Phylogenetic Window Analysis for Detecting Chronological Changes in Natural Selection

Yoshiyuki Suzuki*

Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, 1111 Yata, Mishima-shi, Shizuoka-ken 411-8540, Japan

Abstract: Natural selection operating at the amino acid sequence level can be detected by comparing the rates of synonymous (r_S) and nonsynonymous (r_N) substitutions for the protein-coding nucleotide sequence, where relationships $r_N > r_S$ and $r_N < r_S$ conventionally indicate positive and negative selection, respectively. The direction and magnitude of natural selection operating on a protein may change during evolution because the environmental conditions may vary along with time. Here a phylogenetic window analysis method is proposed for examining the chronological change in natural selection and for detecting natural selection that has operated temporarily in the phylogenetic tree. The phylogenetic window was defined as an interval between two time points in the phylogenetic tree, which was constructed under the assumption of a molecular clock. The total numbers of synonymous and nonsynonymous changes that have occurred for all the parts of branches overlapping with the window were compared to detect natural selection. When this method was applied to the analysis of the intra-host evolution for hypervariable region 1 of hepatitis C virus (HCV), which was known as the major target of humoral immunity, it was found that the pattern of chronological change in natural selection was heterogeneous among patients. The r_N/r_S value was sometimes elevated temporarily, where positive selection was detected, suggesting that the antigenic evolution was punctuated during chronic infection of HCV.

Keywords: Phylogenetic tree, window analysis, natural selection, synonymous substitution, nonsynonymous substitution, hepatitis C virus.

INTRODUCTION

Natural selection operating at the amino acid sequence level can be detected by comparing the rates of synonymous (r_S) (Supplementary Table **S1**) and nonsynonymous (r_N) substitutions for the protein-coding nucleotide sequence [1]. Under the assumption that the synonymous substitution is selectively nearly neutral, relationships $r_N > r_S$ and $r_N < r_S$ conventionally indicate positive and negative selection, respectively. The direction and magnitude of natural selection operating on a protein may change during evolution because the environmental conditions may vary along with time. To detect natural selection operating temporarily, it may be useful to compare r_S and r_N for a specific branch of the phylogenetic tree [2]. However, the numbers of synonymous (c_S) and nonsynonymous (c_N) changes that have occurred for a branch may not be large enough for a statistical test to detect a significant difference between r_S and r_N. The branches where similar selection has operated may be grouped to increase the sensitivity of the test [3, 4], but it is difficult to determine such branches because the environmental condition for each branch is usually unknown.

Similar selection, however, may operate on the contemporary organisms sharing the environmental condition. For example, in the epidemics of human influenza A virus, evolution of hemagglutinin, which is the major target of humoral immunity, has been characterized by long-intervals of antigenic stasis punctuated by short-intervals of antigenic changes [5, 6]. In the former intervals, neutral and occasionally positively selected amino acid substitutions accumulate to provide the basis of antigenic innovations, whereas in the latter intervals, positive selection operates on parallel amino acid substitutions that cause antigenic changes for multiple lineages through epistasis with the substitutions that have accumulated in the former intervals [7]. Similarly, evolution of vesicular stomatitis virus has been characterized by genetic stasis and punctuated equilibrium, which were associated with small and large ecological changes, respectively [8, 9]. In addition, in the chronic infection of hepatitis C virus (HCV), it has been reported that positive selection operating on hypervariable region 1 (HVR1), which consists of the N-terminal 27 amino acid sites of envelope glycoprotein 2 (E2) [10, 11] and is the major target of humoral immunity [12-15], reduced for the entire population in the course of 15.6–21.6 years of follow-up in 5 patients [16], although such a tendency was not found in other patients [17, 18].

In the above studies, the chronological change in natural selection has been inferred mainly based on the comparison of r_S and r_N for branches of the phylogenetic tree. However, since different branches were usually involved in different time intervals, it was difficult to detect natural selection operating for a specific time interval. For this purpose, contemporary parts of the branches across the phylogenetic tree may be grouped to conduct the test of selective neutrality. In the present study, the phylogenetic window analysis method was proposed for examining the chronological change in natural selection and for detecting natural selection that has operated

*Address correspondence to this author at the Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, 1111 Yata, Mishima-shi, Shizuoka-ken 411-8540, Japan; Tel: +81-55-981-6847; Fax: +81-55-981-6848; E-mail: yossuzuk@lab.nig.ac.jp

1874-4044/08

temporarily in the phylogenetic tree. The method was applied to the data of the intra-host evolution for HVR1 of HCV.

MATERIALS AND METHODS

Phylogenetic Window Analysis

The phylogenetic window analysis is intended to estimate the total values of c_S and c_N that have occurred for a specific time interval across the phylogenetic tree and conduct a test of selective neutrality for detecting natural selection. The phylogenetic tree is assumed to be known. To define specific time intervals, the phylogenetic tree is constructed under the assumption of a molecular clock, where the evolutionary rate is estimated by including calibration points or, particularly in the analysis of viral sequences, using the viral strains sampled at different time points [19]. The time scale of the phylogenetic tree is obtained by dividing the branch lengths by the evolutionary rate. In the phylogenetic tree, the ancestral nucleotide sequence at each interior node is inferred by the maximum parsimony (MP) [20, 21] or Bayesian [22, 23] method, and c_S and c_N for each branch are obtained by comparing the nucleotide sequence at its one end with the other [24]. The phylogenetic window is defined as an interval between two time points in the phylogenetic tree, where the length of the time interval corresponds to the window size (w). The window overlaps with some branches of the phylogenetic tree. c_S and c_N for all the parts of branches overlapping with the window are summed to obtain the total numbers of synonymous ($c_{S(W)}$) and nonsynonymous ($c_{N(W)}$) changes that have occurred in the window, respectively. Here c_S and c_N for a part of a branch are obtained simply by fractionating these values for the branch according to the proportion of the overlapping region. The numbers of synonymous (s_S) and nonsynonymous (s_N) sites for the entire sequence are computed as the average of these values for all extant sequences. The null hypothesis of selective neutrality is tested by computing the probability (p) of obtaining the observed or more biased values for $c_{S(W)}$ and $c_{N(W)}$ under the assumption that these values follow a binomial distribution with the probabilities of occurrence of synonymous and nonsynonymous changes given by $s_S/(s_S + s_N)$ and $s_N/(s_S + s_N)$, respectively [24]. Positive and negative selection are inferred when $c_{N(W)}/s_N > c_{S(W)}/s_S$ and $c_{N(W)}/s_N < c_{S(W)}/s_S$ with $p < 0.05$, respectively. r_N/r_S for the window is estimated as $(c_{N(W)}/s_N)/(c_{S(W)}/s_S)$. The phylogenetic tree is scanned by sliding the window with a certain step size (s) for examining the chronological change in natural selection and for detecting natural selection that has operated temporarily. It should be noted that, since multiple substitutions are not corrected for in this method, $c_{S(W)}$ and $c_{N(W)}$ may be underestimated, especially when the branch lengths of the phylogenetic tree are large. Therefore, this method is considered to be suitable for the analysis of closely related sequences. In the present study, however, the degree of underestimation appeared to be negligible for all the data analyzed because the branch lengths were generally small [25].

Sequence Data

The phylogenetic window analysis was applied to the data of the intra-host evolution for HCV. The data consisted of the 5'-terminal 324 nucleotide sites of the E2 gene, which encoded 108 amino acid sites corresponding to positions 384–491 of HCV-1 [26]. The first 27 amino acid sites (positions 384–410) corresponded to HVR1. The nucleotide sequences were derived from 13, 7, 17, 5, and 14 strains serially sampled from patients 1, 2, 3, 4, and 5, respectively, in the course of 15.6–21.6 years of follow-up during chronic infection [16]. The strain names, accession numbers in the International Nucleotide Sequence Database, and isolation years for these sequences are listed in Supplementary Table **S2**.

In a previous study, these sequences have been analyzed for examining the chronological change in natural selection [16]. The phylogenetic tree was constructed for the sequences obtained from each patient under the assumption of a molecular clock, and the r_N/r_S value estimated for each branch was plotted against the time interval between the root of the phylogenetic tree and the middle of the branch. When the plots were superimposed for 5 patients, a negative correlation was observed between r_N/r_S and the time interval. In addition, the r_N/r_S value was found to be greater than 1 not only for HVR1 but also for amino acid positions 457–462.

In the present study, the phylogenetic window analysis was applied to the data obtained from each patient separately for examining whether the r_N/r_S value reduced during chronic infection in every patient. It should be noted that the biological function of amino acid positions 457–462 is unknown. In addition, although both the humoral and cellular immunities against E2 were known to be effective to eliminate HCV [27-31], the latter immunity appeared to exert little effect on driving sequence evolution for E2 [18]. Therefore, only HVR1 (positions 384–410) was used for examining the chronological change in r_N/r_S by the phylogenetic window analysis.

Data Analysis

The multiple alignment of the entire region (324 nucleotide sites) for the total of 56 sequences obtained from 5 patients was made using the computer program CLUSTAL W (version 1.83) [32]. The alignment did not contain any gaps. To determine the position of the root and the topology of the phylogenetic tree for the sequences obtained from each patient, the phylogenetic tree was constructed for the 56 sequences by the neighbor-joining (NJ) method [33] using the 324 nucleotide sites. The evolutionary distance was measured as the p distance, which was known to produce reliable topologies when large numbers of closely related sequences were analyzed [34, 35]. The reliability of each interior branch was assessed by the bootstrap method with 1000 resamplings [36]. MEGA (version 4.0) [37] was used for these analyses.

The branch lengths of the phylogenetic tree for the sequences obtained from each patient were re-estimated under the assumption of a molecular clock. Since the first 81 nucleotide sites encoding HVR1 were examined for the chronological change in r_N/r_S, the molecular clock was not assumed to hold for these sites. Therefore, the remaining 243 nucleotide sites were used for estimating the branch lengths. The model of nucleotide substitution that best fitted these sites was judged by the hierarchical likelihood-ratio test (hLRT) using MODELTEST (version 3.7) [38]. Based on the best fit model, the branch lengths, transition/transversion rate ratio (κ), and rate of nucleotide substitution were estimated under

the assumption of the molecular clock by the maximum likelihood method using TIPDATE (version 1.2) [39]. In addition, the branch lengths and κ were also estimated under the assumption of the rate heterogeneity among branches. The log-likelihood (ln*L*) values obtained under these assumptions were compared by the LRT to test the null hypothesis of the molecular clock. Twice the difference in the ln*L* value was assumed to follow a χ^2 distribution with a degree of freedom of $n - 3$ [39], where n denotes the number of sequences analyzed. The molecular clock was rejected if $p < 0.05$.

The ancestral nucleotide sequence at each interior node of the phylogenetic tree was inferred by the MP method. The values of s_S and s_N were computed by taking into account the κ value estimated above [40]. The phylogenetic window analysis was conducted for the sequences obtained from each patient using 5 years and 0.5 year as w and s ($w = 5$ and $s = 0.5$), respectively, and using $w = 10$ and $s = 1$. In addition, the average pattern of the chronological change in natural selection for 5 patients was examined by summing the $c_{S(W)}$ and $c_{N(W)}$ values for the windows of the same chronological order and averaging the s_S and s_N values for 5 patients.

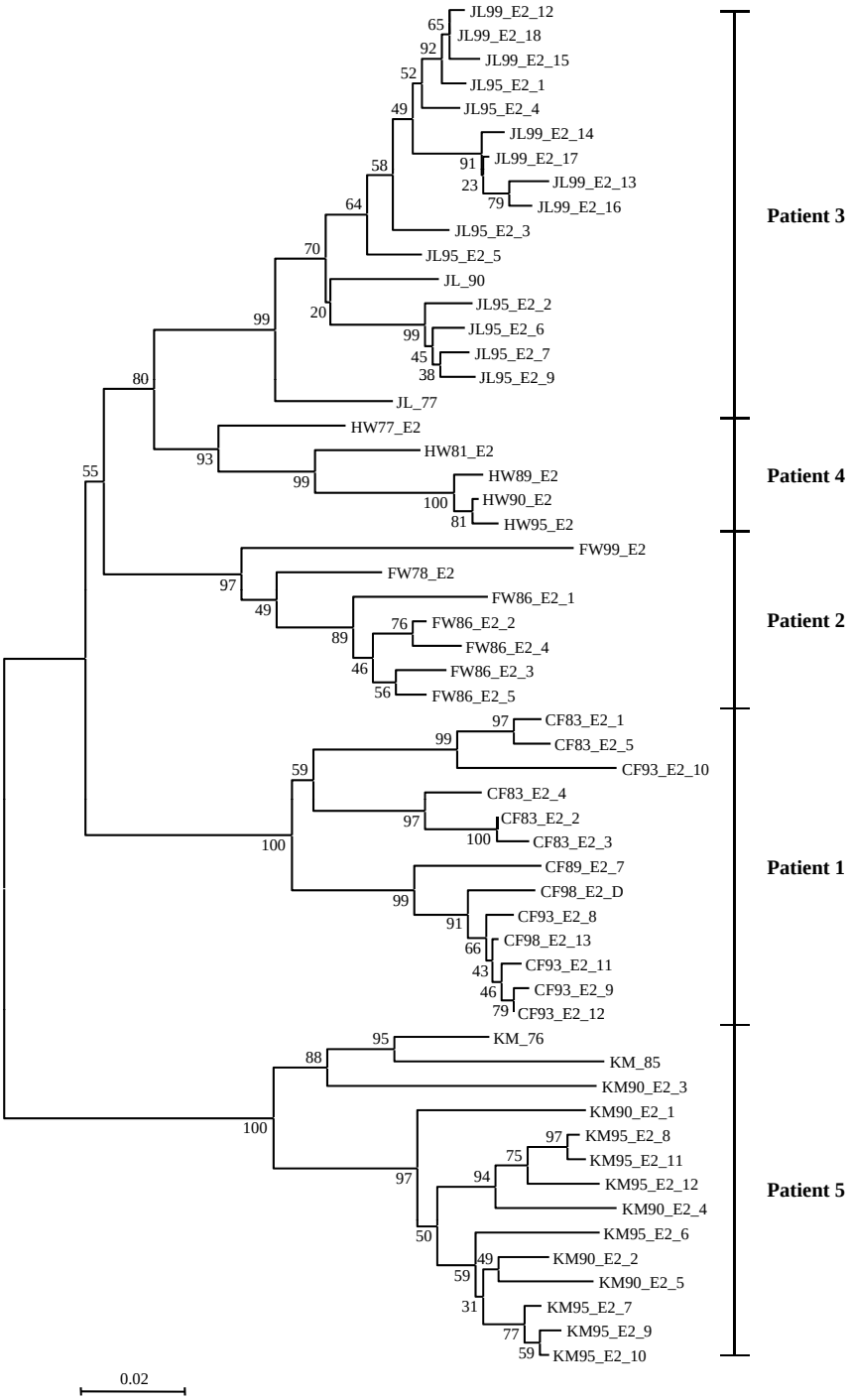

Fig. (1). Phylogenetic tree constructed for the total of 56 HCV strains serially sampled from 5 patients. The bootstrap probability is indicated for each interior branch. The scale bar indicates 0.02 nucleotide substitution per site.

RESULTS

Construction of the Phylogenetic Tree

The phylogenetic tree constructed for the total of 56 HCV strains serially sampled from 5 patients using the 5'-terminal 324 nucleotide sites of the E2 gene is shown in Fig. (**1**). The strains obtained from each patient formed a single cluster, which was supported with a high bootstrap probability (93%–100%). Using the position of the root and the topology of the sub-tree for each patient, the branch lengths were re-estimated with the 243 nucleotide sites, by eliminating the first 81 nucleotide sites that encoded HVR1. The models of nucleotide substitution best fitted to the 243 sites of the sequences obtained from patients 1, 2, 3, 4, and 5 were the model of Kimura [41] with the Γ distribution for the rate heterogeneity among sites (K + Γ), K, K + Γ, K, and the model of Hasegawa *et al*. [42] (HKY) + Γ, respectively (Supplementary Table **S3**). Based on these models, the ln*L* values were obtained under the assumptions of the molecular clock and the rate heterogeneity among branches. When the null hypothesis of the molecular clock was tested for the sequences obtained from each patient, $p > 0.05$ for all patients except for patient 3. However, even for patient 3, $p = 0.0486$, which was not statistically significant when the Bonferroni correction was conducted. These results indicated that the molecular clock could be assumed for the 243 nucleotide sites of the sequences obtained from each patient.

Results of the Phylogenetic Window Analysis

The results of the phylogenetic window analysis conducted for 81 nucleotide sites that encoded HVR1 using the phylogenetic tree constructed for each patient as indicated above were shown in Fig. (**2**). Although the r_N/r_S value for the same chronological region varied to some extent according to the window size assumed ($w = 5$ or $w = 10$), the overall pattern of the chronological change in r_N/r_S was similar between these two cases for each patient (Supplementary Table **S4**). In patient 1, r_N/r_S was initially greater than 1, and decreased to be approximately 1 in the course of chronic infection. The r_N/r_S value was large (25.783) for the last window with $w = 5$. This result was obtained apparently because the c_S value was very small for this window (0.025) due to a statistical error (Supplementary Table **S4**). In fact, positive selection was not detected for this window and r_N/r_S was not large for the corresponding window with $w = 10$. Positive selection was not detected for the entire time period (Fig. **2**; Supplementary Table **S4**). Similarly, in patient 2, r_N/r_S was initially greater than 1 and decreased to be approximately 1. However, the r_N/r_S value slightly increased later, although positive selection was not detected for the entire time period. In contrast to the cases for patients 1 and 2, r_N/r_S was initially close to 1 in patient 3. The r_N/r_S value was elevated in the course of chronic infection, where positive selection was detected for 2 consecutive windows with $w = 5$. However, the r_N/r_S value later decreased to be approximately 1. In patient 4, r_N/r_S was ∞ throughout the entire time period because only the nonsynonymous change was observed ($c_S = 0$ and $c_N > 0$) for the entire phylogenetic tree. Positive selection was detected for 7 consecutive windows in the middle of the phylogenetic tree with $w = 5$, and for 2 consecutive windows at the similar chronological region with $w = 10$. In patient 5, r_N/r_S was initially greater than 1, and decreased to be approximately 1 in the course of chronic infection. However, the r_N/r_S value was elevated in the middle of the phylogenetic tree, where positive selection was detected for 10 consecutive windows with both $w = 5$ and $w = 10$ at the similar chronological region. When the phylogenetic window analysis was conducted by combining the data from 5 patients, r_N/r_S was initially greater than 1, and gradually decreased toward 1. However, positive selection was detected for most of the windows with both $w = 5$ and $w = 10$.

DISCUSSION

Problems in the Phylogenetic Window Analysis

In the present study, the phylogenetic window analysis method was proposed for examining the chronological change in natural selection and for detecting natural selection that has operated temporarily in the phylogenetic tree, by estimating $c_{S(W)}$ and $c_{N(W)}$ and conducting the test of selective neutrality. However, there appeared to be some problems in this method. First, the window size should be large enough to include a sufficient number of nucleotide changes ($c_{S(W)} + c_{N(W)}$), which corresponded to the sample size in a statistical test, for obtaining a significant result. For example, in the analysis of the HCV strains isolated from patient 4, positive selection was detected with $w = 10$ but not with $w = 5$ at the same chronological region of the phylogenetic tree. Since $c_{S(W)} = 0$ for both cases, success or failure in detecting positive selection was determined by whether the sample size ($c_{N(W)}$) was sufficient for obtaining a statistical significance or not, respectively. The window size, however, should not be excessively large, because the effect of natural selection is averaged for the window and natural selection operating only for a short time interval may be obscured. For example, in the analysis of the sequences obtained from patient 3, positive selection was detected with $w = 5$ but not with $w = 10$ at the same chronological region. In fact, the peak of the r_N/r_S value observed with $w = 5$ was obscured with $w = 10$. In addition, the sample size for a window depends on the total length of the parts of branches included in the window. However, since the number of branches may vary along with the phylogenetic tree, the window size required for including a sufficient sample size to obtain a statistical significance may be different among chronological regions of the phylogenetic tree. For example, in the analysis of the sequences obtained from patient 4, positive selection was detected with $w = 10$ only in the middle of the phylogenetic tree, although r_N/r_S was ∞ for the entire time period. These observations suggested that various window sizes should be examined in the analysis [43, 44], as was the case with the present study. Since the sample size varied among the windows and was sometimes insufficient for obtaining a statistical significance, the correction for multiple testing was not adopted in the present study.

Second, c_S and c_N for a part of a branch overlapping with the phylogenetic window were obtained simply by fractionating these values for the branch according to the proportion of the overlapping region, suggesting that the synonymous and nonsynonymous changes were implicitly assumed to be evenly distributed along with the branch. This assumption, however, may be inconsistent with the idea of the phylogenetic window analysis, where the chronological change in

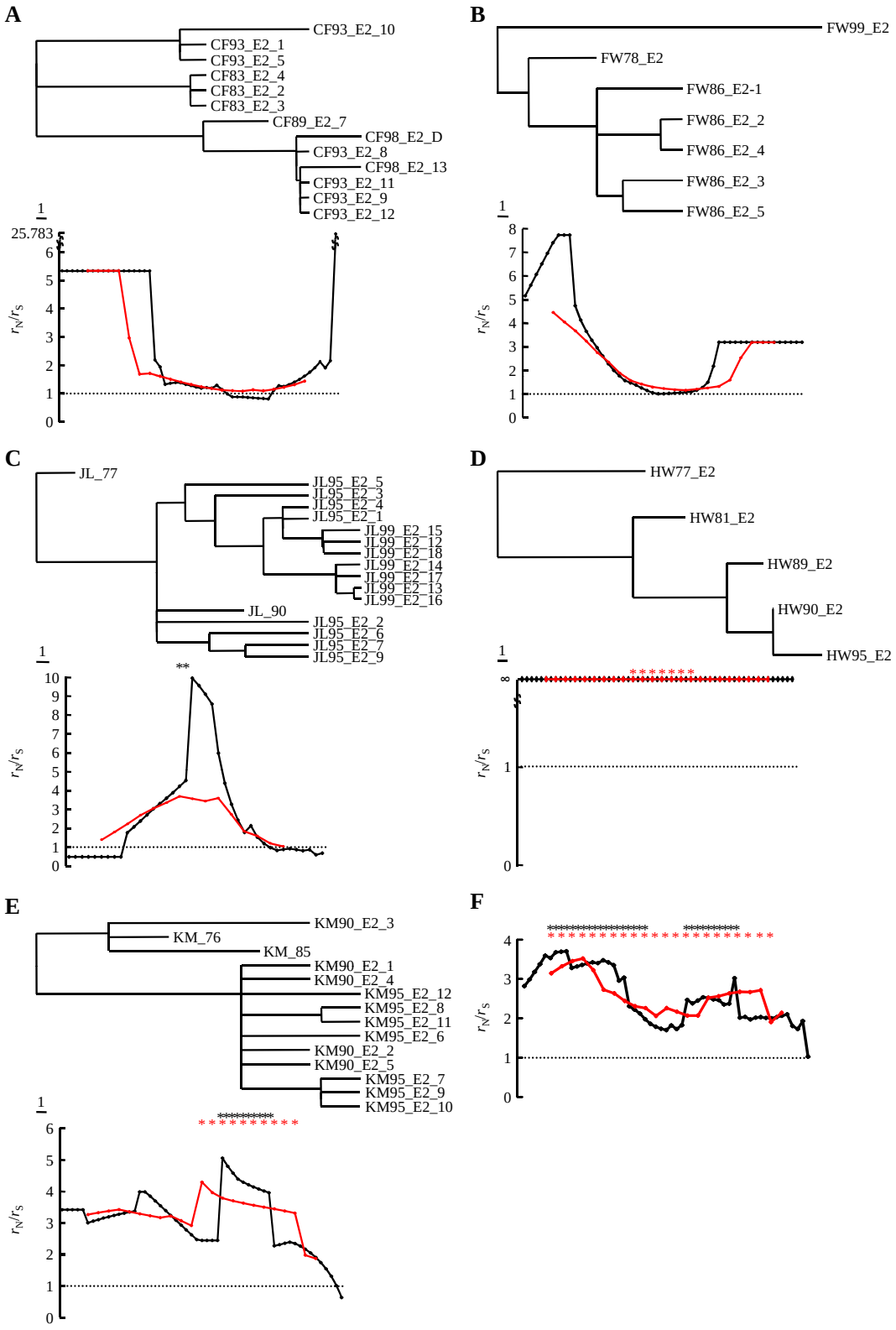

Fig. (2). Results of the phylogenetic window analysis conducted for HVR1 of the HCV strains sampled from patients 1 (A), 2 (B), 3 (C), 4 (D), and 5 (E). The phylogenetic tree was constructed under the assumption of the molecular clock, where the scale bar indicates 1 year. The time scale for the results of the phylogenetic window analysis is matched to that for the phylogenetic tree for each patient. The black and red lines indicate the r_N/r_S values with $w = 5$ and $s = 0.5$ and with $w = 10$ and $s = 1$, respectively. The black and red asterisks indicate the central time points of the windows where positive selection was detected with $w = 5$ and $w = 10$, respectively. The results obtained by combining the data from 5 patients are also indicated (F). The data for the last two windows and the last window were removed for $w = 5$ and $w = 10$, respectively, because $c_{S(W)}$ was 0 and $c_{N(W)}$ was very small (approximately 1) (Supplementary Table **S4**).

r_N/r_S, which may occur at any time point irrespective of the position on the branch, was intended to be detected. Since the effect of natural selection is averaged for each branch, this method may produce conservative results for the extent of chronological change in r_N/r_S and for the detection of natural selection. Nevertheless, this problem may disappear as more sequences become available in the analysis, so that the branch lengths become smaller in the phylogenetic tree, as in the case for influenza A virus [45-47].

Punctuated Evolution of Antigenicity During Chronic Infection of HCV

When the phylogenetic window analysis was applied to the combined data of the intra-host evolution for HVR1 of HCV in 5 patients, it was found that the r_N/r_S value decreased along with time, which was consistent with the previous study [16]. However, positive selection was detected for most of the windows, suggesting that escape mutants from the humoral immunity were generated with a decreasing rate throughout the chronic infection on average. The decrease in the rate of generating escape mutants may have occurred because strain-specific and low titer neutralizing antibodies (nAbs) are elicited in the early stage of HCV infection, whereas cross-reactive and high titer nAbs are elicited later [31, 48]. Alternatively, the humoral immunity against HCV may have diminished in the late stage of chronic infection, so that the selective pressure to generate escape mutants may have reduced [16]. In the present study, however, it was also found that the pattern of chronological change in natural selection was heterogeneous among patients [17, 18]. The r_N/r_S value was sometimes elevated temporarily even in the later stage, where positive selection was detected, indicating that the antigenic evolution was punctuated during chronic infection of HCV. The punctuated evolution of antigenicity, however, may result from different mechanisms according to the driving force of positive selection. If it is assumed that the viral population generally contains antigenic mutants abundantly and positive selection is governed mainly by the change in the environmental condition [49-51], as has been indicated for vesicular stomatitis virus [8, 9], the above observation is considered to reflect the temporary change in the humoral immunity in the patients. On the other hand, if it is assumed that positive selection is governed mainly by *de novo* generation of antigenic mutants that can escape from the humoral immunity, as has been indicated for influenza A virus [5, 6, 52], the above observation is considered to reflect the epistasis among amino acid sites for determining the antigenicity. To distinguish these possibilities, it may be useful to compare the strength of immune responses in the patients as well as the degree of antigenic changes in HCV for the time intervals where positive selection was detected with those for other time intervals.

ACKNOWLEDGEMENTS

The author thanks anonymous reviewers for valuable comments. The present study was supported by KAKENHI 17770007 to Y.S.

REFERENCES

[1] Hughes, A.L.; Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, **1988**, *335*: 167-170.

[2] Zhang, J.; Rosenberg, H.F.; Nei, M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci.* USA, **1998**, *95*: 3708-3713.

[3] Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol*., **1998**, *15*: 568-573.

[4] Yang, Z.; Nielsen, R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol*., **1998**, *46*: 409-418.

[5] Smith, D.J.; Lapedes, A.S.; de Jong, J.C.; Bestebroer, T.M.; Rimmelzwaan, G.F.; Osterhaus, A.D.M.E.; Fouchier, R.A.M. Mapping the antigenic and genetic evolution of influenza virus. *Science*, **2004**, *305*; 371-376.

[6] Wolf, Y.I.; Viboud, C.; Holmes, E.C.; Koonin, E.V.; Lipman, D.J. Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol. Direct*, **2006,** *1*: 34.

[7] Shih, A.C.-C.; Hsiao, T.-C.; Ho, M.-S.; Li, W.-H. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. *Proc. Natl. Acad. Sci. USA*, **2007**, *104*: 6283-6288.

[8] Nichol, S.T.; Rowe, J.E.; Fitch, W.M. Punctuated equilibrium and positive Darwinian evolution in vesicular stomatitis virus. *Proc. Natl. Acad. Sci. USA*, **1993**, *90*: 10424-10428.

[9] Rodriguez, L.L.; Fitch, W.M.; Nichol, S.T. Ecological factors rather than temporal factors dominate the evolution of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*: 13030-13035.

[10] Hijikata, M.; Kato, N.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Shimotohno, K. Hypervariable regions in the putative glycoprotein of hepatitis C virus. *Biochem. Biophys. Res. Commun*., **1991**, *175*: 220-228.

[11] Weiner, A.J.; Brauer, M.J.; Rosenblatt, J.; Richman, K.H.; Tung, J.; Crawford, K.; Bonino, F.; Saracco, G.; Choo, Q.-L.; Houghton, M.; Han, J.H. Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. *Virology*, **1991**, *180*: 842-848.

[12] Weiner, A.J.; Geysen, H.M.; Christopherson, C.; Hall, J.E.; Mason, T.J.; Saracco, G.; Bonino, F.; Crawford, K.; Marion, C.D.; Crawford, K.A.; Brunetto, M.; Barr, P.J.; Miyamura, T.; McHutchinson, J.; Houghton, M. 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.

[13] Farci, P.; Shimoda, A.; Wong, D.; Cabezon, T.; De Gioannis, D.; Strazzera, A.; Shimizu, Y.; Shapiro, M.; Alter, H.J.; Purcell, R.H. 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.

[14] Zibert, A.; Kraas, W.; Meisel, H.; Jung, G.; Roggendorf, M. Epitope mapping of antibodies directed against hypervariable region 1 in acute self-limiting and chronic infections due to hepatitis C virus. *J. Virol*., **1997**, *71*: 4123-4127.

[15] Booth, J.C.; Kumar, U.; Webster, D.; Monjardino, J.; Thomas, H.C. Comparison of the rate of sequence variation in the hypervariable region of E2/NS1 region of hepatitis C virus in normal and hypogammaglobulinemic patients. *Hepatology*, **1998**, *27*: 223-227.

[16] Hanada, K.; Tanaka, Y.; Mizokami, M.; Gojobori, T.; Alter, H.J. A reduction in selective immune pressure during the course of chronic hepatitis C correlates with diminished biochemical evidence of hepatic inflammation. *Virology*, **2007**, *361*: 27-33.

[17] Chen, S.; Wang, Y.-M. Multigene tracking of quasispecies in viral persistence and clearance of hepatitis C virus. *World J. Gastroenterol*., **2005**, *11*: 2874-2884.

[18] Kuntzen, T.; Timm, J.; Berical, A.; Lewis-Ximenez, L.L.; Jones, A.; Nolan, B.; zur Wiesch, J.S.; Li, B.; Schneidewind, A.; Kim, A.Y.; Chung, R.T.; Lauer, G.M.; Allen, T.M. Viral sequence evolution in acute hepatitis C virus infection. *J. Virol*., **2007**, *81*: 11658-11668.

[19] Nei, M. Genetic polymorphism and the role of mutation in evolution. In Evolution of Genes and Proteins.; Nei, M.; Koehn, R.K. Eds.; Sinauer: Sunderland, Mass, **1981**; pp. 165-190.

[20] Fitch, W.M. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool*., **1971**, *20*: 406-416.

[21] Hartigan, J.A. Minimum mutation fits to a given tree. *Biometrics*, **1973**, *29*: 53-65.
[22] Yang, Z.; Kumar, S.; Nei, M. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics*, **1995**, *141*: 1641-1650.
[23] Zhang, J.; Nei, M. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J. Mol. Evol.*, **1997**, *44*: S139-S146.
[24] Suzuki, Y.; Gojobori, T. A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.*, **1999**, *16*: 1315-1328.
[25] Saitou, N. A theoretical study of the underestimation of branch lengths by the maximum parsimony principle. *Sys. Zool.*, **1989**, *38*: 1-6.
[26] Choo, Q.-L.; Richman, K.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coit, D.; Medina-Selby, A.; Barr, P.J.; Weiner, A.; Bradley, D.W.; Kuo, G.; Houghton, M. Genetic organization and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA*, **1991**, *88*: 2451-2455.
[27] Suzuki, Y.; Gojobori, T. Positively selected amino acid sites in the entire coding region of hepatitis C virus subtype 1b. *Gene*, **2001**, *276*: 83-87.
[28] Brown, R.J.P.; Juttla, V.S.; Tarr, A.W.; Finnis, R.; Irving, W.L.; Hemsley, S.; Flower, D.R.; Borrow, P.; Ball, J.K. Evolutionary dynamics of hepatitis C virus envelope genes during chronic infection. *J. Gen. Virol.*, **2005**, *86*: 1931-1942.
[29] Suzuki, Y. New methods for detecting positive selection at single amino acid sites. *J. Mol. Evol.*, **2004**, *59*: 11-19.
[30] Sheridan, I.; Pybus, O.G.; Holmes, E.C.; Klenerman, P. High-resolution phylogenetic analysis of hepatitis C virus adaptation and its relationship to disease progression. *J. Virol.*, **2004**, *78*: 3447-3454.
[31] von Hahn, T.; Yoon, J.C.; Alter, H.; Rice, C.M.; Rehermann, B.; Balfe, P.; McKeating, J.A. Hepatitis C virus continuously escapes from neutralizing antibody and T-cell responses during chronic infection *in vivo*. *Gastroenterology*, **2007**, *132*: 667-678.
[32] Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple-sequence alignment through sequence weighting, position-specific gap penalties, and weight-matrix choice. *Nucleic Acids Res.*, **1994**, *22*: 4673-4680.
[33] Saitou, N.; Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **1987**, *4*: 406-425.
[34] Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*, Oxford University Press: Oxford, New York, **2000**.
[35] Takahashi, K.; Nei, M. Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Mol. Biol. Evol.*, **2000**, *17*: 1251-1258.
[36] Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **1985**, *39*: 783-791.
[37] Tamura, K.; Dudley, J.; Nei, M.; Kumar, S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, **2007**, *24*: 1596-1599.
[38] Posada, D.; Crandall, K.A. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **1998**, *14*: 817-818.
[39] Rambaut, A. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics*, **2000**, *16*: 395-399.
[40] Suzuki, Y. Molecular Evolution of Pathogenic Viruses. PhD thesis. The Graduate University for Advanced Studies, School of Life Science, Department of Genetics, **1999**.
[41] Kimura, M. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol.*, **1980**, *16*: 111-120.
[42] Hasegawa, M.; Kishino, H.; Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.*, **1985**, *22*: 160-174.
[43] Fares, M.A.; Elena, S.F.; Ortiz, J.; Moya, A.; Barrio, E. A sliding window-based method to detect selective constraints in protein-coding genes and its application to RNA viruses. *J. Mol. Evol.*, **2002**, *55*: 509-521.
[44] Suzuki, Y. Three-dimensional window analysis for detecting positive selection at structural regions of proteins. *Mol. Biol. Evol.*, **2004**, *21*: 2352-2359.
[45] Ghedin, E.; Sengamalay, N.A.; Shumway, M.; Zaborsky, J.; Feldblyum, T.; Subbu, V.; Spiro, D.J.; Sitz, J.; Koo, H.; Bolotov, P.; Dernovoy, D.; Tatusova, T.; Bao, Y.; St. George, K.; Taylor, J.; Lipman, D.J.; Fraser, C.M.; Taubenberger, J.K.; Salzberg, S.L. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature*, **2005**, *437*: 1162-1166.
[46] Holmes, E.C.; Ghedin, E.; Miller, N.; Taylor, J.; Bao, Y.; St. George, K.; Grenfell, B.T.; Salzberg, S.L.; Fraser, C.M.; Lipman, D.J.; Taubenberger, J.K. Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol.*, **2005**, *3*: e300.
[47] Suzuki, Y. Natural selection on the influenza virus genome. *Mol. Biol. Evol.*, **2006**, *23*: 1902-1911.
[48] Logvinoff, C.; Major, M.E.; Oldach, D.; Heyward, S.; Talal, A.; Balfe, P.; Feinstone, S.M.; Alter, H.; Rice, C.M.; McKeating, J.A. Neutralizing antibody response during acute and chronic hepatitis C virus infection. *Proc. Natl. Acad. Sci. USA*, **2004**, *101*: 10149-10154.
[49] Eigen, M.; Schuster, P. A principle of natural self-organization. *Naturwissenschaften*, **1977**, *64*: 541-565.
[50] Martell, M.; Esteban, J.; 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.
[51] Fernandez, J.; Taylor, D.; Morhardt, D.R.; Mihalik, K.; Puig, M.; Rice, C.M.; Feinstone, S.M.; Major, M.E. Long-term persistence of infection in chimpanzees inoculated with an infectious hepatitis C virus clone is associated with a decrease in the viral amino acid substitution rate and low levels of heterogeneity. *J. Virol.*, **2004**, *78*: 9782-9789.
[52] Nei, M. The new mutation theory of phenotypic evolution. *Proc. Natl. Acad. Sci. USA*, **2007**, *104*: 12235-12242.

Received: March 14, 2008 Revised: April 10, 2008 Accepted: April 18, 2008

SUPPLEMENTARY MATERIALS

Table S1. List of Abbreviations Used in the Present Paper

Abbreviation	Expansion
r_S (r_N)	Rate of synonymous (nonsynonymous) substitution
c_S (c_N)	Number of synonymous (nonsynonymous) changes
$c_{S(W)}$ ($c_{N(W)}$)	Number of synonymous (nonsynonymous) changes for the phylogenetic window
s_S (s_N)	Number of synonymous (nonsynonymous) sites
w	Window size
s	Step size
p	Probability
n	Number of sequences
κ	Transition/transversion rate ratio
ln*L*	Log-likelihood
hLRT	Hierarchical likelihood-ratio test
MP	Maximum parsimony
NJ	Neighbor-joining
K	Model of Kimura
HKY	Model of Hasegawa et al.
Γ	Γ distribution for the rate heterogeneity among sites
HCV	Hepatitis C virus
HVR1	Hypervariable region 1
E2	Envelope glycoprotein 2
nAb	Neutralizing antibody

Table S2. Strain Names, Accession Numbers in the International Nucleotide Sequence Database, and Isolation Years for the Sequences Obtained from 5 Patients Analyzed in the Present Study [16]

Patient	Strain name	Accession number	Isolation year	Patient	Strain name	Accession number	Isolation year
1	CF83_E2_1	AB272166	1983		JL95_E2_7	AB272200	1995
	CF83_E2_2	AB272167	1983		JL95_E2_9	AB272202	1995
	CF83_E2_3	AB272168	1983		JL99_E2_12	AB272204	1999
	CF83_E2_4	AB272169	1983		JL99_E2_13	AB272205	1999
	CF83_E2_5	AB272170	1983		JL99_E2_14	AB272206	1999
	CF89_E2_7	AB272172	1989		JL99_E2_15	AB272207	1999
	CF93_E2_8	AB272173	1993		JL99_E2_16	AB272208	1999
	CF93_E2_9	AB272174	1993		JL99_E2_17	AB272209	1999
	CF93_E2_10	AB272175	1993		JL99_E2_18	AB272210	1999
	CF93_E2_11	AB272176	1993	4	HW77_E2	AB272187	1977
	CF93_E2_12	AB272177	1993		HW81_E2	AB272188	1981
	CF98_E2_13	AB272178	1998		HW89_E2	AB272189	1989
	CF98_E2_D	AB272179	1998		HW90_E2	AB272190	1990
2	FW78_E2	AB272180	1978		HW95_E2	AB272191	1995
	FW99_E2	AB272181	1999	5	KM_76	AB272212	1976

(Table S2). Contd…..

Patient	Strain name	Accession number	Isolation year	Patient	Strain name	Accession number	Isolation year
	FW86_E2_1	AB272182	1986		KM_85	AB272213	1985
	FW86_E2_2	AB272183	1986		KM90_E2_1	AB272214	1990
	FW86_E2_3	AB272184	1986		KM90_E2_2	AB272215	1990
	FW86_E2_4	AB272185	1986		KM90_E2_3	AB272216	1990
	FW86_E2_5	AB272186	1986		KM90_E2_4	AB272217	1990
3	JL_77	AB272192	1977		KM90_E2_5	AB272218	1990
	JL_90	AB272193	1990		KM95_E2_6	AB272219	1995
	JL95_E2_1	AB272194	1995		KM95_E2_7	AB272220	1995
	JL95_E2_2	AB272195	1995		KM95_E2_8	AB272221	1995
	JL95_E2_3	AB272196	1995		KM95_E2_9	AB272222	1995
	JL95_E2_4	AB272197	1995		KM95_E2_10	AB272223	1995
	JL95_E2_5	AB272198	1995		KM95_E2_11	AB272224	1995
	JL95_E2_6	AB272199	1995		KM95_E2_12	AB272225	1995

Table S3. Results of the Analysis Using the 243 Nucleotide Sites of the E2 Gene for the HCV Strains Serially Sampled from 5 Patients

Patient	Best fit model	ln*L* with clock	ln*L* without clock	*p* value in the LRT	κ	s_S	s_N	Rate (per site per year)
1	K + Γ	–567.824	–562.473	0.381	4.891	24.558	56.442	0.00176
2	K	–480.792	–476.947	0.104	4.991	24.548	56.452	0.00131
3	K + Γ	–575.739	–563.844	0.0486	4.022	24.597	56.403	0.00142
4	K	–440.897	–440.572	0.723	4.057	24.393	56.607	0.00133
5	HKY + Γ	–642.695	–638.245	0.631	5.474	24.261	56.739	0.00161

Table S4. The $c_{S(W)}$, $c_{N(W)}$, and r_N/r_S Values for each Phylogenetic Window

		$w = 5, s = 0.5$				$w = 10, s = 1$			
Patient	Order[a]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	Selection[b]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	Selection
1	1	0.335	4.101	5.334		0.669	8.202	5.334	
	2	0.335	4.101	5.334		0.669	8.202	5.334	
	3	0.335	4.101	5.334		0.669	8.202	5.334	
	4	0.335	4.101	5.334		0.669	8.202	5.334	
	5	0.335	4.101	5.334		1.359	9.252	2.961	
	6	0.335	4.101	5.334		3.823	14.773	1.681	
	7	0.335	4.101	5.334		4.223	16.564	1.707	
	8	0.335	4.101	5.334		4.745	17.573	1.611	
	9	0.335	4.101	5.334		5.136	17.677	1.498	
	10	0.335	4.101	5.334		5.527	17.781	1.400	
	11	0.335	4.101	5.334		5.917	17.885	1.315	
	12	0.335	4.101	5.334		6.308	17.990	1.241	

(Table S4). Contd…..

		$w = 5, s = 0.5$				$w = 10, s = 1$			
Patient	**Order**[a]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**[b]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**
	13	0.335	4.101	5.334		6.699	18.094	1.175	
	14	0.335	4.101	5.334		7.016	17.987	1.116	
	15	0.335	4.101	5.334		6.558	16.594	1.101	
	16	0.335	4.101	5.334		4.328	10.730	1.079	
	17	0.335	4.101	5.334		4.579	11.891	1.130	
	18	0.335	4.101	5.334		4.454	11.192	1.093	
	19	1.025	5.151	2.187		3.996	10.561	1.150	
	20	1.258	5.639	1.950		3.538	9.929	1.221	
	21	3.488	10.672	1.331		3.080	9.298	1.313	
	22	3.689	11.567	1.364		2.622	8.666	1.438	
	23	3.889	12.463	1.394					
	24	4.203	13.369	1.384					
	25	4.410	13.472	1.329					
	26	4.606	13.524	1.278					
	27	4.801	13.576	1.230					
	28	4.997	13.628	1.187					
	29	4.502	12.631	1.221					
	30	4.464	12.194	1.189					
	31	2.429	7.214	1.292					
	32	2.424	6.370	1.143					
	33	2.420	5.527	0.994					
	34	2.300	4.672	0.884					
	35	2.289	4.622	0.879					
	36	2.289	4.622	0.879					
	37	2.215	4.411	0.867					
	38	2.136	4.188	0.853					
	39	2.057	3.964	0.839					
	40	1.978	3.740	0.823					
	41	1.899	3.516	0.806					
	42	1.951	5.026	1.121					
	43	2.160	6.364	1.282					
	44	2.368	6.840	1.257					
	45	2.165	6.570	1.321					
	46	1.936	6.254	1.406					
	47	1.781	6.149	1.502					
	48	1.631	6.057	1.616					
	49	1.481	5.965	1.752					

(Table S4). Contd…..

		$w = 5, s = 0.5$				$w = 10, s = 1$			
Patient	**Order**[a]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**[b]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**
	50	1.331	5.873	1.920					
	51	1.181	5.781	2.129					
	52	0.900	3.956	1.913					
	53	0.463	2.302	2.165					
	54	0.025	1.511	25.783					
2	1	0.508	6.020	5.152		1.563	16.045	4.464	
	2	0.509	6.564	5.606		1.812	16.920	4.060	
	3	0.510	7.108	6.058		2.077	17.590	3.683	
	4	0.511	7.652	6.508		2.372	17.721	3.249	
	5	0.512	8.196	6.957		2.665	16.905	2.759	
	6	0.513	8.741	7.404		2.957	16.088	2.366	
	7	0.514	9.144	7.733		3.654	16.011	1.905	
	8	0.514	9.144	7.733		4.434	16.086	1.577	
	9	0.514	9.144	7.733		4.568	14.970	1.425	
	10	0.931	10.132	4.731		4.515	13.511	1.301	
	11	1.055	10.025	4.133		3.922	11.171	1.238	
	12	1.178	9.919	3.660		3.623	9.925	1.191	
	13	1.302	9.812	3.277		3.307	8.885	1.168	
	14	1.426	9.706	2.961		2.962	8.243	1.210	
	15	1.565	9.394	2.611		2.617	7.601	1.263	
	16	1.711	8.985	2.284		2.271	6.959	1.332	
	17	1.858	8.577	2.008		1.522	5.578	1.593	
	18	2.004	8.169	1.773		0.690	4.044	2.550	
	19	2.150	7.761	1.569		0.504	3.702	3.196	
	20	1.880	6.365	1.472		0.504	3.702	3.196	
	21	1.903	6.063	1.386		0.504	3.702	3.196	
	22	2.086	6.054	1.262					
	23	2.352	6.198	1.146					
	24	2.619	6.343	1.053					
	25	2.870	6.692	1.014					
	26	2.883	6.713	1.013					
	27	2.710	6.392	1.026					
	28	2.538	6.071	1.040					
	29	2.365	5.750	1.057					
	30	2.192	5.429	1.077					
	31	2.020	5.108	1.100					
	32	1.687	4.494	1.158					

(Table S4). Contd…..

		$w = 5, s = 0.5$				$w = 10, s = 1$			
Patient	**Order**[a]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**[b]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**
	33	1.270	3.727	1.276					
	34	0.854	2.960	1.507					
	35	0.438	2.193	2.179					
	36	0.252	1.851	3.196					
	37	0.252	1.851	3.196					
	38	0.252	1.851	3.196					
	39	0.252	1.851	3.196					
	40	0.252	1.851	3.196					
	41	0.252	1.851	3.196					
	42	0.252	1.851	3.196					
	43	0.252	1.851	3.196					
	44	0.252	1.851	3.196					
	45	0.252	1.851	3.196					
	46	0.252	1.851	3.196					
	47	0.252	1.851	3.196					
	48	0.252	1.851	3.196					
	49	0.252	1.851	3.196					
	50	0.252	1.851	3.196					
	51	0.252	1.851	3.196					
3	1	1.304	1.467	0.491		3.063	9.802	1.396	
	2	1.222	1.374	0.491		2.818	11.696	1.810	
	3	1.140	1.282	0.491		2.573	13.208	2.239	
	4	1.058	1.190	0.491		2.337	14.437	2.693	
	5	0.975	1.097	0.491		2.235	15.627	3.048	
	6	0.893	1.005	0.491		2.125	16.345	3.354	
	7	0.821	0.923	0.491		2.015	17.028	3.685	
	8	0.821	0.923	0.491		2.100	17.161	3.563	
	9	0.821	0.923	0.491		2.194	17.294	3.437	
	10	1.800	7.296	1.768		2.323	19.132	3.592	
	11	1.759	8.335	2.066		1.976	12.405	2.738	
	12	1.719	9.375	2.379		2.604	10.974	1.838	
	13	1.678	10.414	2.706		2.535	9.303	1.600	
	14	1.638	11.409	3.038		2.787	7.715	1.207	
	15	1.597	12.110	3.306		3.060	7.325	1.044	
	16	1.557	12.812	3.588					
	17	1.517	13.514	3.886					
	18	1.470	14.202	4.214	*				

(Table S4). Contd…..

		$w = 5, s = 0.5$				$w = 10, s = 1$			
Patient	**Order**[a]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**[b]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**
	19	1.415	14.704	4.533	*				
	20	0.381	8.690	9.952					
	21	0.366	8.009	9.539					
	22	0.351	7.329	9.093					
	23	0.337	6.614	8.563					
	24	0.416	5.686	5.967					
	25	0.503	5.051	4.379					
	26	0.590	4.416	3.262					
	27	0.678	3.781	2.433					
	28	0.772	3.159	1.785					
	29	0.908	4.429	2.126					
	30	1.259	4.412	1.528					
	31	1.610	4.396	1.191					
	32	1.960	4.380	0.974					
	33	2.267	4.360	0.839					
	34	2.147	4.311	0.876					
	35	2.032	4.253	0.913					
	36	2.071	4.094	0.862					
	37	2.109	3.935	0.814					
	38	2.147	4.221	0.857					
	39	2.151	2.897	0.587					
	40	2.123	3.364	0.691					
4	1	0.000	1.843	∞		0.000	3.685	∞	
	2	0.000	1.843	∞		0.000	3.685	∞	
	3	0.000	1.843	∞		0.000	3.685	∞	
	4	0.000	1.843	∞		0.000	3.685	∞	
	5	0.000	1.843	∞		0.000	3.960	∞	
	6	0.000	1.843	∞		0.000	5.439	∞	
	7	0.000	1.843	∞		0.000	6.757	∞	
	8	0.000	1.843	∞		0.000	8.052	∞	
	9	0.000	1.843	∞		0.000	9.347	∞	
	10	0.000	1.843	∞		0.000	10.641	∞	*
	11	0.000	1.843	∞		0.000	11.061	∞	*
	12	0.000	1.843	∞		0.000	11.352	∞	*
	13	0.000	1.843	∞		0.000	11.643	∞	*
	14	0.000	1.843	∞		0.000	11.934	∞	*
	15	0.000	1.843	∞		0.000	11.566	∞	*

(Table S4). Contd…..

		$w = 5, s = 0.5$				$w = 10, s = 1$			
Patient	**Order**[a]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**[b]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**
	16	0.000	1.843	∞		0.000	9.719	∞	*
	17	0.000	1.843	∞		0.000	8.032	∞	
	18	0.000	1.843	∞		0.000	6.369	∞	
	19	0.000	2.117	∞		0.000	4.705	∞	
	20	0.000	2.857	∞		0.000	3.216	∞	
	21	0.000	3.596	∞		0.000	2.628	∞	
	22	0.000	4.267	∞		0.000	2.168	∞	
	23	0.000	4.914	∞		0.000	1.708	∞	
	24	0.000	5.562	∞		0.000	1.249	∞	
	25	0.000	6.209	∞					
	26	0.000	6.857	∞					
	27	0.000	7.504	∞					
	28	0.000	8.151	∞					
	29	0.000	8.524	∞					
	30	0.000	8.059	∞					
	31	0.000	7.465	∞					
	32	0.000	6.940	∞					
	33	0.000	6.438	∞					
	34	0.000	5.936	∞					
	35	0.000	5.434	∞					
	36	0.000	4.932	∞					
	37	0.000	4.430	∞					
	38	0.000	3.873	∞					
	39	0.000	3.042	∞					
	40	0.000	2.583	∞					
	41	0.000	2.253	∞					
	42	0.000	1.924	∞					
	43	0.000	1.594	∞					
	44	0.000	1.264	∞					
	45	0.000	0.934	∞					
	46	0.000	0.604	∞					
	47	0.000	0.275	∞					
	48	0.000	0.074	∞					
	49	0.000	0.174	∞					
	50	0.000	0.274	∞					
	51	0.000	0.374	∞					
	52	0.000	0.474	∞					

(Table S4). Contd…..

		$w = 5, s = 0.5$				$w = 10, s = 1$			
Patient	**Order**[a]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**[b]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**
	53	0.000	0.574	∞					
	54	0.000	0.674	∞					
	55	0.000	0.774	∞					
	56	0.000	0.874	∞					
	57	0.000	0.974	∞					
5	1	0.365	2.919	3.421		1.930	14.771	3.272	
	2	0.365	2.919	3.421		2.038	15.874	3.330	
	3	0.365	2.919	3.421		2.146	16.976	3.382	
	4	0.365	2.919	3.421		2.254	18.079	3.429	
	5	0.365	2.919	3.421		2.362	18.579	3.363	
	6	1.295	9.095	3.002	*	2.470	19.032	3.294	
	7	1.349	9.646	3.056	*	2.578	19.484	3.231	
	8	1.403	10.198	3.107	*	2.686	19.937	3.173	
	9	1.457	10.749	3.153	*	1.810	13.662	3.228	
	10	1.511	11.300	3.197	*	1.810	13.012	3.074	
	11	1.565	11.852	3.237	*	1.810	12.362	2.921	
	12	1.619	12.403	3.275	*	3.121	31.413	4.303	*
	13	1.673	12.955	3.310	*	3.483	32.363	3.973	*
	14	1.727	13.506	3.343	*	3.721	32.995	3.792	*
	15	1.781	14.057	3.374	*	3.949	34.204	3.704	*
	16	0.905	8.433	3.985		4.176	35.462	3.630	*
	17	0.905	8.433	3.985		4.404	36.719	3.565	*
	18	0.905	8.156	3.854		4.632	37.976	3.505	*
	19	0.905	7.830	3.700		4.639	37.432	3.450	*
	20	0.905	7.505	3.547		4.628	36.608	3.383	*
	21	0.905	7.180	3.393		4.616	35.763	3.312	*
	22	0.905	6.855	3.239		3.294	15.216	1.975	
	23	0.905	6.530	3.086		2.921	12.771	1.870	
	24	0.905	6.205	2.932					
	25	0.905	5.880	2.778					
	26	0.905	5.555	2.625					
	27	0.905	5.229	2.471					
	28	0.905	5.182	2.449					
	29	0.905	5.182	2.449					
	30	0.905	5.182	2.449					
	31	0.905	5.182	2.449					
	32	2.035	24.083	5.059					

(Table S4). Contd…..

		$w = 5, s = 0.5$				$w = 10, s = 1$			
Patient	**Order**[a]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**[b]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**
	33	2.216	24.883	4.801					
	34	2.397	25.683	4.581					
	35	2.578	26.483	4.392					
	36	2.702	27.137	4.295					
	37	2.816	27.765	4.216					
	38	2.930	28.394	4.144					
	39	3.044	29.023	4.077					
	40	3.158	29.651	4.015					
	41	3.272	30.280	3.957					
	42	2.255	12.007	2.277					
	43	2.188	11.836	2.313					
	44	2.121	11.664	2.351					
	45	2.054	11.493	2.392					
	46	1.943	10.639	2.342					
	47	1.823	9.667	2.267					
	48	1.703	8.637	2.168					
	49	1.584	7.586	2.048					
	50	1.464	6.535	1.908					
	51	1.345	5.483	1.744					
	52	1.225	4.432	1.547					
	53	1.106	3.381	1.308					
	54	0.986	2.329	1.010					
	55	0.866	1.278	0.631					
1-5[c]	1	2.511	16.349	2.818		7.225	52.506	3.146	*
	2	2.430	16.801	2.993		7.337	56.377	3.326	*
	3	2.349	17.253	3.179		7.465	59.662	3.460	*
	4	2.268	17.705	3.379		7.632	62.125	3.524	*
	5	2.187	18.157	3.594		8.622	64.323	3.230	*
	6	3.037	24.784	3.533	*	11.376	71.676	2.728	*
	7	3.019	25.657	3.679	*	12.471	75.844	2.633	*
	8	3.073	26.208	3.692	*	13.966	78.809	2.443	*
	9	3.127	26.760	3.705	*	13.707	72.950	2.304	*
	10	4.577	34.672	3.280	*	14.174	74.078	2.262	*
	11	4.714	36.156	3.320	*	13.625	64.885	2.062	*
	12	4.851	37.640	3.359	*	15.656	81.654	2.258	*
	13	4.988	39.125	3.395	*	16.025	80.289	2.169	*
	14	5.125	40.564	3.426	*	16.485	78.875	2.071	*

(Table S4). Contd…..

		$w = 5, s = 0.5$				$w = 10, s = 1$			
Patient	**Order**[a]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**[b]	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection**
	15	5.278	41.505	3.404	*	16.183	77.291	2.068	*
	16	4.507	36.174	3.474	*	10.776	62.869	2.526	*
	17	4.613	36.468	3.422	*	10.506	62.219	2.564	*
	18	4.713	36.470	3.350	*	9.776	59.580	2.638	*
	19	5.495	37.563	2.959	*	9.138	56.400	2.672	*
	20	4.424	31.055	3.039	*	8.669	53.455	2.669	*
	21	6.662	35.520	2.308	*	8.200	51.390	2.713	*
	22	7.031	36.072	2.221	*	5.916	26.051	1.906	*
	23	7.482	36.720	2.124	*	2.921	14.480	2.146	
	24	8.143	37.165	1.976		0.000	1.249	∞	
	25	8.688	37.304	1.859					
	26	8.984	37.064	1.786					
	27	9.094	36.482	1.737					
	28	9.211	36.191	1.701					
	29	8.680	36.516	1.821					
	30	8.820	35.276	1.731					
	31	6.963	29.365	1.826					
	32	8.107	46.267	2.471	*				
	33	8.173	44.934	2.380	*				
	34	7.698	43.562	2.450	*				
	35	7.337	42.985	2.536	*				
	36	7.313	42.635	2.524	*				
	37	7.391	42.393	2.483	*				
	38	7.465	42.527	2.466	*				
	39	7.504	40.775	2.352	*				
	40	7.510	41.189	2.374	*				
	41	5.422	37.900	3.026	*				
	42	4.458	20.808	2.020					
	43	4.600	21.644	2.037					
	44	4.741	21.619	1.974					
	45	4.471	20.848	2.019					
	46	4.130	19.348	2.028					
	47	3.856	17.942	2.014					
	48	3.586	16.620	2.006					
	49	3.317	15.576	2.033					
	50	3.047	14.533	2.065					
	51	2.778	13.490	2.102					

(Table S4). Contd…..

		$w = 5, s = 0.5$				$w = 10, s = 1$			
Patient	**Order[a]**	$c_{S(W)}$	$c_{N(W)}$	r_N/r_S	**Selection[b]**	$c_{S(W)}$	$c_{N(W)}$	$r_{N(}r_S$	**Selection**
	52	2.125	8.862	1.805					
	53	1.568	6.257	1.727					
	54	1.011	4.515	1.932					
	55	0.866	2.052	1.026					
	56	0.000	0.874	∞					
	57	0.000	0.974	∞					

[a]Phylogenetic windows are ordered chronologically.
[b]Positively selected windows are indicated with black and red asterisks for $w = 5$ and $w = 10$, respectively.
[c]The $c_{S(W)}$ and $c_{N(W)}$ values for the phylogenetic windows of the same chronological order were summed for patients 1-5.