Observation of positive selection within hypervariable regions of a newly identified DNA virus (SEN virus)[1]

Takeji Umemura[a,b], Yasuhito Tanaka[a], Kendo Kiyosawa[b], Harvey J. Alter[a], J. Wai-Kuo Shih[a,*]

[a] *Department of Transfusion Medicine, Warren Grant Magnuson Clinical Center, National Institutes of Health, Bethesda, MD 20892, USA*
[b] *Second Department of Internal Medicine, Shinshu University School of Medicine, 3-1-1 Asahi, Matsumoto 390-8621, Japan*

Received 21 June 2001; revised 29 November 2001; accepted 29 November 2001

First published online 11 December 2001

Edited by Takashi Gojobori

Abstract To elucidate the evolution of SEN virus (SEN-V), serial sequences of chronically SEN-V-infected patients were analyzed. In the hypervariable regions, non-synonymous substitutions significantly predominated. This could be attributed to positive selection in evading immune surveillance of the hosts and to establish a persistent infection. On the basis of the sequences in the two open reading frames of SEN-V DNA, the rate of synonymous substitutions was 7.32×10^{-4} per site per year. Since this rate is close to RNA viruses and higher than other DNA viruses, the SEN-V might be replicated by machinery with poor or no proofreading function. © 2002 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

Key words: Evolutionary rate; Hypervariable region; Phylogenetic analysis; Positive selection; TT virus; *Circoviridae*

1. Introduction

In 1999, a novel DNA virus designated SEN virus (SEN-V) was discovered in the serum of a human immunodeficiency virus type 1 (HIV-1)-infected injection drug user and named after the initials of this patient [1]. SEN-V is a single-stranded, circular DNA virus of approximately 3800 nucleotides (nt), possibly belonging to the *Circoviridae* family [2] that includes TT virus (TTV) [3], TUS01 [4], SANBAN [5] and YONBAN [6]. Recently, Ukita et al. determined the full sequence of TTV isolates of genotype 12 (TJN01) and genotype 13 (TJN02) from healthy individuals in Japan [7]. Phylogenetic analysis confirmed that SENV-D and TJN01, SANBAN and TJN02, respectively, were the same genotypes [2]. A strong association between two SEN-V variants (SENV-D and SENV-H) infections and transfusion-associated non-A to E hepatitis has been reported [8]. Additionally, 2% of USA blood donors [8] and 20% of healthy Japanese individuals [9] were infected with SENV-D and/or SENV-H. SEN-V transmission can occur by blood transfusion and the relatively high prevalence of SEN-V among drug addicts, and patients with hepatitis C virus (HCV) suggests that this is an important transmission route [8,10,11]. It is not known, however, whether SEN-V causes disease in humans.

Recent studies indicate that the TTV and SEN-V genomes are heterogeneous both among infected individuals and within the same individuals [7,8,11,12]. Nishizawa et al. suggest that the TTV family has three hypervariable regions (HVRs) and the quasispecies nature of circulating TTV may help it escape immune surveillance and establish persistent infection [7,12]. This adaptability of the HVR for viral persistence has been well established for HCV [13,14] and HIV-1 [15].

In the present study we compared the nucleotide sequences of the SENV-D genome obtained from serum samples of a chronically infected patient over an interval of 12 years. By estimating the ratio of synonymous to non-synonymous substitutions in the HVRs, we found that a positive selection was taking place in the HVRs of SENV-D. The rate of synonymous substitutions for open reading frame (ORF)1 and ORF2 of SENV-D was estimated to be 7.32×10^{-4} per site per year on the basis of the regression analysis of phylogenetic branch lengths using SENV-D as an outgroup. The observed mutation rate is closer to that of RNA viruses, such as HCV, than to typical DNA viruses.

2. Materials and methods

2.1. Patients and serum samples

We used sequential serum samples obtained from a patient that was positive for SENV-D DNA by polymerase chain reaction (PCR) as reported previously [8]. The patient (RP) with chronic posttransfusion hepatitis C was persistently infected with SENV-D for 12 years. Three available sera were used for determining the sequence of SEN-V genome: (1) November 30, 1978 at the initial acute phase of the infection; (2) November 22, 1987, 9 years after transfusion; (3) September 19, 1990, 12 years after transfusion. They were designated SENV-D/RP78, SENV-D/RP87, SENV-D/RP90, respectively. Another patient (AB) with posttransfusion non-A to E hepatitis associated with chronic SENV-D infection was assessed. The sequence of SENV-D was determined using two available sera: (1) December 1, 1978 at the initial acute phase of the infection; (2) August 20, 1982, 3.7 years after transfusion. Those sequences that were obtained from patient AB were designated as SENV-D/AB78, SENV-D/AB82, respectively. Those patients did not receive any blood transfusion during the in-

*Corresponding author. Fax: (1)-301-402 1360.
E-mail address: jshih@mail.cc.nih.gov (J.W.-K. Shih).

[1] The sequences reported in this paper have been deposited in the DDBJ, EMBL and GenBank databases (accession numbers AB062295–AB062297, AB072965–AB072969).

Abbreviations: SEN-V, SEN virus; HIV-1, human immunodeficiency virus type 1; TTV, TT virus; HCV, hepatitis C virus; HVR, hypervariable region; ORF, open reading frame; PCR, polymerase chain reaction; d_S, number of synonymous substitutions per potential synonymous site; d_N, number of non-synonymous substitutions per potential non-synonymous site; HBV, hepatitis B virus

0014-5793/02/$22.00 © 2002 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.
PII: S0014-5793(01)03258-6

terval between serum samplings. Samples were stored at −80°C before testing for SEN-V DNA.

2.2. PCR amplification of SEN-V and determination of the nucleotide sequences

Nucleic acids were extracted from 100 μl of serum by QIAamp DNA Blood Mini Kit (Qiagen, Valencia, CA, USA). SENV-D DNA was amplified by PCR in the presence of TaKaRa Ex Taq (Panvera, Madison, WI, USA). PCR amplification of the partial ORF1 sequences was performed for 40 cycles (94°C for 30 s (plus 2 min in the first cycle); 55°C for 60 s; 68°C for 60 s) with the sense primer D1224S (5′-ACGTATCCCAGAGGCAGAAAAT-3′) and the antisense primer D2437AS (5′-ACTTGTACGTCCCGCTTGAATC-3′). In samples of patient RP, PCR amplification of the ORF2 was performed for 40 cycles (94°C for 30 s (plus 2 min in the first cycle); 60°C for 60 s; 68°C for 60 s) with the sense primer D177S (5′-CCGGAGGTGAGTTTACACACCG-3′) and the antisense primer D723AS (5′-CTACCGCGGCGCCTCCTTACTGT-3′).

The size of amplified DNA fragment was confirmed by 1.5% agarose gel electrophoresis followed by ethidium bromide staining. PCR products were purified using the QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA) and directly sequenced by the BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Foster City, CA, USA) using an ABI 310 DNA Sequencer. Total sequences (1578 nt) including ORF1 (1173 nt) and ORF2 (405 nt) were available for analysis.

2.3. Nucleotide sequence data and molecular evolutionary analysis

We collected the nucleotide sequence data of the partial ORF1 region (792 nt) of TJN01 for clones that were periodically isolated from one individual [7]. This individual was called N1 throughout this paper. In this study, we selected one major clone each obtained at three time points over 5.5 years. Those sequences that were obtained from N1 were designated as TJN01/N193, TJN01/N197, and TJN01/N198, respectively.

The DDBJ/EMBL/GenBank accession numbers of SEN-V isolates used in this study are as follows: SENV-D/RP78 (AB062295, AB072967), SENV-D/RP87 (AB062296, AB072968), SENV-D/RP90 (AB062297, AB072969), SENV-D/AB78 (AB072965), SENV-D/AB82 (AB072966), SENV-D (AX025730) [1], TJN01/N193 (AB028670) [7], TJN01/N197 (AB028678), and TJN01/N198 (AB028686). The sequences were aligned with CLUSTAL W (version 1.8). The number of synonymous substitutions per potential synonymous site (d_S) and the number of non-synonymous substitutions per potential non-synonymous site (d_N) for each set of sequences from the different time points were calculated by using the method of Nei and Gojobori [16] incorporating the Jukes–Cantor correction for multiple substitutions, as implemented in the MEGA program version 2.1 [17]. On the basis of the genetic distances, a phylogenetic tree was constructed by the neighbor-joining method [18] and a regression analysis was conducted to estimate the rate of synonymous substitutions in the ORF1 (1173 nt) and ORF2 (405 nt) of SENV-D. Neutrality of each individual was estimated by codon-based tests of selection with 1000 times bootstrap resampling tests in the MEGA program version 2.1 [17,19].

3. Results

To examine whether positive selection is operating on the HVRs, the ratio of overall mean synonymous to non-synonymous substitutions (d_S/d_N) in the HVRs were evaluated in three patients (Table 1). The d_S/d_N ratios lower than 1.0 were observed in samples collected at all time points in RP (0.177; $P=0.002$), AB (0.000; $P=0.001$) and N1 (0.000; $P=0.042$). These data indicated that positive selection was taking place in HVRs of SEN-V within single hosts.

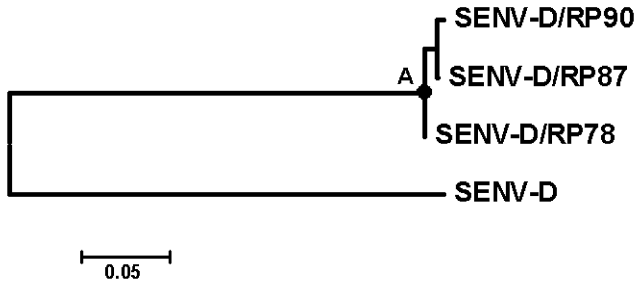

Fig. 1. Phylogenetic tree of the SEN-V sequences from a patient chronically infected SENV-D constructed by neighbor-joining method on the basis of synonymous substitutions per synonymous site of available nucleotide sequences (1578 nt) in ORF1 and ORF2 using SENV-D (AX025730) as the reference sequence. The horizontal bar indicates the substitutions per site and the circle labeled 'A' indicates the ancestor for SENV-D/RP.

A phylogenetic tree of the SEN-V sequences from a patient (RP) with chronically infected SENV-D was constructed using the number of synonymous substitutions per synonymous site of available nucleotide sequences (1578 nt) in ORF1 and ORF2 (Fig. 1). To determine a root of the phylogenetic tree, SENV-D was used as an outgroup. The phylogenetic tree based on the number of synonymous substitutions indicated that the serial sequences formed one cluster and the branch lengths generally increased according to the duration after infection. The branch length of each patient RP isolate relative to a putative common ancestor was determined. These values were plotted against the year the sample was collected, and the resulting relationship was evaluated by regression analysis (Fig. 2). The slope reflected the unit increase in substitutions over time and represented the rate of synonymous substitutions as 7.32×10^{-4} per site per year. All solid circles were well within the 95% confidence limits.

4. Discussion

In the present study, the ratios of d_S/d_N using the sequence data for the HVRs were estimated for SENV-D periodically sampled in three patients. The d_S/d_N ratios were significantly lower than 1.0 in single hosts with SENV-D infection (Table 1). It is probable that the high non-synonymous mutation rate reflects immune pressure exerted as the HVR resulting in the positive selection of persistent variants. Similar immune selection has been found to take place in the V3 loop of gp120 in HIV-1 [20], the *tax* gene of human T-cell leukemia virus-1 [21], and the HVR in the E2 region of HCV [22,23].

The rate of synonymous substitution for the ORF1 and

Table 1
Average synonymous and non-synonymous nucleotide substitutions within different individuals

Sample	Average (S.E.M.)[a]		d_S/d_N	P value[b]
	d_S	d_N		
RP	0.009 (0.007)	0.053 (0.015)	0.177	0.002
AB	0.000 (0.000)	0.063 (0.019)	0.000	0.001
N1	0.000 (0.000)	0.014 (0.007)	0.000	0.042

[a]Calculated by the method of Nei and Gojobori [16] implemented in MEGA [17].
[b]Estimated by codon-based tests of selection with 1000 times bootstrap resampling tests [17,19].

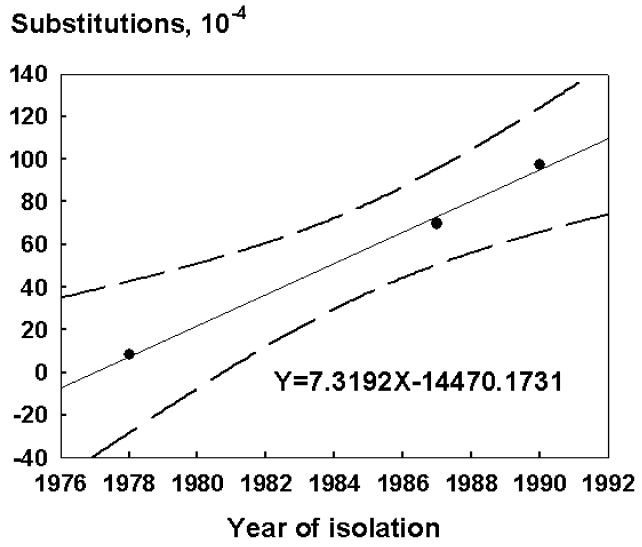

Fig. 2. A regression analysis of phylogenetic branch lengths to determine the rate of synonymous mutations in the ORF1 and ORF2. The branch lengths (solid circle) of the three serial sequences relative to the ancestor A were plotted against time to estimate the rate of accumulation of mutations for patient RP. The solid line represents the linear relationship based on the analysis. The dashed lines indicate the 95% confidence limits of the linear regression line. The synonymous substitution rate of SENV-D was determined as 7.32×10^{-4} per site per year.

ORF2 (1578 nt) of SENV-D (patient RP) was estimated to be 7.32×10^{-4} per site per year (Fig. 2). Interestingly, this value is close to those obtained for RNA viruses that use a polymerase that lacks proofreading function for replication of the virus genome: $3.92–13.08\times10^{-3}$ for HIV-1 [24], $3.59–13.10\times10^{-3}$ for influenza A virus and $0.22–7.51\times10^{-3}$ for HCV [25]. Although hepatitis B virus (HBV) is a DNA virus, it is known that HBV possesses a reverse transcriptase and that it replicates through that enzyme. The rate of substitution for SEN-V is approximately 10-fold higher than that of HBV (4.57×10^{-5}) [26]. Other DNA viruses mutate less frequently; the rates of nucleotide substitution of herpes simplex virus type 1 and human polyomavirus JC were reported to be 3.5×10^{-8} [27] and 4×10^{-7} [28], respectively. Thus, the synonymous substitution rate of SEN-V is close to that of RNA viruses and much higher than that of classic DNA viruses. Recently, the rapid mutation frequencies of maize streak virus, a plant virus with single-stranded circular DNA were reported [29]. The mutation frequencies of this virus, as determined by the number of mutations relative to the consensus divided by the number of nucleotides sequenced, were $3.8–10.5\times10^{-4}$. As previously reported, the mutation rate per base in DNA-based microbes is inversely proportional to genome size. This is consistent with the rapid substitution rate of SEN-V since the *Circoviridae* are among the smallest DNA viruses [30]. More recently, Wilson et al. have reported that some of SEN-V DNA sequences obtained from paired serum samples with a median of 9.3 year intervals were clustered in each group; some subjects had persistent infection of SEN-V [11]. The estimated substitution rates of SENV-D or -H using the clustered paired sequences are probably more than 10^{-4} per site per year in each subject (data not shown). This rapid evolution suggests that SEN-V is replicated by machinery with poor or no proofreading activity via a host DNA polymerase complex, as evidenced by the lack of an identifiable polymerase gene within the SEN-V genome. Alternatively, it is possible that SEN-V is composed of functionally less important genes, such as pseudogenes, that evolve faster than genes that are more functionally relevant [31].

In summary, we have presented that positive selection could operate in the HVRs of SEN-V within single hosts as a mechanism to evade immune surveillance of the host. Our results also suggest that the evidence for the rapid substitution rate of the SEN-V in a chronically infected patient followed for 12 years.

References

[1] Primi, D., Fiordalisi, G., Mantero, J.L., Mattioli, S., Sottini, A., Bonelli, F., Vaglini, L., Olivero, P., Dal Corso, A. and Bonelli, M. (2000) International patent number WO0028039 (http://ep.espacenet.com/).
[2] Tanaka, Y., Primi, D., Wang, R.Y., Umemura, T., Yeo, A.E., Mizokami, M., Alter, H.J. and Shih, J.W. (2001) J. Infect. Dis. 183, 359–367.
[3] Nishizawa, T., Okamoto, H., Konisi, K., Yoshizawa, H., Miyakawa, Y. and Mayumi, M. (1997) Biochem. Biophys. Res. Commun. 241, 92–97.
[4] Okamoto, H., Takahashi, M., Nishizawa, T., Ukita, M., Fukuda, M., Tsuda, F., Miyakawa, Y. and Mayumi, M. (1999) Virology 259, 428–436.
[5] Hijikata, M., Takahashi, K. and Mishiro, S. (1999) Virology 260, 17–22.
[6] Takahashi, K., Hijikata, M., Samokhvalov, E.I. and Mishiro, S. (2000) Intervirology 43, 119–123.
[7] Ukita, M., Okamoto, H., Nishizawa, T., Tawara, A., Takahashi, M., Iizuka, H., Miyakawa, Y. and Mayumi, M. (2000) Arch. Virol. 145, 1543–1559.
[8] Umemura, T., Yeo, A.E.T., Sottini, A., Moratto, D., Tanaka, Y., Wang, R.Y.H., Shih, J.W.K., Donahue, P., Primi, D. and Alter, H.J. (2001) Hepatology 33, 1303–1311.
[9] Umemura, T., Yeo, A.E.T., Shih, J.W.K., Matsumoto, A., Orii, K., Tanaka, E., Kiyosawa, K., Primi, D. and Alter, H.J. (2000) Hepatology 32, 381A.
[10] Umemura, T., Alter, H.J., Tanaka, E., Yeo, A.E.T., Shih, J.W-K., Orii, K., Matsumoto, A., Yoshizawa, K. and Kiyosawa, K. (2001) J. Infect. Dis. 184, 1246–1251.
[11] Wilson, L.E., Umemura, T., Astemborski, J., Ray, S.C., Alter, H.J., Strathdee, S., Vlahov, D. and Thomas, D.L. (2001) J. Infect. Dis. 184, 1315–1319.
[12] Nishizawa, T., Okamoto, H., Tsuda, F., Aikawa, T., Sugai, Y., Konishi, K., Akahane, Y., Ukita, M., Tanaka, T., Miyakawa, Y. and Mayumi, M. (1999) J. Virol. 73, 9604–9608.
[13] Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S. and Shimotohno, K. (1991) Biochem. Biophys. Res. Commun. 175, 220–228.
[14] Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J.C., Strazzera, A., Chien, D.Y., Munoz, S.J., Balestrieri, A., Purcell, R.H. and Alter, H.J. (2000) Science 288, 339–344.
[15] Meyerhans, A., Cheynier, R., Albert, J., Seth, M., Kwok, S., Sninsky, J., Morfeldt-Manson, L., Asjo, B. and Wain-Hobson, S. (1989) Cell 58, 901–910.
[16] Nei, M. and Gojobori, T. (1986) Mol. Biol. Evol. 3, 418–426.
[17] Kumar, S., Tamura, K. and Nei, M. (1993) Comput. Appl. Biosci. 10, 189–191.
[18] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 18, 414–423.
[19] Felsenstein, J. (1985) Evolution 39, 783–791.
[20] Yamaguchi, Y. and Gojobori, T. (1997) Proc. Natl. Acad. Sci. USA 94, 11264–11269.
[21] Niewiesk, S. and Bangham, C.R.M. (1996) J. Mol. Evol. 42, 452–458.
[22] Smith, D.B. and Simmonds, P. (1997) J. Mol. Evol. 45, 238–246.
[23] McAllister, J., Casino, C., Davidson, F., Power, J., Lawlor, E., Yap, P.L., Simmonds, P. and Smith, D.B. (1998) J. Virol. 72, 4893–4905.

[24] Gojobori, T., Moriyama, E.N. and Kimura, M. (1990) Proc. Natl. Acad. Sci. USA 87, 10015–10018.
[25] Ina, Y., Mizokami, M., Ohba, K. and Gojobori, T. (1994) J. Mol. Evol. 38, 50–56.
[26] Orito, E., Mizokami, M., Ina, Y., Moriyama, E.N., Kaneshima, N., Yamamoto, M. and Gojobori, T. (1989) Proc. Natl. Acad. Sci. USA 86, 7059–7062.
[27] Sakaoka, H., Kurita, K., Iida, Y., Takada, S., Umene, K., Kim, Y.T., Ren, C.S. and Nahmias, A.J. (1994) J. Gen. Virol. 75, 513–527.
[28] Hatwell, J.N. and Sharp, P.M. (2000) J. Gen. Virol. 81, 1191–1200.
[29] Isnard, M., Granier, M., Frutos, R., Reynaud, B. and Peterschmitt, M. (1998) J. Gen. Virol. 79, 3091–3099.
[30] Drake, J.W. (1991) Proc. Natl. Acad. Sci. USA 88, 7160–7164.
[31] Li, W-H., Gojobori, T. and Nei, M. (1981) Nature 292, 237–239.