Journal of Viral Hepatitis, 2010, **17,** 229–235 doi:10.1111/j.1365-2893.2009.01193.x

Hepatitis B virus: origin and evolution

S. M. Jazayeri,[1] S. M. Alavian[2] and W. F. Carman[3] [1]*Hepatitis B Molecular Laboratory, Department of Virology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran;* [2]*Baqiyatallah University of Medical Sciences, Baqiyatallah Research Centre for Gastroenterology and Liver Disease, Tehran, Iran; and* [3]*West of Scotland Specialist Virology Centre, Glasgow, UK*

Received January 2009; accepted for publication August 2009

SUMMARY. The pathogenesis of hepatitis B virus (HBV) is complex and it appears that molecular variants play a role in this process. HBV undergoes numerous rounds of error prone production within an infected host. The resulting quasispecies are heterogeneous and in the absence of archaeological records of past infection, the evolution of HBV can only be inferred indirectly from its epidemiology and by genetic analysis. This review gathered the controversies about the HBV origin and factors influencing its quasispecies. Also, it provided some evidence on how HBV genotypes correlated with human history and patterns of migration. It is our belief that this topic deserves further attention and thus it is likely that more critical research work will be performed to elucidate the unknown mechanisms and processes in this area.

Keywords: HBV genotypes, HBV origin, hepatitis B virus evolution.

HEPATITIS B VIRUS (HBV) EVOLUTION

The origin(s) of viruses cannot be known with certainty. PCR and other sensitive molecular techniques can reveal some viral genome sequences from the relatively recent past, but very ancient viral genomes will remain a matter for speculation. Comparative sequence analysis suggests that both RNA and DNA viruses have deep, archaic evolutionary roots both for genome structural organization and with regard to certain genomic and protein domains. Both DNA and RNA viruses can emerge and evolve by a variety of mechanisms including mutation, recombination and reassortment [1]. Nucleotide substitutions in viral genomes can have several effects, including evasion of vaccine-induced or natural immunity, drug resistance, and changes in pathogenicity, alteration in tissue or species tropism, and viral persistence.

Escape mutation is the process by which amino-acid substitution at one or more positions alters an epitope to the extent that the virus can persist in the presence of an adequate immune response to the initial epitope. As recognition of the first epitope leads to the destruction of infected cells or neutralization of free virus, viruses that encode for the expression of a mutated epitope (or cells that present it on their surface) will survive. Viruses that are replicating more rapidly are more likely to develop mutations and for them to be selected [2]. A comparison of the rates of synonymous and non-synonymous substitution in viral genomes can elucidate, to some extent, whether there has been selection at the amino-acid level. One would expect the rate of synonymous to be less than the rate of non-synonymous substitution if there is selective pressure exerted on this viral gene [3].

Abbreviations: DHBV, duck hepatitis B virus; HBV, hepatitis B virus; ORF, open reading frames.

Correspondence: Seyed Mohammad Jazayeri, Hepatitis B Molecular Laboratory, Department of Virology, School of Public Health, Tehran University of Medical Sciences, PO Box: 14155-6446 Tehran, Iran. E-mail: jazayerism@tums.ac.ir

HBV MUTATION RATE

Although the reverse transcriptase activity of HBV may be responsible for the high rate of nucleotide substitution of the virus compared to that of other DNA viruses, replication of the HBV genome may not always depend upon reverse transcription and thus, the frequency of mutation of the HBV genome during replication may not be as high as that of a retroviral genome [4]. This is due to the fact that the mutation rate of HBV is influenced by some related factors. Firstly, the genome is very compact with overlapping open reading frames (ORFs), which limits the number of viable mutations [5]. In a virus with overlapping ORFs, many mutations will generate non-viable virions though some mutations may give the virus a replication advantage [6]. The second constraint is imposed by the need to conserve the direct repeats, promoters and other cis-acting elements, which are involved in replication of the genome. Thirdly, HBV replication takes place inside the nucleocapsid and, unlike retroviruses, this process involves only one copy of the RNA pregenome, limiting the chance of homologous

recombination [5,7]. Despite these constraints, up to 12% of nucleotides may vary between isolates of HBV [8]. The net result is a higher mutation rate in unconstrained parts of the genome. One such region is the pre-S, and especially pre-S2 [4]. Considering the rate of synonymous substitution is higher than non-synonymous substitution for all HBV-ORFs [9], it has been suggested that HBV genome variation is constrained by amino-acid changes [10]. This led to the hypothesis that each protein or each domain of an ORF has its own functional and structural constraints which conserve the amino-acid sequence over the evolutionary time. The assumption is that the selective force on each ORF works independently of the region of gene overlap and that base exchanges in the overlapping regions are therefore the result of dual selection by the two overlapping ORFs [11]. The appearance of mutations in overlapping but unrelated viral genes (for example S and P genes) may produce HBV mutants with altered antigenicity and/or replication and a natural history that may be distinctly different to wild type HBV [6].

The rate of replication and mutation is an important factor in the genetic evolution of a virus. Some DNA viruses, such as hepadnaviruses, which include a reverse transcription step in their replication cycle, show rates of evolution in the range of 10^{-4}–10^{-5} nucleotide substitutions per site per year, which is approaching the values of some RNA viruses. HBV exhibits a mutation rate more than 10-fold higher than other DNA viruses. The estimated rate of evolution for HBV is $<2 \times 10^{-4}$ base substitutions per site per year, which reflects a highly dynamic process with a large production of virus [12]. This value is intermediate between DNA and RNA viruses [5]. Although the rate of synonymous substitutions for HBV is 10^4 times higher than that of a host genome, it is 10^{-2} less than that of retroviral genes [10]. Nevertheless, this rate provides support for the hypothesis that HBV evolved from a retrovirus or retrovirus-like progenitor through a process of deletion [13,14].

The HBV genome is extremely stable unless exposed to host immune responses, as exemplified by the completely conserved nucleotide sequence over a 20- to 35-year period in HBeAg-positive asymptomatic carriers with very high levels of virus replication. In contrast, mutations are seen in most, if not all, HBeAg-negative carriers and are distributed over all regions of the viral genome [15]. Some of these amino-acid substitutions can be found as minor species at earlier time points suggesting that the selection process is slow and that multiple strains can co-exist [16]. Some observations have also shown that if conditions do favour the emergence of a variant (for example in the case of a poor match vaccine between immunity and the virus), then the time-scale of emergence is likely to be in the order of decades. As HBV persists in individuals infected for long periods spreading the infection relatively slowly, a variant will only emerge slowly [17].

SELECTION OF THE FITTEST STRAIN

The term 'fixation' refers to random mutations not being lost but becoming incorporated into sequences that undergo further rounds of replication and become progeny virions [18]. There are two ways by which fixation occurs: one is the occurrence of mutations that do not affect fitness, which has been shown in cell culture in which viral populations gradually shift their consensus sequence while exposed to a constant *in vitro* environment. Lenhoff *et al.* [19] generated a cytopathic mutant of duck hepatitis B virus (DHBV) (G133E) in the pre-S protein of DHBV. Inclusion of this mutant into susceptible ducklings resulted in enhanced viral replication, increased the pool of viral cccDNA, and caused hepatocyte destruction. The liver damage caused by G133E DHBV subsided over time resulting in mild chronic hepatitis similar to that observed in wild type virus-infected birds and coincided with a reduction in viral replication to wild type virus levels in the liver. Further, they identified at least one non-cytopathic revertant from the serum of G133E-infected birds after recovery, suggesting that acute liver injury could result from infection with a cytopathic hepadnavirus, but such viruses may be rapidly replaced by non-cytopathic variants during persistent infection. In other words, these cytopathic viruses are not as replication fit as the wild type in the context of persistent infection [20].

The alternative mode of mutant fixation is positive selection of fitter virus as driven by antibodies, antiviral agents or differences in cell biology. In the natural course of HBV infection, cellular and humoral immune pressure against virus-specific proteins may select the fittest viral strains [21]. The type and number of mutations that accumulate in an individual genome are either a marker of the duration and/or severity of the liver disease, or the type and the intensity of immune responses [22]. Selection by the host is a major force for evolutionary change within a virus population [23]. Changes in host selection pressure may greatly affect substitution rates in HBV, with lower rates of changes in those individuals who continue to produce HBeAg compared to those who have cleared it [24–27]. In numerous studies, investigators have shown that during the HBeAg-positive active replication phase of chronic infection, despite the virus being capable of mutating because of the poor proof-reading ability of the reverse transcriptase, sequences able to translate HBeAg remain dominant. Possibly, HBeAg-producing sequences have an intracellular advantage over HBeAg-non-producers. Because of the dominant replication efficiency of HBeAg-producing strains and thus tolerance, few cells appear to be destroyed before the elimination phase, therefore, there appears to be no selection pressure and thus HBeAg-minus mutants are lost or remain as minor populations. However, during the elimination phase of chronic disease, there is the added factor of positive selection pressure, and hepatocytes with HBeAg-non-producing strains become dominant. At the molecular level, this is explained

by the presence of pre-C mutants [28]. Moreover, it has been shown that once the resistant virus become predominant in the viral quasispecies obtained after treatment with lamivudine, drug removal led to the rapid replacement of the resistant virus by the wild type [29]. Other studies showed that the pre-C and core promoter mutations were replaced by wild type during long term therapy, but with continuing therapy, mutations reappeared independent of viral breakthrough [30,31]. Selection is likely to occur if the immune response is incomplete (by using vaccines or monoclonal antibody immunotherapy) [12].

HBV ORIGIN

Attempts have been made to relate these observations to HBV evolution. Much of our knowledge about human hepatitis has relied heavily upon information derived from infection (natural and experimental) of non-human primates during the last 50 years. However, there seem to be some difficulties due to the relationship between the human genotypes A–E and G (and even more divergent genotype F) to each other and to other primate-associated genotypes. The origin of HBV in humans is as confusing as that of the hepadnaviruses from other primates. Different theories have been proposed by investigators on HBV origins based on the hypothesis that the numbers of nucleotide and amino-acid substitutions over time, the molecular clocks, are indicators of viral evolution [3]. A primate origin for HBV infection was proposed by MacDonald *et al.* This theory based on the finding of variants in chimpanzees [32], woolly monkeys [33] and orangutans [34], suggested that these viruses co-evolved with their primate hosts over periods of 10–35 million years. This hypothesis has been supported by the observations that areas of high HBV prevalence in humans are those in which contact with, and cross-species transmission from primates are most likely (South America, Sub Saharan Africa and Southeast Asia). Indeed, certain HBV genotypes are specific to these three areas (F, E, B/C, respectively). Moreover, the mixture of HBV genotypes found outside these areas, such as Europe and North America, may have resulted from much more recent epidemic spread [35]. In contrast, a recent emergence hypothesis for HBV infection indicated that the current wide distribution of HBV in apes must have arisen through several cross-species or subspecies transmission in the relatively recent past [36].

Based on observation of Norder *et al.* that most of the dendrograms obtained from gibbon and chimpanzee strains represented early lineages, assumptions were made that these viruses were indigenous to their respective hosts and not recent acquisitions from man (genotype F). Therefore, they suggested that either genotype F represents an early cross-species transfer from a non-ape primate to man, or that a hepadnavirus of a common ancestor to man and apes gave rise to two viral lineages [37]. Thus, they proposed that the evolutionary history of HBV corresponds to the spread of anatomically modern humans as they migrated from Africa 100 000 years ago [38,39] and different genotypes infecting humans evolved since this dispersal. However, this hypothesis does not explain the origin of the various non-human primate viruses which are interspersed among the human genotypes in the phylogenetic tree. The phylogenetic tree of the various primate HBV variants in no way reflects the phylogeny of the host species, as would be expected for co-speciation [40]. For example the presence of genotype F in native American populations is inconsistent with the presence of genotype B and C in their genetically nearest relatives, the Mongoloid Northeast Asian. Indeed, there is little relationship between HBV genotype distribution with any of the other human population groups (Southeast Asian, Caucasians, and African population) [40]. Alternatively, the HBV genotypes may have evolved later than, and independent of, human migration [22].

According to the finding that HBV showed a nucleotide substitution rate of 2.1×10^{-5} substitutions per site per year over a mean observation period of 22 years, Orito *et al.* proposed that the human genotypes of HBV would have originated from a common ancestor approximately 3000 years ago [10]. In this study, they showed that three major clusters of HBV (birds, mammals and humans) diverged from their common origin in the same order as that of host evolution. They concluded that the evolution of the hepadnavirus family was independent of host-species divergence and for HBV in humans this has taken place much more recently than has divergence of humans.

Alternatively, a New World origin for HBV was proposed by Bollyky *et al.* who suggested that HBV originated from the Americas and spread into the Old World over the last 400 years after contact from Europeans during colonization; a genotype F origin. Further, they considered the possibility that if the virus originally entered the Americas from Asia, this may have required a higher rate of nucleotide substitution as it adapted to this naïve human population [41]. However, the main problem for this hypothesis is the observation of the widespread distribution of HBV in Old World primate species. A remarkable example is a shared genotype of HBV infecting West African chimpanzees [32,42], which showed approximately 11% divergence from the human genotypes A–E. This finding was based on analysis of mutations in the C-terminus region of the core protein (which is well conserved among hepadnaviruses) between human genotypes E/F and the chimpanzee one. Interestingly, HBV-E/F and the non-human primate hepadnaviruses had a common motif within 20 nucleotides upstream from the stop codon for the core gene, whereas, HBV-A/B/C/D genotypes contained a different motif at this site [42]. It has been revealed that sequences in wild-born Old World primates from Africa and Southeast Asia were unrelated to five human HBV strains (A–E); the conclusion is that the virus was not acquired from humans, and all the Old World non-human primate HBVs were on a common

ancestral branch [43]. This finding together with the observation that the closest relative of the woolly monkey HBV is genotype F [33], led to the speculation that chimpanzees have their own hepadnavirus, which resembles the human hepadnavirus (genotype F) [42,43].

At present, the problems associated with each of these hypotheses for the origin of HBV prevent a definitive conclusion. Resolution of these issues requires more extensive sequence analysis of HBV in poorly sampled areas as well as combined human and primate studies together with utilizing models of DNA substitution which better describe the process of viral evolution [41].

HBV-GENOTYPES EVOLUTION

The genetic variability of HBV is observed both as the evolution of genotypes (and thus subtypes), i.e. a divergence of the viral genome in the carrier population, and as the emergence of mutations in each infected subject [44]. Subtypes of HBV have largely evolved separately with their hosts over time and random mixing of different subtypes within individuals has been rare [11]. Okamoto *et al.* [14] observed that a 54-year-old patient was chronically infected with three different clones of HBV. They suggested that the clones would have evolved from a common ancestor virus during 54 years after the primary infection rather than infection by three different strains of HBV (triple infection). Further, they suggested that following a small-dose infection, only one genotype of virus, predominant in the donor, would most likely infect the recipient. In another study the predominant HBV genotype was shown to be quite stable over a period of some 30 years [45].

HBV has been classified into eight genotypes, from A–H and there are some hints that the outcome of the disease and the response to therapy might be correlated to these genomic groups (although not as serious as HCV).

Hepatitis B virus genotypes show a characteristic geographic distribution with a proposed association with human migration. Several scientific fields are employed for the study of human population history, including archaeology, linguistics, anthropology and recently, genetics. Thanks to the introduction of modern technology into genetics in the past two decades, humans have been facing some very interesting findings about their history. Besides mitochondrial DNA sequence analysis, persistent viruses have opened a new window into this area, especially HBV.

Figure 1 shows the prevalence of HBV genotypes throughout the world. It is of interest that genotype is clearly linked to migration. For example, the ancestors of Eskimos migrated from Southeast Asia to the North and passed through the Behring channel to Alaska, which partially explains why there was as high a prevalence of HBsAg positivity in this area as in Southeast Asia [46].

Second, in some studies from South America, genotypes E and A, which are the dominant types in Africans, were found in an area with genotype F/H dominancy. They originate from those with African descent who came into South America during the slave trading period a few hundred years ago. In our own research [47], (Basuni, unpublished data) in the Pacific region (which is an endemic area with genotype C dominancy) we compared HBV surface and core genes with the ones from Southeast Asian patients and from international databases. The gradient of nucleotide and amino acid variations from west to east in our study were most consistent with the hypothesis of migration of Polynesian people from Southern China through Melanesia and Fiji and their radiation across the Pacific to fill the Polynesian triangle in different times. We also found an interesting association that supported the migration history of Southeast Asian ancestors southwards and their colonization of the Pacific islands.

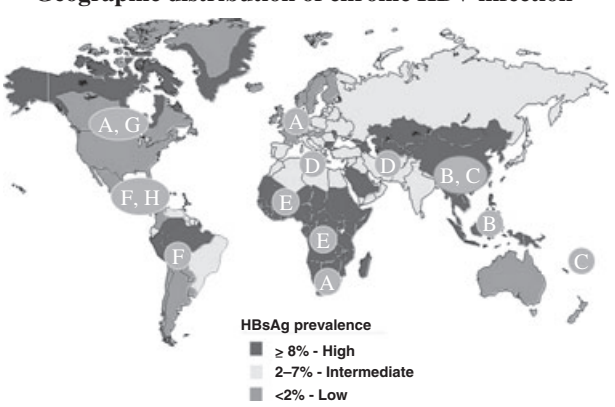

Fig. 1 Geographic distribution of chronic HBV infection and HBV genotypes A to H. Only dominant genotypes shown in circles.

HBV GENOTYPE D EVOLUTION, AS AN EXAMPLE

Genotype D is the most prevalent and the most distributed HBV genotype. It is found in Western populations, the Indian subcontinent, the Middle East and North Africa. Genotype D contains four subgroups (D1–D4) and two subtypes (ayw2 and ayw3). We have collected all the available data from India, Bangladesh (Jazayeri, unpublished data), Turkey, Pakistan, and Iran. We considered the relationship between the potential genotypes in these countries and their evolution. We also considered the samples from East to West. Our unpublished data on 66 Bangladeshi isolates revealed four genotypes (A, B, C and D) with genotype D predominant. In India (Fig. 2) genotype D was dominant (67%), but there were other genotypes: A (22%), C (8%) and recombinants (3%).

In Pakistanis (Fig. 3) 62% were genotype D, A (14%), C (6%), other genotypes (4%) and recombinants (10%). However, in Turkey (Fig. 4), authors found genotype D as the

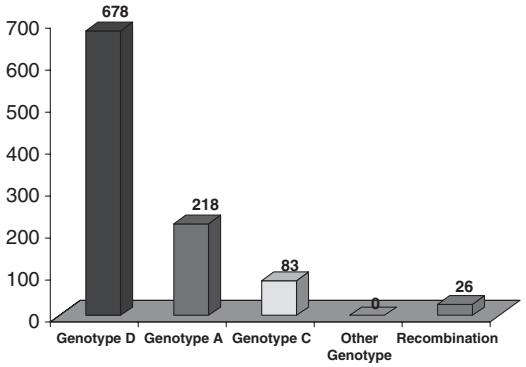

Fig. 2 Prevalence of HBV genotypes in India.

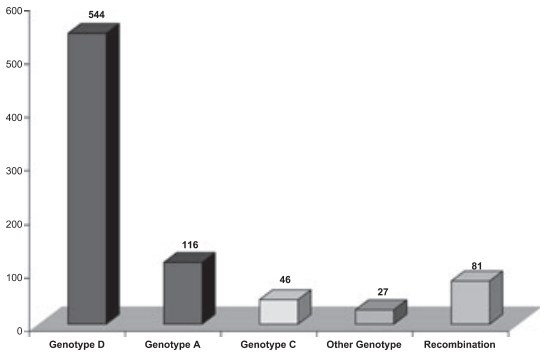

Fig. 3 Prevalence of HBV genotypes in Pakistan.

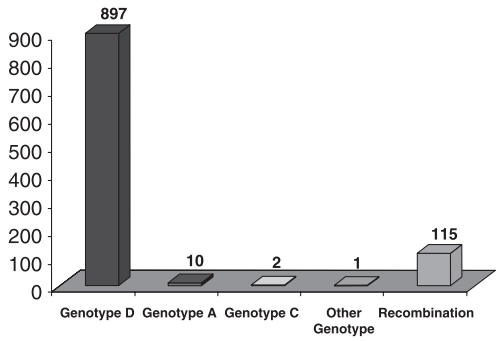

Fig. 4 Prevalence of HBV genotypes in Turkey.

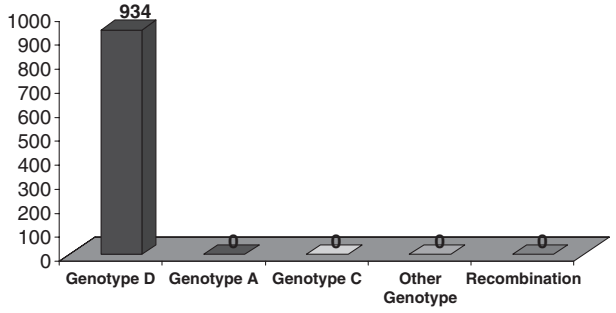

Fig. 5 Prevalence of HBV genotypes in Iran.

dominant genotype (87%), 1.3% both genotypes A and C and recombinants 12%. In Iran (Fig. 5), interestingly, no other genotype than D has been found.

According to archaeological and anthropological findings, the ancestors of Caucasians (Arians) firstly colonized the North of the Caspian Sea. Because of difficulties in agriculture and climate change, they migrated in three directions: one group moved west towards Europe, another group moved south to Iran (and established the ancient Persian Empire) and the last group migrated to India.

It might be that those people who acquired the genotype D virus before their migration, then transmitted the virus generation to generation after their migration. This is why the dominant genotype in India, Iran and most parts of Europe is D.

It is likely that selection of HBV genotypes A–H had been occurring in different parts of the world (perhaps related to immune pressure based largely on HLA types). After colonization of infected people with certain genotypes, their importation to this area occurred from intermixing of people. Further evolution occurred, giving some of these sequences a distinctive motif, and some genetic recombination between genotypes also occurred, which led to the heterogenous pattern in parts of this area (like Bangladesh, India and Pakistan). In some areas, isolation of people in the absence of intermixing with other genotypes let to a homologous pattern (like Iran and Turkey).

Due to lack of reliable data from other genotype D-dominant countries in The Middle East and North Africa, the analysis of whole data regarding genotype D evolution in the world is not conclusive. In addition, the sample size in some studies from such countries is quite small.

Future studies are needed to carry out work on the prevalence of HBV genotypes in the neighbouring countries of this region and comparing data to the European sequences which might lead to more precise conclusions.

REFERENCES

1 Holland J, Domingo E. Origin and evolution of viruses. *Virus Genes* 1998; 16: 13–21.
2 Carman WF. Infection associated with medical intervention: hepatitis viruses and HGV. *Br Med Bull* 1998; 54(3): 731–748.
3 Mizokami M, Orito E. Molecular evolution of hepatitis viruses. *Nippon Shokakibyo Gakkai Zasshi* 1999; 96: 1033–1043.
4 Kidd-Ljunggren K, Simonsen O. Reappearance of hepatitis B 10 years after kidney transplantation. *N Engl J Med* 1999; 341: 127–128.
5 Morozov V, Pisareva M, Groudinin M. Homologous recombination between different genotypes of hepatitis B virus. *Gene* 2000; 260: 55–65.
6 Torresi J. The virological and clinical significance of mutations in the overlapping envelope and polymerase genes of hepatitis B virus. *J Clin Virol* 2002; 25: 97–106.

7 Sugauchi F, Orito E, Ichida T *et al.* Hepatitis B virus of genotype B with or without recombination with genotype C over the precore region plus the core gene. *J Virol* 2002; 76: 5985–5992.

8 Carman WF. Molecular variants of hepatitis B virus. *Clin Lab Med* 1996; 16: 407–428.

9 Yang Z, Lauder IJ, Lin HJ. Molecular evolution of the hepatitis B virus genome. *J Mol Evol* 1995; 41: 587–596.

10 Orito E, Mizokami M, Ina Y *et al.* Host-independent evolution and a genetic classification of the hepadnavirus family based on nucleotide sequences. *Proc Natl Acad Sci USA* 1989; 86: 7059–7062.

11 Kodama K, Ogasawara N, Yoshikawa H, Murakami S. Nucleotide sequence of a cloned woodchuck hepatitis virus genome: evolutional relationship between hepadnaviruses. *J Virol* 1985; 56: 978–986.

12 Domingo E, Mas A, Yuste E *et al.* Virus population dynamics, fitness variations and the control of viral disease: an update. *Prog Drug Res* 2001; 57: 77–115.

13 Miller RH, Robinson WS. Common evolutionary origin of hepatitis B virus and retroviruses. *Proc Natl Acad Sci USA* 1986; 83: 2531–2535.

14 Okamoto H, Imai M, Kametani M, Nakamura T, Mayumi M. Genomic heterogeneity of hepatitis B virus in a 54-year-old woman who contracted the infection through materno-fetal transmission. *Jpn J Exp Med* 1987; 57: 231–236.

15 Hannoun C, Horal P, Lindh M. Long-term mutation rates in the hepatitis B virus genome. *J Gen Virol* 2000; 81: 75–83.

16 Carman WF. Molecular variants of hepatitis B virus. *Clin Lab Med* 1996; 16: 407–428.

17 Wilson JN, Nokes DJ, Carman WF. The predicted pattern of emergence of vaccine-resistant hepatitis B: a cause for concern? *Vaccine* 1999; 17: 973–978.

18 Carman W, Thomas H, Domingo E. Viral genetic variation: hepatitis B virus as a clinical example. *Lancet* 1993; 341: 349–353.

19 Lenhoff RJ, Luscombe CA, Summers J. Competition in vivo between a cytopathic variant and a wild-type duck hepatitis B virus. *Virol* 1998; 251: 85–95.

20 Bartholomeusz A, Locarnini S. Hepatitis B virus mutants and fulminant hepatitis B: fitness plus phenotype. *Hepatology* 2001; 34: 432–435.

21 Liu CJ, Kao JH, Shau WY, Chen PJ, Lai MY, Chen DS. Naturally occurring hepatitis B surface gene variants in chronic hepatitis B virus infection: correlation with viral serotypes and clinical stages of liver disease. *J Med Virol* 2002; 68: 50–59.

22 Gunther S, Fischer L, Pult I, Sterneck M, Will H. Naturally occurring variants of hepatitis B virus. *Adv Virus Res* 1999; 52: 25–137.

23 Carman WF, Thomas HC, Zuckerman AJ, Harrison TJ. Molecular Variants. Hong Kong: Churchill Livingston, 1998.

24 Carman WF. Variation in the core and X genes of hepatitis B virus. *Intervirology* 1995; 38: 75–88.

25 Okumura A, Takayanagi M, Aiyama T *et al.* Serial analysis of hepatitis B virus core nucleotide sequence of patients with acute exacerbation during chronic infection. *J Med Virol* 1996; 49: 103–109.

26 Bozkaya H, Ayola B, Lok AS. High rate of mutations in the hepatitis B core gene during the immune clearance phase of chronic hepatitis B virus infection. *Hepatology* 1996; 24: 32–37.

27 Bozkaya H, Akarca US, Ayola B, Lok AS. High degree of conservation in the hepatitis B virus core gene during the immune tolerant phase in perinatally acquired chronic hepatitis B virus infection. *J Hepatol* 1997; 26: 508–516.

28 Carman WF, Jacyna MR, Hadziyannis S *et al.* Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. *Lancet* 1989; 2: 588–591.

29 Abdelhamed AM, Kelley CM, Miller TG, Furman PA, Isom HC. Rebound of hepatitis B virus replication in HepG2 cells after cessation of antiviral treatment. *J Virol* 2002; 76: 8148–8160.

30 Cho SW, Hahm KB, Kim JH. Reversion from pre core/core promoter mutants to wild type HBV during the course of lamivudine therapy. *Hepatology* 2000; 32: 1163–1169.

31 Suzuki F, Suzuki Y, Tsubota A *et al.* Mutations of polymerase, precore and core promoter gene in hepatitis B virus during 5-year lamivudine therapy. *J Hepatol* 2002; 37: 824–830.

32 MacDonald DM, Holmes EC, Lewis JC, Simmonds P. Detection of hepatitis B virus infection in wild-born chimpanzees (*Pan troglodytes verus*): phylogenetic relationships with human and other primate genotypes. *J Virol* 2000; 74: 4253–4257.

33 Lanford RE, Chavez D, Brasky KM, Burns III RB, Rico-Hesse R. Isolation of a hepadnavirus from the woolly monkey, a New World primate. *Proc Natl Acad Sci USA* 1998; 95: 5757–5761.

34 Warren KS, Heeney JL, Swan RA, Heriyanto, Verschoor EJ. A new group of hepadnaviruses naturally infecting orangutans (*Pongo pygmaeus*). *J Virol* 1999; 73: 7860–7865.

35 Simmonds P. Reconstructing the origins of human hepatitis viruses. *Philos Trans R Soc Lond B Biol Sci* 2001; 356: 1013–1026.

36 Starkman SE, MacDonald DM, Lewis JC, Holmes EC, Simmonds P. Geographic and species association of hepatitis B virus genotypes in non-human primates. *Virol* 2003; 314: 381–393.

37 Norder H, Ebert JW, Fields HA, Mushahwar IK, Magnius LO. Complete sequencing of a gibbon hepatitis B virus genome reveals a unique genotype distantly related to the chimpanzee hepatitis B virus. *Virol* 1996; 218: 214–223.

38 Norder H, Courouce AM, Magnius LO. Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. *Virol* 1994; 198: 489–503.

39 Magnius LO, Norder H. Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. *Intervirology* 1995; 38: 24–34.

40 Simmonds P. The origin and evolution of hepatitis viruses in humans. *J Gen Virol* 2001; 82: 693–712.

41 Bollyky PL, Holmes EC. Reconstructing the complex evolutionary history of hepatitis B virus. *J Mol Evol* 1999; 49: 130–141.

42 Takahashi K, Brotman B, Usuda S, Mishiro S, Prince AM. Full-genome sequence analyses of hepatitis B virus (HBV) strains recovered from chimpanzees infected in the wild: implications for an origin of HBV. *Virol* 2000; 267: 58–64.
43 Robertson BH. Viral hepatitis and primates: historical and molecular analysis of human and nonhuman primate hepatitis A, B, and the GB-related viruses. *J Viral Hepat* 2001; 8: 233–242.
44 Lindh M, Hannoun C, Dhillon AP, Norkrans G, Horal P. Core promoter mutations and genotypes in relation to viral replication and liver damage in East Asian hepatitis B virus carriers. *J Infect Dis* 1999; 179: 775–782.
45 Blackberg J, Kidd-Ljunggren K. Occult hepatitis B virus after acute self-limited infection persisting for 30 years without sequence variation. *J Hepatol* 2000; 33: 992–997.
46 McMahon BJ, Lanier AP, Wainwright RB. Hepatitis B and hepatocellular carcinoma in Eskimo/Inuit population. *Int J Circumpolar Health* 1998; 57(Suppl. 1): 414–419. (Review).
47 Jazayeri MS, Basuni AA, Cooksley G, Locarnini S, Carman WF. Hepatitis B virus genotypes, core gene variability and ethnicity in the Pacific region. *J Hepatol* 2004; 41(1): 139–146.