Cancer Chemother Pharmacol (2004) 54 (Suppl 1): S83–S86
DOI 10.1007/s00280-004-0892-0

Masashi Mizokami · Yasuhito Tanaka

Molecular evolutionary analysis predicts the incidence of hepatocellular carcinoma in the United States and Japan

Published online: 13 August 2004

Abstract Long-term serial serum samples containing hepatitis C virus (HCV) from the US and Japan were molecularly clocked to determine the time-origin of the HCV epidemic. Based on the molecular clock that held significantly, it is estimated that HCV genotype 1 first appeared in the US around 1910 whereas in Japan HCV surfaced before 1882. By regression analyses and coalescent theory, widespread dissemination of HCV in the US population occurred during the 1960s, approximately 30 years later than in Japan. Currently, the prevalence of hepatocellular carcinoma (HCC) is strikingly higher in Japan than in the US. If HCC is a function of increased HCV exposure time, the molecular clock predicts that the burden of HCC in the US will accelerate in the next two to three decades.

Keywords Hepatitis C virus · Hepatocellular carcinoma · Molecular clock · USA · Japan

Introduction

In the US from 1988 to 1994, the prevalence of antibody to hepatitis C virus (anti-HCV) was 1.8% and the calculated prevalence of viremia was 1.3%. This prevalence led to the estimate that approximately 2.7 million Americans were chronically infected with HCV [3] and that this infection would result in approximately 10,000 deaths per year from cirrhosis and hepatocellular carcinoma (HCC) [5]. Despite these statistics, it has been observed that the prevalence of HCC in the US is well below that reported in Japan [10, 13, 18]. The differences in the HCC prevalence between Japan and the US may be due to various genetic and/or environmental influences, but one important factor may be that the HCV endemic has been present in Japan for a longer time. To investigate the hypothesis that a longer duration of the HCV endemic accounts for the differing HCC burden between the US and Japan, the constant evolutionary rate of the virus over time, i.e. a molecular clock of HCV evolution, was established. By retrospective extrapolation, the molecular clock provides insight into both the time when a virus enters a population ("divergence time") and the time when it expands within that population ("spread time").

This work was presented at the 19th Bristol-Myers Squibb Nagoya International Cancer Treatment Symposium, "State of the Arts for Digestive Organs", 14–15 November 2003, Nagoya, Japan.

M. Mizokami (✉) · Y. Tanaka
Department of Clinical Molecular Informative Medicine,
Nagoya City University Graduate School of Medical Sciences,
Kawasumi, Mizuho, Nagoya 467-8601, Japan
E-mail: mizokami@med.nagoya-cu.ac.jp
Tel.: +81-52-8538289
Fax: +81-52-8420021

Long-term serial samples

Previous studies have estimated the rate of synonymous and non-synonymous substitutions in the HCV genome [1, 9, 14, 15]. However, these earlier molecular clock estimates should be regarded as preliminary as most: were based on relatively short observation times; analyzed only short regions of the HCV genome; examined a small number of samples; and/or used simplified models of nucleotide evolution. Additionally, the extrapolation of substitution rates based on pairwise comparisons can give misleading estimates of the molecular clock and the divergence time of HCV. To address these issues, this study utilized serial samples obtained from individuals under long-term medical follow-up ranging up to 20 years [2], and enhanced accuracy of the determination by sequencing long segments of the HCV genome. The analysis used 36 serial samples from 10 US patients with HCV genotype 1a, and the analysis was supplemented by single samples from 11 additional US genotype 1a patients [19]. The interval between the first and last sample in the 10 subjects for whom serial samples were available ranged from 7 to 21.6 years (mean ± SD, 15.1 ± 5.4 years).

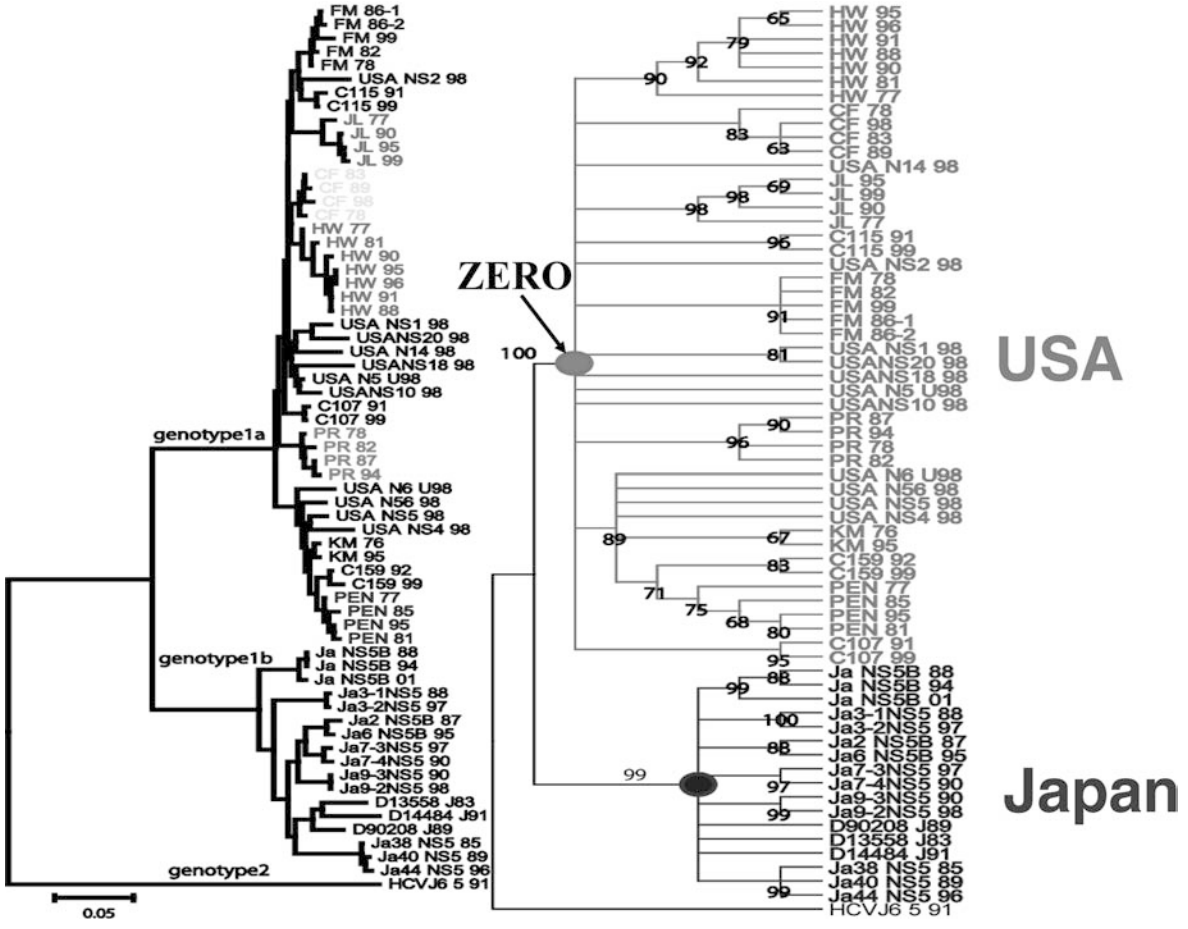
Molecular clock of HCV

Two basic approaches were used to determine the divergence time. Each approach measured long sequences (1778 nt) in combined genomic regions as well as in the shorter NS5B region, and tested multiple serial samples from the majority of the subjects. Based on the phylogenetic tree (Fig. 1), linear regression showed that although the figures for molecular clocks differed somewhat from patient to patient, for each individual patient the molecular clock remained within a well-defined range ($1.70–2.83\times10^{-3}$ per site per year, $P<0.001$ to $P=0.045$), and that the mean figure for the molecular clocks for all patients was also within well-defined parameters. Figure 2 shows the molecular clock within an individual patient (patient H) in the NS5B region. Using several of the models, regression analyses with the entire US cohort indicated the mean evolutionary rate of all codon substitutions within the combined genomic regions was 0.67×10^{-3} ($0.53–0.79\times10^{-3}$) per site per year and the rate of synonymous substitutions was 1.32×10^{-3} ($1.00–1.74\times10^{-3}$) per site per year [19].

Fig. 1 A phylogenetic tree of HCV NS5B region. The tree-enforced molecular clock (*right*) was constructed by PAUP (phylogenetic analysis using parsimony) on the basis of the NJ (neighbor-joining) tree (*left*). The distances from the ancestral sequence of the most recent common ancestor (*ZERO*) to each strain were estimated for evolutionary analyses. Modified with permission from Tanaka et al. [19]. Copyright 2002, National Academy of Sciences, USA

Divergence time of HCV

The constant molecular clock of HCV can be used to determine the divergence time of HCV in the US and Japan. Given a phylogenetic tree and assuming a stable and linear molecular clock, one can plot the total branch length from the tips of the branches to the ancestral node against the year of sampling and then fit a regression. Using this approach, the divergence time of the most recent common ancestor of genotype 1a in the US was estimated to be approximately 1910 (Fig. 3). This estimate lies well within the 95% significance data (1894–1912) of the divergence time estimated by most models [19]. It is possible that HCV was introduced into the US population during the Spanish-American war in 1898 to 1900, when soldiers infected with the virus returned to the US from endemic areas. Additionally, the divergence time of the most recent common ancestor of this blood-borne virus in the US is consistent with the introduction of modern blood transfusion practice following the discovery of blood types by Landsteiner in 1900.



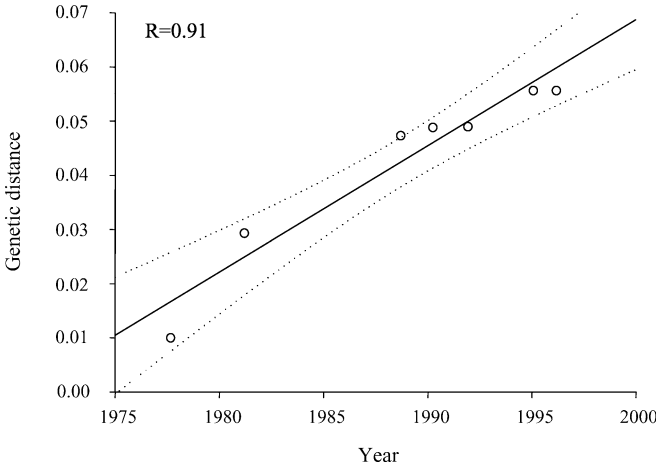


Fig. 2 Molecular clock within an individual in the NS5B region. Using the Pamilo-Bianchi-Li model, linear regression showed that the molecular clock in patient H remained within a well-defined range (rate 2.17×10^{-3} per site per year, $P=0.005$). Modified with permission from Tanaka et al. [19]. Copyright 2002, National Academy of Sciences, USA

Based on similar molecular evolutionary rates between genotype 1a and 1b in our analysis, the divergence time of the most recent common ancestor of HCV genotype 1b in Japan was estimated to be prior to 1882, i.e. approximately 30 years earlier than in the US.

Spread time of HCV strains between the US and Japan

Another approach to the study of HCV evolution is the use of coalescent theory which details the history of changes in effective number of HCV infections inferred from a phylogenetic tree reconstructed on nucleotide sequences of the virus genome [16, 17]. Using the figure for the mean molecular clock based on US genotype 1a sequences, the effective number of HCV infections in the US proportional to the actual population number was estimated and compared with the corresponding Japanese population infected with HCV. This analysis indicated that the growth of the US HCV genotype 1a-infected population occurred around 1960, at least 30 years later than the widespread introduction of genotype 1b into the Japanese population (Fig. 4) [19]. The spread of HCV in Japan may be linked to two distinct events: the widespread treatment of schistosomiasis with intravenous antimony sodium tartrate which began in 1923 [8], and the use of intravenous stimulants during and after World War II. Of note, the spread of genotype 1b in the Japanese population began to decrease around 1995, whereas the HCV genotype 1a in the US is still growing exponentially (Fig. 4) [4, 11]. The growth time of the HCV-infected population in the US is consistent with the spread time estimated by individual molecular clocks.

These analyses thus indicate that both the divergence time and the spread time for HCV in Japan occurred decades before these same events in the US. If the difference in HCC prevalence between the US and Japan simply reflects the longer duration of the endemic in Japan, the calculated time disparity would predict an increase in the disease burden of HCC in the US over the next two to three decades.

An increase in the HCC prevalence has already been reported in the US over the past two decades, with younger age groups mainly affected [6]. El-Serag and Mason [7] reported a threefold increase in the age-adjusted incidence of HCV-associated HCC (2.3 per 100,000 between 1993 and 1995 to 7.0 per 100,000

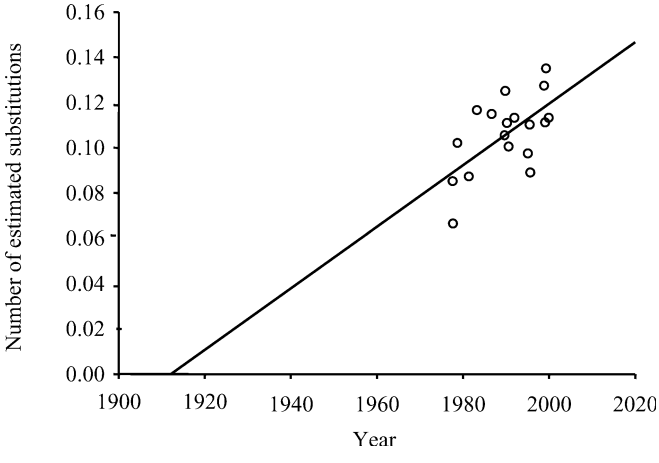


Fig. 3 Divergence time of HCV genotype 1a in combined regions. The regression analysis with serial samples of US HCV genotype 1a indicated the mean evolutionary rates. The divergence time of the most recent common ancestor of US genotype 1a was estimated to have occurred around 1910. Modified with permission from Tanaka et al. [19]. Copyright 2002, National Academy of Sciences, USA

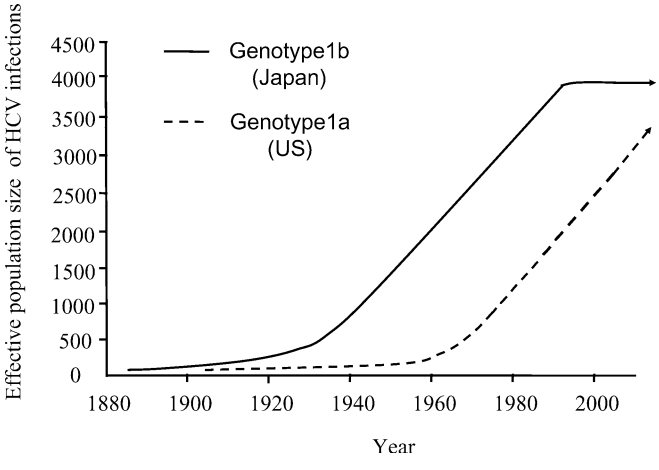


Fig. 4 Effective population size of HCV in the US relative to the population size in Japan over the past century. The growth of the US HCV genotype 1a population was estimated to have occurred around 1960, at least 30 years later than the spread time of genotype 1b population in Japan. Modified with permission from Tanaka et al. [19]. Copyright 2002, National Academy of Sciences, USA

between 1996 and 1998), whereas age-adjusted rates for primary liver cancer due to HBV and alcoholic cirrhosis remained stable. More recently, Nair et al [12] have reported a marked increase in US deaths related to HCC over the past three decades. Such data support the hypothesis that HCC incidence is related to HCV exposure time within a population.

Conclusion

Analysis of long-term serial samples from individual subjects produced a unique and reproducible molecular clock estimate for the HCV strain infecting each individual. The mean molecular clock suggests that the most recent common ancestor of HCV in the US occurred around 100 years ago and that it was widely transmitted in the 1960s. In contrast, in Japan, the recent common ancestry point occurred more than 100 years ago and the virus was widely disseminated approximately 30 years earlier than in the US. As Japan has a much higher HCC prevalence than that of the US [10, 13, 18], it is predicted that an increase in HCC prevalence will occur in the US over the next two to three decades.

Acknowledgements We would like to thank Dr. H.J. Alter and Dr. J.W.K. Shih from the Department of Transfusion Medicine, Warren Grant Magnuson Clinical Center, National Institute of Health, USA, and Dr. T. Gojobori and Dr. K. Hanaka from the Center for Information Biology, National Institute of Genetics, Japan.

References

1. Abe K, Inchauspe G, Fujisawa K (1992) Genomic characterization and mutation rate of hepatitis C virus isolated from a patient who contracted hepatitis during an epidemic of non-A, non-B hepatitis in Japan. J Gen Virol 73:2725
2. Alter HJ, Purcell RH, Shih JW, Melpolder JC, Houghton M, Choo QL, Kuo G (1989) Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. N Engl J Med 321:1494
3. Alter MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, Gao F, Moyer LA, Kaslow RA, Margolis HS (1999) The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. N Engl J Med 341:556
4. Armstrong GL, Alter MJ, McQuillan GM, Margolis HS (2000) The past incidence of hepatitis C virus infection: implications for the future burden of chronic liver disease in the United States. Hepatology 31:777
5. Center for Disease Control and Prevention (1998) Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. Morbidity Mortality Weekly Rep 47:RR-19
6. El-Serag HB, Mason AC (1999) Rising incidence of hepatocellular carcinoma in the United States. N Engl J Med 340:745
7. El-Serag HB, Mason AC (2000) Risk factors for the rising rates of primary liver cancer in the United States. Arch Intern Med 160:3227
8. Iida F, Iida R, Kamijo H, Takaso K, Miyazaki Y, Funabashi W, Tsuchiya K, Matsumoto Y (1999) Chronic Japanese schistosomiasis and hepatocellular carcinoma: ten years of follow-up in Yamanashi Prefecture, Japan. Bull World Health Organ 77:573
9. 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
10. Kiyosawa K, Sodeyama T, Tanaka E, Gibo Y, Yoshizawa K, Nakano Y, Furuta S, Akahane Y, Nishioka K, Purcell RH (1990) Interrelationship of blood transfusion, non-A, non-B hepatitis and hepatocellular carcinoma: analysis by detection of antibody to hepatitis C virus. Hepatology 12:671
11. Moriya T, Koyama T, Tanaka J, Mishiro S, Yoshizawa H (1999) Epidemiology of hepatitis C virus in Japan. Intervirology 42:153
12. Nair S, Shivakumar KS, Thuluvath PJ (2002) Mortality from hepatocellular and biliary cancers: changing epidemiological trends. Am J Gastroenterol 97:167
13. Nishioka K, Watanabe J, Furuta S, Tanaka E, Iino S, Suzuki H, Tsuji T, Yano M, Kuo G, Choo QL (1991) A high prevalence of antibody to the hepatitis C virus in patients with hepatocellular carcinoma in Japan. Cancer 67:429
14. Ogata N, Alter HJ, Miller RH, Purcell RH (1991) Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. Proc Natl Acad Sci U S A 88:3392
15. Okamoto H, Kojima M, Okada S, Yoshizawa H, Iizuka H, Tanaka T, Muchmore EE, Peterson DA, Ito Y, Mishiro S (1992) Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability. Virology 190:894
16. Pybus OG, Rambaut A, Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. Genetics 155:1429
17. Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH (2001) The epidemic behavior of the hepatitis C virus. Science 292:2323
18. Shiratori Y, Shiina S, Imamura M, Kato N, Kanai F, Okudaira T, Teratani T, Tohgo G, Toda N, Ohashi M (1995) Characteristic difference of hepatocellular carcinoma between hepatitis B- and C- viral infection in Japan. Hepatology 22:1027
19. Tanaka Y, Hanada K, Mizokami M, Yeo AE, Shih JW, Gojobori T, Alter HJ (2002) Inaugural Article: A comparison of the molecular clock of hepatitis C virus in the United States and Japan predicts that hepatocellular carcinoma incidence in the United States will increase over the next two decades. Proc Natl Acad Sci U S A 99:15584