J. Microbiol. Biotechnol. (2000), **10**(6), 823–828

JOURNAL OF
MICROBIOLOGY
AND
BIOTECHNOLOGY

Phylogenetic Analysis of Hepatitis B Virus Genome Isolated from Korean Patient Serum

KIM, SEON-YOUNG[1], HYEN-SAM KANG[2], AND YEON-SOO KIM[1]*

[1]*Laboratory of Cell Biology, Korea Research Institute of Bioscience and Biotechnology, 52 Eoundong, Yusong, Taejon 305-333, Korea*
[2]*Department of Microbiology, College of Natural Sciences, Seoul National University, Seoul 151-742, Korea*

Received: July 27, 2000
Accepted: October 1, 2000

Abstract The complete nucleotide sequence of hepatitis B virus DNA isolated from Korean patient serum was determined and characterized, and its phylogenetic relation was then investigated. The viral genome was 3,215 base pairs long and included four well known open reading frames (i.e. surface antigens, core antigens, X protein, and DNA polymerase). The sequence of the surface antigen showed that the HBV genome under investigation, designated HBV315, was characteristic of subtype *adr*. A phylogenetic analysis using the total genome sequence revealed that HBV315 was grouped into genomic group C together with isolates from Japan, China, Thailand, Polynesia, and New Caledonia. The mean percent similarity between HBV315 and other HBV isolates in genomic group C was 97.25%, and that with other genomic groups ranged from 86.16% to 91.25%. The predicted amino acid sequences of HBV315 were compared with two closely related subtype *adr* isolates, M38636 and D12980. The results showed that the X gene product was identical in the three strains, while there were significant amino acid sequence differences between HBV315 and M38636 in the Pre-S1 and Pre-S2 regions.

Key words: HBV, subtype *adr*, genomic group C, phylogenetic analysis

The Hepatitis B virus (HBV), a member of the hepadnavirus family, is a circular, partially double-stranded virus of approximately 3,200 nt. This genome has four open reading frames (ORFs) encoding the envelope (preS1, preS2, and the surface antigen HbsAg), core (preCore precursor protein, HbeAg, and HbcAg), polymerase, and X protein [18]. Hepadnaviruses are classified into two genera representing isolates from birds (Avihepadnavirus) and rodents and humans (Orthohepadnavirus). The avihepadnaviruses include the isolates from ducks and heron, while rodent viruses have been isolated from the woodchuck, ground squirrel, and arctic ground squirrel [26]. In primates, hepadnaviruses have been isolated from humans, chimpanzees, gibbons, orangutans, and woolly monkeys [14, 15, 17, 28]. Genotypically, HBV genomes have been classified into 6 groups, designated A-F, based on an intergroup divergence of 8% or more in the complete nucleotide sequence, and this genotypic grouping shows a characteristic geographical phenotype [19]. Group A genomes are found predominantly in northern Europe and sub-Saharan Africa. Groups B and C genomes are confined to the Far East. Group D genomes are found in the Mediterranean and the near and middle east, while group E is found in West Africa. Group F originates from aboriginal populations in Australia [28]. In addition, a recent report has also suggested a seventh group designated as group G [26].

HBV produces a chronic infection that affects a substantial proportion of the world population and results in death from chronic liver disease and hepatocellular carcinoma. About 5% of the human population is believed to be infected with HBV, and most of the Asia region (including Korea) remains as a high endemic area (infection ratio over 8%), thereby warranting extensive study of this virus [18]. There are many hepatitis and hepatocellular carcinoma patients in Korea [5, 10, 23], and among the many etiological agents, HBV is the most significant. In this regard, securing an entire molecular clone of HBV is highly valuable for studying various aspects of HBV biology. Previously, one full genome sequence of HBV from Korean patient serum was reported, which indicated a mutant subtype *adr* HBV, 3,213 bp long [11, 25]. However, since no systematic or phylogenetic study was performed in this aspect, its information content is limited. Although there have been a few phylogenetic studies using HBV isolates from Korean patients, these studies were restricted to only partial sequences of HBV [2, 13]. Accordingly, it would

*Corresponding author
Phone: 82-42-860-4270; Fax: 82-42-860-4597;
E-mail: kimys@mail.kribb.re.kr

appear that the current study is the first time that a phylogenetic analysis has been performed using full genome sequences of HBV isolated from a Korean patient.

About 16 years ago, the current authors cloned HBV from the serum of a Korean patient [12], and recently the complete nucleotide sequence was obtained. The HBV clone was then characterized, and a phylogenetic analysis performed with other HBV genomes to identify the phylogenetic position of the cloned isolate. This paper reports on the characteristics of the HBV clone and the results of the phylogenetic analysis.

MATERIALS AND METHODS

HBV Strains and DNA Sequencing

The HBV strain (HBV315) sequenced and characterized in this study was previously cloned and reported [12]. The whole HBV genome was sequenced using an ABI automated sequencer with seven sequencing primers designed successively (Table 1). The whole sequence was registered in the GenBank (accession number AF286594).

Genome Analysis and Phylogenetic Analysis

For the phylogenetic analysis, 26 other HBV sequences were retrieved from the GenBank and analyzed together (Table 1). A total of 27 sequences were aligned using the CLUSTALW program and adjusted manually to maximize the homology. The molecular phylogenetic analysis was performed using PAUP 4.02 [6, 27]. Most of the parsimonious trees were found by a heuristic search using the following options [Starting tree(s) obtained via stepwise addition, Simple addition sequence, TBR branch-swapping algorithm, Branches collapsed (creating polytomies) if maximum branch length is zero, and 'MulTrees' option in effect]. The woodchuck HBV isolate was used as an outgroup taxon to root the tree. Five-hundred bootstrap resamplings were performed to identify the statistical support for each branch. Sequence similarities between HBV sequences were obtained using PAUP 4.0. A neighbor-joining tree based on the pairwise distances was also calculated in PAUP4.0, as described above.

Table 1. Primers used in the sequencing of HBV315.

	Primer sequence (5'→3')	Position*
Primer 1	TTTCCGAGAGAGGACAACAG	1,359-1,340
Primer 2	AAACATTCCTTGAGCTTTAG'	947-928
Primer 3	CGATAGCCAGGACAAATTGG	372-353
Primer 4	ACCCTGGCCCGAATGCTCCC	3,046-3,027
Primer 5	AAGAATAAAGCCCAGTAAAG	2,499-2,480
Primer 6	TCCACAGAAGCTCCAAATTC	1,941-1,922
Primer 7	CACCTTTACCCCGTTGCCC	1,138-1157

*Positions are designated according to the numbering of Kim *et al.* [11].

Open Reading Frame Analysis and Comparison

The four reading frames of the HBV315 strain were analyzed using the DNASIS program (DNASIS 6.0, Hitachi), and compared with two other closely related *adr* strains [strain HPBCG (GenBank accession D12980) and strain hpbcgadr (GenBank accession M38636)].

RESULTS AND DISCUSSION

Complete Nucleotide Sequence and Genome Analyses

The construction of the recombinant plasmid pHBV315 was previously reported [12], and its detailed map is shown in Fig. 1. The HBV315 genome was digested at the *Bam*HI site located at position 1,400 and inserted into the *Bam*HI site in pBR322. The complete nucleotide sequence of the cloned HBV315 was 3,215 bp, which is typical of human HBV. The regulatory elements were located at the same position as in other HBV isolates. The 11-bp direct repeats (DR1 and DR2) were located at positions 1,590-1,600 and 1,824-1,834, respectively, the extended TATA box at position 2,776-2,782, and the extended poly-A addition signal at 1,916-1,921.

The detailed open reading frame analysis is shown in Fig. 2 in comparison with two closely related HBV strains (M38636 and D12980). HBV315 was coded for four open reading frames (P, X, C, and S) as in other HBV isolates. The X gene of HBV315, nucleotides 1,374-1,838, encode the X protein (154 amino acids), which is a trans-activator. The C gene, nucleotides 1,814-2,449, encode pre-C and core regions composed of 29 amino acids (nucleotides 1,814 to 1,900) and 183 amino acids (nucleotides 1901-2,452), respectively. The large surface antigen of HBV315, composed of pre-S2 (2,848 to 3,204, and 119 amino acids),

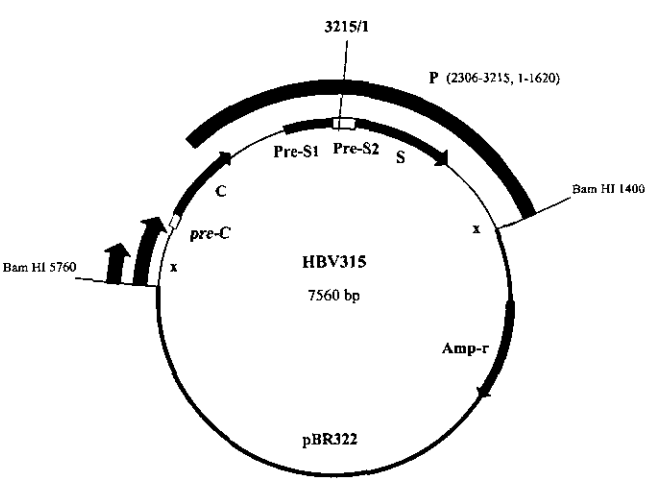

Fig. 1. The map of plasmid pHBV315. C, S, P, and X represent the 4 genes of HBV.
The nucleotide positions of the HBV genome are shown in parentheses, and the numbering was adopted from Kim *et al.* [11].

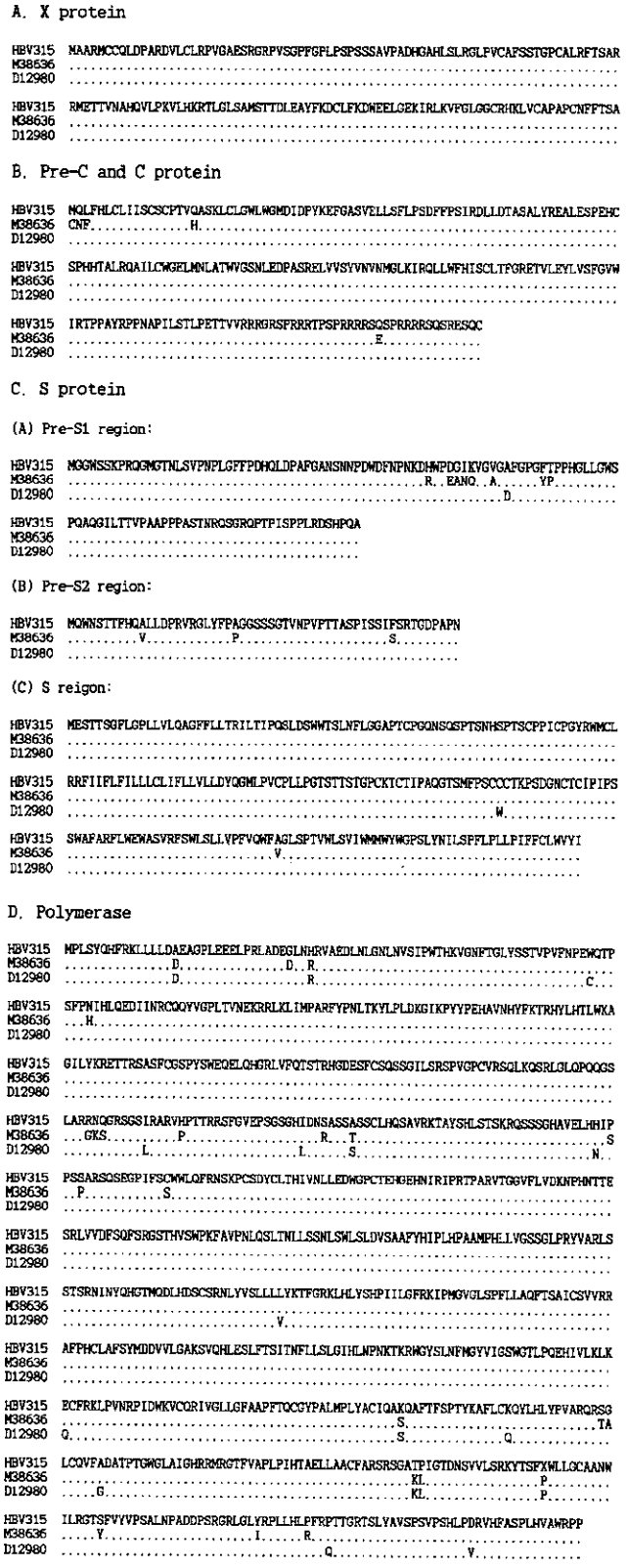

Fig. 2. Comparison of amino acid sequences of HBV315 with two closely related isolates (D12980 and M38636).
A, X protein; B, pre-C and C protein; C, pre-S1, pre-S2, and S antigen; and D, polymerase.

pre-S1 (3,205 to 154, and 55 amino acids), and S gene (155 to 835, and 226 amino acids) regions, encoded 400 amino acids. The P gene (nucleotides 2,307 to 1,623) encoded 813 amino acids, which was the DNA polymerase.

Phylogenetic Analysis

In the phylogenetic analysis, the data matrix consisted of 27 taxa with 3,256 characters. Among the 3,256 characters, 1,533 characters were found to be constant, 892 of the variable characters were parsimonious uninformative, and 831 characters were parsimony-informative. The heuristic search produced the three most parsimonious trees with 3,891 steps, CI (consistency index 0.630), RI (retention index 0.626), RC (rescaled consistency index 0.394), and HI (homoplasy index 0.370) (Fig. 3). The woodchuck HBV strain (WHV2, GenBank accession J02442) was used as an outgroup taxon to root the trees. The neighbor-joining algorithm produced a similar dendrogram with a parsimony-based phylogram (data not shown).

In the parsimony analysis, the HBV315 strain was clustered with other *adr* type isolates from Korea, Japan, Thailand, China, Polynesia, and New Caledonia, all belonging to genomic group C (bootstrap support 78%). The most closely related strain to HBV315 was HPBCG (GenBank accession D12980), isolated from Japan. The sequence

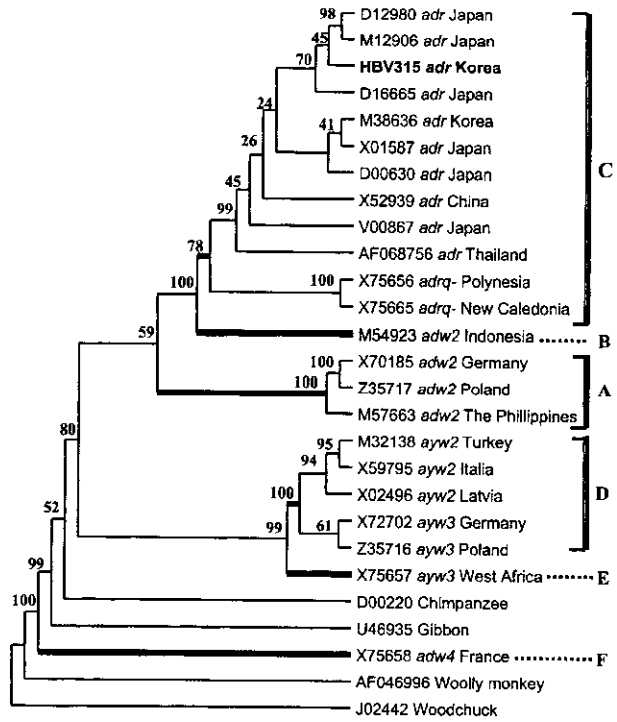

Fig. 3. Most parsimonious tree inferred from aligned nucleotide sequence of 27 taxa.
The Woodchuck squirrel HBV was used as an outgroup taxon to root the tree. This is one of the three most parsimonious trees (tree length 3,891 steps, CI 0.630, RI 0.626, RC 0.394, and HI 0.370). The statistical supports from 500 bootstrap resamplings are shown on each branch.

homology between HBV315 and D12980 was 98.63%, differing only in 44 base pairs. HBV315 and another Korean isolate, hpbcgadr (GenBank accession M38636), showed a 97.32% similarity, with 86 base pair differences. Genomic group C is characterized by the *adr* serotype and by its geographical distribution in East Asia and the Pacific region. In the genomic group C, isolates from East Asia (Korea, Japan, China, and Thailand) clustered together with 99% bootstrap support, and two isolates from Pacific region (Polynesia and New Caledonia) were clustered to the East Asian isolates branch with 78% bootstrap support. The mean intergroup similarities between HBV315 and other groups are shown in Table 2. Thus, HBV315 showed a 91.25% similarity with group A, 91.04% with group B, 89.98% with group D, 90.07 with group E, and 86.16% with group F. The similarity between HBV315 and the Chimpanzee isolate was 90%, and the similarity between HBV315 and the Gibbon HBV was 90.16%.

In the phylogenetic tree, 6 HBV genomic groups (A to F) were clearly identified as shown previously [19]. The Chimpanzee isolate D00220 and Gibbon isolate U46935 clustered inside the human HBV genomic group F isolate X75658, while the woolly monkey isolate AF046996 branched outside the human genomic group [14, 15].

Comparison of HBV315 with M38636 and D12980

Figure 2 shows a comparison of the amino acids sequences between HBV315, M38636, and D12980. The X gene product showed 100% identity between these isolates (Fig. 2A) [9]. In the C gene product, HBV315 was 100% identical with D12980, yet differed from M38686 by five amino acids (Fig. 2B). The different amino acids were centered on the amino terminal region of the pre-C region, and this was attributed to a frameshift mutation in M38686 as previously reported [11, 25].

The S gene product consists of three parts. In the Pre-S1 region, HBV315 only differed from D12980 by one amino acid (99.2% similarity), and differed from M38686 by 8 amino acids (93.3% similarity). The variable region was centered on 51 to 68 amino acids, which were not involved in the hepatocyte binding [21]. In the Pre-S2 region, HBV315 and D12980 were identical, while M38636 was different by three amino acids (96.1% similarity) [8]. In the S region, HBV315 and D12980 were identical again,

Table 2. Strains used in comparison with the HBV315.

Clone/Strain	Subtype	Genotype	Origin	Accession No.
pFDW294	*adw2*	A	The Philippines	M57663
HBVGen2	*adw2*	A	Poland	Z35717
pMND122	*adw2*	B	Indonesia	M54923
pRTP299	*adw2*	B	Indonesia	X75656
hpbhbva	*adr*	C	Japan	D16665
HBV315	*adr*	C	Korea	AF286594
pHBV1-1	*adr*	C	Japan	M12906
hpbcgadr	*adr*	C	Korea	M38636
pHBr330	*adr*	C	Japan	V00867
pBRHBadr4	*adr*	C	Japan	X01587
hbvprex	*adr*	C	China	X52939
HMA	*adrq-*	C	New Caledonia	X75665
NA	*adr*	C	Thailand	AF068756
HPBADRC	*adr*	C	Japan	D00630
HPBCG	*adr*	C	Japan	D12980
HBValpha1	*ayw2*	D	Turkey	M32138
pHBV320	*ayw2*	D	Latvia	X02496
aywmut	*ayw2*	D	Italia	X59795
A938	*ayw2*	D	Germany	X70185
pPYW796	*ayw3*	D	Germany	X72702
HBVGen1	*ayw3*	D	Poland	Z35716
Bas	*ayw3*	E	West Africa	X75657
For	*adw4*	F	France	X75658
adw/LSH	chimpanzee			D00220
Gibbon HV	Gibbon			U46935
WMHBV	woolly monkey			AF046996
WHV2	woodchunk			J02442

Table 3. Mean percentage of similarities in nucleotide sequences of complete genomes between HBV315 and HBV clones in group C or clones in other groups.

HBV clones	HBV315
D12980	98.63
M12906	98.26
D16665	97.02
M38636	97.32
X01587	97.42
D00630	97.54
X52939	97.20
V00867	97.99
AF068756	96.80
X75656	96.14
X75665	95.46
A	91.25
B	91.04
C	97.25
D	89.95
E	90.07
F	86.16
Chi-HBV	90.00
Gib-HBV	90.16
WM-HBV	78.31
WC-HBV	61.87

while M38686 differed by two amino acids (99.1% similarity) (Fig. 2C) [20, 22].

In the polymerase, HBV315 differed from D12980 by 17 amino acids (97.9% similarity), and from M38686 by 22 amino acids (97.5% similarity). Accordingly, the most variant region was the Pre-S1 region, and the most invariant region was the X region.

Future Application as a DNA Vaccine

The phylogenetic analysis showed that HBV315 belonged to genomic group C and serotype *adr*, which is the major type found in Eastern Asia including Korea. Based on the open reading frame analysis, HBV315 was shown to express four HBV genes (P, X, C, and S) normally without nonsense mutation or frameshift mutation, thereby rendering it as a valuable source for vaccine development. HBV infection remains an important worldwide health problem and the prospect for controlling further infection depends on the availability of safe, effective, and affordable vaccines. Although currently available antigen-based vaccines are safe and effective, they are quite expensive to produce for worldwide use, especially for underdeveloped countries. Furthermore, the emergence of diverse mutant HBV escaping existing vaccines necessitates the development of new vaccines [1, 29]. Recently, DNA vaccines have been successfully developed and proved to be as effective as antigen-based vaccines [3, 4, 7, 16, 24]. In this regard, the HBV315 clone we have studied may be a good starting point for the development of various DNA vaccines, especially for use in Asia. As shown in Fig. 1, the entire S region and C region can be easily obtained by a PCR. For example, by PCR amplification of the S region and subsequent cloning into appropriate mammalian expression vectors, a useful DNA vaccine against HBV subtype *adr* can easily be manufactured. Furthermore, site-directed mutagenesis can be applied to make various mutant DNA vaccines, including Gly-145-Arg, Met-133-The, Met-133-Leu, Gly-130-Asn, Ile-126-Thr, or Ile-126-Thr mutants [1], that have been resisting current vaccines.

REFERENCES

1. Chen, W. N. and C. J. Oon. 1999. Human hepatitis B virus mutants: Significance of molecular changes. *FEBS Lett.* **453:** 237-242.
2. Cho, S. W., Y. J. Shin, K. B. Hahm, J. H. Jin, Y. S. Kim, J. H. Kim, and H. J. Kim. 1999. Analysis of the precore and core promoter DNA sequence in liver tissues from patients with hepatocellular carcinoma. *J. Korean Med. Sci.* **14:** 424-430.
3. Chow, Y. H., W. L. Huang, W. K. Chi, Y. D. Chu, and M. H. Tao. 1997. Improvement of hepatitis B virus DNA vaccines by plasmids coexpressing hepatitis B surface antigen and interleukin-2. *J. Virol.* **71:** 169-178.
4. Geissler, M., V. Bruss, S. Michalak, B. Hockenjos, D. Ortmann, W. B. Offensperger, J. R. Wands, and H. E. Blum. 1999. Intracellular retention of hepatitis B virus surface proteins reduces interleukin-2 augmentation after genetic immunizations. *J. Virol.* **73:** 4284-4292.
5. Han, B. K., B. Y. Lee, M. K. Min, and K. H. Jung. 1998. Expression and characterization of recombinant E2 protein of Hepatitis C Virus by insect cell/Baculovirus expression system. *J. Microbiol. Biotechnol.* **8:** 361-368.
6. Hong, S. G., J. S. Chun, J. S. Nam, Y. D. Park, and K. S. Bae. 2000. Phylogenetic analysis of genus Sporobolomyces based on partial sequences of 26S rDNA. *J. Microbiol. Biotechnol.* **10:** 363-366.
7. Hui, J., M. Mancini, G. Li, Y. Wang, P. Tiollais, and M. L. Michel. 1999. Immunization with a plasmid encoding a modified hepatitis B surface antigen carrying the receptor binding site for hepatocytes. *Vaccine* **17:** 1711-1718.
8. Kidd-Ljunggren, K., A. M. Courouce, M. Oberg, and A. H. Kidd. 1994. Genetic conservation within subtypes in the hepatitis B virus pre-S2 region [published erratum appears in *J. Gen. Virol.* 1995 Mar; 76(Pt 3): 727]. *J. Gen. Virol.* **75(Pt 6):** 1485-1490.
9. Kidd-Ljunggren, K., M. Oberg, and A. H. Kidd. 1995. The hepatitis B virus X gene: Analysis of functional domain variation and gene phylogeny using multiple sequences. *J. Gen. Virol.* **76(Pt 9):** 2119-2130.
10. Kim, H. S., Y. J. Chung, Y. J. Jeon, and S. H. Lee. 1999. Large-scale culture of Hepatitis A Virus in human diploid MRC-5 cells and partial purification of the viral antigen for use as a vaccine. *J. Microbiol. Biotechnol.* **9:** 386-392.

11. Kim, K.-T., Sang Won Hyun, Yong Sok Kim, and Hyune Mo Rho. 1988. Complete nucleotide sequence of Hepatitis B Virus (subtype adr). *Korean Biochem. J.* **21:** 319-331.
12. Kim, Y.-S. and Hyen-Sam, Kang. 1984. Cloning and expression of Hepatitis B Virus surface antigen gene. *Korean Biochem. J.* **17:** 70-79.
13. Koh, K. C., H. S. Lee, and C. Y. Kim. 1995. Association of the core clustering mutations (codon 21-34) and the severity of chronic hepatitis B in Korean patients. *Korean J. Intern. Med.* **10:** 87-93.
14. Lanford, R. E., D. Chavez, K. M. Brasky, R. B. Burns, 3rd, and R. Rico-Hesse. 1998. Isolation of a hepadnavirus from the woolly monkey, a New World primate. *Proc. Natl. Acad Sci. USA* **95:** 5757-5761.
15. Lanford, R. E., D. Chavez, R. Rico-Hesse, and A. Mootnick. 2000. Hepadnavirus infection in captive gibbons. *J. Virol.* **74:** 2955-2959.
16. Loirat, D., Z. Li, M. Mancini, P. Tiollais, D. Paulin, and M. L. Michel. 1999. Muscle-specific expression of hepatitis B surface antigen: No effect on DNA-raised immune responses. *Virology* **260:** 74-83.
17. MacDonald, D. M., E. C. Holmes, J. C. Lewis, and P. Simmonds. 2000. Detection of hepatitis B virus infection in wild-born chimpanzees (Pan troglodytes verus): Phylogenetic relationships with human and other primate genotypes. *J. Virol.* **74:** 4253-4257.
18. Mahoney, F. J. 1999. Update on diagnosis, management, and prevention of hepatitis B virus infection. *Clin. Microbiol. Rev.* **12:** 351-366.
19. Norder, H., A. M. Courouce, and L. O. Magnius. 1994. Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. *Virology* **198:** 489-503.
20. Norder, H., A. M. Courouce, and L. O. Magnius. 1993. Complete nucleotide sequences of six hepatitis B viral genomes encoding the surface antigen subtypes ayw4, adw4q-, and adrq- and their phylogenetic classification. *Arch. Virol.* Suppl. **8:** 189-199.
21. Norder, H., B. Hammas, S. Lofdahl, A. M. Courouce, and L. O. Magnius. 1992. Comparison of the amino acid sequences of nine different serotypes of hepatitis B surface antigen and genomic classification of the corresponding hepatitis B virus strains. *J. Gen. Virol.* **73(Pt 5):** 1201-1208.
22. Ohba, K., M. Mizokami, T. Ohno, K. Suzuki, E. Orito, J. Y. Lau, Y. Ina, K. Ikeo, and T. Gojobori. 1995. Relationships between serotypes and genotypes of hepatitis B virus: Genetic classification of HBV by use of surface genes. *Virus Res.* **39:** 25-34.
23. Park, C. H., Y. H. Kee, J. H. Lee, J. H. Oh, J. C. Park, and H. J. Myung. 1999. Purification and characterization of recombinant Hepatitis C Virus replicase. *J. Microbiol. Biotechnol.* **9:** 881-884.
24. Prange, R. and M. Werr. 1999. DNA-mediated immunization to hepatitis B virus envelope proteins: preS antigen secretion enhances the humoral response. *Vaccine* **17:** 617-623.
25. Rho, H. M., K. Kim, S. W. Hyun, and Y. S. Kim. 1989. The nucleotide sequence and reading frames of a mutant hepatitis B virus subtype adr. *Nucleic Acids Res.* **17:** 2124.
26. Stuyver, L., S. De Gendt, C. Van Geyt, F. Zoulim, M. Fried, R. F. Schinazi, and R. Rossau. 2000. A new genotype of hepatitis B virus: Complete genome and phylogenetic relatedness. *J. Gen. Virol.* **81:** 67-74.
27. Swofford, D. L. 1998. *PAUP*. Phylogenetic Analysis Using Parsimony* (* and other methods), 4.0 ed. Sinauer Associates, Sunderland, MA, U.S.A.
28. Warren, K. S., J. L. Heeney, R. A. Swan, Heriyanto, and E. J. Verschoor. 1999. A new group of hepadnaviruses naturally infecting orangutans (Pongo pygmaeus). *J. Virol.* **73:** 7860-7865.
29. Wu, L., J. W. He, X. Yao, H. M. Li, and Y. M. Wen. 1999. A novel hepatitis B virus variant S 129 (Gln → Leu): Lack of correlation between antigenicity and immunogenicity. *J. Med. Virol.* **59:** 424-430.