The encoding reactions had a final volume of 15 µl, including 0.5 µl of multiplex PCR products, 20 mM Tris-HCl, pH 7.6, 25 mM potassium acetate, 10 mM magnesium acetate, 10 mM DTT, 1 mM NAD, 0.1% Triton X-100 (1× Taq DNA ligase buffer) with 0.33 nM of each probe and 5 U Taq DNA ligase (New England BioLabs, Ipswich, MA, USA). Encoding reactions were conducted using a TGradient or PTC-225 under the following conditions: 95°C for 5 min, followed by 58°C for 15 min. The reaction was stopped by holding the temperature at 10°C.

The labeling reactions had a final volume of 12 µl, including 6 µl of ligation products, 0.5 µM each labeled primer (Alexa555-cED-1 and Alexa647-cED-2), 2.5 nM each D1 primer (D1_i), 50 mM KCl, 2 mM Mg^{2+}, 0.1 mM DTT, 0.2 mM each dNTP (N = A, G, C), 0.1 mM [^{3}H]-dTTP, 0.25 mg/ml activated salmon sperm DNA (1× *Ex Taq* Buffer) and 0.05 U of *Ex Taq*™ polymerase (TaKaRa, Shiga, Japan). Labeling reactions were conducted using a TGradient or PTC-225 under the following conditions: first held at 95°C for 1 min, followed by 30 cycles of 95°C for 30 s, 55°C for 6 min and 72°C for 30 s. The reaction was stopped by holding the temperature at 10°C. Total running times for labeling using TGradient and PTC-225 were 3 h 49 min 48 s and 4 h 8 min 48 s, respectively.

In the detection step, a hybridization mixture was prepared by mixing 6.25 µl of labeling products with 8.75 µl of hybridization buffer containing 0.5× SSC, 0.1% SDS, 15% formamide, 1 mM EDTA and 3.125 fmol of hybridization control (Alexa555-labeled D1_100 and Alexa647-labeled D1_100). The hybridization control was prepared for ensuring the hybridization step. Ten microliters of hybridization mixture was applied to each block on the universal DNA chip. Hybridization was carried out for 30 min at 37°C in a hybridization oven (ThermoStat plus; Eppendorf, Ham, Germany). After hybridization, glass slides were washed in washing buffer (0.1× SSC, 0.1% SDS) by shaking at 60 rpm for 3 min. Glass slides were consecutively washed in distilled water by shaking at 60 rpm for 1 min and then dried up by centrifugation at 500× g for 1 min. Hybridization images were scanned at photomultiplier voltages of 400 V for Alexa555 and 480 V for Alexa647 using a commercially available DNA chip scanner and fluorescence image analysis was performed using commercially available software (GenePix 4000B unit and GenePix Pro 4.1 software package; Molecular Devices, Sunnyvale, CA, USA).

Labeling with Kapa 2GFast HotStart DNA polymerase

The labeling reactions with Kapa 2GFast HotStart DNA polymerase had a final volume of 12 µl, including 6 µl of ligation products, 0.5 µM each labeled primer (Alexa555-cED-1 and Alexa647-cED-2), 2.5 nM each D1 primer (D1_i), 1.5× KAPA2G Buffer (including 2.25 mM Mg^{2+}), an additional 2.25 mM Mg^{2+} (final concentration of Mg^{2+}: 4.5 mM), 0.2 mM dNTPs and 0.4 U of Kapa 2GFast HotStart DNA polymerase. Labeling reactions were conducted using a TGradient or PTC-225 under the following conditions: first held at 95°C for 1 min, followed by 30 cycles of 95°C for 15 s, 55°C for 120 s and 72°C for 5 s. The reaction was stopped by holding the temperature at 10°C. The total running times for labeling using TGradient and PTC-225 were 1 h 29 min 48 s and 1 h 48 min 34 s, respectively.

Results

Singleplex PCR using 192 pairs of locus-specific primers

Singleplex PCR was conducted under the same reaction condition with multiplex PCR using 25 ng of genomic DNA to ensure target amplicon detection and to confirm the emergence of extra bands (unexpected amplicons). Singleplex PCR with 192 pairs of locus-specific primers revealed that most of the primer pairs are able to achieve sensitive detection (concentration of target amplicon ≥5 nM) and specific amplification without extra bands (concentration of unexpected amplicons <2 nM) except for 14 pairs of primers; low sensitivity (<5 nM) for 5 pairs of primers (61, 99, 102, 189 and 191) and low specificity with extra bands (≥2 nM) for 9 pairs of primers (40, 56, 62, 70, 91, 106, 149, 173 and 174) (Figure 2 and Table S2). Five pairs among the 9 low-specific primer pairs with extra bands (62, 70, 149, 173 and 174) resulted from heteroduplex formation of target amplicons during polyacrylamide gel electrophoresis. Despite the presence of extra bands, the remaining 4 pairs of low-specific primers had a target amplicon with a detectable concentration ≥5 nM.

Validation of efficacy of 192-plex PCR by 96-plex genotyping with the DigiTag2 assay

The DigiTag2 assay enables the simultaneous analysis of 96 target SNPs in: (1) multiplex PCR with locus-specific primers to amplify target genomic regions including target SNPs; (2) multiple oligonucleotide ligation assay with locus-specific probes to determine the genotype of each SNP; and (3) hybridization to the universal DNA chip tethered with probe sequences identical to D1_i (23-mer) (Figure 1) [10]. The validity of 192-plex PCR was assessed with 96 individual DNAs (population control samples) by comparing two sets of 96-plex genotype calls acquired from 96-plex PCR with those from 192-plex PCR (Table 1).

Conversion rate shows the proportion of successfully genotyped SNPs with fewer than 3 undetected samples after excluding low-quality genotyping data, which had more than 5 undetected SNPs in a total of 96 SNPs. However, the composition of failed SNPs in genotyping was not identical, and the conversion rate showed no differences between 192-plex PCR and 96-plex PCR. For the 1st set of 96 SNPs, 7 SNPs among 10 failed SNPs were matched between 192-plex PCR and 96-plex PCR, and for the 2nd set, 8 SNPs among the 9 failed SNPs were matched. The average call rate for successfully genotyped SNPs was over 99.79% for both sets of 96-plex genotyping, even if 192-plex PCR products were adopted for target preparation. Reproducibility was determined by independent genotyping with 96 individuals twice. As a consequence, four discordant genotype calls were observed in the duplicated genotyping data. Concordance of genotype calls between 192-plex PCR and 96-plex PCR was determined using 6,290 genotype calls for the 1st set and 7,884 genotype calls for the 2nd set. Consequently, 14,171 out of 14,174 genotype calls were matched by comparison with 83 SNPs for the 1st set and 86 SNPs for the 2nd set. In total, 3 discordant genotype calls were observed (Figure 3).

Short-acting multiplex PCR by use of Kapa 2GFast HotStart DNA polymerase

Kapa 2GFast HotStart DNA polymerase was employed to perform multiplex PCR with the locus-specific primers for target preparation in genotyping with the DigiTag2 assay. To optimize reaction conditions with Kapa 2GFast HotStart DNA polymerase, singleplex PCR was conducted using 25 ng of genomic DNA with three randomly chosen pairs of locus-specific primers. The designed amplicon sizes for the three pairs of primers were 501 bp, 671 bp and 492 bp. We performed singleplex PCR using a two-step protocol (denature and extension steps) with varied extension periods (15 s, 30 s, 60 s and 120 s) and with varied Mg^{2+} concentrations (3.0 mM and 4.5 mM) (Figure 4). The most sensitive detection and highest levels of amplification for the three

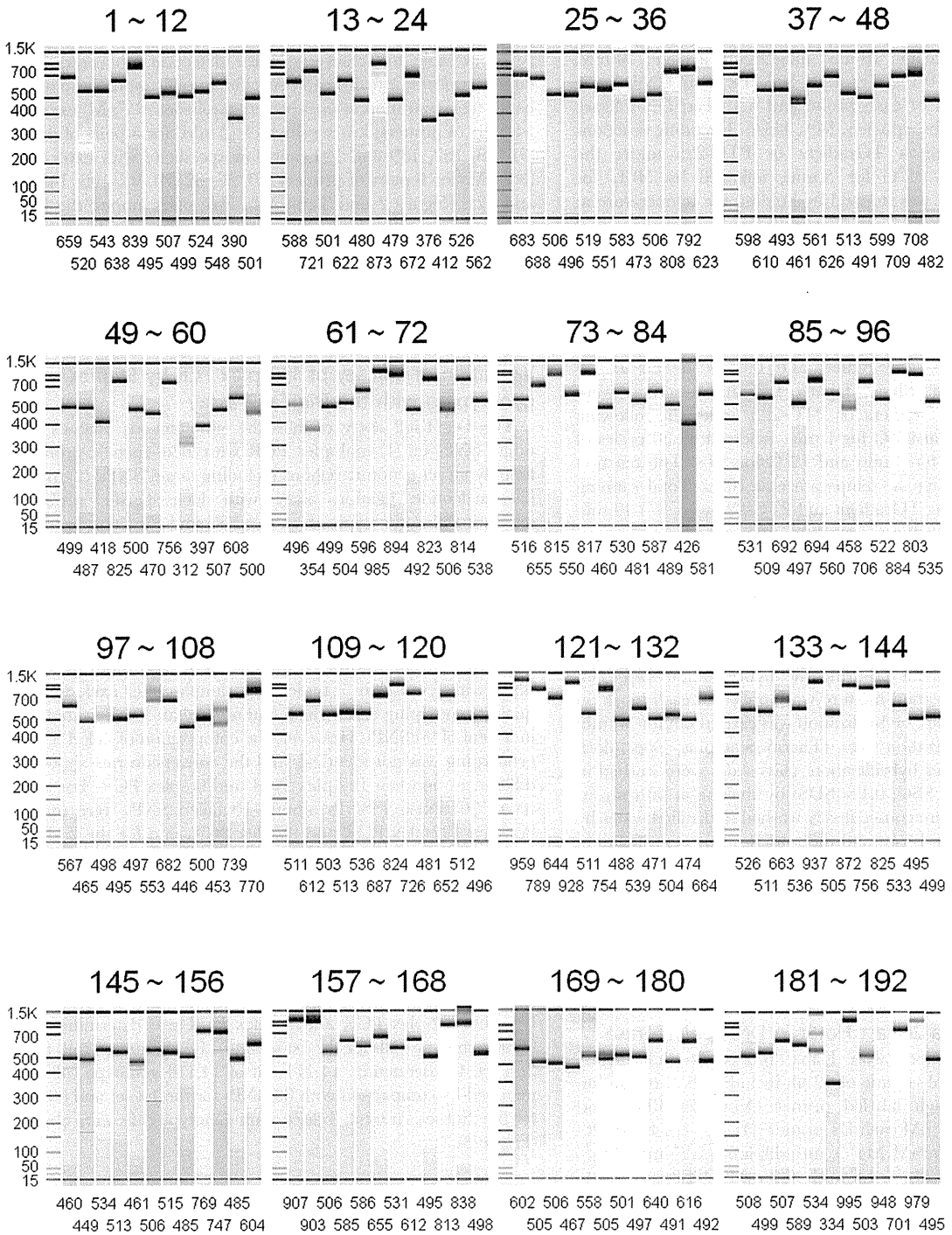

Figure 2. Electropherogram of singleplex PCR products with 192 pairs of locus-specific primers. The designed amplicon size is depicted below each lane.
doi:10.1371/journal.pone.0029967.g002

pairs of primers were observed with 120 s for the extension period and 4.5 mM for the Mg^{2+} concentration. The total running time for multiplex PCR with locus-specific primers was less than 2 hours, which is about 3 h 30 min shorter than the previous protocol (see MATERIALS AND METHODS).

The total running time of the DigiTag2 assay was markedly reduced when the labeling step was also conducted using Kapa 2GFast HotStart DNA polymerase instead of *Ex Taq* polymerase. When the DigiTag2 assay was conducted with Kapa 2GFast HotStart DNA polymerase for multiplex PCR and labeling step, the total running time of the assay was about 7 hours, which is about 6 hours shorter than the previously used protocol in combination with QIAGEN Multiplex PCR Kit for multiplex PCR and *Ex Taq* polymerase for the labeling step.

Table 1. Validation of efficacy of 192-plex PCR by 96-plex genotyping.

		192-plex PCR	96-plex PCR
1st set	Conversion rate	86/96 SNP	86/96 SNP
	Call rate	99.84% (7,728/7,740 genotype)	99.81% (6,695/6,708 genotype)
	reproducibility	99.99% (7,288/7,289 genotype)	100% (6,121/6,121 genotype)
	concordance	99.98% (6,289/6,290 genotype)	
2nd set	Conversion rate	87/96 SNP	87/96 SNP
	Call rate	99.79% (8,074/8,091 genotype)	99.79% (8,161/8,178 genotype)
	reproducibility	99.97% (7,792/7,794 genotype)	99.99% (7,712/7,713 genotype)
	concordance	99.97% (7,882/7,884 genotype)	

doi:10.1371/journal.pone.0029967.t001

Table 2 summarizes the quality parameters (conversion rate, call rate, reproducibility and concordance) when genotyping was conducted with 192-plex PCR or 96-plex PCR by use of Kapa 2GFast HotStart DNA polymerase. The conversion rate was slightly decreased when multiplex PCR was conducted in 192-plex form. However, the conversion rates were better than those observed when multiplex PCR was conducted with the QIAGEN Multiplex PCR Kit. The composition of failed SNPs in genotyping was not consistent for the 1st set of 96 SNPs, in which 4 SNPs were matched between 192-plex PCR and 96-plex PCR. For the 2nd set, a total of 8 failed SNPs in the 96-plex PCR were completely matched to those in the 192-plex PCR. When the composition of failed SNPs were compared between Kapa 2GFast HotStart DNA polymerase and QIAGEN Multiplex PCR Kit, the 1st set had 5 matched SNPs in a total of 8 failed SNPs for 192-plex PCR, and 4 matched SNPs in 5 failed SNPs for 96-plex PCR. From the 2nd set, 5 SNPs in a total of 9 failed SNPs were matched when 192-plex PCR was conducted and 4 SNPs in a total of 8 failed SNPs were matched when 96-plex PCR was conducted. The average call rate for successfully genotyped SNPs was over 99.76% for both sets of 96-plex genotyping, even if 192-plex PCR products were adopted for target preparation. The reproducibility was 100% for the 2nd set; however, three discordant genotype calls were observed for the 1st set. With regard to the concordance of genotype calls between 96-plex PCR and 192-plex PCR, only one discordant genotype call was observed in the comparison for the 1st set, and no discordant genotype calls were observed in the 2nd set.

Table 3 shows the concordance rate in comparison with the genotype calls by the use of Kapa 2GFast HotStart DNA polymerase or QIAGEN Multiplex PCR Kit for multiplex PCR. For the 1st set, there were 4 discordant genotype calls with 96-plex PCR and 8 discordant genotype calls with 192-plex PCR. For the 2nd set of 96 SNPs, there was one discordant genotype call in genotyping with 96-plex PCR and 192-plex PCR.

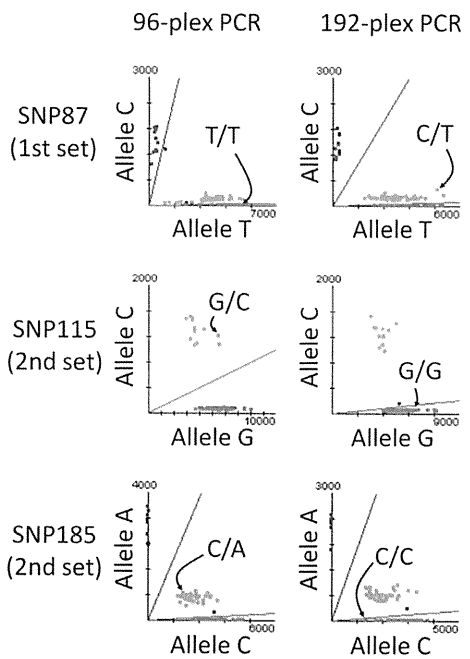

Figure 3. Scatter plots for three SNPs with 3 discordant genotypes. Scatter plots in genotyping with 192-plex PCR and 96-plex PCR are depicted side-by-side. The genotypes of discordant samples are indicated in the scatter plots by arrows.
doi:10.1371/journal.pone.0029967.g003

Discussion

The locus specific primers sufficiently worked in a multiplex form under the same reaction conditions without any optimization processes, either 96-plex PCR or 192-plex PCR. We also found that either 96-plex PCR or 192-plex PCR could be accomplished within two hours through the use of Kapa 2GFast HotStart DNA polymerase. The total running time of the DigiTag2 assay was shortened by 6 hours over the original 13-hour long protocol using Kapa 2GFast HotStart DNA polymerase for both multiplex PCR and the labeling step. The quality parameters (conversion rate, call rate, reproducibility and concordance) observed in genotyping with the new protocol were the same as those observed in the original protocol using QIAGEN Multiplex PCR Kit for multiplex PCR and *Ex Taq* polymerase for the labeling step. The DigiTag2 assay worked with a conversion rate of over 93.2% (179 / 192 SNPs), average call rate of over 99.80% (16,789/16,823 genotypes) and reproducibility of over 99.99% (16,135/16,136 genotypes) using 96-plex PCR under the new protocol. The composition of successfully genotyped SNPs was different when the genotype calls were acquired using the different polymerases (Kapa 2GFast HotStart DNA polymerase and QIAGEN Multiplex PCR Kit), which would result from a varying amplification bias in multiplex PCR. We also found that 192-plex PCR with locus-specific primers worked in 96-plex genotyping with the DigiTag2 assay, giving the same quality parameter data as those observed in genotyping with 96-plex PCR. However, the

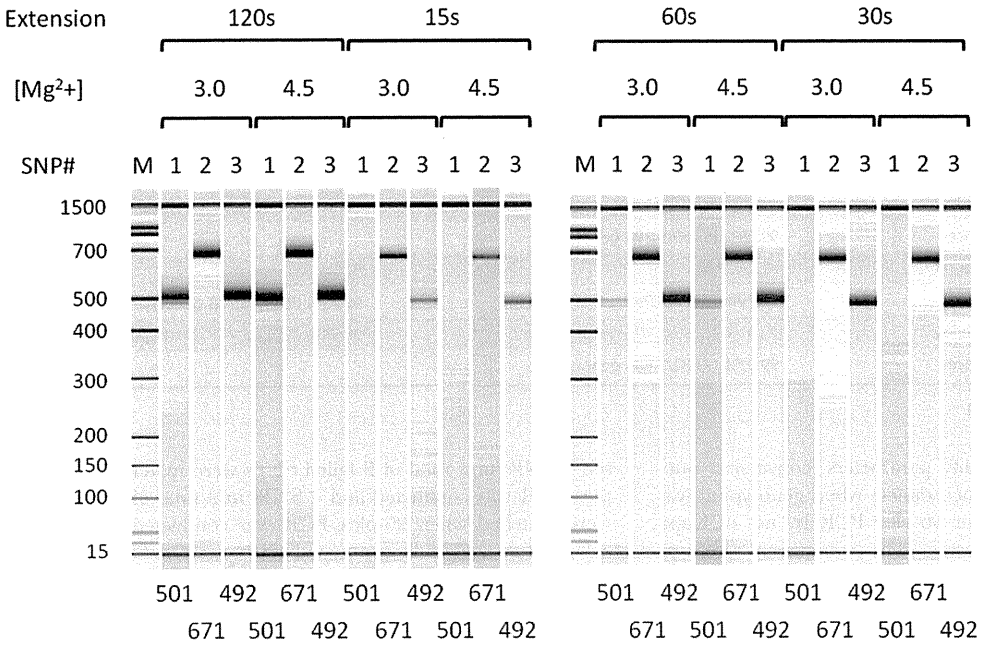

Figure 4. Electropherogram of singleplex PCR products using Kapa 2GFast HotStart DNA polymerase. Singleplex PCR was performed with varied extension periods (15 s, 30 s, 60 s and 120 s) and with varied Mg^{2+} concentrations (3.0 mM and 4.5 mM) using three pairs of locus-specific primers. The designed amplicon size is depicted below each lane.
doi:10.1371/journal.pone.0029967.g004

composition of successfully genotyped SNPs was not consistent between 192-plex PCR and 96-plex PCR, which may be explained by changing the interactions between primer pairs in 192-plex PCR and in 96-plex PCR. The composition of successful SNPs was not consistent when using different polymerases or multiplex systems in the multiplex PCR, which casts some shadows on the reliability of the assay. Regardless of the existing shadows, indeed, 96-plex and 192-plex PCR work with a high conversion rate in genotyping over 93.2%. To clear the existing shadows, it is necessary to continuously accumulate genotyping data.

In this study, fifteen discordant genotype calls were in total observed in the comparison of genotype calls with: i) duplicated genotyping data; ii) genotyping data by use of 192-plex PCR and 96-plex PCR; and iii) genotyping data with different types of polymerases (Table S3). Table S3 shows the genotype calls acquired 8 times under different conditions. All fifteen discordant genotype calls were analyzed with direct sequencing, of which 13 genotype calls were determined. In 8 of 15 discordant genotype calls, the genotype calls were completely different depending on the type of polymerase. The genotype calls acquired using Kapa 2GFast HotStart DNA polymerase were 100% concordant (6 of 6) with those acquired by direct sequencing. This suggests that SNP allelic bias in PCR amplification readily occurred with the QIAGEN Multiplex PCR Kit; however, the error rate in genotyping was only 0.04% (6 out of 14,886 genotypes). The remaining 7 discordant genotype calls were randomly observed in 1 out of 8 different conditions. This shows that the random error rates were almost equal in the genotype data acquired with both types of polymerases (4 out of 62,227 genotypes for QIAGEN Multiplex PCR Kit and 3 out of 66,008 genotypes for Kapa 2GFast HotStart DNA polymerase).

Among the five low-sensitivity primer pairs found on singleplex PCR (61, 99, 102, 189 and 191), no amplicons were detected by primer pair 189 and low concentrations (<5 nM) of amplicon were detected by the 4 other primer pairs (Table S2). Therefore, the SNP189 failed in genotyping, independently of the type of

Table 2. Validation of efficacy of 192-plex and 96-plex PCR with Kapa 2GFast HotStart DNA polymerase.

		192-plex PCR	96-plex PCR
1st set	Conversion rate	88/96 SNP	91/96 SNP
	Call rate	99.84% (8,259/8,272 genotype)	99.76% (8,443/8,463 genotype)
	reproducibility	99.97% (8,069/8,071 genotype)	99.99% (8,339/8,340 genotype)
	concordance	99.99% (7,982/7,983 genotype)	
2nd set	Conversion rate	87/96 SNP	88/96 SNP
	Call rate	99.91% (8,171/8,178 genotype)	99.83% (8,346/8,360 genotype)
	reproducibility	100% (7,705/7,705 genotype)	100% (7,796/7,796 genotype)
	concordance	100% (8,161/8,161 genotype)	

doi:10.1371/journal.pone.0029967.t002

Table 3. Concordance of genotype calls between Kapa 2GFast HotStart DNA polymerase and QIAGEN Multiplex PCR Kit.

		Kapa 2G	QIAGEN
1st set	96-plex PCR	99.94% (6,513/6,517 genotype)	
	192-plex PCR	99.89% (7,441/7,449 genotype)	
2nd set	96-plex PCR	99.99% (7,778/7,779 genotype)	
	192-plex PCR	99.99% (7,700/7,701 genotype)	

doi:10.1371/journal.pone.0029967.t003

polymerase and multiplicity in multiplex PCR (192-plex or 96-plex). However, the SNP191, which was amplified by primer pair 191, was successfully genotyped only when the QIAGEN Multiplex PCR Kit was used for the multiplex PCR. The concentration of amplicon amplified by primer pair 99 was the same as the 2.8 nM observed with the amplicon amplified by primer pair 191. SNP99, which was amplified by primer pair 99, was successfully genotyped independently of polymerase type and multiplicity in multiplex PCR (192-plex or 96-plex). These results suggest that the sensitivity in genotyping with Kapa 2GFast HotStart DNA polymerase was lower than the previously used protocol with QIAGEN Multiplex PCR Kit. These results would be explained by a biased amplification with the shortened protocol using Kapa 2GFast HotStart DNA polymerase, which tends to lead to a consequent biased genotyping. However, the investigated number of primer pairs would not be sufficient to decide the sensitivity in genotyping; therefore, it is necessary to continuously accumulate genotyping data. As the investigated number of primer pairs was only 192 (384 primers) in this study, melting temperature of each primer and the number of potential amplicons predicted by the MFE primer software were strongly associated with low sensitivity and low specificity in an amplification, respectively (multiple regression analysis, $P = 1.26 \times 10^{-37}$ and $P = 1.52 \times 10^{-21}$, respectively).

Through the use of Kapa 2GFast HotStart DNA polymerase, the genotype calls for 96 SNPs can be acquired in about 7 hours by the DigiTag2 assay. The genotyping platform with high conversion rate plays an important role for the replication studies to identify the disease associated genes from candidate loci found in the GWAS (genome-wide association study). The DigiTag2 assay with an improved protocol will be an efficient platform for screening an intermediate number of SNPs (tens to hundreds of sites) in the replication studies. Because of limitations in the variation of DNA coded numbers (DCNs), 192-plex genotyping is not available for the current DigiTag2 assay. However, 192-plex PCR can save genomic DNA samples and time for target preparation. Moreover, 192-plex PCR is also available for direct-sequencing and other PCR-based assays to amplify the target regions from genomic DNA.

Supporting Information

Table S1 Sequence information of 192 pairs of locus specific primer.
(XLSX)

Table S2 Results of singleplex PCR with 192 pairs of locus specific primer.
(XLSX)

Table S3 The 15 discordant genotype calls in 8 different conditions.
(XLSX)

Acknowledgments

We would like to thank M. Takasu for technical support, and H. Adachi, N. Tabei and J. Fujimiya (Dynacom Co., Ltd.) for assistance with primer and probe design.

Author Contributions

Conceived and designed the experiments: NN KT. Performed the experiments: YM MS. Analyzed the data: NN YM MS. Contributed reagents/materials/analysis tools: NN YM MS. Wrote the paper: NN KT.

References

1. Deter J, Gala M, Charbonnel N, Cosson JF (2009) Characterization and PCR multiplexing of polymorphic microsatellite loci in the whipworm *Trichuris arvicolae*, parasite of arvicoline rodents and their cross-species utilization in *T. muris*, parasite of murines. Mol Biochem Parasitol 167: 144–146.
2. Hosseini-Maaf B, Hellberg Å, Chester MA, Olsson ML (2007) An extensive polymerase chain reaction–allele-specific polymorphism strategy for clinical ABO blood group genotyping that avoids potential errors caused by null, subgroup, and hybrid alleles. Transfusion 47: 2110–2125.
3. Goguet de la Salmonière YO, Kim CC, Tsolaki AG, Pym AS, Siegrist MS, et al. (2004) High-throughput method for detecting genomic-deletion polymorphisms. J Clin Microbiol 42: 2913–2918.
4. Landegren U, Nilsson M (1997) Locked on target: strategies for future gene diagnostics. Ann Med 29: 585–590.
5. Pemov A, Modi H, Chandler DP, Bavykin S (2005) DNA analysis with multiplex microarray-enhanced PCR. Nucleic Acids Res 33: e11.
6. Meuzelaar LS, Lancaster O, Pasche JP, Kopal G, Brookes AJ (2007) MegaPlex PCR: a strategy for multiplex amplification. Nat Methods 4: 835–837.
7. Sun Z, Chen Z, Hou X, Li S, Zhu H, et al. (2008) Locked nucleic acid pentamers as universal PCR primers for genomic DNA amplification. Plos One 3: e3701.
8. Dahl F, Gullberg M, Stenberg J, Landegren U, Nilsson M (2005) Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments. Nucleic Acids Res 33: e71.
9. Krjutškov K, Andreson R, Mägi R, Nikopensius T, Khrunin A, et al. (2008) Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays. Nucleic Acids Res 36: e75.
10. Nishida N, Tanabe T, Takasu M, Suyama A, Tokunaga K (2007) Further development of multiplex single nucleotide polymorphism typing method, the DigiTag2 assay. Anal Biochem 364: 78–85.
11. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, et al. (2009) Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 41: 1105–1109.
12. Miyagawa T, Kawashima M, Nishida N, Ohashi J, Kimura R, et al. (2009) Variant between CPT1B and CHKB associated with susceptibility to narcolepsy. Nat Genet 40: 1324–1328.
13. Qu W, Shen Z, Zhao D, Yang Y, Zhang C (2009) MFEprimer: multiple factor evaluation of the specificity of PCR primers. Bioinformatics 25: 276–278.

J Gastroenterol (2011) 46:117–124
DOI 10.1007/s00535-010-0307-4

ORIGINAL ARTICLE—LIVER, PANCREAS, AND BILIARY TRACT

A population-based cohort study for the risk factors of HCC among hepatitis B virus mono-infected subjects in Japan

Atsunori Kusakabe · Yasuhito Tanaka · Manami Inoue · Fuat Kurbanov · Kanako Tatematsu · Shunsuke Nojiri · Takashi Joh · Shoichiro Tsugane · Masashi Mizokami

Received: 20 May 2010 / Accepted: 2 August 2010 / Published online: 7 September 2010

Abstract
Background There have only been a few prospective studies investigating risk factors associated with the development of hepatocellular carcinoma (HCC) among chronic hepatitis B patients all over the world, and no study has been conducted in Japanese population.
Methods A population-based cohort consisting of 19393 subjects (middle aged or older) with over 13 years' follow-up was investigated in Japan.
Results Of 19393 subjects, 479 had hepatitis B virus (HBV) mono-infection (2.5%). During the 245923 person-years' follow-up (average follow-up period 12.7 years), 13 cases of newly diagnosed HCC were documented in the HBV mono-infected group. Several factors at baseline (male, smoking, alanine aminotransferase, the positivity of HBe antigen and HB core-related antigen, the proportion of HBV DNA $\geq$ 5 log copies/mL, T1753V mutation, and A1762T/G1764A double mutation) were significantly associated with HCC among HBV mono-infected subjects. Multivariate-adjusted Cox hazard model showed that A1762T/G1764A (hazard ratio 7.05 [95% confidence interval (CI) 1.03–48.12, $P = 0.046$]) was the only independent risk factor for the development of HCC. Kaplan–Meier method also showed that the probability of HCC occurrence-free was significantly lower in HBV mono-infected subjects with A1762T/G1764A double mutation than those without these mutations.
Conclusion HBV mono-infected subjects with A1762T/G1764A double mutation could be at high risk of HCC development during the natural course of HBV infection.

Keywords Hepatitis B virus · Hepatocellular carcinoma · Risk factor · Japan

This article is authored in cooperation with the JPHC Study Group, members of which are listed in the Appendix.

The online version of the editorial referring to this article can be found under doi:10.1007/s00535-010-0314-5.

A. Kusakabe · S. Nojiri · T. Joh
Department of Gastroenterology and Metabolism, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan

Y. Tanaka (✉) · F. Kurbanov · K. Tatematsu
Department of Virology and Liver Unit, Nagoya City University Graduate School of Medical Sciences, Kawasumi, Mizuho, Nagoya 467-8601, Japan
e-mail: ytanaka@med.nagoya-cu.ac.jp

M. Inoue · S. Tsugane
Epidemiology and Prevention Division, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan

M. Mizokami
The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan

Introduction

Hepatitis B virus (HBV) infection is one of the most common viral infections among humans and chronic hepatitis B (CHB) infection affects 350–400 million people worldwide [1]. Furthermore, CHB infection is the most serious etiology of hepatocellular carcinoma (HCC) in Asia [2], and the second most in Japan [3]. In fact, it was reported that individuals who happened to have HBV chronic infections have higher risk of developing HCC than uninfected ones [2]. Hence, it is very important to identify the risk factors associated with the development of HCC among CHB patients.

The mechanisms of oncogenesis of HBV remain obscure, but several factors have been identified to be associated with a high risk of developing of HCC among CHB patients. In regard to virological factors, hepatitis B e antigen (HBeAg) positivity, high serum HBV DNA levels, HBV genotype (genotype C more than genotype B), pre-core mutations (G1896A), and core promoter mutations (A1762T/G1764A) have been reported [4–11]. In addition, other risk factors such as host factors (male, increasing age) [12] or disease factors (alanine aminotransferase (ALT) levels, cirrhosis) [12, 13] have been mentioned. Among these factors, A1762T/G1764A double mutation has been thought to be the most convincing factor associated with the development of HCC among CHB patients [9, 10, 14–16]. But most of these risk factors were identified by cross-sectional or case-cohort studies with relatively limited numbers of study population, and it has been unclear which factor is the most important in developing HCC. A prospective study for the development of HCC associated with HBV infection is required to solve these problems, but there have been only a few prospective studies about this from Asia [17–19].

In the present report, we describe results of the first large-scale population-based cohort study carried out in Japan to identify the risk factors for the development of HCC among HBV mono-infected Japanese patients.

Methods

Study population and baseline survey

The Japan Public Health Center-based Prospective Study (JPHC Study) was launched in 1993–1994 in the Cohort II among registered Japanese inhabitants aged 40–69 years at the time of the baseline survey in 6 prefectural public health center areas. Details of the study design have been described elsewhere [20, 21]. A part of one public health center area was excluded since its study population was defined differently to the others. Initially, we defined a population-based cohort of 68975 subjects after exclusion of ineligible subjects.

At baseline, a self-administered questionnaire survey on various lifestyle factors was conducted (response rate 82%). Blood was also provided voluntarily by 39% of the respondents during health checkup provided by the respective local government. Clinical laboratory examinations such as ALT and γ-glutamyl transpeptidase (γ-GTP) were carried out on these samples by using commercial assays.

For the present analysis, we restricted subjects to those who responded to the questionnaire and for whom blood samples were available. We further excluded those with a history of HCC and those with missing data on HBV and hepatitis C virus (HCV) infection status. Finally, a total of 19393 individuals were included in the present analysis.

Subjects were checked for hepatitis B surface antigen (HBsAg) and antibody to hepatitis C virus (anti-HCV). Patients with HBsAg alone were classified into the HBV mono-infected group. All anti-HCV positive subjects were excluded from the study to eliminate the confounding effect of HCV infection on the HCC occurrence.

A baseline self-administered questionnaire survey on several lifestyle factors (body mass index (BMI), alcohol intake, and smoking) was conducted. Alcohol intake exceeding 80 g/day was defined as excessive.

Follow-up and identification of HCC

Enrolled subjects were followed from the baseline survey for average 12.7 years (until 31 December 2006). Residence status and survival were confirmed through the residential registry. Residence and death registration are required in Japan by law and the registries are believed to be complete. Inspection of the resident registry is legally sanctioned by resident registration law.

The occurrence of HCC was determined by notification from the participating hospitals in the study areas and by data linkage with population-based cancer registries. Death certificates were used as a supplementary information source to confirm the cause of death. The site of origin and histological type were coded by using the *International Classification of Diseases for Oncology, 3rd edn.* (ICD-O-3; C22.0) [22]. Through this procedure, a total of 110 newly diagnosed HCC cases were identified among the 19393 subjects during the follow-up period.

Serological analysis

HBsAg, anti-HBs, and anti-HCV in sera were determined by commercial enzyme immunoassay (Axcysm, Abbott Japan, Tokyo, Japan; or Lumipulse forte, Fujirebio, Tokyo, Japan). HBV core-related antigen (HBcrAg), which correlates with HBV DNA in serum, was measured in serum, using a chemiluminescence enzyme immunoassay (CLEIA) as described previously (detection limit 1.0 kU/mL) [23]. In brief, HBcrAg comprises HBcAg and HBeAg, which share the first 149 amino acids encoded by the core gene. Using the specific enzyme immunoassay within the common region, this assay is able to detect HBcAg and HBeAg even in anti-HBc or anti-HBe antibody-positive specimens [23].

Quantification of serum HBV DNA

HBV DNA parts spanning the S gene [nt 427–606] were amplified by real-time detection polymerase chain reaction (RTD-PCR) using the previously described protocol [24],

with slight modification [25]. The detection limit of this assay was 100 copies/mL.

Sequencing of HBV genome

To investigate HBV genotypes and specific mutations, nucleic acids were extracted from serum samples (100 μL) using the QIAamp DNA extraction kit (Qiagen, Hilden, Germany) and subjected to polymerase chain reaction (PCR) for amplifying genomic areas bearing enhancer II/core promoter/pre-core regions [nt 1628–2364], as described previously [26]. Amplicons were sequenced directly with use of the ABI Prism Big Dye ver. 3.0 kit in the AMI 3100 DNA automated sequencer (Applied Biosystems, Foster City, CA, USA). All sequences were analyzed in both forward and backward directions. HBV genotypes were determined by molecular evolutionary analysis. Reference sequences were retrieved from the DDBJ/EMBL/GenBank database and aligned by CLUSTAL X, and then genetic distances were estimated with the 6-parameter method in the Hepatitis Virus Database (http://s2as02.genes.nig.ac.jp/) [27]. Based on obtained distances, phylogenetic trees were constructed by the neighbor-joining (NJ) method with the mid-point rooting option. To confirm the reliability of the phylogenetic trees, bootstrap resampling tests were performed 1000 times.

Statistical analysis

Person-years of follow-up from the date of baseline survey were calculated until the date of diagnosis of HCC, date of death, relocation from the public health center area, or 31 December 2006, whichever occurred first.

Statistical differences were evaluated by Mann–Whitney U test, Fisher's exact probability test, and χ^2 test, where appropriate. Differences were considered to be statistically significant at $P < 0.05$. Multivariate-adjusted Cox hazard model were utilized to sort out independent risk factors for the development of HCC among HBV mono-infected subjects. In addition, Kaplan–Meier method and log-rank test were used to compare the probability of HCC occurrence-free between HBV mono-infected subjects with and without this independent factor. STATA Software ver. 8.0 (Stata-Corp LP, College Station, TX, USA) was employed for all analyses.

Results

Baseline characteristics in HBV mono-infected subjects with and without HCC

Of 19393 subjects, 479 had HBV mono-infection (2.5%), 1051 had HCV mono-infection (5.4%), and 20 (0.1%) had co-infection of both HBV and HCV. During the 245923 person-years' follow-up (average follow-up period 12.7 years), a total of 110 cases of newly diagnosed HCC (73 men and 37 women) were documented. Of these, 13 subjects had HBV mono-infection (11 men and 2 women), 78 had HCV mono-infection, 2 had co-infection, and 17 had neither HBV nor HCV. The summary of the population studied is shown in Fig. 1.

The characteristics of these HBV mono-infected subjects with and without HCC are shown in Table 1. The proportion of male was significantly higher in the HCC group (11/13 (85%) vs. 209/466 (45%); $P < 0.005$). As to lifestyles, the proportion of smoking was significantly higher in the HCC group (7/13 (54%) vs. 92/466 (20%); $P < 0.005$), but there were no significant differences in BMI and alcohol intake.

In regard to virological factors, the positivity of HBeAg and HBcrAg was significantly higher in the HCC group (3/13 (23%) vs. 14/466 (3%); $P < 0.001$, and 7/13 (54%) vs. 99/466 (21%), $P < 0.005$, respectively). Furthermore, both serum HBcrAg and HBV DNA levels were significantly higher in the HCC group (Table 1). The proportion of patients with HBV DNA $\geq$ 5 log copies/mL, which was reported as a risk factor of HCC development, was also significantly higher in the HCC group (6/13 (46%) vs. 39/466 (8%), $P < 0.001$). The proportion of HBV genotype C2/Ce tended to be higher in the HCC group, but there was no significant difference. As to HBV specific mutations, T1753V mutation and A1762T/G1764A double mutation was significantly more frequent in the HCC group (6/13 (46%) vs. 78/421 (19%); $P < 0.05$, and 11/13 (87%) vs. 142/421 (34%), $P < 0.001$, respectively) (Table 1). Figure 2

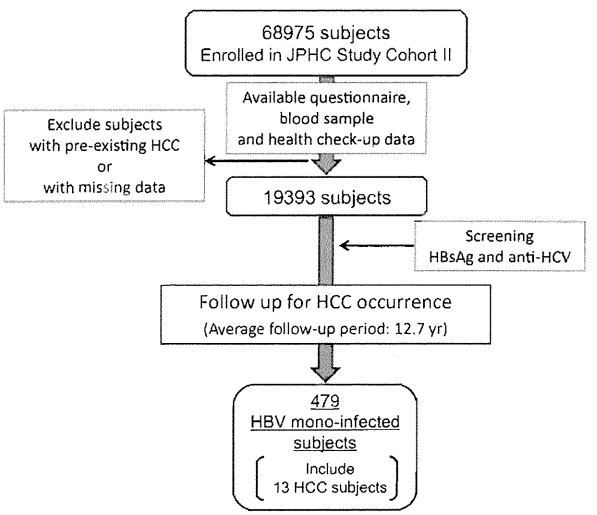

Fig. 1 The summary of the population in this study. Of 479 HBV mono-infected subjects, 13 had HCC

Table 1 Characteristics of HBV mono-infected subjects with and without HCC

	HCC group ($n = 13$)	Non-HCC group ($n = 466$)	P value
Age (years)	58.8 ± 6.3	55.1 ± 8.5	NS
Male	11 (85%)	209 (45%)	<0.005
BMI	22.3 ± 2.9	23.4 ± 3.0	NS
Alcohol	0	36 (8%)	NS
Smoking	7 (54%)	92 (20%)	<0.005
ALT (IU/L)	44.7 ± 30.0	23.8 ± 20.4	<0.001
γ-GTP (IU/L)	31.7 ± 16.2	23.2 ± 25.2	NS
HBeAg positive	3 (23%)	14 (3%)	<0.001
HBcrAg (kU/mL)	39276 ± 121639	6486 ± 47987	<0.05
HBcrAg positive	7 (54%)	99 (21%)	<0.005
HBV DNA (log copies/mL)[a]	6.1	4.1	<0.001
HBV DNA ≥ 5 log copies/mL	6 (46%)	39 (8%)	<0.001
Genotype B1/Bj	4 (31%)	264 (57%)	NS (0.0637)
Genotype C2/Ce	9 (69%)	202 (43%)	NS (0.0637)
C1653T	6 (46%)	116/421 (28%)	NS
T1753 V	6 (46%)	78/421 (19%)	<0.05
A1762T/G1764A	11 (87%)	142/421 (34%)	<0.001
G1896A	11 (87%)	348/421 (83%)	NS

NS not significant

[a] Median value

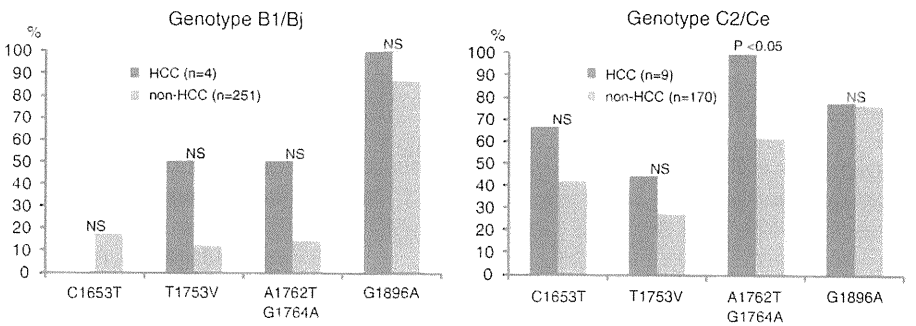

Fig. 2 The proportion of HBV specific mutations between HCC and non-HCC group when divided into each HBV genotype (B1/Bj or C2/Ce). In the HBV genotype C2/Ce, only A1762T/G1764A double mutation was significantly higher in HCC group ($P < 0.05$)

shows the proportion of HBV specific mutations between HCC and non-HCC group when divided into each HBV genotype (B1/Bj or C2/Ce). Only A1762T/G1764A double mutation was significantly higher in HCC group than non-HCC with genotype C2/Ce ($P < 0.05$).

Factors independently associated with the development of HCC in HBV mono-infected subjects

To identify the independent risk factors associated with the development of HCC, the following factors were evaluated by using multivariate-adjusted Cox hazard model among HBV mono-infected subjects: age, gender, BMI, alcohol intake, smoking, ALT, γ-GTP, HBeAg positivity, HBcrAg positivity, HBV DNA ≥ 5 log copies/mL, genotype (C2/Ce or not), and HBV specific mutations (C1653 V, T1753 V, A1762T/G1764A and G1896A). As a result, only A1762T/G1764A double mutation (hazard ratio 7.05 [95% CI 1.03–48.12, $P = 0.046$]) was an independent risk factor for the development of HCC among HBV mono-infected subjects (Table 2).

Kaplan–Meier method also showed that the probability of HCC occurrence-free was lower in HBV mono-infected subjects with A1762T/G1764A double mutation than without these mutations (log-rank: $P = 0.0001$) (Fig. 3).

Discussion

This is the first large-scale prospective population-based cohort study to identify the risk factors for HCC development among Japanese CHB patients on the basis of the prospective cohort study. To date, several cross-sectional or case-cohort studies identified factors associated with a high risk of developing of HCC among CHB patients; however, few prospective studies were carried out on this

Table 2 Independent risk factors associated with HCC among HBV mono-infected subjects by multivariate-adjusted Cox hazard model

Factors	Hazard ratio	95% confidence interval	*P* value
Age (years)			
<57[a]	1		
≥57	1.42	0.38–5.35	0.603
Gender			
Female	1		
Male	2.38	0.38–15.04	0.357
BMI			
<23[a]	1		
≥23	0.60	0.15–2.31	0.454
Alcohol			
Absent	1		
Present	–	–	
Smoking			
Absent	1		
Present	2.84	0.63–12.76	0.172
ALT (IU/L)			
<18[a]	1		
≥18	0.36	0.05–2.81	0.332
γ-GTP (IU/L)			
<16[a]	1		
≥16	4.41	0.71–27.26	0.110
HBeAg			
Negative	1		
Positive	1.15	0.10–13.74	0.913
HBcrAg			
Negative	1		
Positive	3.68	0.48–28.27	0.211
HBV DNA			
<5 log copies/mL	1		
≥5 log copies/mL	3.66	0.56–23.94	0.176
Genotype			
Non-Ce	1		
Ce	0.67	0.10–4.36	0.678
Mutations			
C1653T			
Absent	1		
Present	2.05	0.52–8.09	0.308
T1753V			
Absent	1		
Present	1.18	0.26–5.44	0.830
A1762T/G1764A			
Absent	1		
Present	7.05	1.03–48.12	0.046
G1896A			
Absent	1		
Present	2.69	0.40–18.27	0.312

[a] Median value

subject [17–19]. Fang et al. first reported that in men aged over 30 years, A1762T/G1764A double mutation was an etiological risk factor of HCC after 36 months of follow-up in a prospective study [17]. No large-scale population based cohort study in Japan was carried out to identify the risk factors for the development of HCC and analyze the clinical impact of HBV mono-infection and/or the other risk factors on the incidence of HCC among CHB patients in Japan. To address this issue, we elucidated the risk factors for HCC development among Japanese patients in this cohort study.

Most of the subjects enrolled in this cohort study were considered as healthy individuals and the proportion of HBV mono-infected subjects among them (2.5%) nearly corresponds to the estimated HBsAg prevalence in middle-aged or older population in Japan (about 1.5%) [28]. During the follow-up period of about 13 years, 2.7% of HBV mono-infected subjects in this cohort developed HCC. In univariate analysis, the following factors were significantly associated with the development of HCC: male, smoking, high ALT, HBeAg positivity, high levels of serum HBcrAg (the positivity of serum HBcrAg), high levels of serum HBV DNA (the proportion of HBV DNA ≥ 5 log copies/mL), T1753V mutation, and A1762T/G1764A double mutation. In multivariate-adjusted Cox hazard model, however, only A1762T/G1764A double mutation remained as an independent risk factor for the development of HCC among HBV mono-infected subjects, and this was also confirmed by further survival time analysis.

However, note that there were some limitations in this study: (1) As a result of the relatively small number of HBV-infected subjects (479), the statistical power might be not so strong. (2) Most of the HBV mono-infected subjects in this study had no medical therapy for HBV infection at entry because they seemed to be healthy HBV carriers, but the use of antiviral therapy was unknown during the follow-up period. (3) The analyses such as lifestyle, clinical data, and HBV specific mutations were based on only one time-point at entry, so the changes of these factors during the follow-up could not be assessed. To confirm our results strictly, further investigations to eliminate these limitations seem to be necessary.

It has been previously reported by a number of studies that A1762T/G1764A double mutation was associated with the development of HCC [4–11]. However, these results were cross-sectional or case-control studies and their observation required confirmation in prospective population-based study. Several mechanisms of hepatocarcinogenesis have been hypothesized in the context of the A1762T/G1764A double mutation: enhancement of HBV virulence by increasing the host immune response [29, 30], increasing viral replication [31–33], or altering the coding

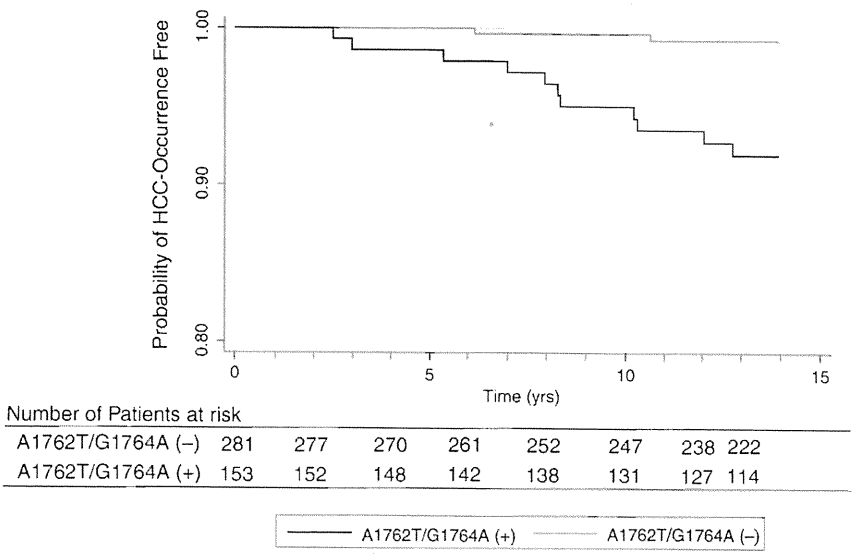

Fig. 3 Kaplan–Meier survival curve for the probability of HCC occurrence in HBV mono-infected subjects with and without A1762T/G1764A double mutation. The probability of HCC occurrence-free was lower in HBV mono-infected subjects with A1762T/G1764A double mutation than without this mutation (log-rank: $P = 0.0001$)

region for the X antigen [9, 34]. In this prospective cohort study, the A1762T/G1764A double mutation was the only independent risk factor of HCC development that is consistent with the results of Fang's prospective study in Guangxi [17], and others [18, 19].

High serum HBV DNA levels ($\geq 10^4$ copies/mL) have been reported to be associated with the development of HCC [18, 35–37]. We also measured serum HBV DNA levels and confirmed that the level of serum HBV DNA was significantly higher in the HCC group. Furthermore, the proportion of serum HBV DNA $\geq$ 5 log copies/mL was also significantly higher in the HCC group. HBcrAg levels in serum were closely correlated with serum HBV DNA levels in HBV patients without antiviral therapy [23]. However, as the majority of the HBV mono-infected subjects enrolled in this study seemed to be inactive carriers, the overall levels of both serum HBV DNA and serum HBcrAg were low. This might have affected the result of the multivariate analysis.

In several studies, stronger association between HBV genotype C and HCC was reported in comparison with HBV genotype B [38–40]. This might be due to higher frequency of A1762T/G1764A double mutation in the genotype C [41, 42]. In our cohort study there was no association between HBV genotype C2/Ce and the development of HCC in spite of the involvement of A1762T/G1764A double mutation (Table 1; Fig. 2).

In conclusion, the risk factors associated with the development of HCC were identified among Japanese HBV carriers in this prospective population-based cohort study. Only A1762T/G1764A double mutation in HBV BCP/X region could be associated with HCC development.

Appendix

Members of the Japan Public Health Center-based Prospective Study (JPHC Study, principal investigator: S. Tsugane) Group are:

S. Tsugane, M. Inoue, T. Sobue, and T. Hanaoka, National Cancer Center, Tokyo; J. Ogata, S. Baba, T. Mannami, A. Okayama, and Y. Kokubo, National Cardiovascular Center, Osaka; K. Miyakawa, F. Saito, A. Koizumi, Y. Sano, I. Hashimoto, T. Ikuta, and Y. Tanaba, Iwate Prefectural Ninohe Public Health Center, Iwate; Y. Miyajima, N. Suzuki, S. Nagasawa, Y. Furusugi, and N. Nagai, Akita Prefectural Yokote Public Health Center, Akita; H. Sanada, Y. Hatayama, F. Kobayashi, H. Uchino, Y. Shirai, T. Kondo, R. Sasaki, Y. Watanabe, Y. Miyagawa, and Y. Kobayashi, Nagano Prefectural Saku Public Health Center, Nagano; Y. Kishimoto, E. Takara, T. Fukuyama, M. Kinjo, M. Irei, and H. Sakiyama, Okinawa Prefectural Chubu Public Health Center, Okinawa; K. Imoto, H. Yazawa, T. Seo, A. Seiko, F. Ito, F. Shoji and R. Saito, Katsushika Public Health Center, Tokyo; A. Murata, K. Minato, K. Motegi, and T. Fujieda, Ibaraki Prefectural Mito Public Health Center, Ibaraki; K. Matsui, T. Abe, M. Katagiri, and M. Suzuki, Niigata Prefectural Kashiwazaki and Nagaoka Public Health Center, Niigata; M. Doi, A. Terao, Y. Ishikawa, and T. Tagami, Kochi Prefectural Chuo-higashi Public Health Center, Kochi; H. Sueta, H. Doi, M. Urata, N. Okamoto, and F. Ide, Nagasaki Prefectural Kamigoto Public Health Center, Nagasaki; H. Sakiyama, N. Onga, H. Takaesu, and M. Uehara, Okinawa Prefectural Miyako Public Health Center, Okinawa; F. Horii, I. Asano, H. Yamaguchi, K. Aoki, S. Maruyama, M. Ichii, and M. Takano, Osaka Prefectural Suita Public

Health Center, Osaka; Y. Tsubono, Tohoku University, Miyagi; K. Suzuki, Research Institute for Brain and Blood Vessels Akita, Akita; Y. Honda, K. Yamagishi, S. Sakurai, and N. Tsuchiya, Tsukuba University, Ibaraki; M. Kabuto, National Institute for Environmental Studies, Ibaraki; M. Yamaguchi, Y. Matsumura, S. Sasaki, and S. Watanabe, National Institute of Health and Nutrition, Tokyo; M. Akabane, Tokyo University of Agriculture, Tokyo; T. Kadowaki, Tokyo University, Tokyo; M. Noda and T. Mizoue, National Center for Global Health and Medicine, Tokyo; Y. Kawaguchi, Tokyo Medical and Dental University, Tokyo; Y. Takashima and M. Yoshida, Kyorin University, Tokyo; K. Nakamura, Niigata University, Niigata; S. Matsushima and S. Natsukawa, Saku General Hospital, Nagano; H. Shimizu, Sakihae Institute, Gifu; H. Sugimura, Hamamatsu University, Shizuoka; S. Tominaga, Aichi Cancer Center Research Institute, Aichi; H. Iso, Osaka University, Osaka; M. Iida, W. Ajiki, and A. Ioka, Osaka Medical Center for Cancer and Cardiovascular Disease, Osaka; S. Sato, Chiba Prefectural Institute of Public Heath, Chiba; E. Maruyama, Kobe University, Hyogo; M. Konishi, K. Okada, and I. Saito, Ehime University, Ehime; N. Yasuda, Kochi University, Kochi; S. Kono, Kyushu University, Fukuoka.

References

1. Lee WM. Hepatitis B virus infection. N Engl J Med. 1997;337:1733–45.
2. Beasley RP, Hwang LY, Lin CC, Chien CS. Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22 707 men in Taiwan. Lancet. 1981;2:1129–33.
3. 17th Nationwide follow-up study of primary hepatoma. Liver Cancer Study Group of Japan. 2006; in Japanese.
4. Yang HI, Lu SN, Liaw YF, You SL, Sun CA, Wang LY, et al. Hepatitis B e antigen and the risk of hepatocellular carcinoma. N Engl J Med. 2002;347:168–74.
5. Chen CJ, Yang HI, Su J, Jen CL, You SL, Lu SN, et al. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. JAMA. 2006;295:65–73.
6. Tsubota A, Arase Y, Ren F, Tanaka H, Ikeda K, Kumada H. Genotype may correlate with liver carcinogenesis and tumor characteristics in cirrhotic patients infected with hepatitis B virus subtype adw. J Med Virol. 2001;65:257–65.
7. Chan HL, Hui AY, Wong ML, Tse AM, Hung LC, Wong VW, et al. Genotype C hepatitis B virus infection is associated with an increased risk of hepatocellular carcinoma. Gut. 2004;53:1494–8.
8. Tong MJ, Blatt LM, Kao JH, Cheng JT, Corey WG. Precore/basal core promoter mutants and hepatitis B viral DNA levels as predictors for liver deaths and hepatocellular carcinoma. World J Gastroenterol. 2006;12:6620–6.
9. Baptista M, Kramvis A, Kew MC. High prevalence of 1762(T) 1764(A) mutations in the basic core promoter of hepatitis B virus isolated from black Africans with hepatocellular carcinoma compared with asymptomatic carriers. Hepatology. 1999;29:946–53.
10. Kao JH, Chen PJ, Lai MY, Chen DS. Basal core promoter mutations of hepatitis B virus increase the risk of hepatocellular carcinoma in hepatitis B carriers. Gastroenterology. 2003;124:327–34.
11. Liu CJ, Chen BF, Chen PJ, Lai MY, Huang WL, Kao JH, et al. Role of hepatitis B viral load and basal core promoter mutation in hepatocellular carcinoma in hepatitis B carriers. J Infect Dis. 2006;193:1258–65.
12. Sherman M. Hepatocellular carcinoma: epidemiology, risk factors, and screening. Semin Liver Dis. 2005;25:143–54.
13. Yuen MF, Yuan HJ, Wong DK, Yuen JC, Wong WM, Chan AO, et al. Prognostic determinants for chronic hepatitis B in Asians: therapeutic implications. Gut. 2005;54:1610–4.
14. Fang ZL, Ling R, Wang SS, Nong J, Huang CS, Harrison TJ. HBV core promoter mutations prevail in patients with hepatocellular carcinoma from Guangxi, China. J Med Virol. 1998;56:18–24.
15. Fang ZL, Yang J, Ge X, Zhuang H, Gong J, Li R, et al. Core promoter mutations (A(1762)T and G(1764)A) and viral genotype in chronic hepatitis B and hepatocellular carcinoma in Guangxi, China. J Med Virol. 2002;68:33–40.
16. Hsia CC, Yuwen H, Tabor E. Hot-spot mutations in hepatitis B virus X gene in hepatocellular carcinoma. Lancet. 1996;348:625–6.
17. Fang ZL, Sabin CA, Dong BQ, Ge LY, Wei SC, Chen QY, et al. HBV A1762T, G1764A mutations are a valuable biomarker for identifying a subset of male HBsAg carriers at extremely high risk of hepatocellular carcinoma: a prospective study. Am J Gastroenterol. 2008;103:2254–62.
18. Yuen MF, Tanaka Y, Fong DY, Fung J, Wong DK, Yuen JC, et al. Independent risk factors and predictive score for the development of hepatocellular carcinoma in chronic hepatitis B. J Hepatol. 2009;50:80–8.
19. Yuan JM, Ambinder A, Fan Y, Gao YT, Yu MC, Groopman JD. Prospective evaluation of hepatitis B 1762(T)/1764(A) mutations on hepatocellular carcinoma development in Shanghai, China. Cancer Epidemiol Biomarkers Prev. 2009;18:590–4.
20. Tsugane S, Sobue T. Baseline survey of JPHC study–design and participation rate. Japan Public Health Center-based Prospective Study on Cancer and Cardiovascular Diseases. J Epidemiol. 2001;11:S24–9.
21. Inoue M, Yoshimi I, Sobue T, Tsugane S. Influence of coffee drinking on subsequent risk of hepatocellular carcinoma: a prospective study in Japan. J Natl Cancer Inst. 2005;97:293–300.
22. World Health Organization. In: International classification of diseases for oncology, 3rd edn. Geneva: WHO; 2000.
23. Kimura T, Rokuhara A, Sakamoto Y, Yagi S, Tanaka E, Kiyosawa K, et al. Sensitive enzyme immunoassay for hepatitis B virus core-related antigens and their correlation to virus load. J Clin Microbiol. 2002;40:439–45.
24. Abe A, Inoue K, Tanaka T, Kato J, Kajiyama N, Kawaguchi R, et al. Quantitation of hepatitis B virus genomic DNA by real-time detection PCR. J Clin Microbiol. 1999;37:2899–903.
25. Tanaka Y, Hasegawa I, Kato T, Orito E, Hirashima N, Acharya SK, et al. A case–control study for differences among hepatitis B virus infections of genotypes A (subtypes Aa and Ae) and D. Hepatology. 2004;40:747–55.
26. Sugauchi F, Mizokami M, Orito E, Ohno T, Kato H, Suzuki S, et al. A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. J Gen Virol. 2001;82:883–92.
27. Shin-I T, Tanaka Y, Tateno Y, Mizokami M. Development and public release of a comprehensive hepatitis virus database. Hepatol Res. 2008;38:234–43.
28. Report of hepatitis epidemiology group. Ministry of Health, Labour and Welfare. 2006; in Japanese.

29. Hunt CM, McGill JM, Allen MI, Condreay LD. Clinical relevance of hepatitis B viral mutations. Hepatology. 2000;31:1037–44.
30. Takahashi K, Aoyama K, Ohno N, Iwata K, Akahane Y, Baba K, et al. The precore/core promoter mutant (T1762A1764) of hepatitis B virus: clinical significance and an easy method for detection. J Gen Virol. 1995;76(Pt 12):3159–64.
31. Buckwold VE, Xu Z, Chen M, Yen TS, Ou JH. Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication. J Virol. 1996;70:5845–51.
32. Tang H, Raney AK, McLachlan A. Replication of the wild type and a natural hepatitis B virus nucleocapsid promoter variant is differentially regulated by nuclear hormone receptors in cell culture. J Virol. 2001;75:8937–48.
33. Yu X, Mertz JE. Distinct modes of regulation of transcription of hepatitis B virus by the nuclear receptors HNF4alpha and COUP-TF1. J Virol. 2003;77:2489–99.
34. Kidd-Ljunggren K, Oberg M, Kidd AH. The hepatitis B virus X gene: analysis of functional domain variation and gene phylogeny using multiple sequences. J Gen Virol. 1995;76(Pt 9):2119–30.
35. Yu MW, Yeh SH, Chen PJ, Liaw YF, Lin CL, Liu CJ, et al. Hepatitis B virus genotype and DNA level and hepatocellular carcinoma: a prospective study in men. J Natl Cancer Inst. 2005;97:265–72.
36. Liu CJ, Chen BF, Chen PJ, Lai MY, Huang WL, Kao JH, et al. Role of hepatitis B virus precore/core promoter mutations and serum viral load on noncirrhotic hepatocellular carcinoma: a case-control study. J Infect Dis. 2006;194:594–9.
37. Ohata K, Hamasaki K, Toriyama K, Ishikawa H, Nakao K, Eguchi K. High viral load is a risk factor for hepatocellular carcinoma in patients with chronic hepatitis B virus infection. J Gastroenterol Hepatol. 2004;19:670–5.
38. Orito E, Mizokami M, Sakugawa H, Michitaka K, Ishikawa K, Ichida T, et al. A case-control study for clinical and molecular biological differences between hepatitis B viruses of genotypes B and C. Japan HBV Genotype Research Group. Hepatology. 2001;33:218–23.
39. Fujie H, Moriya K, Shintani Y, Yotsuyanagi H, Iino S, Koike K. Hepatitis B virus genotypes and hepatocellular carcinoma in Japan. Gastroenterology. 2001;120:1564–5.
40. Sakugawa H, Nakasone H, Nakayoshi T, Orito E, Mizokami M, Yamashiro T, et al. Preponderance of hepatitis B virus genotype B contributes to a better prognosis of chronic HBV infection in Okinawa, Japan. J Med Virol. 2002;67:484–9.
41. Yuen MF, Sablon E, Tanaka Y, Kato T, Mizokami M, Doutreloigne J, et al. Epidemiological study of hepatitis B virus genotypes, core promoter and precore mutations of chronic hepatitis B infection in Hong Kong. J Hepatol. 2004;41:119–25.
42. Yuen MF, Tanaka Y, Mizokami M, Yuen JC, Wong DK, Yuan HJ, et al. Role of hepatitis B virus genotypes Ba and C, core promoter and precore mutations on hepatocellular carcinoma: a case control study. Carcinogenesis. 2004;25(9):1593–8.

Research Article

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | JOURNAL OF HEPATOLOGY

Evolution of hepatitis B genotype C viral quasi-species during hepatitis B e antigen seroconversion

Shuang Wu[1], Fumio Imazeki[1,*], Fuat Kurbanov[2], Kenichi Fukai[1], Makoto Arai[1], Tatsuo Kanda[1], Yutaka Yonemitsu[1], Yasuhito Tanaka[2], Masashi Mizokami[3], Osamu Yokosuka[1]

[1]*Department of Medicine and Clinical Oncology, Graduate School of Medicine, Chiba University, Chiba, Japan;* [2]*Department of Virology, Liver Unit, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan;* [3]*Research Center for Hepatitis and Immunology, International Medical Center of Japan Kounodai Hospital, Ichikawa, Japan*

Background & Aims: Although the evolution of viral quasi-species may be related to the pathological status of disease, little is known about this phenomenon in hepatitis B, particularly with respect to hepatitis B e antigen (HBeAg) seroconversion.

Methods: Nucleotide sequences of the hepatitis B virus (HBV) *X/precore/core* region was analyzed at five time-points in four groups of chronic hepatitis B patients, interferon-induced seroconverters (IS, $N = 9$), interferon non-responders (IN, $N = 9$), spontaneous seroconverters (SS, $N = 9$), and non-seroconverters (SN, $N = 9$) followed during 60 months on an average. Only patients with genotype C were studied.

Results: Analysis of 1800 nucleotide sequences showed that there was no statistical difference between the nucleotide genetic distances of seroconverters (IS and SS; 6.9×10^{-3} substitutions (st)/site and 6.7×10^{-3} st/site, respectively) and those of non-seroconverters (IN and SN; 5.3×10^{-3} st/site and 3.8×10^{-3} st/site, respectively) before seroconversion. Compared to non-seroconverters (IN and SN; 5.1×10^{-3} st/site and 5.9×10^{-3} st/site, respectively), the sequence diversity of seroconverters (IS and SS; 10.9×10^{-3} st/site and 9.9×10^{-3} st/site, respectively) was significantly higher after seroconversion ($p < 0.05$), and was higher in seroconverters after seroconversion than before seroconversion ($p < 0.05$), while this changed very little in non-seroconverters during the observation period. Phylogenetic trees showed greater complexity in secoconverters than non-seroconverters. Parsimony-based estimation of the direction of sequence change between descendants and ancestors before HBeAg seroconversion, revealed higher frequencies of transversional A to T substitution in seroconverters (0.06 *vs.* 0.02, $p = 0.0036$) that coincided with the dynamics of quasi-species possessing A1762T mutation.

Conclusions: The distinctly greater viral diversity in HBeAg seroconverters after seroconversion could be related to escape mutants resulting from stronger selection pressure.

Introduction

Hepatitis B virus (HBV) is a major human pathogen which can cause severe hepatic disease, including chronic hepatitis, cirrhosis (LC), and hepatocellular carcinoma (HCC). Quasi-species comprises a complex and dynamic distribution of non-identical but related genomes [1]. The evolution of viral quasi-species has been reported as important in the pathogenesis of RNA viruses such as hepatitis C virus [2–6] and human immunodeficiency virus [7–10], but little is known about HBV. HBV is a hepatotropic, non-cytopathic DNA virus replicated by an error-prone polymerase through an RNA intermediate. Because of this feature, the replication of HBV lacks fidelity. This results in a complex distributions of genomes with naturally-acquired mutations or mutations selected by either antiviral therapy or the immune response of the host. HBV quasi-species have not been subjected to detailed investigation, especially in the context of hepatitis B e antigen (HBeAg) seroconversion (SC), an immunologically mediated event. Whether there is a causal relationship between HBV seroconversion and HBV quasi-species remains unclear. HBV-related disease is known to be mediated both virologically and immunologically. Several studies have depicted the dynamic evolution of HBV quasi-species during lamivudine resistance or multiple drug resistance. This highlights the importance of HBV molecular evolution in revealing the mechanism of drug resistance [11,12]. HBV-specific cytotoxic T-cells play a significant role in the control of replication of HBV, which has been well documented in the literature [13–16].

Precore/core protein is the target of immunologically mediated HBeAg seroconversion. When the *precore/core* gene in HBV DNA is transcribed and translated, HBeAg is produced and secreted into the circulation [17,18]. But the synthesis and secretion of HBeAg are aborted by the emergence of a point mutation from G to A at nucleotide (nt)1896 (G1896A). Convincing lines of evidence have indicated a close association between HBeAg/anti-HBe seroconversion and the emergence of *precore* and *core* promoter mutations [19,20].

Keywords: Chronic hepatitis B; Quasi-species; Hepatitis B e antigen seroconversion.

Received 15 December 2009; received in revised form 4 June 2010; accepted 7 June 2010; available online 25 August 2010

* Corresponding author. Address: Department of Medicine and Clinical Oncology, Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuo-Ku, Chiba 260-8670, Japan. Tel.: +81 43 226 2083; fax: +81 43 226 2088.

E-mail address: imazekif@faculty.chiba-u.jp (F. Imazeki).

Abbreviations: SC, seroconversion; ALT, alanine aminotransferase; CHB, chronic hepatitis B; HBV, hepatitis B virus; IFN, interferon; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; IS, interferon-induced HBeAg seroconverters; IN, IFN non-responders; SS, spontaneous seroconverters; SN, non-seroconverters.

Research Article

The purpose of this study was to elucidate the evolution of HBV quasi-species during HBeAg seroconversion. The results might help us to better understand the pathogenic mechanisms of HBV. We selected patients with well-characterized clinical phenotypes and compared their viral diversity based on the nucleotide sequences of the *X*/precore/core region. *Precore* and *core* promoter mutations were also investigated in detail before and after HBeAg seroconversion.

Materials and methods

Patients

Sera from 36 chronic hepatitis B patients with well-characterized clinical follow-up for >5 years were selected from a chronic hepatitis B database (77 seroconverters and 67 non-seroconverters) at Chiba University Hospital. Only patients with genotype C (subtype C2) were studied to ensure that differences found in viral evolution were not due to genotypic variation. Nine patients in each group were selected randomly if they fulfilled the following criteria and had sufficiently long follow-up. The index group comprised patients with documented HBeAg seroconversion (spontaneous seroconverters, SS), with serum at the following time-points relative to HBeAg seroconversion: time-point I (−25.2 ± 6.2/months), time-point II (−11.6 ± 2.7/months), time-point III (1 ± 2.3/months), time-point IV (12.5 ± 3.3/months), and time-point V (25 ± 3.6 months). Untreated control patients included those who were followed for a similar period of time and were persistently HBeAg positive (non-seroconverters, SN), and they were matched for average age of seroconversion and time-point intervals of the SS group. A second index group of patients with interferon (IFN)-induced HBeAg seroconversion (IFN seroconverters, IS), with serum at the following time-points relative to HBeAg seroconversion: time-point I (−24.3 ± 3.1/months), time-point II (−11.2 ± 1.9/months), time-point III (1 ± 1.2/months), time-point IV (12.7 ± 1.7/months), and time-point V (25.4 ± 2.2/months). Control patients were persistently HBeAg-positive despite IFN therapy (IFN non-responders, IN). Controls were matched for the average age of seroconversion, sex and time-point intervals of the IS group.

HBeAg seroconversion was defined as the loss of HBeAg and the development of anti-HBe. The serial serum samples in this study were taken at five time-points for each patient, as described above. This study was approved by the Ethics Committee of Chiba University Hospital.

Serological examination

HBsAg, HBeAg and anti-HBe were determined by enzyme-linked immunosorbent assay (ELISA; Abbott Laboratory, Chicago, IL). HBV genotype was determined from the patients' sera by ELISA (HBV genotype EIA; Tokushu-Meneki Laboratory, Tokyo, Japan), based on the method described by Usuda et al. [21]. Serum HBV DNA levels were monitored using the Roche Amplicor Monitor test (Roche Diagnostics, Tokyo, Japan), which has a lower detection limit of 2.6 $\log_{10}$ copies/ml, at each time-point.

Cloning and sequencing

Total DNA was extracted from 200 μl of each serum sample using QIAamp DNA blood mini kits (Qiagen, Hilden, Germany) according to the manufacturer's instructions and eluted in 200 μl distilled water. Because HBeAg seroconversion is associated with a decrease in HBV DNA levels, nested PCR was performed for all the samples. The primers for the first round of PCR were 5′-TCG CAT GGA GAC CAC CGT GA-3′ (sense, nt1604–1623) and 5′-ATA GCT TGC CTG AGT GC-3′ (antisense, nt 2076–2060). The primers for the second round of PCR were 5′-CAT AAG AGG ACT CTT GGA CT-3′ (sense, nt 1653–1672) and 5′-GGA AAG AAG TCA GAA GGC-3′ (antisense, nt 1974–1957).

Amplification was performed with 2 μl of DNA template (extracted DNA from serum samples for the first round PCR and the first round PCR products for the second round PCR) in 50 μl reaction under the following conditions: an initial 2 min of denaturation at 94 °C and 36 cycles of 94 °C denaturation for 1 min, annealing at either 54 °C or 52 °C for 1 min, in the first and second round respectively, and 72 °C extension for 1 min. The last cycle was followed by a final extension at 72 °C for 10 min. A 473-base pairs fragment (nt 1604–2076) containing the *X/precore/core* region was amplified.

PCR reactions were followed by cloning using TOPO® TA cloning kits (Invitrogen, Carlsbad, CA). All PCR products were purified with QIAquick PCR Purification Kit (Qiagen, Hilden, Germany), then cloned into the TOPO vector, and transformed into *Escherichia coli*. At least 15 clones per one cloning for samples from PCR reactions proceeded subsequent to the electrophoretic size separation on 1.2% agarose gel. Ten positive clones per cloning for samples from each PCR reaction were sequenced using BigDye® Terminator and a 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA). The cloning PCR and sequencing primers were M13-forward, 5′-GTA AAA CGA CGG CCA GT-3′, and M13-reverse, 5′-GGA AAC AGC TAT GAC CAT G-3′.

Sequence analysis

The DNAPARS program from PHYLIP v3.65 package, implemented in Simmonic Sequence Editor version 1.5 [22], was used for sequence analysis. To evaluate quasi-species-based evolution of HBV strains in chronic patients, sequences of clones ($N = 10$) isolated at each time-point ($N = 5$) from individual patients ($N = 36$) were subjected to alignment and used to generate one parsimonious ancestral sequence. Maximum nucleotide composition distances were evaluated pair-wise between the ancestral sequence and the sequences of each of the 10 clones with a mean value estimated for each patient at a given time-point (*MEGA* version 4 [23]). All patients were categorized into four groups with respect to seroconversion status and the mean distance value for each group was calculated for each time-point.

The differences in genetic distance among clinical groups and time-points, and diversity at each time-point, were analyzed using ANOVA (analysis of variance). Student's *t*-test was also performed to determine the average of genetic diversities in non-seroconverters. All graphical data are presented as means ± standard deviation (SD). Results were considered statistically significant at $p < 0.05$. The statistical analysis was performed with SPSS (2004; SPSS Inc., Tokyo, Japan).

Construction of phylogenetic trees

To examine the evolution of the viral sequence and whether this evolution was elicited by quasi-species or mutagenesis, phylogenetic trees were constructed using the Neighbor-Joining (NJ) model with the Simmonic Sequence Editor version 1.5, based on the genomic sequences of HBV. Moreover, to investigate viral genetic features possibly associated with seroconversion, sequences isolated at time-points 1 and 2 were further analyzed phylogenetically. Neighbor-Joining trees were constructed at time-points 1 and 2 (Fig. S1 and S2, respectively) using all groups of sequences.

Results

Baseline clinical characteristics of the patients and sequential levels of serum ALT and HBV DNA

The clinical and laboratory characteristics of all patients are listed in Table 1. The levels of alanine aminotransferase (ALT) and HBV DNA over time are illustrated in Fig. 1A and B, respectively. Serum ALT levels, a marker of hepatocyte damage, normalized after seroconversion and, for all groups except the interferon non-responders, were <40 IU/L at the end-point of observation. HBV DNA loads decreased markedly in seroconverters (<3 $\log_{10}$ copies/ml, $p < 0.0001$) but changed very little in non-seroconverters. It is noteworthy that, at the second year after seroconversion, serum HBV DNA loads increased in interferon-induced seroconverters compared to spontaneous seroconverters, without statistical significance ($p^H = 0.1087$) (Fig. 1B).

Viral nucleotide sequence diversity

Viral sequence diversity, phylogenetic trees, and mutation pattern based on 1800 HBV nucleotide sequences from clones of the *X/precore/core* region, were analyzed among selected patients.

Table 1. Baseline clinical features of patients.

	IFN Seroconverters (IS)	IFN Non-seroconverters (IN)	Spontaneous Seroconverters (SS)	Spontaneous Non-seroconverters (SN)
Age (y)	40 ± 8	40 ± 11	29 ± 10	34 ± 6
Male : Female	6:3	8:1	5:4	7:2
HBV DNA (log_{10} copies/ml)	6.8 ± 0.9	6.8 ± 1.0	6.8 ± 1.2	7.1 ± 0.8
ALT (IU/L)	88.3 ± 48.6	94.3 ± 144.4	89.8 ± 71.4	67.6 ± 48.7

Note: The IFN-induced group (seroconverters and non-responders) was older than the spontaneous group (seroconverters and non-responders). Males were the majority in all groups. Baseline serum HBV DNA and ALT levels are similar among the four groups. Data are shown as mean ± SD.

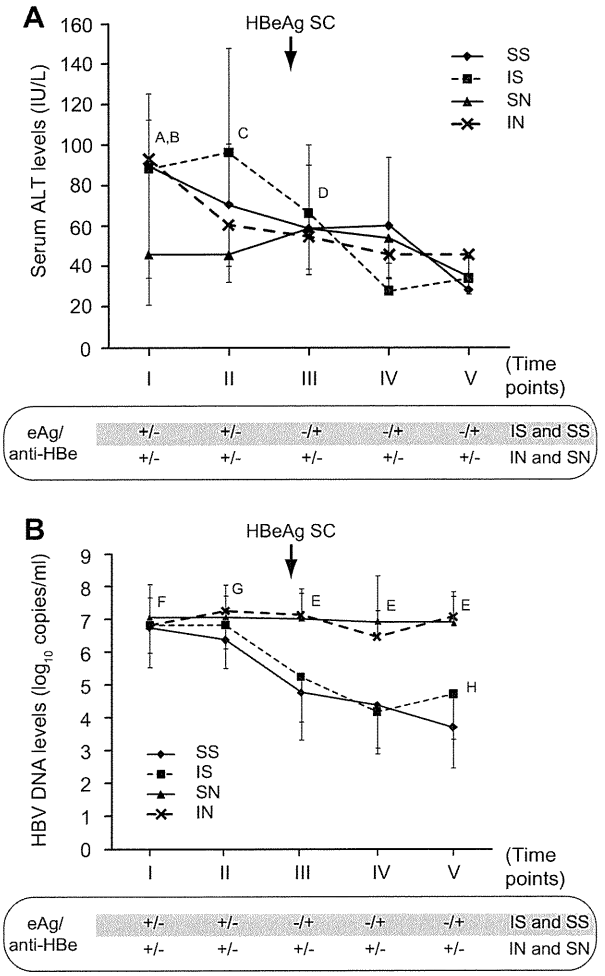

Fig. 1. Serum ALT and DNA levels in the four groups. The group of spontaneous seroconverters (SS) is a solid line diamond, IFN-induced seroconverters (IS) is a broken line square, IFN non-responders (IN) is a broken line asterisk, and non-seroconverters controls (SN) is a solid line triangle. (A) $p^A = 0.0234$ comparing time-point I with time-point IV, $p^B = 0.0028$ comparing time-point I with time-point V, $p^C = 0.007$ comparing time-point II with time-point V, $p^D = 0.0068$ comparing time-point III with time-point V. (B) $p^E < 0.0001$ comparing seroconverters with non-seroconverters, $p^F < 0.0001$ comparing time-point I with III, IV, V, $p^G < 0.0001$ comparing time-point II with the other time-points, $p^H = 0.1087$ at time-point V in seroconverters.

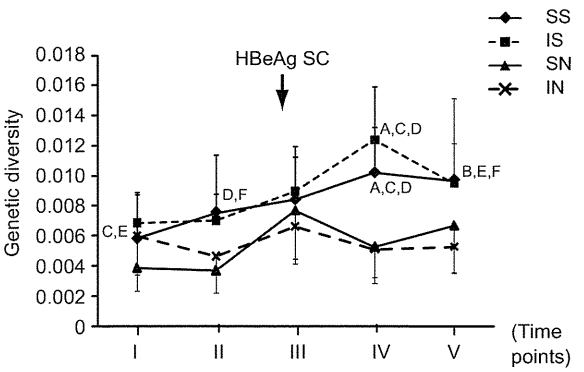

Fig. 2. Viral genetic diversity in the four groups. The group of spontaneous seroconverters (SS) is a solid line diamond, IFN-induced seroconverters (IS) is a broken line square, IFN non-responders (IN) is a broken line asterisk and non-seroconverters controls (SN) is a solid line triangle. $p^A < 0.0001$ comparing seroconverters with non-seroconverters at time-point IV, $p^B = 0.0301$ comparing seroconverters with non-seroconverters at time-point V, $p^C = 0.0013$ and $p^D = 0.0025$ comparing I and II with time-point IV in seroconverters. $p^E = 0.0121$ and $p^F = 0.021$ comparing time-points I and II with V in seroconverters.

Striking differences in nucleotide sequence diversity were revealed between seroconverters and non-seroconverters before and after seroconversion (Fig. 2). The nucleotide sequence diversity of seroconverters was similar to that of non-seroconverters before seroconversion. Analysis of genetic distance showed that the viral sequence diversity of seroconverters was significantly greater than that of non-seroconverters after seroconversion (Fig. 2, $p^A < 0.0001$ at time-point IV, $p^B = 0.0301$ at time-point V) and was greater in seroconverters after seroconversion than before (Fig. 2, $p^C = 0.0013$ and $p^D = 0.0025$), while almost no changes were observed in non-seroconverters during the observation period.

It is noteworthy that, in interferon-induced seroconverters at the last time-point of observation, the nucleotide sequence diversity was less, although this increased clearly at the first year after seroconversion. This tendency of reversed change at the last two time-points was also seen in HBV DNA loads (Fig. 1B), namely, increase or decrease of the genetic diversity accompanied by decrease or increase of the viral load in interferon-induced seroconverters. On the other hand, the nucleotide sequence diversity increased continuously in spontaneous seroconverters, accompanied by a concurrent decrease of viral loads (Fig. 1B) during the follow-up period. Amino acid sequence diversity had an almost

Research Article

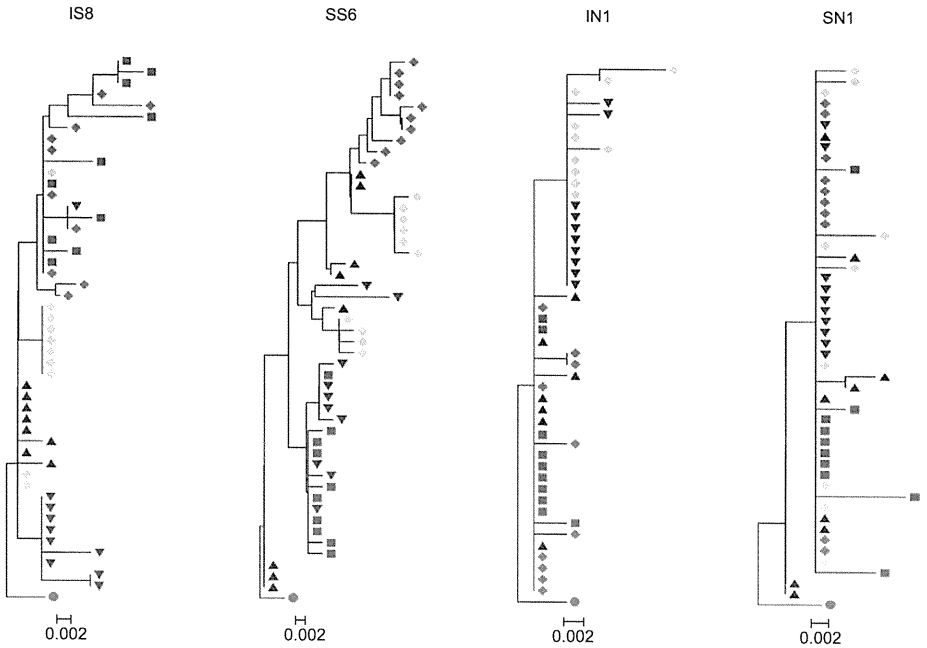

Fig. 3. Representative Neighbor-Joining phylogenetic trees of HBV sequences for each clinical group showing complex trees in seroconverters. HBV *X/precore/core* sequences from time-points I (purple filled triangle), II (blue filled inverted triangle), III (green filled square), IV (red filled diamond) and V (sky blue filled diamond) serum samples are analyzed phylogenetically and their positions are displayed on the trees. A sequence retrieved from the time-point I (red dot) of each group as outgroup in the trees, respectively. Scale bar represents 0.002% genetic variation. Seroconversion patients (IS, SS) show relatively complex branching patterns, forming clusters over time. With the pressure of seroconversion, the genetic diversity increased. In contrast, patients without seroconversion (IN, SN) were simply branching patterns and the genetic diversity in these patients changed very little over time.

identical pattern to that of DNA nucleotide sequence diversity (data not shown).

Construction of phylogenetic trees

Phylogenetic trees were complex for secoconverters and comparatively simple for non-seroconverters. In seroconverters (IS and SS), the arrangement and branch lengths of the trees were consistently more complex and longer than those for non-seroconverters. The genetic diversity was great after seroconversion in seroconverters (IS and SS) and less in non-seroconverters (IN and SN) (Fig. 3).

To investigate viral genetic features possibly associated with seroconversion, sequences isolated at time-points 1 and 2 (before seroconversion) were further analyzed phylogenetically. Trees were reconstructed using Neighbor-Joining, ML (data not shown), and PAML methods (data not shown). In general, no clusters were seen to be supported by robust bootstrap values for any group or particular patient quasi-species. This indicates that the region of the HBV genome studied does not contain patterns of variability sufficient for robust phylogenetic relation reconstruction. However, variability of branch lengths in the tree indicated that seroconversion patient groups exhibit greater diversity of the quasi-species compared to patients without seroconversion. This is in agreement with the genetic distance plot (Fig. 2), showing greater deviation from the mean values observed in patients with seroconversion. The IN group exhibited least deviation on the distance plot (Fig. 2) and shortest branch lengths on the trees (Fig. 3).

Interclonal differences of the quasi-species

To investigate whether a particular mutation pattern of evolution of the quasi-species is associated with seroconversion, we further analyzed the sequence changes in all patients at time-points 1 and 2, corresponding to the time before seroconversion. Parsimony-based ancestral sequences were generated using the Simmonic Sequence Editor. Aligned sequences of time-points 1 and 2 from a single patient were used as the input. Frequencies of changes in 12 types of mutations, including 4 transitions (CT, TC, AG, and GA) and 8 transversions (AT, TA, AC, CA, CG, GC, GT, and TG) were evaluated between generated descendants and ancestral sequences for each clone of the patient. Statistical *t*-test comparison of mean values of nucleotide changes between seroconversion and non-seroconversion groups is summarized in Table 2 and Supplementary Table 1.

Analysis of sequence changes indicated a higher frequency of transversional A to T in spontaneous seroconverters (SS *vs.* SN = 0.06 *vs.* 0.02, $p = 0.04$) and IFN-induced seroconverters (IS *vs.* IN = 0.05 *vs.* 0.01, $p = 0.05$) and A to C changes in IFN-induced seroconverters (IS *vs.* IN = 0.025 *vs.* 0.006, $p = 0.04$) before seroconversion. Comparison of seroconversion groups (SS and IS) indicated a higher frequency of transversional A to T mutation pattern ($p = 0.003$, Table 2) and the trend of G to A mutation is higher in seroconversion groups (SS and IS) (Table 2). Subsequently, alignments of the clones were generated. Visual inspection of the alignments indicated variation in the ratio of A1762T mutation in clones isolated from each patient at time-points 1 and 2 (Fig. 4). In contrast to non-seroconverters, seroconverters

Table 2. *t*-test comparison of mean values of nucleotide changes between seroconversion and non-seroconversion groups.

	Seroconversion (n = 18)	Non-seroconversion (n = 18)	*p*
CT	0.117033	0.103750	0.637023
TC	0.156706	0.201328	0.155252
AG	0.125483	0.148372	0.498916
GA	0.196722	0.124511	0.073433
AT	0.061194	0.022128	0.003665
TA	0.049372	0.045417	0.778612
AC	0.027944	0.012550	0.145158
CA	0.017128	0.011094	0.523868
CG	0.009439	0.007744	0.835337
GC	0.018167	0.014894	0.748267
GT	0.009839	0.019217	0.272185
TG	0.041783	0.035528	0.731324

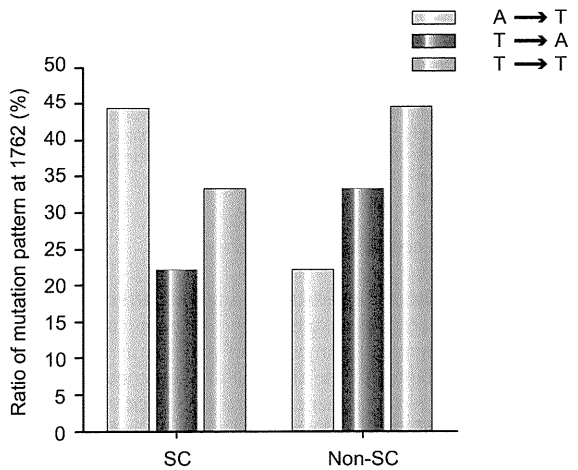

Fig. 4. The evolution of the *core* promoter mutation (A1762T) between seroconversion and control groups from time-point I to II. SC indicates seroconversion and non-SC, non-seroconversion. Alignment of the clones was carried out and the frequency of A1762T mutation in clones isolated from each patient at time-points 1 and 2 was determined. Subsequently, the evolutionary ratio of mutation from time-point I to II was calculated.

showed a higher frequency of A to T mutation pattern in the core promoter region from time-point I to II.

Core promoter (A1762T/G1764A) and precore (G1896A) mutations

Given that the *core* promoter/*precore* mutations influenced virus replication and HBeAg seroconversion, we analyzed the sequential change of *core* promoter (A1762T/G1764A)/*precore* (G1896A) mutations over time (Table 3). After seroconversion, patients with more than 50% *precore* mutant clone had higher HBV DNA loads than those with less than 50% of *precore* mutant clone (*precore* wild type) virus at time-point V [5.4 ± 1.3 (n = 5) *vs.* 3.8 ± 1.1 (n = 13), p = 0.0185] and 8 patients with a HBV DNA load less than 4.0 $\log_{10}$ copies/ml had all *precore* wild-type virus at time-point V (Table 3). Clinical progress of these patients was investigated over 10 years as median (range 1–20 years) after HBeAg seroconversion. HCC developed in 3 of 5 patients with *precore* mutant virus, compared to 1 of 13 patients with *precore* wild-type virus at time-point V (p = 0.017). On the other hand, 3 patients with ASC had all *precore* wild-type virus at time-point V (Table 3).

Discussion

In this study, analysis of 1800 nucleotide sequences from 36 HBV carriers showed that the viral diversity of seroconverters (IS and SS) after seroconversion was significantly greater than that of non-seroconverters (IN and SN) (Fig. 2, p <0.05) and was higher after seroconversion than before, in the seroconverters (Fig. 2, p <0.05). Phylogenetic analysis also generated complex trees for seroconverters and relatively simple trees for non-seroconverters. Analysis on interclonal differences in the quasi-specie showed a higher frequency of transversional A to T mutation pattern in seroconverters that coincided with the A1762T core promoter mutation. These findings suggested that HBeAg seroconversion involves dynamic shifts of the serum HBV quasi-species.

Osiowy et al. [24] examined viral quasi-species in eight HBeAg-negative patients at two time-points 25 years apart and obtained the evolutionary rate. Their results suggested that HBV diversity may be generated more rapidly than those estimated previously [25–29]. The higher evolutionary rate may be related to the seroconversion event driving quasi-specie complexity and diversification [24]. Our phylogenetic study showed that viral quasi-species populations appear to be replaced by new populations arising from a different clade after seroconversion.

Increased immune responses are accompanied by the reduction of viral loads and stronger immune pressure induces the selection of escape mutations, which leads to greater viral diversity [30]. According to this scenario, in our study, non-seroconverters have a high viral load and low quasi-species diversity and they obviously have a weak immune response.

Lim et al. [31] reported that viral genetic diversity in genotype B CHB patients was 2.4-fold greater in HBeAg seroconverters (spontaneous or IFN-induced) than in non-seroconverters before seroconversion. In this study of genotype C CHB patients, the nucleotide genetic distance was 1.49-fold greater in seroconverters (IS and SS) than in non-seroconverters before seroconversion but there was no statistical difference. This discrepancy might be due to the smaller region for analysis of genetic distance in our study than that of Lim et al. Another interpretation is that the host's immune response to the selection of mutant virus might differ between genotype B and genotype C. The natural course of CHB and the response to treatment could be affected by HBV genotype and there are some lines of evidence that indicate that the prevalence rates of *precore* and *core* promoter mutations vary among patients infected with HBV strains of different genotypes [32–34].

T-test comparison of mean values of nucleotide changes (Table 2) and linear logistic regression univariate analysis of mutations associated with seroconversion between seroconverters and non-seroconverters (data not shown) indicated a variation in the AT mutation pattern in the former (p = 0.003 and p = 0.006, respectively). This coincided with differences in the

Research Article

Table 3. *Core* promoter and *precore* mutations in seroconverters (IS and SS).

Patients	CP (ntA1762T/G1764A) (percent)			PC (ntG1896A) (percent)			DNA Loads (log_{10} copies/ml)			Histological diagnosis
	I	III	V	I	III	V	I	III	V	
IS1	100	100	100	0	0	70	5.7	3.8	4.8	CHB
IS2	100	100	100	90	100	90	7.6	7.2	7.6	HCC
IS3	70	100	100	10	0	10	6.5	5.2	5.5	CHB
IS4	90	100	10	0	10	0	7.6	6.2	3.3	CHB
IS5	100	40	20	0	0	0	7.6	3.6	4.1	CHB
IS6	70	90	90	20	10	90	5.7	4.1	4.5	HCC
IS7	100	100	90	0	10	10	7.2	3.1	3.4	LC
IS8	80	100	60	0	0	60	7.6	4.0	4.5	CHB
IS9	100	100	10	0	0	0	6.0	4.5	4.8	HCC
SS1	0	60	0	80	0	80	7.6	4.2	5.4	HCC
SS2	80	100	90	10	90	10	6.6	7.6	5.9	ASC
SS3	100	90	60	10	0	0	6.5	4.3	2.8	ASC
SS6	30	100	10	0	0	10	3.9	4.4	4.1	CHB
SS7	80	100	100	0	0	0	7.6	2.8	2.6	ASC
SS8	0	100	90	0	20	0	7.6	5.4	3.6	CHB
SS9	0	80	20	0	10	0	7.6	4.0	2.6	CHB
SS10	50	20	40	0	0	40	7.3	3.9	2.6	CHB
SS11	100	100	100	0	0	0	6.1	6.3	3.8	CHB

IS: interferon induced seroconverter; SS: spontaneous seroconverter; ASC: asymptomatic carriers; CHB: chronic hepatitis B; LC: cirrhosis; HCC: hepatocellular carcinoma.

ratio of T1762A quasi-species between seroconverters and non-seroconverters, indicating that it might be a marker preceding seroconversion in HBV/genotype C-infected patients as reported previously [35–37].

HBeAg seroconversion is an incomplete marker of immune control, although most patients experience some clinical benefit from it [38,39]. Previous studies have shown that the average rate of spontaneous HBeAg seroconversion in patients with chronic hepatitis B is about 10% per year [40,41]. HBeAg seroconversion associated with incomplete viral suppression may result in the emergence of the *precore* mutant and attendant chronic sequelae. Mutations in the *precore* and *core* promoter regions of the HBV genome have been reported in many HBeAg-negative CHB patients. Longitudinal studies found that the A1896 mutation emerges or is selected around the time of HBeAg seroconversion, and high *precore* mutant ratios have been associated with persistent hepatitis after anti-HBe seroconversion [42]. Patients who continued to have high HBV DNA titres after HBe seroconversion had a lower genetic heterogeneity but more often had the *precore* mutant.

The limitations of this study were, the small size of study group, only 10 clones per sample, and a small region for analysis of genetic distance. In addition, the *X/precore/core* region is a highly conserved region, investigation of another region of the HBV genome, such as the polymerase, might help us to better understand the evolution of quasi-species of HBV.

In conclusion, the distinctly greater viral diversity after seroconversion in HBeAg seroconverters could be related to increased HBV-specific T-cell responses and escape mutants which arise from selective pressure caused by host immune activity. Long-term follow-up is required to determine whether hepatitis B viral diversity decreases or remains at a high level. Further study will be needed to elucidate the relationship between seroconversion and viral quasi-species in relation to antiviral therapy.

Conflict of interest

The authors who have taken part in this study declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

References

[1] Domingo E, Holland JJ. RNA virus mutations and fitness for survival. Annu Rev Microbiol 1997;51:151–178.

[2] Arataki K, Kumada H, Toyota K, Ohishi W, Takahashi S, Tazuma S, et al. Evolution of hepatitis C virus quasispecies during ribavirin and interferon-alpha-2b combination therapy and interferon-alpha-2b monotherapy. Intervirology 2006;49:352–361.

[3] Laskus T, Wilkinson J, Gallegos-Orozco JF, Radkowski M, Adair DM, Nowicki M, et al. Analysis of hepatitis C virus quasispecies transmission and evolution in patients infected through blood transfusion. Gastroenterology 2004;127:764–776.

[4] Pawlotsky JM. Hepatitis C virus genetic variability: pathogenic and clinical implications. Clin Liver Dis 2003;7:45–66.

[5] Pellerin M, Lopez-Aguirre Y, Penin F, Dhumeaux D, Pawlotsky JM. Hepatitis C virus quasispecies variability modulates nonstructural protein 5A transcriptional activation, pointing to cellular compartmentalization of virus–host interactions. J Virol 2004;78:4617–4627.

[6] Fan X, Mao Q, Zhou D, Lu Y, Xing J, Xu Y, et al. High diversity of hepatitis C viral quasispecies is associated with early virological response in patients undergoing antiviral therapy. Hepatology 2009;50:1765–1772.

[7] da Silva J. The evolutionary adaptation of HIV-1 to specific immunity. Curr HIV Res 2003;1:363–371.

[8] Edwards CT, Holmes EC, Pybus OG, Wilson DJ, Viscidi RP, Abrams EJ, et al. Evolution of the human immunodeficiency virus envelope gene is dominated by purifying selection. Genetics 2006;174:1441–1453.

[9] Lipsitch M, O'Hagan JJ. Patterns of antigenic diversity and the mechanisms that maintain them. J R Soc Interface 2007;4:787–802.
[10] Ross HA, Rodrigo AG. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. J Virol 2002;76:11715–11720.
[11] Villet S, Pichoud C, Villeneuve JP, Trepo C, Zoulim F. Selection of a multiple drug-resistant hepatitis B virus strain in a liver-transplanted patient. Gastroenterology 2006;131:1253–1261.
[12] Yim HJ, Hussain M, Liu Y, Wong SN, Fung SK, Lok AS. Evolution of multi-drug resistant hepatitis B virus during sequential therapy. Hepatology 2006;44:703–712.
[13] Bertoletti A, Maini M, Williams R. Role of hepatitis B virus specific cytotoxic T cells in liver damage and viral control. Antiviral Res 2003;60:61–66.
[14] Huang CF, Lin SS, Ho YC, Chen FL, Yang CC. The immune response induced by hepatitis B virus principal antigens. Cell Mol Immunol 2006;3:97–106.
[15] Panther E, Spangenberg HC, Neumann-Haefelin C, Rosler K, Blum HE, von Weizsacker F, et al. The role of the virus specific T-cell response in acute and chronic HBV and HCV infection. Z Gastroenterol 2004;42:39–46.
[16] Tan AT, Koh S, Goh V, Bertoletti A. Understanding the immunopathogenesis of chronic hepatitis B virus: an Asian prospective. J Gastroenterol Hepatol 2008;23:833–843.
[17] Bruss V, Gerlich WH. Formation of transmembraneous hepatitis B e-antigen by cotranslational in vitro processing of the viral precore protein. Virology 1988;163:268–275.
[18] Garcia PD, Ou JH, Rutter WJ, Walter P. Targeting of the hepatitis B virus precore protein to the endoplasmic reticulum membrane: after signal peptide cleavage translocation can be aborted and the product released into the cytoplasm. J Cell Biol 1988;106:1093–1104.
[19] Carman WF, Jacyna MR, Hadziyannis S, Karayiannis P, McGarvey MJ, Makris A, et al. Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. Lancet 1989;2:588–591.
[20] Okamoto H, Tsuda F, Akahane Y, Sugai Y, Yoshiba M, Moriyama K, et al. Hepatitis B virus with mutations in the core promoter for an e antigen-negative phenotype in carriers with antibody to e antigen. J Virol 1994;68:8102–8110.
[21] Usuda S, Okamoto H, Iwanari H, Baba K, Tsuda F, Miyakawa Y, et al. Serological detection of hepatitis B virus genotypes by ELISA with monoclonal antibodies to type-specific epitopes in the preS2-region product. J Virol Meth 1999;80:97–112.
[22] Simmonds P. Recombination and selection in the evolution of picornaviruses and other Mammalian positive-stranded RNA viruses. J Virol 2006;80: 11124–11140.
[23] Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24: 1596–1599.
[24] Osiowy C, Giles E, Tanaka Y, Mizokami M, Minuk GY. Molecular evolution of hepatitis B virus over 25 years. J Virol 2006;80:10307–10314.
[25] 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.
[26] Hannoun C, Horal P, Lindh M. Long-term mutation rates in the hepatitis B virus genome. J Gen Virol 2000;81:75–83.
[27] Simmonds P. The origin and evolution of hepatitis viruses in humans. J Gen Virol 2001;82:693–712.
[28] Fares MA, Holmes EC. A revised evolutionary history of hepatitis B virus (HBV). J Mol Evol 2002;54:807–814.
[29] Locarnini S. Molecular virology and the development of resistant mutants: implications for therapy. Semin Liver Dis 2005;25 (Suppl. 1):9–19.
[30] Stumpf MP, Pybus OG. Genetic diversity and models of viral evolution for the hepatitis C virus. FEMS Microbiol Lett 2002;214:143–152.
[31] Lim SG, Cheng Y, Guindon S, Seet BL, Lee LY, Hu P, et al. Viral quasi-species evolution during hepatitis Be antigen seroconversion. Gastroenterology 2007;133:951–958.
[32] Chan HL, Hussain M, Lok AS. Different hepatitis B virus genotypes are associated with different mutations in the core promoter and precore regions during hepatitis B e antigen seroconversion. Hepatology 1999;29:976–984.
[33] 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.
[34] Blackberg J, Kidd-Ljunggren K. Genotypic differences in the hepatitis B virus core promoter and precore sequences during seroconversion from HBeAg to anti-HBe. J Med Virol 2000;60:107–112.
[35] Lindh M, Gustavson C, Mardberg K, Norkrans G, Dhillon AP, Horal P. Mutation of nucleotide 1, 762 in the core promoter region during hepatitis B e seroconversion and its relation to liver damage in hepatitis B e antigen carriers. J Med Virol 1998;55:185–190.
[36] Kim YS, Kim SI, Hwang SG, Kim JO, Cho JY, Lee JS, et al. Diversity of core promoter mutations in immune clearance phase of chronic HBV infection. Eur J Gastroenterol Hepatol 1999;11:821–825.
[37] Liu CJ, Chen PJ, Lai MY, Kao JH, Chen DS. Evolution of precore/core promoter mutations in hepatitis B carriers with hepatitis B e antigen seroreversion. J Med Virol 2004;74:237–245.
[38] Hsu YS, Chien RN, Yeh CT, Sheen IS, Chiou HY, Chu CM, et al. Long-term outcome after spontaneous HBeAg seroconversion in patients with chronic hepatitis B. Hepatology 2002;35:1522–1527.
[39] Hui CK, Leung N, Shek TW, Yao H, Lee WK, Lai JY, et al. Sustained disease remission after spontaneous HBeAg seroconversion is associated with reduction in fibrosis progression in chronic hepatitis B Chinese patients. Hepatology 2007;46:690–698.
[40] Liaw YF, Chu CM, Su IJ, Huang MJ, Lin DY, Chang-Chien CS. Clinical and histological events preceding hepatitis B e antigen seroconversion in chronic type B hepatitis. Gastroenterology 1983;84:216–219.
[41] Lok AS, Lai CL, Wu PC, Leung EK, Lam TS. Spontaneous hepatitis B e antigen to antibody seroconversion and reversion in Chinese patients with chronic hepatitis B virus infection. Gastroenterology 1987;92:1839–1843.
[42] Chu CM, Yeh CT, Lee CS, Sheen IS, Liaw YF. Precore stop mutant in HBeAg-positive patients with chronic hepatitis B: clinical characteristics and correlation with the course of HBeAg-to-anti-HBe seroconversion. J Clin Microbiol 2002;40:16–21.