Volume 315, number 2, 201–203 FEBS 11968 January 1993

Base transitions and base transversions seen in mutations among various types of the hepatitis C viral genome

Torahiko Tanaka, Nobuyuki Kato, Makoto Hijikata and Kunitada Shimotohno

Virology Division, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104, Japan

Received 18 November 1992

We counted base transitions (Ts) and transversions (Tv) in mutations among various types of hepatitis C virus (HCV) sequences. The Ts/Tv ratio was useful for distinguishing the (geno)type of the compared HCV sequences, since the ratio was 4.0 or more among sequences of the same type, while it was less than 2.0 among sequences of different types. This observation reflects the evolutionary pathway of HCV in which Tv accumulate among distant genomes by mechanisms including multiple base substitutions.

Hepatitis C virus; Genotype; Mutation; Base transition; Base transversion; Multiple base substitution

1. INTRODUCTION

Hepatitis C virus (HCV) is a major causative agent for non A, non B viral hepatitis in Japan and other countries. Although the particle of HCV is unidentified, entire HCV genome sequences have been reported by researchers in Japan [1–5], Taiwan [6] and the USA [7,8]. The genome is single positive strand RNA of about 9.5 kb with reported nucleotide (NT) variations of up to 33% of the total sequences. At present, at least 6 (geno)types of HCV are distinguishable based upon sequence diversity [9]. Here we studied the manner of NT variation among HCV isolates and found that base transversions (Tv) accumulated between distant HCV isolates.

2. EXPERIMENTAL

Base transitions (Ts) are NT changes from purine to purine (A↔G) or pyrimidine to pyrimidine (C↔T). Base transversions (Tv) are NT changes between purine and pyrimidine (A↔C, A↔T, G↔C and G↔T). Ts and Tv were counted using a computer program developed by Dr. N. Munakata (Radiobiology Division, National Cancer Center Research Institute).

According to our proposal [9], the six HCV genotypes are designated as HCV-I, -II, -III, -IV, -V and -VI. The sequences of the following HCV isolates for which whole genomic sequences are available were used in the present study: HCV-I_1 [7], HCV-I_H [8], HCV-II_J [1], HCV-II_{JT} [2], HCV-II_{BK} [3], HCV-II_{J33} (Cho et al. unpublished), HCV-II_T [6], HCV-III_{J6} [4] and HCV-IV_{J8} [5]. For HCV-V and -VI, only partial sequences in the NS5 region (NT 8273–8612 of HCV- II_{JT}) were used for analysis (HCV-V_{T1}, -V_{T7}, VI_{T9}, -VI_{T10}) [9] since the whole genomic sequences are not yet available. When comparing sequences of different lengths, deleted or inserted NT sequences were eliminated (when compared to HCV-II, HCV-I has an insertion of 3 NT, while HCV-III and -IV both have insertions and deletions totalling 84 NT and 15 NT, respectively).

Correspondence address: K. Shimotohno, Virology Division, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104, Japan.

3. RESULTS

The sequences of HCV-I_1, -II_J, -III_{J6} and -IV_{J8} were compared to that of HCV-II_{JT}, and the numbers of Ts and Tv in each functional region (genomic region) were shown in Fig. 1. As seen, Ts predominated between HCV-II_{JT} and -II_J, i.e. within the same genotype. On the other hand, the ratio of Ts to Tv was about 1 between HCV-II_{JT} and -I_1, while between HCV-II_{JT} and -III_{J6} or -IV_{J8}, Tv accounted for more than half of the NT variations. Only in the 5′ untranslated region, the most conserved region in the HCV genome, Ts were predominant. We next examined the Ts/Tv ratios among representative sequences of the 6 types of HCV (Table I). The values obtained from partial sequences showed no apparent discrepancy to those obtained from whole genomic sequences. The ratio was 4.0 or more within the same genotype, but only 1.0–2.0 between HCV-I and -II, HCV-III and -IV, or HCV-V and -VI. In other comparisons, the ratio was less than 1.0. The relationship between NT diversity and the rate of Tv showed a sigmoidal curve that would saturate at around 60% Tv (Fig. 2). The results indicated that the more distant the genomes, the higher the rate of Tv. In contrast to the above observations, the Ts/Tv ratio in hypervariable region, which resided in putative envelope protein region (NT 1512–1571 of HCV-II_{JT}) [11], was less than 1.0 even within the same genotype with one minor exception (between HCV-I_1 and -I_H, Ts/Tv = 11.0). We could derive no specific rules for a relationship between NT changes (Ts or Tv) and amino acid substitutions.

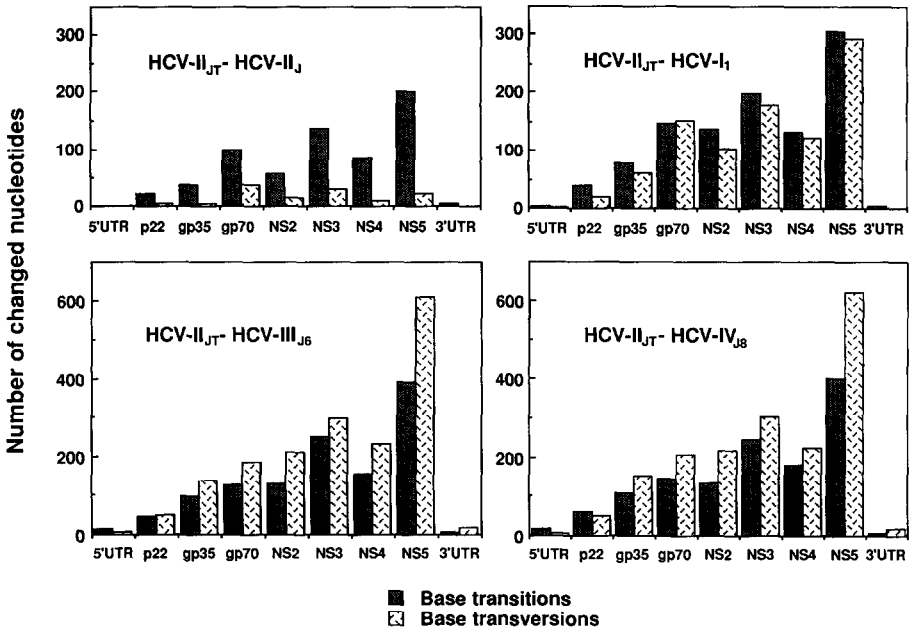

Fig. 1. Base transitions and transversions in each genomic region of HCV in comparisons of the same or different types of HCV genomes. The genomic regions were defined as in [2], except that NS4a and 4b were combined and designated as NS4. 5′ UTR and 3′ UTR denote the 5′ and 3′ untranslated regions, respectively.

4. DISCUSSION

We have demonstrated here that the Ts/Tv ratio is useful for distinguishing the HCV genotypes of compared sequences, even when only partial sequences are available. This is based on the fact that the rate of Tv is higher among distant genomes. The accumulation of Tv seems to reflect the manner of evolution of the HCV genome. When the incidence of Tv in one mutation is constant and low, which occurs due to the features of HCV RNA-replicase, the increase in Tv is attributable to an increase in the number of positions in which multiple base substitutions occur. The incidence of Tv in one mutation is estimated as 10–20% by extrapolating the curve in Fig. 2; hence one Tv occurs for each 5–10 mutations at the same position (multiple base substitutions). Within the same genotype, NT diversity is less than 10% and the rate of Tv was 20% or less. In this situation, the number of NT receiving multiple base substitutions seems few. Between HCV-I and -II or between HCV-III and -IV, NT diversity is about 20% and the rate of Tv is about 45%. Considered simply, the increase in the rate of Tv (about 25% of total base exchanges) is caused by multiple base substitutions, which means that a significant number of mutations are forced to occur in the same position(s). For more distant genomes, the rate of Tv is further increased. However, it must be noted that the rate of Tv saturates at a certain level, since if Tv occur twice at the same position, the final result appears as Ts. The situation for HCV-I and -III, -IV, or HCV-II and -III, -IV seems to be near this stage. In the hypervariable region, multiple base substitutions occur even within close genomes probably by an especially high frequency of mutation limited to this region. There may be some mechanisms for a high incidence of mutation and/or a low pressure of selection in this region.

Thus, once NT diversity reaches a certain level (about 10%), new mutations appear to be permitted preferentially at positions that have already received mutations. This shows a pressure of selection, which reduces the number of mutations in the genome. Further information about the manner of mutation may help to clarify the mechanisms for the selection and molecular evolution of the HCV genome. HCV may provide a good model for the study of gene evolution.

Table I

Ratios of base transitions to transversions (Ts/Tv) in mutations among 6 types of HCV

	I	II	III	IV	V	VI
I	5.4	1.1	0.7	0.8	0.7*	0.7*
II		5.2	0.7	0.7	0.7*	0.7*
III			24*	1.3	0.6*	0.6*
IV				6.0*	0.7*	0.8*
V					8.5*	1 8*
VI						4.0*

The values were obtained using whole genomic sequences or partial sequences in the NS5 region (*). Sequences compared are as follows: HCV-I_1 [7], -I_H [8], -II_J [1], -II_{JT} [2], -III_{J6} [4], -III_{K2a} [10], -IV_{J8} [5], -IV_{K2b} [10], -V_{T1} [9], -V_{T7} [9], -VI_{T9} [9], -VI_{T10} [9]. HCV-III_{K2a} and -IV_{K2b} are partial sequences of the same region as HCV-V and -VI in [9].

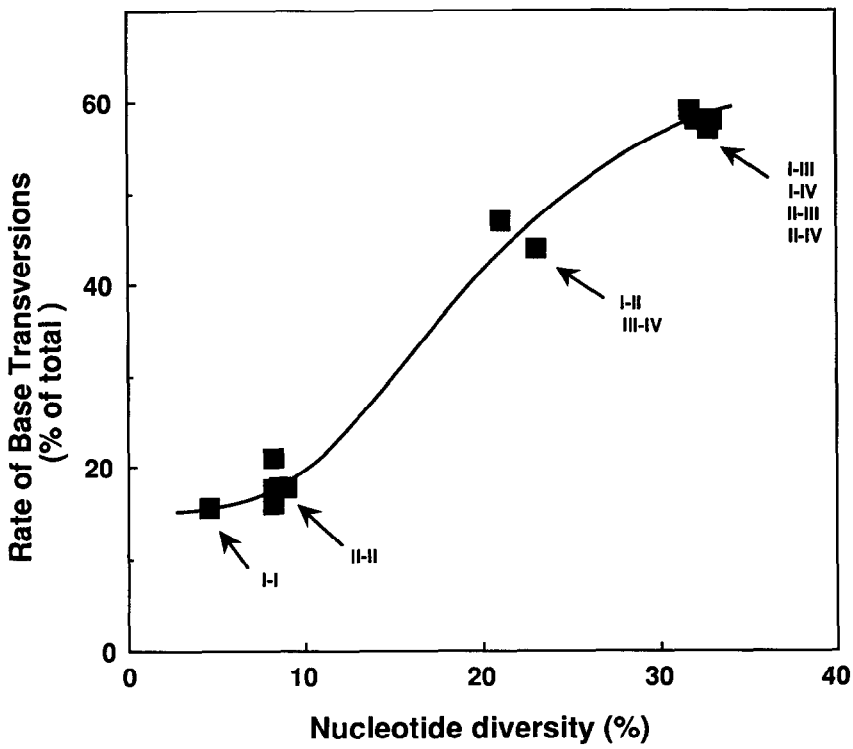

Fig. 2. Relationship between nucleotide diversity and the rate of base transversions.

Acknowledgements: This study was supported by grants-in-aid for Cancer Research and for a Comprehensive 10-Year Strategy of Cancer Control from the Ministry of Health and Welfare of Japan.

REFERENCES

[1] Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T. and Shimotohno, K. (1990) Proc. Natl. Acad. Sci. USA 87, 9524–9528.

[2] Tanaka, T., Kato, N., Nakagawa, M., Ootsuyama, Y., Cho, M.-J., Nakazawa, T., Hijikata, M., Ishimura, Y. and Shimotohno, K. (1992) Virus Res. 23, 39–53.

[3] Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E., Andoh, T., Yoshida, I. and Okayama, H. (1991) J. Virol. 65, 1105–1113.

[4] Okamoto, H., Okada, S., Sugiyama, Y., Kurai, K., Iizuka, H., Machida, A., Miyakawa, Y. and Mayumi, M. (1991) J. Gen. Virol. 72, 2697–2704.

[5] Okamoto, H., Kurai, K., Okada, S., Yamamoto, K., Iizuka, H., Tanaka, T., Fukuda, S., Tsuda, F. and Mishiro, S. (1992) Virology 188, 331–341.

[6] Chen, P.-J., Lin, M.-H., Tai, K.-F., Liu, P.-C., Lin, C.-J. and Chen, D.-S. (1992) Virology 188, 102–113.

[7] Choo, Q.-L., Richman, K.H., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, A., Barr, P.J., Weiner, A.J., Bradley, D.W., Kuo, G. and Houghton, M. (1991) Proc. Natl. Acad. Sci. USA 88, 2451–2455.

[8] Inchauspe, G., Zebedee, S., Lee, D.-H., Sugitani, M., Nasoff, M. and Prince, A.M. (1991) Proc. Natl. Acad. Sci. USA 88, 10292–10296.

[9] Mori, S., Kato, N., Yagyu, A., Tanaka, T., Ikeda, Y., Petchclai, B., Chiewsilp, P., Kurimura, T. and Shimotohno, K. (1992) Biochem. Biophys. Res. Commun. 183, 334–342.

[10] Enomoto, N., Takada, A., Nakao, T. and Date, T. (1990) Biochem. Biophys. Res. Commun. 170, 1021–1025.

[11] Kato, N., Ootsuyama, Y., Tanaka, T., Nakagawa, M., Nakazawa, T., Muraiso, K., Ohkoshi, S., Hijikata, M. and Shimotohno, K. (1992) Virus Res. 22, 107–123.