNA and complexity analysis of 33 cloned cDNAs by gel shift

The 1-kb HCV cDNA product was ligated into vector pT-adv and used to transform *Escherichia coli* TOP 10F' competent cells (TA Cloning kit; CLONTECH Laboratories, Inc.). Transformants were detected according to the manufacturer's protocol, and cloning efficiency was >90%.

Then 80 g/L polyacrylamide gel electrophoresis was carried out with the addition of 150 g/L urea to increase the resolution. For each subject, the gel shift patterns of 33 cloned cDNAs were examined by amplifying a 470-bp sequence spanning E1 gene and 570-bp sequence including HVR1, responsively, and by a nonradioactive method that detected distinct variants within a sample by using a combination of heteroduplex analysis (HDA) and single-stranded conformational polymorphism (SSCP) on a single gel (SSCP+HDA)[19]. Sequences obtained from the serial passage were analyzed by a divergent variant from the acute-phase sample from each subject. A clonotype is defined as two or more cloned cDNAs that have indistinguishable patterns of electrophoretic migration by SSCP+HDA. The complexity of the quasispecies was characterized by the clonotype ratio, calculated as the number of clonotypes divided by 33, the number of cloned cDNAs examined[20]. The clonotype ratio therefore varied from 0.03 (homogenous) to 1.

Nucleotide sequencing

To examine each subject's quasispecies for signature sequences (motifs uniquely shared by a group of sequences) and for evaluations in the sensitivity of the SSCP+HDA method, a subset of cloned cDNAs was identified. For each subject, at least two cloned cDNAs were selected for sequencing based on gel shift patterns: one from the majority clonotype, another from each clonotype consisting of the cloned cDNAs with the largest heteroduplex gel shift. Sequences were positively determined from the M13 reverse primer and negatively from T7 promoter binding sites of plasmid clones by using a PRISM 377 DNA Sequencer (version 3.3; Applied Biosystems, Inc., Foster City, CA). Sequences were assembled by using the ESEE3s program, and primer sequences were removed prior to analysis.

Phylogenetic analysis

DNA distance matrices were calculated by using the DNADIST program, maximum-likelihood, with a transition-to-transversion ratio of 4.25[25], and phylogenetic trees were generated by the Neighbor-joining program with random addition. Subtype reference sequences used for phylogenetic analysis had the following GenBank accession numbers: 1a, M62321; 1b, D10934; 2a, D00944; 2b, D10988; 3a, D17763; 4a, Y11604; 5a, Y13184; 6a, Y12083.

Nonsynonymous substitutions per potential nonsynonymous site (dN) and synonymous substitutions per potential synonymous site (dS) were calculated by the method of Nei and Gojobori[26].

Statistical analysis

Quantitative values were compared using the Student's *t* test, the Kruskal-Wallis test or the analysis of the variance when necessary. *P* values lower than 0.05 were considered statistically significant. All statistical calculations were performed by using the SPSS for Windows, version 8.0 software package.

RESULTS

Clonotypes detected by SSCP+HDA method

The SSCP+HDA used in this investigation showed that each specimen contained a swarm of distinct but related variants represented by clonotypes. The number of clonotypes within the E1 region varied from 2 to 6 per sample, and within the E2/HVR1 region changed from 3 to 21 per sample. A subset of clonotypes always persisted in E1 during serial passages from each subject. In E2/HVR1 derived from five individuals, the subjects with self-limited viremia had the persistence of clonotypes during serial passages, whereas acute and chronic samples from persistent-infected cases shared no clonotype (Figure 1).

The quasispecies complexity was examined by assessing 33 cDNA clones from each specimen using the clonotype ratio of E2/HVR1. Serum samples collected from chronic phase of infection had higher quasispecies complexity than those collected from acute phase of infection (Figure 2), but no trends were observed as clonotype ratio values changed with the changes in circulating viral load (Figure 3).

Representative sequence analysis

Using SSCP+HDA to select representative cloned cDNAs, we identified 26 distinct cloned cDNAs for sequencing (Figure 4). The sequences of single variants from two subjects (A and B) were 1 022 bp and the sequences from two other subjects (C and F) were 1 021 bp, which had a 1-bp deletion, and those of subject E was 1 019 bp, which had a 3-bp deletion. To determine the genetic identity of cDNA clones of the same clonotype, two representative sequences representing the majority clonotype were compared for each specimen. No two-cloned cDNAs identified as being distinct by SSCP+HDA analysis had identical sequences, underscoring that the SSCP+HDA method was both highly sensitive and specific in detecting differences among cDNA clones, as previously reported. For each majority clonotype, which of the two sequences was free

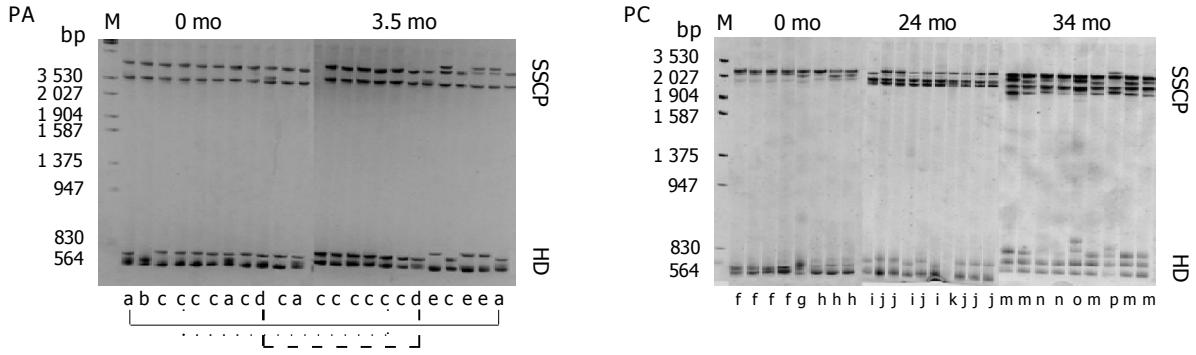


Figure 1 Analysis by SSCP+HDA of the E2 region (including the HVR1) of HCV in two HCV-positive IDUs. Clonotypes (groups of electrophoretically indistinguishable cloned cDNAs) were assigned sequential letter designations (a-p). As can be seen, the number of clonotypes in patient with clearance of viremia (PA) E did not differ significantly after antibody seroconversion compared with the pre-antibody seroconversion sample, and the clonotypes composition remained largely unchanged during the follow-up. In contrast, the number of clonotypes increased in patient with persistent viremia (PC) C, and the clonotypes composition showed constant evolution during the follow-up.

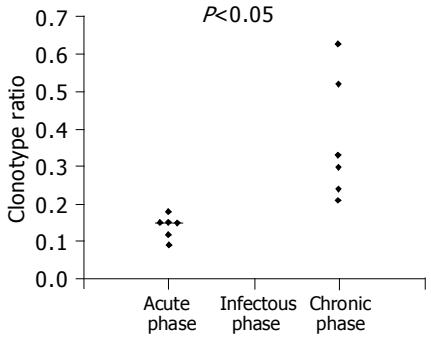


Figure 2 Clonotype ratio and outcome. Clonotype ratio values obtained for the samples from five individuals at each time point were calculated as described in Materials and methods.

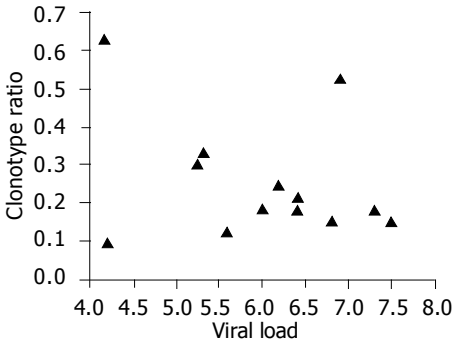


Figure 3 Correlation between HCV clonotype ratio and serum HCV RNA load. No correlation was found between them (r = 0.2).

of sporadic substitutions, one was used in all subsequent analyses to represent that clonotype (Table 2).

Figure 5 displays a dendrogram illustrating E2 sequence diversity in individuals from the two study groups. Sequences from each individual tended to cluster tightly, segregated away from clusters of sequences from other individuals in the same genotype.

Genetic evolution of viral quasispecies and outcome of acute infection

The genetic distances (genetic diversity) of the HCV quasispecies were assessed by examining viral sequences spanning the envelope genes both within and outside the HVR1 for intrahost evolutionary analysis (Table 3). Results of these analyses revealed a difference in virus evolution according to the outcome of the disease. During serial passages, individuals who cleared viremia did not change significantly in the genetic diversity, whereas those with persistent viremia showed a marked increase in diversity. When the genetic diversity both in the complete E1 and the E2 regions outside HVR1, based on the analysis of 102 predicted amino acids, was analyzed, the viral diversity was

Table 2 Primers for amplification of E1 and E2 regions of HCV genome[1]

Region	Designation	Sequence	5′ to 3′ polarity	Position of 5′ base[2]
C/E2	P1	GCAACAGGGAAYYTDCCCGGTTGCTC	Outer sense	837–862
	P2	TTCATCCASGTRCAVCCRAACCA	Outer antisense	2 020–1 998
	P3	CCATGTGCTCYTTYTCTATCTTC	Inner sense	855–874
	P4	GTTTAAARCARTACACYGGRCCRCANAC	Inner antisense	1 882–1 857
C/E1	P3	CCATGTGCTCYTTYTCTATCTTC	Sense	855–874
	P5	GTRGGBGACCAGTTCATCATCAT	Antisense	1 333–1 311
E2	P6	GGGAYATGATGATGAACTGGTC	Sense	1 306–1 327
	P4	GTTTAAARCARTACACYGGRCCRCANAC	Antisense	1 882–1 857

[1]Degenerate bases are indicated with standard codes of the international union of Pure and Applied Chemistry. [2]Position of 5′ base relative to the HCV genomic sequence of the HCV J6 strain.

Table 3 Comparison of changes in genetic diversity of viral strains in individuals with self -limited and persistent viremia[1]

Patient group	Number of patients	Time points[2]	Interval (wk)	Change in genetic diversity (×10^{-2})		
				E1	HVR1	E2
Clearance	2	B *vs* A	3.75±0.35	0.08±0.20	0.30±0.71	–0.37±0.04
Persistence	3	B *vs* A	12.67±9.86	–0.20±0.14	3.22±0.68[a]	0.17±0.11
		C *vs* A	20.0±12.16	0.06±0.23	4.10±2.35	0.21±0.39
		B *vs* C	7.33±3.03	–0.26±0.11	–0.89±1.81	0.15±0.71

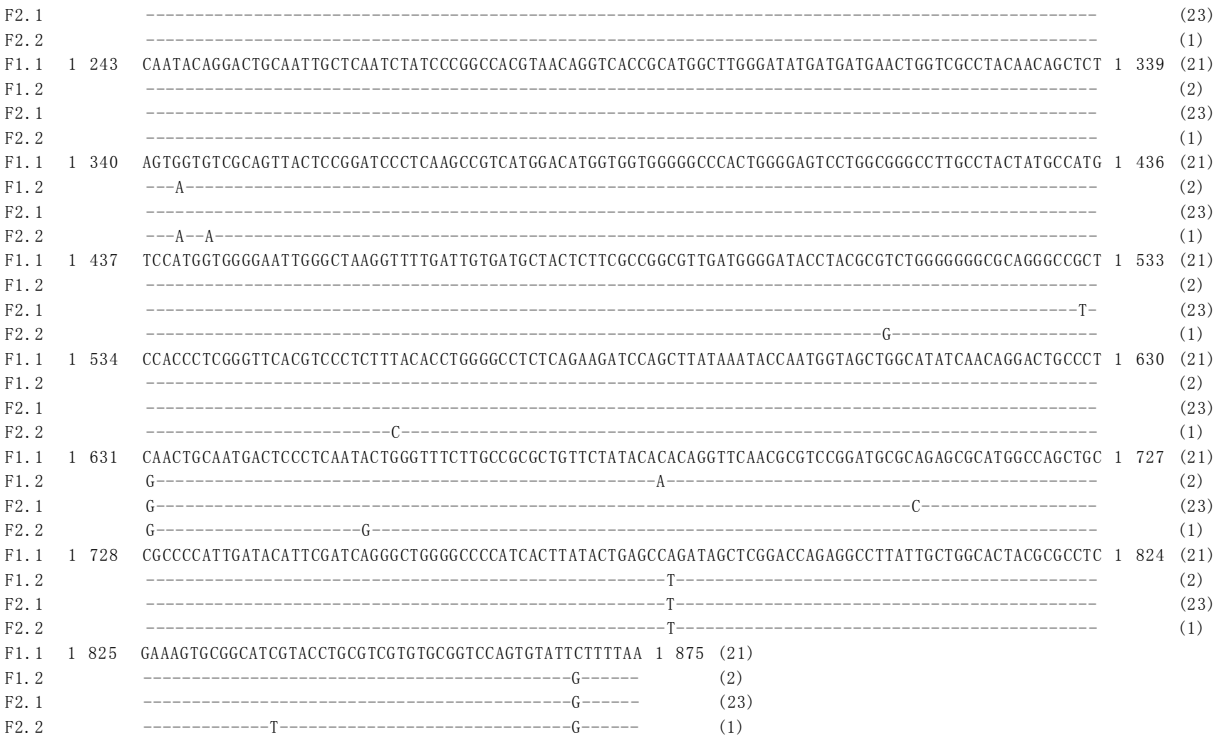
[1]Genetic diversity was assessed by using the DNADIST program, maximum-likelihood. The data are represented as mean±SE. Negative values indicate a reduction in genetic diversity of viral strains; positive values indicate an increase in genetic diversity of viral strains. [2]The first time point (A) corresponds to the baseline, the second (B) and the third (C) after antibody seroconversion. [a]P<0.05 time points B *vs* A of self-limited *vs* persistent viremia by unpaired Welch's t test.

Nucleotide sequences of envelope region at 3.5-mo serial samples from the subject E

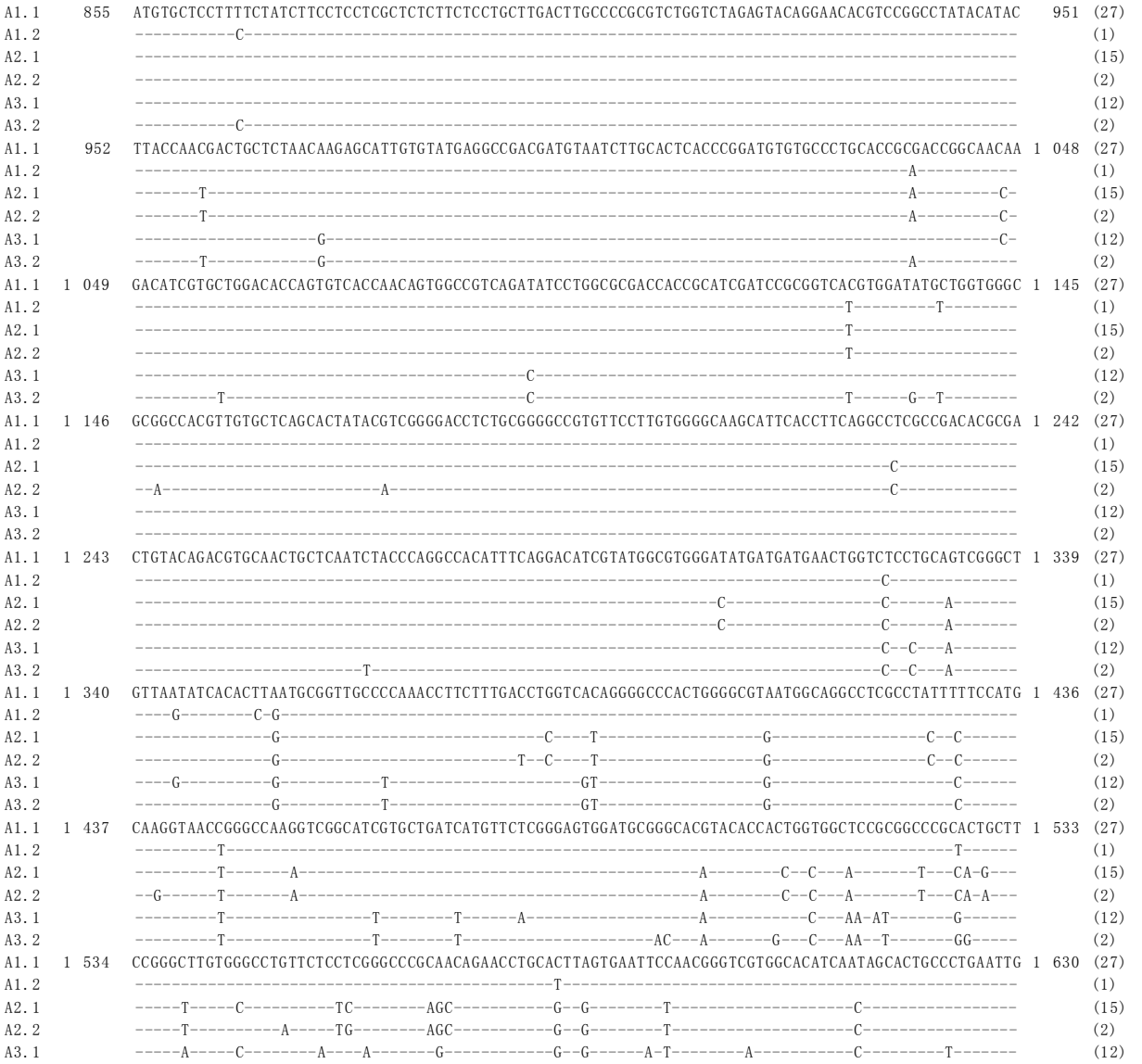
```
E1.1     855  ATGTGCTCTTTCTCTATCTTCCTCCTAGCCCTGCTTTCCTGCCTGACTGTGCCCGCTTCAGCCTACCAAGTGCGCAACTCCACGGGGCTTTATCATG 951    (26)
E1.2          --------C-------------------------------------------------------------------------------------          (4)
E2.1          --------C-------------------------------------------------------------------------------------          (29)
E2.2          --------C--T----------------------------------------------------------------------------------          (3)
E1.1     952  TCACCAATGACTGCCCTAACTCGAGCATTGTGTACGAGGCAGCTGATGCCATCCTGCACACTCCGGGGTGCGTCCCTTGCGTTCGCGAGGGTAACAC 1 048 (26)
E1.2          ---------------------------------------------------------------------------------------------          (4)
E2.1          ---------------------------------------------------------------------------------------------          (29)
E2.2          ---------------------------------------------------------------------------------------------          (3)
E1.1  1 049   CTCGAGGTGTTGGGTGGCGGTGACCCCCACGGTGGCCACCAGGGATGGCAAACTCCCCACAACGCAGCTTCGACGTCACATCGATCTGCTTGTCGGG 1 145 (26)
E1.2          ---------------------------------------------------------------------------------------------          (4)
E2.1          ---------------------------------------------------------------------------------------------          (29)
E2.2          ---------------------------------------------------------------------------------------------          (3)
E1.1  1 146   AGTGCCACCTTCTGCTCGGCTCTTTACGTGGGGGACTTGTGCGGGTCCGTCTTTCTTGTCGGTCAGCTGTTCACCTTCTCTCCCAGGCGCCACTGGA 1 242 (26)
E1.2          ---------------------------------------------------------------------------------------------          (4)
E2.1          ---------------------------------------------------------------------------------------------          (29)
E2.2          ------------------------------------------------------------------------------------------A----          (3)
E1.1  1 243   CGACGCAAGACTGCAATTGTTCTATGTATCCCGGCCATATAACGGGTCATCGCATGGCATGGGATGTGATGATGAACTGGTCCCCTACGACGGCATT 1 339 (26)
E1.2          ------------------------------------------------------------A--------------------------------          (4)
E2.1          ---------------------------------------------------------------------------------------------          (29)
E2.2          ------------------------------------------------------------A--------------------------------          (3)
E1.1  1 340   GGTAGTGGCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGATATGATCGCTGGTGCCCACTGGGGAGTTCTAGCGGGCATAGCGTATTTCTCCATG 1 436 (26)
E1.2          ---------------------------------------------------------------------------------------------          (4)
E2.1          ---------------------------------------------------------------------------------------------          (29)
E2.2          ---------------------------------------------------------------------------------------------          (3)
E1.1  1 437   GTGGGGAACTGGGCGAAGGTCGTGGTGGTGCTGCTGCTATTCGCCGGCGTTGACGCGGATACCTACGTCACCGGGGGAACTGCCGGGCACACCGTAT 1 533 (26)
E1.2          ------------------------------------------------------------------------------------G----------          (4)
E2.1          ---------------------------------------------------------------------------------------------          (29)
E2.2          ---------------------------------------------------------------------------------------------          (3)
E1.1  1 534   CGAGACTCAGCAGACTGCTCTCACCGGGCGCAAAGCAGAATATCCAGCTGATCAACTCCAACGGCAGCTGGGACATCAATAGGACAGCCCTGAACTG 1 630 (26)
E1.2          --G----------G-----------------------------------------------------------C---------------------          (4)
E2.1          --G----------G-----------------------------------------------------------C---------------------          (29)
E2.2          --G-----------------------------------------------------------------------C---------------------          (3)
E1.1  1 631   TAACGACAGCCTCAACACCGGCTGGATAGCAGGGCTCTTTTACCACTACAAATTCAACTCTTCAGGCTGCCCCGAGAGAATGGCCAGCTGTCATCCT 1 727 (26)
E1.2          ---------------------------------------------------------------------------------------------          (4)
E2.1          ---------------------------------------------------------------------------------------------          (29)
E2.2          ---------------------------------------------------------------------------------------------          (3)
E1.1  1 728   CTTACCGATTTTGCCCAGGGCTGGGGCCCTATCGGGTACGCCAATGGAAGCGGCCCCGACCATCGCCCCTACTGCTGGCACTACCCCCCAAGACCTT 1 824 (26)
E1.2          ---------------------------------------------------------------------------------------------          (4)
E2.1          ------------------------------------------------------------------------------------------T----          (29)
E2.2          ---------------------------------------------------------------------------------------------          (3)
E1.1  1 825   GTGGTATTGTTCCGGCACAGAGTGTCTGTGGCCCGGTGTACTGTTTTAA 1 873 (26)
E1.2          ----------G--------------G-----------------------        (4)
E2.1          ----------G--------------G-----------------------        (29)
E2.2          ----------G-------G------G-----------------------        (3)
```

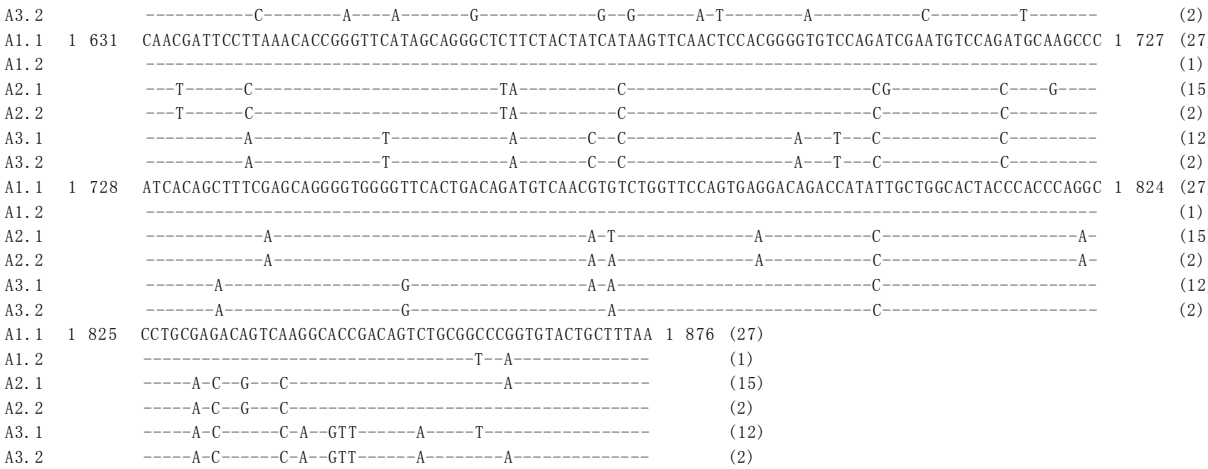
Nucleotide sequences of envelope region at 2-mo serial samples from the subject F

```
F1.1     855  ATGTGCTCTTTCTCTATCTTCCTTTTAGCCTTGCTATCCTGTTTGACCACCCCAGCTTCCGCTTACGAAGTGCGTAACGTGTCCGGGATATACCATG 951    (21)
F1.2          ---------------------------------------------------------------------------------------------          (2)
F2.1          ---------------------------------------------------------------------------------------------          (23)
F2.2          ---------------------------------------------------------------------------------------------          (1)
F1.1     952  TCACGAACGACTGCTCCAACTCAAGCATTGTGTATGAGGCAGCGGACCTGATCATGCATACCCCTGGGTGCGTGCCCTGCGTTCGGGAAGGCAACTC 1 048 (21)
F1.2          ---------------------------------------------------------------------------------------------          (2)
F2.1          ---------------------------------------------------------------------------------------------          (23)
F2.2          ---------------------------------------------------------------------------------------------          (1)
F1.1  1 049   CTCCCGTTGCTGGGTAGCGCTCACTCCCACGCTCGCGGCCAGGAACGCCACGATCCCCACTGCGACAGTACGACGGCATGTCGATCTGCTCGTTGGG 1 145 (21)
F1.2          ---------------------------------------------------------------------------------------------          (2)
F2.1          ---------------------------------------------------------------------------------------------          (23)
F2.2          ---------------------------------------------------------------------------------------------          (1)
F1.1  1 146   GCGGCTGCTTTCTCTTCCGCCATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCTTGTCTCTCAGCTGTTCACCTTCTCGCCTCGCCGGTATGAGA 1 242 (21)
F1.2          ---------------------------------------------------------------------------------------------          (2)
```

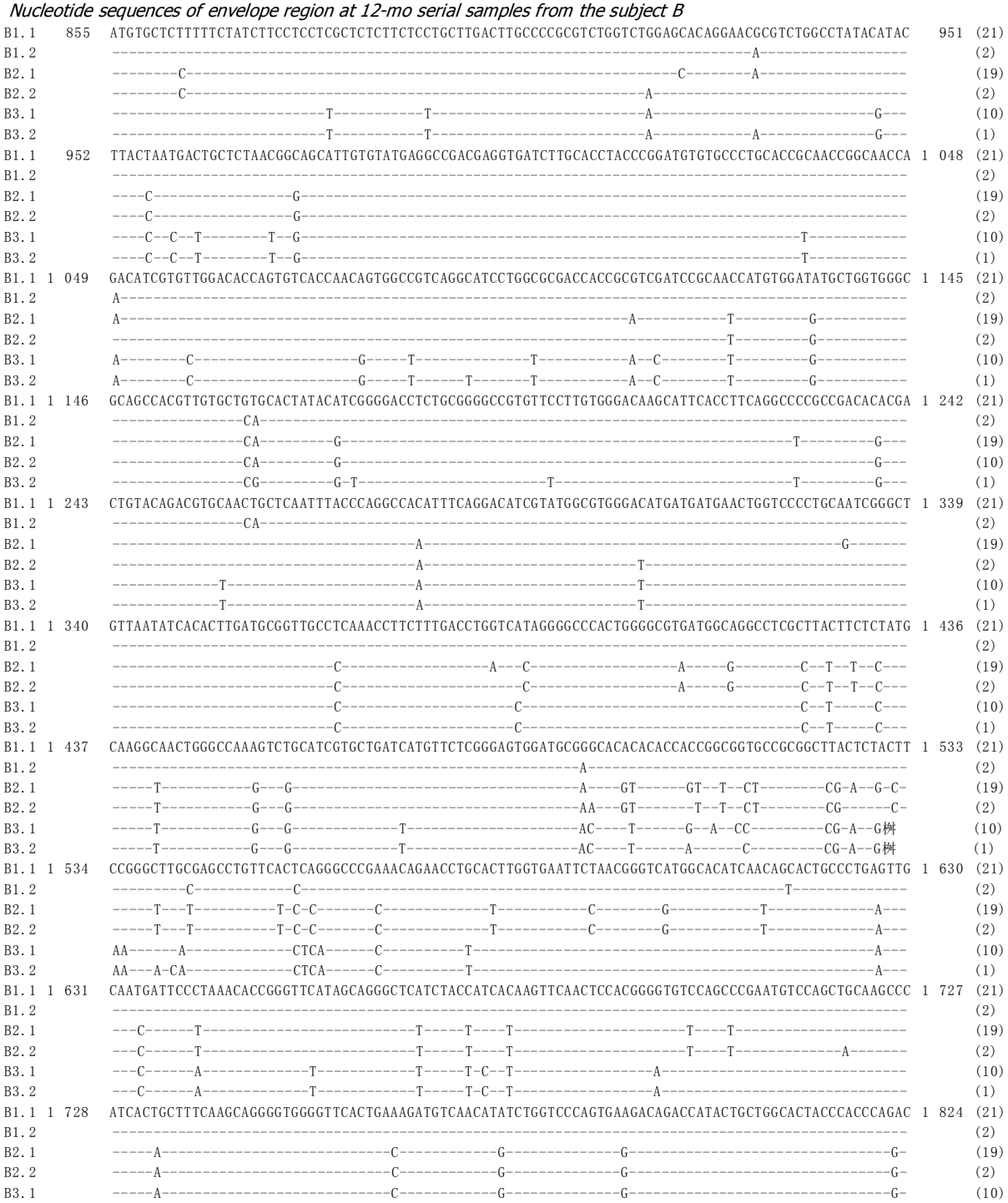


Nucleotide sequences of envelope region at 14-mo serial samples from the subject A





Nucleotide sequences of envelope region at 12-mo serial samples from the subject B



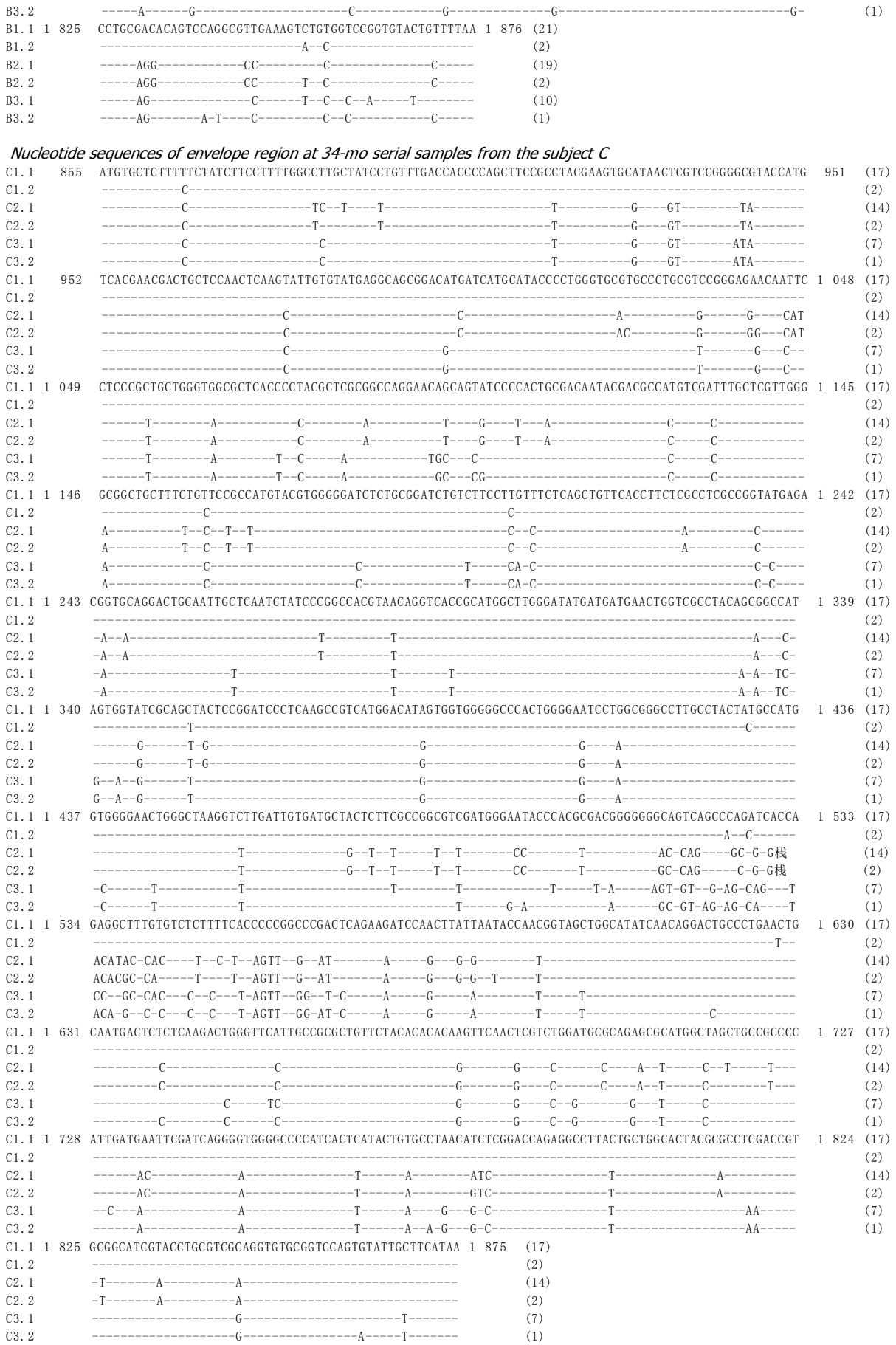


Figure 4 Alignment of nucleotide sequences of E1 and the 5' half of E2 including HVR1 for the majority sequences and each sequence from each subject A, B, C, E and F at different time points. In the first column, alphabetical labels indicate different samplings, while in the last column the numbers indicate the number of cloned cDNAs (out of 33 assessed for each sample) with SSCP+HD pattern. Dashes indicate identity to the nucleotide at the position in the first sequence.

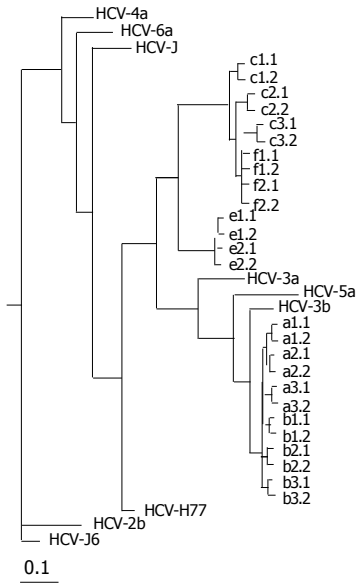


Figure 5 Unrooted tree showing the diversity of 384-nt E2 sequences from subjects. Letters represent individuals who cleared viremia (e and f) and those with persistent viremia (a-c), and lowercase numbers indicate the different clones obtained. The number and line at the bottom denote the proportion of nucleotides substituted for a given horizontal branch length. Dendrograms were produced using the Neighbor-joining program.

consistently lower than within HVR1 in individuals with persistent viremia. To investigate whether the different patterns of viral variation were due to positive selection, dN and dS both within and outside HVR1 were measured by comparing the sequences obtained at each time point from each subject with the sequence of the first time point (Figure 6). For individuals with persistent viremia, the mean number of dN within HVR1 predominated and gradually increased, compared to that in the E1 and E2 regions outside HVR1. However, these differences did not reach significance. By contrast, both dN and dS for the E1 and E2 regions including HVR1 was consistently lower in individuals with clearance of viremia. These data indicated that HCV persistence was associated with genetic evolution of the viral quasispecies and positive selection of HVR1 by the host immune system.

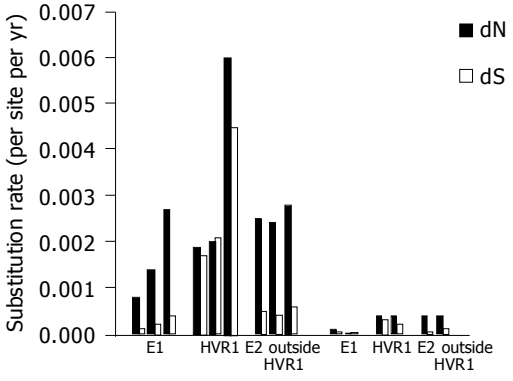


Figure 6 Nonsynonymous rate and synonymous rates of mutations for the E1, HVR1 and E2 regions outside HVR1 from five individuals.

DISCUSSION

This study, using PCR clones derived from two regions of the HCV envelope sequence E1, and E2/HVR1, which possessed different degrees of nucleotide sequence variability, showed that the SSCP+HDA might be applicable to the investigation of HCV genetic diversity at the intrahost level. Sequence differences in a relatively large number of clones from a subgenomic fragment (33 per sample) could be screened during a single procedure. Clones yielding differing gel migratory positions were rapidly identified to be processed further for nucleotide sequencing analyses if required. In each individual there was a major variant (most commonly observed) examined in E1, accompanied by minor variants, which were nearly always found in subsequent specimens in serial infection. The persistence of some variants through serial passage might indicate that E1 had functional constraints on genetic variation, such as RNA secondary structure or binding sites factors that regulated replication or translation[20]. However, two different patterns of the evolution of quasispecies in E2 emerged during the acute phase of HCV infection: in subjects with clearance of viremia, the number of viral variants within quasispecies was reduced and the quasispecies composition remained largely unchanged during the follow-up, while in subjects with persistent viremia, the number of viral variants increased during persistent infection and quasispecies showed constant evolution thereafter, suggesting that the dynamics of viral quasispecies during HCV primary infection might contribute to the outcome of HCV infection. These data showed that the interplay between different HCV strains and hosts might result in different, perhaps unique, quasispecies compositions.

The assessment of quasispecies diversity, as opposed to complexity, requires sequencing procedures to be carried out. These are particularly necessary to define sequence changes in the minority variants. Sequencing of inserted E1 and E2/HVR1 clones derived from all five subjects showed that variation in sequences of the minority variants involved single-nucleotide substitutions from majority variant, accounting for the tight clustering of sequences seen in a dendrogram. These data were consistent with quasispecies evolution from a single HCV founder strain and again pointed to the rarity of multiple HCV carriage in IDUs[27]. Despite the vulnerability of IDUs to HCV multiple transmission through a variety of unknown routes, we found no evidence of mixed HCV genotype infection, neither did we find higher quasispecies complexity nor genetic distance values in the first samples from the subjects with persistent viremia compared to the subjects with clearance of viremia. These findings demonstrated that the multiple HCV transmission events hypothesized to occur in IDUs, for various reasons, might not be associated with an increase in genetic complexity.

From a mechanistic perspective, variation with the HCV genome is assumed to be caused by random mutation and selection of variants, which are most fit to propagate in a given host. For example, in the immunocompetent host, antibodies directed against envelope gene product appear to play an important role in shaping quasispecies repertoires[28-30]. However, in protein-coding regions, multiple forces affect the balance between fixation of silent

(synonymous) mutations *vs* those that alter amino acid sequence (nonsynonymous). Synonymous mutations are often thought to represent a molecular clock, independent of external pressure and expected to occur at a rate proportional to the organism's reproductive rate, whereas nonsynonymous mutations are selected by immune pressure[31]. We also evaluated the E1, HVR1, and E2 regions outside HVR1 for synonymous and nonsynonymous substitutions. As shown in Figure 3, the rates of synonymous substitutions predominated and increased over time in the E1 and E2 regions. In contrast to the rates of substitutions found in the E1 and E2 regions outside HVR1, the rate of nonsynonymous substitutions predominated in the HVR1 region. A higher accumulation rate of nonsynonymous substitutions seen in individuals with persistent viremia was correlated with the greater genetic diversity, being consistent with a positive selection for change within HVR1. Published data from acute hepatitis C showed that viral genetic diversity in resolving patients was considerably less than in those who progressed to chronic disease, and crucially, that nonsynonymous substitutions in HVR1 were more common in rapid and slow progressors, compared with the resolving cases[9]. The data of this study confirmed further that this region might be under selective pressure by the host immune system.

No correlation between the quasispecies complexity and serum viral RNA levels demonstrated in this study does not support the concept that viral quasispecies arise as a consequence of the limited fidelity of HCV replication. Although our data demonstrates the importance of immune pressure in the evolution of HCV quasispecies, selection of quasispecies in hepatocytes, both *in vivo* and *in vitro*, may result from the replication of a small subset of viruses due either to random sampling or to selection of only a few fit variants[32,33]. Direct competition between virus strains, resulting in interference preventing simultaneous continuous infection by closely related variants, could also be possible.

In summary, the current study demonstrates the adaptation of the SSCP+HDA method or characterizing and tracking HCV quasispecies by analyzing multiple regions of the HCV genome in individuals with different outcomes of acute hepatitis C. This approach allows a larger number of patients and a larger proportion of the HCV genome to be analyzed than prior longitudinal studies of quasispecies diversity. We conclude that HCV persistence may be associated with a complexity of quasispecies and positive selection of HVR1 by the host immune system, and we postulate that both host and viral factors can play important roles in the pathogenesis of chronic hepatitis C in human populations. Further research is necessary to determine the extent to which the observed results are due to a different type of immune response in immunosuppressed patients compared with the immunocompetent host, a replicative advantage of quasispecies populations for certain HCV subgenotypes, and/or tropism of quasispecies variants for hepatic *vs* nonhepatic compartments.

REFERENCES

1 **Di Bisceglie AM.** Hepatitis C. *Lancet* 1998; **351**: 351-355
2 **Grakoui A,** Wychowski C, Lin C, Feinstone SM, Rice CM. Expression and identification of hepatitis C virus polyprotein cleavage products. *J Virol* 1993; **67**: 1385-1395
3 **Smith DB,** Simmonds P. Characteristics of nucleotide substitution in the hepatitis C virus genome: constraints on sequence change in coding regions at both ends of the genome. *J Mol Evol* 1997; **45**: 238-246
4 **Kato N,** Ootsuyama Y, Tanaka T, Nakagawa M, Nakazawa T, Muraiso K, Ohkoshi S, Hijikata M, Shimotohno K. Marked sequence diversity in the putative envelope proteins of hepatitis C viruses. *Virus Res* 1992; **22**: 107-123
5 **Martell M,** Esteban JI, Quer J, Genesca J, Weiner A, Esteban R, Guardia J, Gomez J. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J Virol* 1992; **66**: 3225-3229
6 **Steinhauer DA,** Holland JJ. Rapid evolution of RNA viruses. *Annu Rev Microbiol* 1987; **41**: 409-433
7 **Weiner AJ,** Geysen HM, Christopherson C, Hall JE, Mason TJ, Saracco G, Bonino F, Crawford K, Marion CD, Crawford KA. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. *Proc Natl Acad Sci USA* 1992; **89**: 3468-3472
8 **Pawlotsky JM.** Hepatitis C virus infection: virus/host interactions. *J Viral Hepat* 1998; **5 Suppl 1**: 3-8
9 **Farci P,** Shimoda A, Coiana A, Diaz G, Peddis G, Melpolder JC, Strazzera A, Chien DY, Munoz SJ, Balestrieri A, Purcell RH, Alter HJ. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* 2000; **288**: 339-344
10 **Farci P,** Strazzera R, Alter HJ, Farci S, Degioannis D, Coiana A, Peddis G, Usai F, Serra G, Chessa L, Diaz G, Balestrieri A, Purcell RH. Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome. *Proc Natl Acad Sci USA* 2002; **99**: 3081-3086
11 **Kao JH,** Chen PJ, Lai MY, Wang TH, Chen DS. Quasispecies of hepatitis C virus and genetic drift of the hypervariable region in chronic type C hepatitis. *J Infect Dis* 1995; **172**: 261-264
12 Kato N, Ootsuyama Y, Sekiya H, Ohkoshi S, Nakazawa T, Hijikata M, Shimotohno K. Genetic drift in hypervariable region 1 of the viral genome in persistent hepatitis C virus infection. *J Virol* 1994; **68**: 4776-4784
13 **Manzin A,** Solforosi L, Petrelli E, Macarri G, Tosone G, Piazza M, Clementi M. Evolution of hypervariable region 1 of hepatitis C virus in primary infection. *J Virol* 1998; **72**: 6271-6276
14 **Farci P,** Alter HJ, Wong DC, Miller RH, Govindarajan S, Engle R, Shapiro M, Purcell RH. Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated *in vitro* neutralization. *Proc Natl Acad Sci USA* 1994; **91**: 7792-7796
15 **Farci P,** Shimoda A, Wong D, Cabezon T, De Gioannis D, Strazzera A, Shimizu Y, Shapiro M, Alter HJ, Purcell RH. Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. *Proc Natl Acad Sci USA* 1996; **93**: 15394-15399
16 **Manzin A,** Solforosi L, Petrelli E, Macarri G, Tosone G, Piazza M, Clementi M. Evolution of hypervariable region 1 of hepatitis C virus in primary infection. *J Virol* 1998; **72**: 6271-6276
17 **McAllister J,** Casino C, Davidson F, Power J, Lawlor E, Yap PL, Simmonds P, Smith DB. Long-term evolution of the hypervariable region of hepatitis C virus in a common-source-infected cohort. *J Virol* 1998; **72**: 4893-4905
18 **Ni YH,** Chang MH, Chen PJ, Lin HH, Hsu HY. Evolution of hepatitis C virus quasispecies in mothers and infants infected through mother-to-infant transmission. *J Hepatol* 1997; **26**: 967-974
19 **Lee JH,** Stripf T, Roth WK, Zeuzem S. Non-isotopic detection of hepatitis C virus quasispecies by single strand conformation polymorphism. *J Med Virol* 1997; **53**: 245-251
20 **Kreis S,** Whistler T. Rapid identification of measles virus

strains by the heteroduplex mobility assay. *Virus Res* 1997; **47**: 197-203

21 **Maynard JH,** Upadhyaya M. High-throughput screening for the detection of unknown mutations: improved productivity using heteroduplex analysis. *Biotechniques* 1998; **25**: 648-651

22 **Wang YM,** Ray SC, Laeyendecker O, Ticehurst JR, Thomas DL. Assessment of hepatitis C virus sequence complexity by electrophoretic mobilities of both single-and double-stranded DNAs. *J Clin Microbiol* 1998; **36**: 2982-2989

23 **Ray SC,** Wang YM, Laeyendecker O, Ticehurst JR, Villano SA, Thomas DL. Acute hepatitis C virus structural gene sequences as predictors of persistent viremia: hypervariable region 1 as a decoy. *J Virol* 1999; **73**: 2938-2946

24 **Hadziyannis E,** Fried MW, Nolte FS. Evaluation of Two Methods for Quantitation of Hepatitis C Virus RNA. *Mol Diagn* 1997; **2**: 39-46

25 **Roth WK,** Lee JH, Ruster B, Zeuzem S. Comparison of two quantitative hepatitis C virus reverse transcriptase PCR assays. *J Clin Microbiol* 1996; **34**: 261-264

26 **Park YS,** Lee KO, Oh MJ, Chai YG. Distribution of genotypes in the 5′ untranslated region of hepatitis C virus in Korea. *J Med Microbiol* 1998; **47**: 643-647

27 **Smith DB,** Simmonds P. Characteristics of nucleotide substitution in the hepatitis C virus genome: constraints on sequence change in coding regions at both ends of the genome. *J Mol Evol* 1997; **45**: 238-246

28 **Nei M,** Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986; **3**: 418-426

29 **Harris KA,** Teo CG. Diversity of hepatitis C virus quasispecies evaluated by denaturing gradient gel electrophoresis. *Clin Diagn Lab Immunol* 2001; **8**: 62-73

30 **Kato N,** Sekiya H, Ootsuyama Y, Nakazawa T, Hijikata M, Ohkoshi S, Shimotohno K. Humoral immune response to hypervariable region 1 of the putative envelope glycoprotein (gp70) of hepatitis C virus. *J Virol* 1993; **67**: 3923-3930

31 **Weiner AJ,** Geysen HM, Christopherson C, Hall JE, Mason TJ, Saracco G, Bonino F, Crawford K, Marion CD, Crawford KA. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. *Proc Natl Acad Sci USA* 1992; **89**: 3468-3472

32 **Yun Z,** Barkholt L, Sonnerborg A. Dynamic analysis of hepatitis C virus polymorphism in patients with orthotopic liver transplantation. *Transplantation* 1997; **64**: 170-172

33 **Rumin S,** Berthillon P, Tanaka E, Kiyosawa K, Trabaud MA, Bizollon T, Gouillat C, Gripon P, Guguen-Guillouzo C, Inchauspe G, Trepo C. Dynamic analysis of hepatitis C virus replication and quasispecies selection in long-term cultures of adult human hepatocytes infected *in vitro*. *J Gen Virol* 1999; **80 (Pt 11)**: 3007-3018

Science Editor Kumar M and Ma JY **Language Editor** Elsevier HK