

Slow Evolutionary Rate of GB Virus C/Hepatitis G Virus

Yoshiyuki Suzuki,¹ Kazuhiko Katayama,² Shuetsu Fukushi,² Tsutomu Kageyama,² Akira Oya,² Hirofumi Okamura,³ Yasuhito Tanaka,⁴ Masashi Mizokami,⁴ Takashi Gojobori¹

¹ Center for Information Biology, National Institute of Genetics, Mishima, Japan

² Basic Research Division, BioMedical Laboratories, Inc., Saitama, Japan

³ Third Department of Internal Medicine, Kyorin University School of Medicine, Tokyo, Japan

⁴ Second Department of Medicine, Nagoya City University Medical School, Nagoya, Japan

Received: 2 June 1998 / Accepted: 8 August 1998

Abstract. With the aim of elucidating evolutionary features of GB virus C/hepatitis G virus (GBV-C/HGV), molecular evolutionary analyses were conducted using the entire coding region of this virus. In particular, the rate of nucleotide substitution for this virus was estimated to be less than 9.0×10^{-6} per site per year, which was much slower than those for other RNA viruses. The phylogenetic tree reconstructed for GBV-C/HGV, by using GB virus A (GBV-A) as outgroup, indicated that there were three major clusters (the HG, GB, and Asian types) in GBV-C/HGV, and the divergence between the ancestor of GB- and Asian-type strains and that of HG-type strains first took place more than 7000–10,000 years ago. The slow evolutionary rate for GBV-C/HGV suggested that this virus cannot escape from the immune response of the host by means of producing escape mutants, implying that it may have evolved other systems for persistent infection.

Key words: GBV-C/HGV — GBV-A — Phylogenetic tree — Substitution rate — Divergence time

Introduction

GB virus C/hepatitis G virus (GBV-C/HGV) was discovered as a putative agent of non-A-E hepatitis (Simons et

al. 1995a; Linnen et al. 1996), although disease association of this virus remains to be clarified. The genome of this virus is a positive-stranded RNA, in which nine genes (E1, E2, p7, NS2, NS3, NS4a, NS4b, NS5a, and NS5b) are encoded as a single long open reading frame (Erker et al. 1996). The genomic organization and sequence of GBV-C/HGV suggested that it was a member of the family *Flaviviridae*, though it seemed to lack the nucleocapsid (core) protein (Simons et al. 1995a; Leary et al. 1996; Linnen et al. 1996).

The phylogenetic analyses for GBV-C/HGV have shown that there were three major clusters in this virus, and they were named the HG, GB, and Asian types (Mukaide et al. 1997). In addition, the phylogenetic analysis for viruses belonging to the family *Flaviviridae* suggested that GB virus A (GBV-A) was the most closely related virus to GBV-C/HGV (Zuckerman 1996).

From the evolutionary point of view, it is of importance to estimate the evolutionary rate of GBV-C/HGV, particularly for elucidating the evolutionary origin and history of this virus. The analysis of the sequence variability for GBV-C/HGV isolated from various locations in the world indicated that the genomic sequence of this virus was highly conserved compared with that of hepatitis C virus (HCV) (Okamoto et al. 1997), suggesting that the evolutionary rate for GBV-C/HGV might be slower than for HCV.

Masuko et al. (1996) and Nakao et al. (1997) estimated the rate of nucleotide substitution for this virus by dividing the proportion of nucleotide difference between two sequences obtained from single patient, by the dif-

ference in their sampling times. Then the rate was estimated to be $(0.8-1.9) \times 10^{-3}$ (Masuko et al. 1996) and 3.9×10^{-4} (Nakao et al. 1997) per site per year, indicating that GBV-C/HGV evolved with an extremely high rate. In fact, it was a rate similar to or slightly slower than that of HCV ($(0.22-7.51) \times 10^{-3}$ (Ina et al. 1994)).

However, we have recently reported that GBV-C/HGV may have originated from Africa and was transmitted along with human migrations, which began about 100,000 years ago, by a phylogenetic analysis of nucleotide sequences for the NS3 and NS5a regions (Tanaka et al. 1998). Although we did not mention the evolutionary rate of GBV-C/HGV in that report, we noted that the rate of nucleotide substitution should be of the order of 10^{-6} to 10^{-7} per site per year, with the assumption that GBV-C/HGV diverged 100,000 years ago. Thus, the rate of nucleotide substitution for GBV-C/HGV might be much slower than the value obtained in the above estimation.

For obtaining the correct rate of nucleotide substitution for GBV-C/HGV, the estimation by Masuko et al. (1996) and Nakao et al. (1997) had two serious problems in their methodology. First, they did not make correction for multiple substitutions in the comparison of nucleotide sequences. However, this might have a small effect on the estimation, because the sequence divergence between nucleotide sequences compared was relatively small (Nei 1987). The second problem was much more severe. In their estimation, it was implicitly assumed that the virus from the earlier serum sample was the direct ancestor of the virus from the later sample. However, this assumption should not always hold, particularly when polymorphism had already existed in the viral sequence of the earlier serum sample. If this was the case, the assumption may result in overestimation of the substitution rate, because sequences compared may have diverged before the sampling time of the earlier serum.

In this study, the rate of nucleotide substitution for GBV-C/HGV was estimated by reconstructing phylogenetic trees, avoiding the above two problems, with the entire coding region of this virus. The results obtained supported our idea of a slow evolutionary rate for GBV-C/HGV. Moreover, a phylogenetic analysis of GBV-C/HGV using GBV-A as outgroup was conducted to investigate the evolutionary history of GBV-C/HGV.

Materials and Methods

Sequence Data

The sequence data of the entire coding region for GBV-C/HGV and GBV-A were collected from the international DNA databanks (DDBJ/EMBL/GenBank) with accession numbers AB003288-AB003293 (Takahashi et al. 1997), AB008342, AF006500, D87255 (Shao et al. 1996), D87262, D87263 (Nakao et al. 1997), D90600, D90601 (Okamoto et al. 1997), U36380 (Leary et al. 1996), U44402, U45966 (Lininen et al. 1996), U63715 (Erker et al. 1996), U75356, AB008335,

AB008336, and D87708-D87714 (Katayama et al. 1998) for GBV-C/HGV and U22303 (Simons et al. 1995b) and U94421 (Leary et al. 1997) for GBV-A.

The data for GBV-C/HGV included two pairs of sequences which were obtained from two patients at different times. These patients were called patients A and B throughout this paper. From patient A, D87714 and AB008335 were obtained, where D87714 was sampled 4.9 years earlier than AB008335 (Katayama et al. 1998). D87262 and D87263 were obtained from patient B, where D87262 was sampled 8.4 years earlier than D87263 (Nakao et al. 1997).

The route of viral transmission for patient A was completely unknown, because patient A had never received blood transfusions (Katayama et al. 1998). Patient B was considered to be infected through blood transfusions, according to the medical history (Nakao et al. 1997). These patients did not receive any blood transfusions during the interval period of serum samplings.

Data Analysis

Nucleotide sequences were aligned with each other, using the computer program CLUSTAL W (Thompson et al. 1994). The evolutionary distance for the entire coding region between different GBV-C/HGV isolates was estimated by the one-parameter method (Jukes and Cantor 1969) and the method of Nei and Gojobori (1986), in order to estimate the variance of branch lengths in the phylogenetic tree by the method of Nei and Jin (1989). Note that the entire coding region consisted, in total, of 8340 nucleotide sites excluding gaps. The phylogenetic tree was reconstructed by the neighbor-joining method (Saitou and Nei 1987) with 10,000 times of bootstrap resampling (Felsenstein 1985).

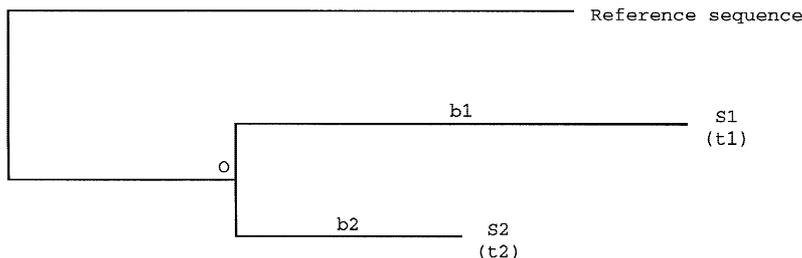
To estimate the rate of nucleotide substitution for GBV-C/HGV, the reference sequence was taken into account, in addition to the sequences which were derived from a single host. Let us designate the two sequences derived from a single host as S1 and S2, their sampling times as t_1 and t_2 , their last common ancestor as O, and the branch lengths from S1 to O and S2 to O as b_1 and b_2 , respectively (Fig. 1). The rate was calculated using $(b_1 - b_2)/(t_1 - t_2)$ (Li et al. 1988). The method of Nei and Jin (1989) was used for estimating the variance of branch lengths, which was then used for estimating the variance of rates.

The evolutionary origin and history of GBV-C/HGV was investigated by reconstructing a phylogenetic tree for GBV-C/HGV by using GBV-A as outgroup. The entire coding regions for GBV-C/HGV and GBV-A, which consisted of a total of 6567 nucleotide sites excluding gaps, were used for this purpose.

Results

The phylogenetic tree reconstructed for the entire coding region of GBV-C/HGV indicated that there were three major clusters in GBV-C/HGV: they were the HG, GB, and Asian types, as proposed by Mukaide et al. (1997) (Fig. 2). The geographical region where these strains were obtained was biased; namely, 21 out of 27 sequences were derived from Japan. However, the 27 sequences included all genotypes which have been reported all over the world. Therefore, these sequences were considered to be representatives of the GBV-C/HGV sequences disseminated worldwide.

When we focused our attention on the sequences that were obtained from a single patient to estimate the rate of nucleotide substitution, it was found that the branch length from D87714 to the common ancestor was longer than that from AB008335 (Fig. 2). Since the serum for



$$\text{Rate of nucleotide substitution} \quad \left(\frac{\text{site}}{\text{year}} \right) = \frac{b1 - b2}{t1 - t2}$$

Fig. 1. Method for estimating the rate of nucleotide substitution for GBV-C/HGV. The rate was estimated by dividing the difference in branch lengths from the sequences obtained from the single host to their common ancestor, by the difference in their sampling times.

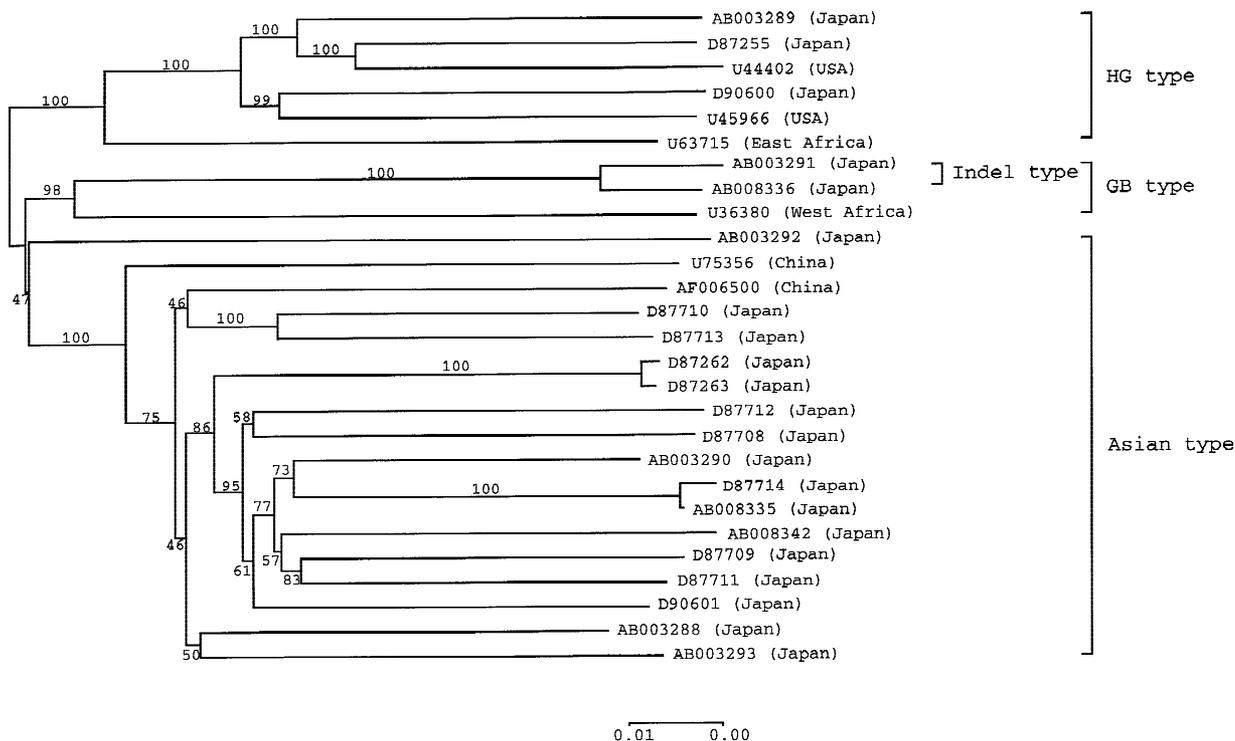


Fig. 2. The phylogenetic tree reconstructed for the entire coding region (8340 nucleotides) of GBV-C/HGV. The geographical origins of the isolates were indicated in parentheses. There were three major clusters (the HG, GB, and Asian types) in GBV-C/HGV, with se-

quences which had an extra 12 amino acids in the NS5a protein designated as Indel type. The number on each branch indicated the bootstrap probability for the clusters supported by that branch.

D87714 was sampled 4.9 years earlier than that for AB008335, the rate of nucleotide substitution was estimated as a negative value $(-7.1 \pm 1.5) \times 10^{-4}$ per site per year. The same situation was observed for D87262 and D87263, where D87262 was sampled 8.4 years earlier than D87263, with the rate of $(-5.7 \pm 7.7) \times 10^{-5}$. These results indicated the possibility that the ancestral sequences of AB008335 and D87263 have remained almost-unchanged during the total of 13.3 years.

It was possible that the negative values might be obtained from incorrect estimation of the branch length in the phylogenetic tree, which could be derived from the following reasons; incorrect topology of the tree; selective pressure disturbing the constancy of the rate; and

some peculiar genes with abnormal modes of evolution. To investigate whether these three possibilities actually took place, we conducted the following analyses.

First, we estimated the rate of nucleotide substitution for each patient adopting each of the other sequences as a reference sequence, in order to exclude the influence of the topology from estimation. For patient A, a negative value was obtained in all cases using 25 reference sequences. For patient B, however, four sequences, AB003291, AB008336, U36380, and U63715, supported a positive rate $((1.7-10.3) \times 10^{-5})$, but still at a much slower rate than calculated previously (Masuko et al. 1996; Nakao et al. 1997). It should be noted, however, that the sequences which were closely related to those

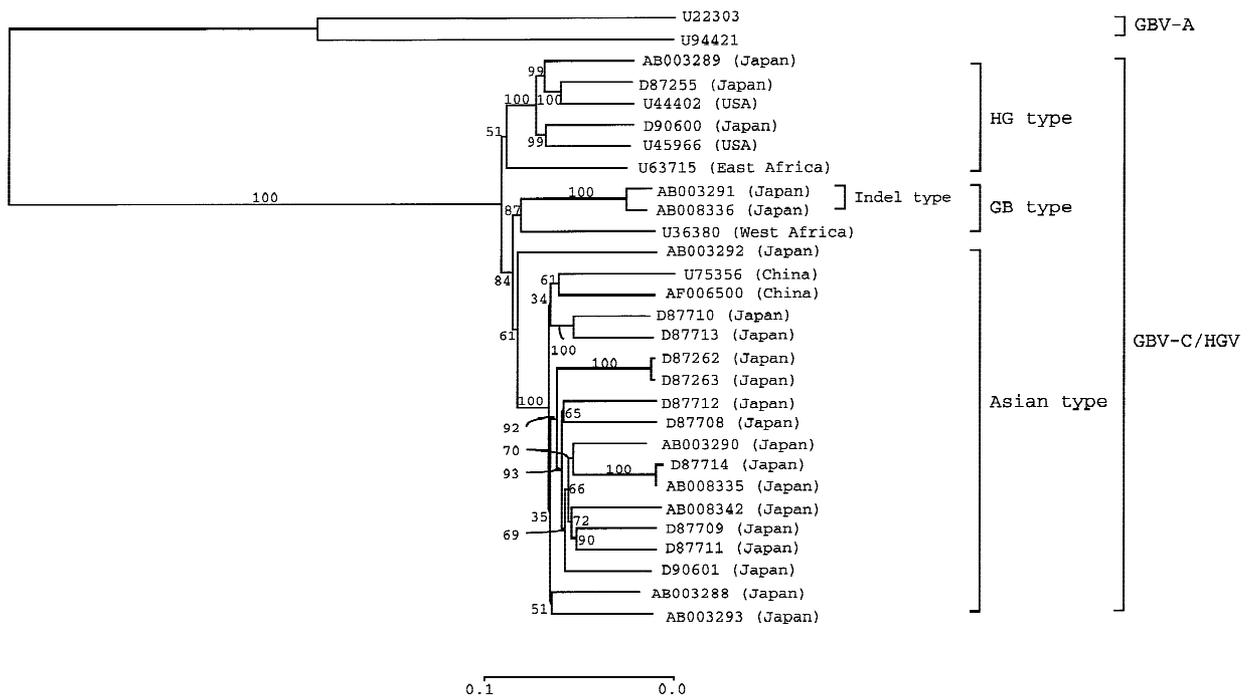


Fig. 3. The phylogenetic tree reconstructed for the entire coding region (6567 nucleotides) of GBV-C/HGV with GBV-A used as outgroup. See the legend to Fig. 2 for more information.

from patients A and B, namely, the sequences belonging to the Asian type (Fig. 2), all supported a negative rate. Thus, the negative rates estimated from Fig. 2 might not be artifacts due to the incorrect topology of the phylogenetic tree.

Second, we estimated the rates of synonymous and nonsynonymous substitutions for GBV-C/HGV. If selective pressure disturbed the constancy of the rate, the effect on the rate of nonsynonymous substitution should be stronger than that of synonymous substitution, because selection operates, in general, more severely on the amino acid sequence level. The rates of synonymous substitution for patients A and B were estimated to be $(-8.2 \pm 3.4) \times 10^{-4}$ and $(-3.7 \pm 2.7) \times 10^{-4}$, respectively, whereas those of nonsynonymous substitution were $(-6.6 \pm 1.6) \times 10^{-4}$ and $(0.6 \pm 4.3) \times 10^{-5}$, respectively. For both patients, the rate of synonymous substitution had larger absolute values of negative sign than that of nonsynonymous substitution, indicating the possibility that selective pressure was not the cause of negative values of the rate.

Third, we estimated the rate of nucleotide substitution for each gene, to investigate whether some genes had peculiar rates of nucleotide substitution. No gene supported a positive rate for patient A, whereas, for patient B, the sign depended upon the gene $(-14.2\text{--}26.8) \times 10^{-5}$). In the latter case, however, no statistically significant difference was observed in the rate between any pair of genes, indicating that the difference was possibly derived from statistical fluctuations.

Summarizing these results, it was concluded that the

ancestral sequences of AB008335 and D87263 have remained almost-unchanged in patients A and B, respectively. Therefore, it seemed impossible to estimate definitely the rate of nucleotide substitution for GBV-C/HGV from the presently available data. However, we could estimate the upper limit of the rate by the following manner (Orito et al. 1989). In principle, the rate of nucleotide substitution should be a positive value. If we assumed that only one nucleotide substitution took place in the entire coding region of GBV-C/HGV having 8340 nucleotides during the total of 13.3 years, the rate was estimated to be 9.0×10^{-6} per site per year ($1/8340/13.3$). In practice, however, no substitution was observed. Therefore, the rate of nucleotide substitution for GBV-C/HGV should be less than 9.0×10^{-6} per site per year.

To investigate the evolutionary history of GBV-C/HGV, a phylogenetic tree was reconstructed for GBV-C/HGV by using GBV-A as outgroup. Similarly to the phylogenetic tree that was reconstructed without GBV-A, the sequences of GBV-C/HGV were divided into three major clusters: the HG, GB, and Asian types (Fig. 3) (Mukaide et al. 1997). Moreover, the divergence between the ancestor of GB- and Asian-type strains and that of HG-type strains first took place (Fig. 3). That was supported by a reasonably high bootstrap probability (84%) for the branch, indicating the clustering of the GB and Asian types (Fig. 3). Assuming the rate of nucleotide substitution for GBV-C/HGV to be less than 9.0×10^{-6} per site per year, the divergence time of GBV-C/HGV was estimated to be more than 7000–10,000 years ago.

Table 1. Comparison of the rate of nucleotide substitution for GBV-C/HGV with that of other RNA viruses and mammals

Species	Rate (/site/year)	Reference
GBV-C/HGV	$<9.00 \times 10^{-6}$	This study
Hepatitis C virus	$(0.22-7.51) \times 10^{-3}$	Ina et al. (1994)
Hepatitis D virus	$(0.35-1.64) \times 10^{-3}$	Krushkal and Li (1995)
HIV-1 ^a	$(3.92-13.08) \times 10^{-3}$	Gojobori et al. (1990)
Influenza A virus	$(3.59-13.10) \times 10^{-3}$	Gojobori et al. (1990)
Mammals	$(0.56-3.94) \times 10^{-9}$	Li et al. (1985)

^a Human immunodeficiency virus type 1.

Discussion

It was concluded that the nucleotide sequence of GBV-C/HGV have remained almost-unchanged during a total of 13.3 years. It was highly possible that D87714 and AB008335, and D87262 and D87263, were derived from the constituents of polymorphism which had already existed at the sampling time of the earlier serum in patients A and B, respectively. Indeed, it was known, from the medical history, that patient B had received 4 U of blood transfusions before the sampling of the earlier serum, and the polymorphism was observed in that serum, including a minor mutation which became dominant 8.4 years later (Nakao et al. 1997).

In the present study, the rate of nucleotide substitution for GBV-C/HGV was estimated to be less than 9.0×10^{-6} per site per year. It was clear that the rates estimated previously (Masuko et al. 1996; Nakao et al. 1997) were overestimated by about 1000 times. This difference in the rate of nucleotide substitution might give us a totally different feature of evolutionary history for GBV-C/HGV. In our previous study, it was reported that GBV-C/HGV may have originated in Africa, and was transmitted along with human migrations which began about 100,000 years ago (Tanaka et al. 1998). The result in the present study was consistent with the above hypothesis, because the divergence time of GBV-C/HGV was estimated to be more than 7000–10,000 years ago. If this divergence time were true, GBV-C/HGV may be useful for clarifying the migration pattern of humans, as was the case for human T-cell lymphotropic virus types I and II (for review, see Suzuki and Gojobori 1998). This was because GBV-C/HGV can easily be transmitted vertically (Hino et al. 1998), and the parenteral transmission due to the blood product was the relatively recent event. The observation that the genotypes of GBV-C/HGV were strongly correlated with their geographic distribution also supported that idea (Muerhoff et al. 1996).

The rate of nucleotide substitution for GBV-C/HGV appeared to be much slower than for other RNA viruses (Table 1). Some possible mechanisms could be considered from this observation. First, the RNA-dependent RNA polymerase of GBV-C/HGV may have a higher

fidelity compared with those of other RNA viruses. Second, the strong functional constraint may be operating at the nucleotide sequence level of this virus. At any rate, such a slow rate indicates that GBV-C/HGV cannot infect the host persistently by means of producing escape mutants against immune responses of the host. This hypothesis is supported by the observation that patients positive for anti-E2 antibody and those positive for GBV-C/HGV RNA in their sera were well segregated (Pilot-Matias et al. 1996a). Moreover, it has been known that a hypervariable region was not detected in the genomic sequence of GBV-C/HGV (Erker et al. 1996) and that most individuals seropositive for GBV-C/HGV produced antibodies against only a single antigen (Pilot-Matias et al. 1996b). On the other hand, it has been reported that some patients infected with GBV-C/HGV did not appear to mount an immune response against this virus (Pilot-Matias et al. 1996b), so that GBV-C/HGV could establish persistent infection in humans (Linnen et al. 1996). The viral particle of GBV-C/HGV has been found to be covered with lipoproteins, and that was considered to be responsible for the lack of an immune response in the host (Sato et al. 1996). Therefore, GBV-C/HGV may have evolved its own system for persistent infection in the host.

It has been known that there was a polymorphism in the length of the NS5a protein for GBV-C/HGV, due to an indel of 12 amino acids (Takahashi et al. 1997). The isolates which had a longer NS5a protein were called Indel type, and the others were called non-Indel type (Tanaka et al. 1998). In our previous study, we proposed that the Indel type first diverged from other GBV-C/HGV, and the indel was derived from a deletion rather than an insertion (Tanaka et al. 1998). In the present study, however, the divergence between the ancestor of GB- and Asian-type strains and that of HG-type strains first took place, in which the indel was considered to be derived from an insertion, from the viewpoint of parsimonious principle (Fig. 3). This was consistent with the fact that a 12-amino acid sequence, which was similar to the extra 12 amino acids observed in the Indel-type strains, existed just downstream of the indel site in all GBV-C/HGV isolates (Tanaka et al. 1998). This was because, if we assumed that the indel was derived from an insertion, the above observation was also explained with smallest number of evolutionary steps. Therefore, it was speculated that the ancestral sequence of extant GBV-C/HGV strains may be the non-Indel type.

The discrepancy between our previous (Tanaka et al. 1998) and present results was probably derived from the difference in the alignment and sequence length used for the phylogenetic analysis. In fact, we excluded from the present analysis almost all of the NS5a region which was used previously, because we failed to make a reliable alignment for that region with the present sequence set.

Moreover, 6567 nucleotide sites were used for reconstructing the phylogenetic tree for GBV-C/HGV and GBV-A in the present study, while a much smaller number of sites (489 and 462 nucleotide sites in the NS3 and NS5a regions, respectively) was used in the previous study. It has been reported for GBV-C/HGV that, for phylogenetic analysis, a larger number of sites produced a more accurate result, and the entire coding region produced a more reliable phylogenetic tree than the regions which were used in our previous study (Smith et al. 1997). Thus, it seemed likely that the phylogenetic tree reconstructed in the present study was more reliable than the previous one.

In the present study, we could not estimate the rate of nucleotide substitution for GBV-C/HGV definitely, because of the short interval period for serum samplings compared with the slow evolutionary rate of this virus. Nevertheless, we successfully showed that the rate was slower than 9.0×10^{-6} per site per year, indicating that it was approximately 1000 times slower than the rate currently believed. However, to obtain an unambiguous result for the rate of nucleotide substitution for GBV-C/HGV, it was necessary to use more sequence data which were sampled with several decades of time interval.

Acknowledgments. We are grateful to Dr. Naoko Takezaki at the National Institute of Genetics, Japan, for providing computer programs in our phylogenetic analyses. We thank Dr. Silvana Gaudieri at the National Institute of Genetics, Japan, for valuable comments on the manuscript. This study was supported, in part, by grants from the Ministry of Education, Science, Sport, and Culture of Japan.

References

- Erker JC, Simons JN, Muerhoff S, Leary TP, Chalmers ML, Desai SM, Mushahwar IK (1996) Molecular cloning and characterization of a GB virus C isolate from a patient with non-A-E hepatitis. *J Gen Virol* 77:2713–2720
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791
- Gojobori T, Moriyama EN, Kimura M (1990) Molecular clock of viral evolution, and the neutral theory. *Proc Natl Acad Sci USA* 87: 10015–10018
- Hino K, Moriya T, Ohno N, Takahashi K, Hoshino H, Ishiyama N, Katayama K, Yoshizawa H, Mishiro S (1998) Mother-to-infant transmission occurs more frequently with GB virus C than hepatitis C virus. *Arch Virol* 143:65–72
- Ina Y, Mizokami M, Ohba K, Gojobori T (1994) Reduction of synonymous substitutions in the core protein gene of hepatitis C virus. *J Mol Evol* 38:50–56
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) *Mammalian protein metabolism*. Academic Press, New York, pp. 21–132
- Katayama K, Kageyama T, Fukushi S, Hoshino F, Kurihara C, Ishiyama N, Okamura H, Oya A (1998) Full-length GBV-C/HGV genomes from nine Japanese isolates: characterization by comparative analysis. *Arch Virol* 143:1063–1075
- Krushkal J, Li W-H (1995) Substitution rates in hepatitis delta virus. *J Mol Evol* 41:721–726
- Leary TP, Muerhoff AS, Simons JN, Pilot-Matias TJ, Erker JC, Chalmers ML, Schlauder GG, Dawson GJ, Desai SM, Mushahwar IK (1996) Sequence and genomic organization of GBV-C: A novel member of the *Flaviviridae* associated with human non-A-E hepatitis. *J Med Virol* 48:60–67
- Leary TP, Desai SM, Erker JC, Mushahwar IK (1997) The sequence and genomic organization of a GB virus A variant isolated from captive tamarins. *J Gen Virol* 78:2307–2313
- Li W-H, Luo C-C, Wu C-I (1985) Evolution of DNA sequences. In: MacIntyre RJ (ed) *Molecular evolutionary genetics*. Plenum Press, New York, pp. 1–94
- Li W-H, Tanimura M, Sharp PM (1988) Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol Biol Evol* 5:313–330
- Linnen J, Wages J, Zhang-Keck ZY, Fry KE, Krawczynski KZ, Alter H, Koonin E, Gallagher M, Alter M, Hadziyannis S, Karayiannis P, Fung K, Nakatsuji Y, Shih JWK, Young L, Piatak M, Hoover C, Fernandez J, Chen S, Zou JC, Morris T, Hyams KC, Ismay S, Lifson JD, Hess G, Fong SKH, Thomas H, Bradley D, Margolis H, Kim JP (1996) Molecular cloning and disease association of hepatitis G virus: A transfusion-transmissible agent. *Science* 271: 505–508
- Masuko K, Mitsui T, Iwano K, Yamazaki C, Okuda K, Meguro T, Murayama N, Inoue T, Tsuda F, Okamoto H, Miyakawa Y, Mayumi M (1996) Infection with hepatitis GB virus C in patients on maintenance hemodialysis. *N Engl J Med* 334:1485–1490
- Muerhoff AS, Simons JN, Leary TP, Erker JC, Chalmers ML, Pilot-Matias TJ, Dawson GJ, Desai SM, Mushahwar IK (1996) Sequence heterogeneity within the 5'-terminal region of the hepatitis GB virus C genome and evidence for genotypes. *J Hepatol* 25:379–384
- Mukaide M, Mizokami M, Orito E, Ohba K, Nakano T, Ueda R, Hikiji K, Iino S, Shapiro S, Lahat N, Park YM, Kim BS, Oyunsuren T, Rezieg M, Al-Ahdal MN, Lau JYN (1997) Three different GB virus C/hepatitis G virus genotypes: phylogenetic analysis and a genotyping assay based on restriction fragment length polymorphism. *FEBS Lett* 407:51–58
- Nakao H, Okamoto H, Fukuda M, Tsuda F, Mitsui T, Masuko K, Iizuka H, Miyakawa Y, Mayumi M (1997) Mutation rate of GB virus C/hepatitis G virus over the entire genome and in subgenomic regions. *Virology* 233:43–50
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- Nei M, Jin L (1989) Variances of the average numbers of nucleotide substitutions within and between populations. *Mol Biol Evol* 6: 290–300
- Okamoto H, Nakao H, Inoue T, Fukuda M, Kishimoto J, Iizuka H, Tsuda F, Miyakawa Y, Mayumi M (1997) The entire nucleotide sequences of two GB virus C/hepatitis G virus isolates of distinct genotypes from Japan. *J Gen Virol* 78:737–745
- Orito E, Mizokami M, Ina Y, Moriyama EN, Kameshima N, Yamamoto M, Gojobori T (1989) Host-independent evolution and a genetic classification of the hepadnavirus family based on nucleotide sequences. *Proc Natl Acad Sci USA* 86:7059–7062
- Pilot-Matias TJ, Carrick RJ, Coleman PF, Leary TP, Surowy TK, Simons JN, Muerhoff S, Buijk SL, Chalmers ML, Dawson GJ, Desai SM, Mushahwar IK (1996a) Expression of the GB virus C E2 glycoprotein using the Semliki Forest virus vector system and its utility as a serologic marker. *Virology* 225:282–292
- Pilot-Matias TJ, Muerhoff AS, Simons JN, Leary TP, Buijk SL, Chalmers ML, Erker JC, Dawson GJ, Desai SM, Mushahwar IK (1996b) Identification of antigenic regions in the GB hepatitis viruses GBV-A, GBV-B, and GBV-C. *J Med Virol* 48:329–338
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sato K, Tanaka T, Okamoto H, Miyakawa Y, Mayumi M (1996) Association of circulating hepatitis G virus with lipoproteins for a lack

- of binding with antibodies. *Biochem Biophys Res Commun* 229:719–725
- Shao L, Shinzawa H, Ishikawa K, Zhang X, Ishibashi M, Misawa H, Yamada N, Togashi H, Takahashi T (1996) Sequence of hepatitis G virus genome isolated from a Japanese patient with non-A-E hepatitis: Amplification and cloning by long reverse transcription-PCR. *Biochem Biophys Res Commun* 228:785–791
- Simons JN, Leary TP, Dawson GJ, Pilot-Matias TJ, Muerhoff AS, Schlauder GG, Desai SM, Mushahwar IK (1995a) Isolation of novel virus-like sequences associated with human hepatitis. *Nat Med* 1:564–569
- Simons JN, Pilot-Matias TJ, Leary TP, Dawson GJ, Desai SM, Schlauder GG, Muerhoff AS, Erker JC, Buijk SL, Chalmers ML, van Sant CL, Mushahwar IK (1995b) Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proc Natl Acad Sci USA* 92:3401–3405
- Smith DB, Cuceanu N, Davidson F, Jarvis LM, Mokili JLK, Hamid S, Ludlam CA, Simmonds P (1997) Discrimination of hepatitis G virus/GBV-C geographical variants by analysis of the 5' non-coding region. *J Gen Virol* 78:1533–1542
- Suzuki Y, Gojobori T (1998) The origin and evolution of human T-cell lymphotropic virus types I and II. *Virus Genes* 16:69–84
- Takahashi K, Hijikata M, Hino K, Mishiro S (1997) Entire polyprotein-ORF sequences of Japanese GBV-C/HGV isolates: Implications for new genotypes. *Hepatology Res* 8:139–148
- Tanaka Y, Mizokami M, Orito E, Ohba K, Kato T, Kondo Y, Mboudjeka I, Zekeng L, Kaptue L, Bikandou B, M'Pele P, Takehisa J, Hayami M, Suzuki Y, Gojobori T (1998) African origin of GB virus C/hepatitis G virus. *FEBS Lett* 423:143–148
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Zuckerman AJ (1996) Alphabet of hepatitis viruses. *Lancet* 346:558–559