Tracing the Evolution of Hepatitis C Virus in the United States, Japan, and Egypt By Using the Molecular Clock

MASASHI MIZOKAMI and YASUHITO TANAKA
Department of Clinical Molecular Informative Medicine, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan

The molecular clock has been a very powerful tool in looking back at the epidemic spread of HCV infection in the United States (US) and Japan, as well as in Egypt. This analysis estimates that the growth of the US HCV genotype 1a (HCV-1a)–infected population occurred around 1960, at least 30 years later than the widespread introduction of HCV-1b into the Japanese population. In Japan, the estimated effective number of HCV infections indicated a rapid exponential growth in the 1920s among patients with schistosomiasis, which coincides with injection treatment for schistosomiasis since 1921 in previously schistosomiasis-endemic areas. In Egypt, the spread of HCV-4a would have increased exponentially during the 1940s through 1980, which was also consistent with the duration of intravenous antimony campaigns for the treatment of shistosomiasis in that country. The implications are that Japan has set the model for HCV-related HCC, and that the high HCC incidence in Japan might be replicated by the rest of the world as their HCV-infected population ages and the duration of HCV infection approaches that currently observed in Japan.

Chronic HCV infection is usually clinically mild, but more than 20% of patients progress to severe chronic hepatitis and cirrhosis, occasionally culminating in hepatocellular carcinoma (HCC) during the course of 2–3 decades. Because the time lag between HCV infection and cancer development is several decades,[1] it is important to estimate the demographic history of HCV infection to predict the future burden of disease.

Recently, the molecular clock theory has been successfully applied to estimate molecular evolutionary rate in long-term serial serum samples obtained from HCV-infected patients in the United States (US) and Japan; a 30-year lag in HCV spread time was thereby demonstrated between these countries.[2] Insofar as a long duration of HCV infection is a critical determinant for the development of HCC, the molecular clock predicted that the incidence of HCC will increase in the US during the next 2–3 decades and approach the high rates currently observed in Japan.

In this review, HCV genotype 1a (HCV-1a) strains in the US, HCV-1b in Japan,[2,3] and HCV-4a in Egypt[4] were analyzed by a coalescent-based approach by using principles of both population genetics and mathematical epidemiology,[5] and the HCV spread times in each country are mentioned.

Methods

Subjects

Long-term serial HCV-1a sequences in the US were obtained from our previous study.[2] One hundred thirty-one previously reported HCV-1b sequences including 64 from *Schistosoma japonicum*–positive sera and 67 from *S japonicum*–negative sera in Japan[3] and 47 previously reported HCV-4a sequences from Egypt[4] were used in this study.

Molecular Evolutionary Analyses

A reconstructed tree was built on the NS5B sequence of 339 nucleotide by a heuristic maximum-likelihood topology search with stepwise addition and the nearest neighbor-interchange algorithms. Tree likelihood scores were calculated by using HKY85+G method with the molecular clock enforced by PAUP version 4.0b8 (Sinauer Associates, Inc, Sunderland, MA). As estimates of the demographic history, a nonparametric function $N(t)$, known also as the skyline plot, was obtained by transforming coalescent intervals of an observed genealogy into a piecewise plot that represents an effective number of infections through time.[5,6] A parametric maximum-likelihood was estimated by several models with the computer software Genie v3.5 (Oxford Evolutionary Biology, Oxford, England) to build a statistical framework for inferring the demographic history of a population on phylogenies reconstructed on sampled DNA sequences.[6] This model assumes a continuous epidemic process in which the viral transmission parameters remain constant through time. Model fitting was evaluated by likelihood ratio tests of the parametric maximum-likelihood estimates.[7] In this study, we used substitution rate estimated

Abbreviations used in this paper: HCC, hepatocellular carcinoma; HCV-1a, HCV genotype 1a; US, United States.

1542-3565/05/$30.00
PII: 10.1053/S1542-3565(05)00705-6

previously, 5.8×10^{-4} substitutions per site per year for the NS5B region.[2]

Results

Molecular Clock of Hepatitis C Virus

On the basis of the phylogenetic tree (Figure 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, and that the mean figure for the molecular clocks for all patients was also within well-defined parameters. The regression analyses with the entire US cohort indicated that the mean evolutionary rate of all codon substitutions within the combined genomic regions was 0.67 (0.53–0.79) $\times 10^{-3}$ per site per year, and the rate of synonymous substitutions was 1.32 (1.00–1.74) $\times 10^{-3}$ per site per year (Figure 2).[2]

Divergence Time of Hepatitis C Virus Strains in the United States

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. With this approach, the divergence time of the most recent common ancestor of HCV-1a in the US was estimated to be approximately 1910 (Figure 2).[2] It is possible that HCV was introduced into the US population during the Spanish-American War in 1898–1900 when soldiers infected with the virus returned to the US from endemic areas. In addition, 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 after the discovery of blood types by Landsteiner in 1900.

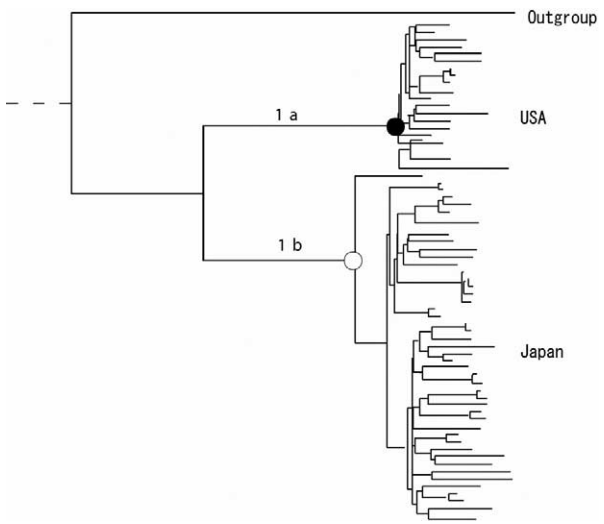

Figure 1. A phylogenetic tree within the NS5B region constructed by the neighbor-joining method. *Closed circle* indicates the most recent common ancestor of HCV-1a in the US, and *open circle* indicates the most recent common ancestor of HCV-1b in Japan.

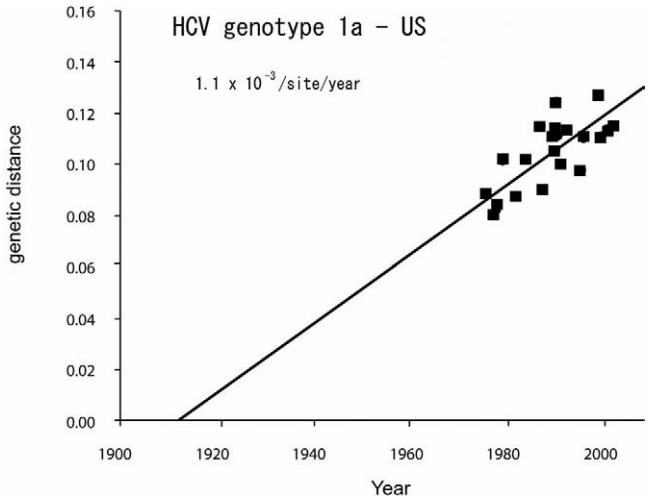

Figure 2. Estimating the most recent common ancestor of HCV-1a in the US on the basis of data collected during the last 3 decades. A regression analysis within the NS5B region was performed to estimate a mean molecular clock. The regression analysis with serial samples of US HCV-1a indicated the mean evolutionary rates of synonymous substitutions (P-B-L Model) (1.1×10^{-3} per site per year, $P < .05$). We estimated the time associated with zero branch length, ie, the time of the ancestral sequence, by using the above mean molecular clock by regression analysis. The divergence time of the most recent common ancestor of US HCV-1a was estimated to have occurred around 1910.

Spread Time of Hepatitis C Virus Strains Between the United States and Japan

Another approach to studying HCV evolution is the use of coalescent theory, which details the history of changes in virus effective population size inferred from a phylogenetic tree reconstructed by nucleotide sequences of the virus genome.[5,6] This analysis estimates that the growth of the US HCV-1a–infected population occurred around 1960, at least 30 years later than the widespread introduction of HCV-1b into the Japanese population (Figure 3).[2] The spread of HCV in Japan might be linked to 2 distinct occurrences, the widespread treatment of schistosomiasis with intravenous antimony sodium tartrate since 1921[8] and the use of intravenous stimulants during and after World War II. Of note, the spread of HCV-1b in the Japanese population started to decrease around 1995, whereas HCV-1a in the US is still growing exponentially (Figure 3).[9,10] The exponential increase of HCV-1a in the US might depend on recent discovery of HCV. 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