FEBS Letters 450 (1999) 66–71 FEBS 21936

Hepatitis B virus genotype assignment using restriction fragment length polymorphism patterns

Masashi Mizokami[a,b,*], Tatsunori Nakano[b], Etsuro Orito[b], Yasuhito Tanaka[b], Hiroshi Sakugawa[c], Motokazu Mukaide[d], Betty H. Robertson[e]

[a] *Department of Blood Transfusion, Nagoya City University Medical School, Nagoya, Japan*
[b] *Second Department of Medicine, Nagoya City University Medical School, Kawasumi, Mizuho, Nagoya 467-8601, Japan*
[c] *First Department of Medicine, University of the Ryukyus, Okinawa, Japan*
[d] *Center of Molecular Biology and Cytogenetics, SRL, Inc., Tokyo, Japan*
[e] *Hepatitis Branch, Division of Viral and Rickettsial Diseases, Center for Disease Control and Prevention, Atlanta, GA, USA*

Received 15 March 1999

Abstract Hepatitis B virus (HBV) is classified into genotypes A–F, which is important for clinical and etiological investigations. To establish a simple genotyping method, 68 full-genomic sequences and 106 S gene sequences were analyzed by the molecular evolutionary method. HBV genotyping with the S gene sequence is consistent with genetic analysis using the full-genomic sequence. After alignment of the S sequences, genotype specific regions are identified and digested by the restriction enzymes, *Hph*I, *Nci*I, *Alw*I, *Ear*I, and *Nla*IV. This HBV genotyping system using restriction fragment length polymorphism (RFLP) was confirmed to be correct when the PCR products of the S gene in 23 isolates collected from various countries were digested with this method. A restriction site for *Ear*I in genotype B was absent in spite of its presence in all the other genotypes and genotype C has no restriction site for *Alw*I. Only genotype E is digested with *Nci*I, while only genotype F has a restriction site for *Hph*I. Genotype A can be distinguished by a single restriction enzyme site for *Nla*IV, while genotype D digestion with this enzyme results in two products that migrates at 265 and 186 bp. This simple and accurate HBV genotyping system using RFLP is considered to be useful for research on HBV.

Key words: Hepatitis B virus; Genotyping; Restriction fragment length polymorphism; Restriction enzyme; S gene

1. Introduction

Hepatitis B virus (HBV) is a partially double-stranded DNA virus, which is a major causative agent of acute and chronic hepatitis, liver cirrhosis and hepatocellular carcinoma in Asian, African and Southern European countries. The HBV genome has a compact genetic organization with four open reading frames, C, S, X and P. The P gene covers more than 70% of the complete genome and overlaps the entire pre-S and S genes and X and C genes partially [1]. Comparison of complete HBV genome sequences has revealed genetic clusters which have been classified as six different HBV genotypes designated A–F [2,3]. An HBV genotype has been defined as sequences which have a similarity over the whole genome of about 92% [2]. Since HBV genotypes predominate in certain geographic areas, there has been speculation about the relationship between HBV genotypes, development of chronic liver disease and the use of HBV sequences for anthropological studies [4].

The aims of this study are to confirm the classification of HBV genotypes by molecular phylogenetic analysis with the full-genomic sequences and sequences of the S gene, and to establish a simple HBV genotyping system using restriction fragment length polymorphism (RFLP) analysis of the S gene region.

2. Materials and methods

2.1. Phylogenetic tree of the full genome and S gene of HBV

First, 68 full-genomic HBV sequences were retrieved from the DNA database (DDBJ/GenBank) and aligned to maximize homologies. Genetic distances were estimated using the six-parameter method [5] and a phylogenetic tree was constructed using the neighbor-joining (N-J) method [6]. The S genes obtained from these 68 full-genomic sequences were then analyzed separately using the same method to evaluate the reliability of this region alone. Finally, these 68 S region sequences, 11 additional S region sequences from DDBJ/GenBank and 23 newly determined S region sequences were used to construct a phylogenetic tree using the same method.

To confirm the reliability of the phylogenetic tree, bootstrap resampling tests were performed 1000 times [7].

2.2. Determination of the sequences of the S gene

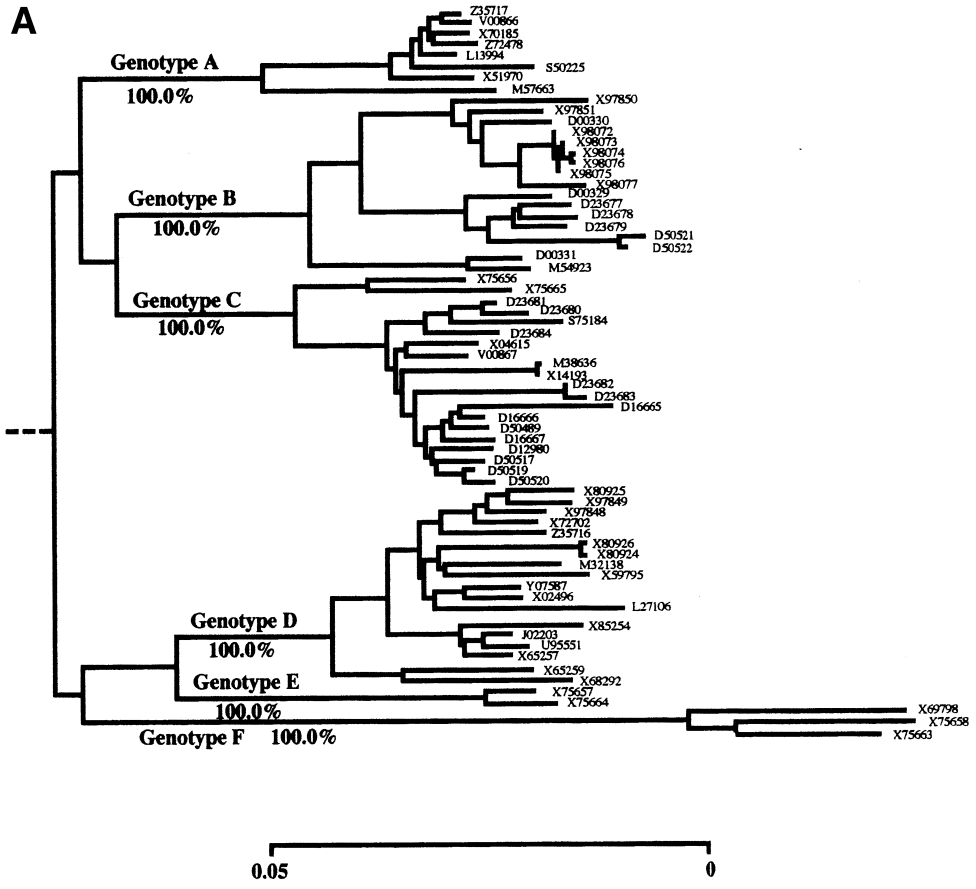
Twenty-three serum samples were evaluated from patients with chronic HBV infection in diverse geographic locations including the Altai in Siberia, Central America, China, Ghana, Japan, the Kamchatka peninsula (Siberia), Mongolia, Saudi Arabia and Vietnam. To amplify HBV DNA by polymerase chain reaction (PCR), DNA extracted from the serum was added to the following amplification mixture: 2.5 μl of 10×EX-Taq polymerase buffer (Takara Shuzo Co., Japan), 2.5 μl of 2.5 mM deoxyribonucleotide triphosphate, 0.1 μl of EX-Taq (5 units, Takara Shuzo Co., Japan), 10 pmol of sense (HBV11; 5′-GGGTCACCATATTCTTGGGAACAAGAKCTAC-3′)

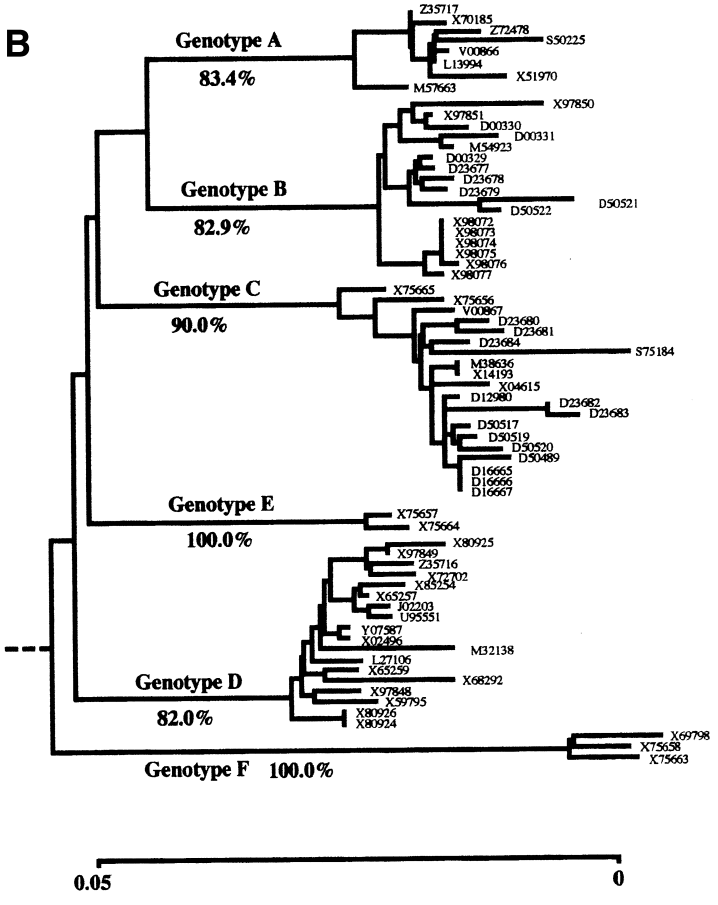
→

Fig. 1. A: Phylogenetic tree of the full-genomic sequences constructed by the neighbor-joining method using 68 full-genomic sequences retrieved from DDBJ/GenBank. The names on the tree indicate the accession number of each sequence. On the bootstrap analysis for evaluation of the statistical reliability of the tree, each genotype indicates 100%. The horizontal bar indicates the number of nucleotide substitutions per site. B: Phylogenetic tree of 68 sequences of the S gene is also constructed with 68 sequences of the S gene derived from the 68 full-genomic sequences. The topology of this tree is the same as that of the full-genomic tree and the reliabilities of genotypes A, B, C, D, E and F are 83.4%, 82.9%, 90.0%, 100%, 82.0% and 100%, respectively. The horizontal bar indicates the number of nucleotide substitutions per site.

*Corresponding author. Fax: (81) (52) 853-8682.
E-mail. mizokami@med.nagoya-cu.ac.jp

0014-5793/99/$20.00
PII: S0014-5793(99)00471-8


A
Genotype A
100.0%
Genotype B
100.0%
Genotype C
100.0%
Genotype D
100.0%
Genotype E
100.0%
Genotype F 100.0%
Z35717
V00866
X70185
Z72478
L13994
S50225
X51970
M57663
X97850
X97851
D00330
X98072
X98073
X98074
X98076
X98075
X98077
D00329
D23677
D23678
D23679
D50521
D50522
D00331
M54923
X75656
X75665
D23681
D23680
S75184
D23684
X04615
V00867
M38636
X14193
D23682
D23683
D16665
D16666
D50489
D16667
D12980
D50517
D50519
D50520
X80925
X97849
X97848
X72702
Z35716
X80926
X80924
M32138
X59795
Y07587
X02496
L27106
X85254
J02203
U95551
X65257
X65259
X68292
X75657
X75664
X69798
X75658
X75663
0.05
0


B
Genotype A
83.4%
Genotype B
82.9%
Genotype C
90.0%
Genotype E
100.0%
Genotype D
82.0%
Genotype F 100.0%
Z35717
X70185
Z72478
S50225
V00866
L13994
X51970
M57663
X97850
X97851
D00330
D00331
M54923
D00329
D23677
D23678
D23679
D50522
D50521
X98072
X98073
X98074
X98075
X98076
X98077
X75665
X75656
V00867
D23680
D23681
D23684
S75184
M38636
X14193
X04615
D12980
D23682
D23683
D50517
D50519
D50520
D50489
D16665
D16666
D16667
X75657
X75664
X80925
X97849
Z35716
X72702
X85254
X65257
J02203
U95551
Y07587
X02496
M32138
L27106
X65259
X68292
X97848
X59795
X80926
X80924
X69798
X75658
X75663
0.05
0

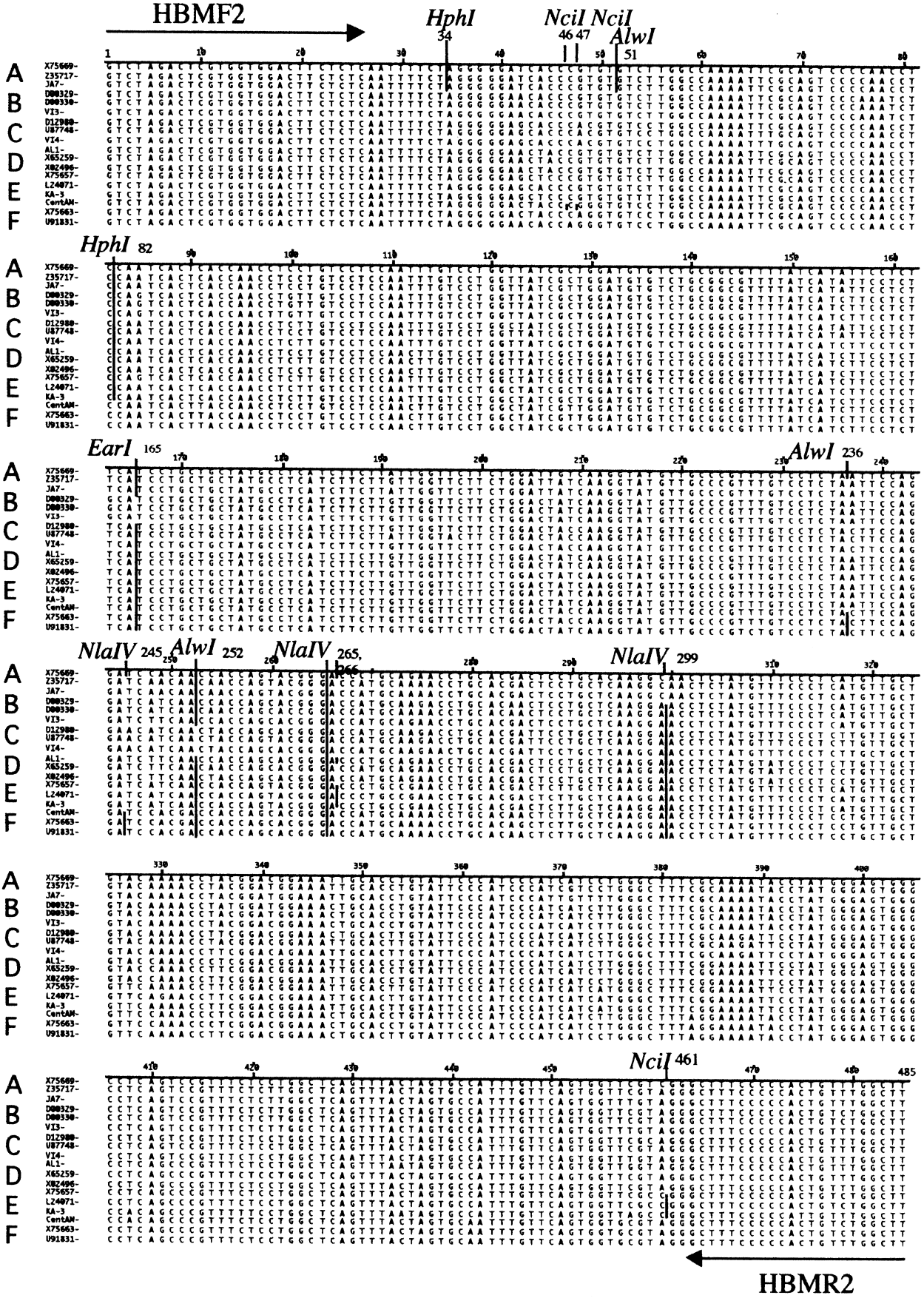


Fig. 2. Alignment of the sequences of the S gene. The alignment of the sequences of the S gene is shown with representative sequences of each genotype and the isolates in this study. The restriction enzyme recognition sites are indicated.

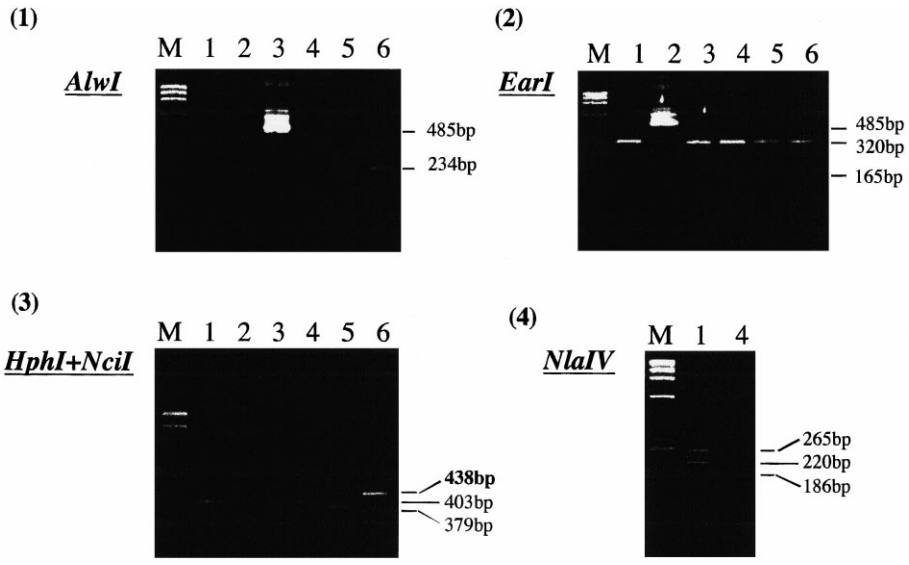


Fig. 3. RFLP pattern of the products of the S gene. Lanes 1–6 indicate HBV genotypes A–F, respectively. 1: Only genotype C was digested by *Alw*I. 2: Only genotype B was not digested by *Ear*I. 3: By *Hph*I and *Nci*I, genotypes E and F were digested, resulting in fragments of 379 bp and 438 bp, respectively, although the other genotypes were 403 bp long. 4: By *Nla*IV, genotype A was digested into 220 and 265 bp fragments while genotype D was 186 bp.

and antisense primers (HBV22; 5′-CAATWCKYTGACANACTTTCCAATCARTWGG-3′) in a 25 μl reaction volume. The amplification profile was 2 min at 96°C, followed by 25 cycles at 94°C for 15 s (denaturation), 45 s at 60°C (annealing) and 45 s at 72°C (extension), and was performed in a 96-well cycler (GeneAmp 9600, Perkin-Elmer Cetus, Norwalk, CT, USA). 1 μl of the first-round PCR product was then added to a second-round PCR mixture with the same composition but with a set of inner sense (HBMF1; 5′-YCCTGCTGGTGGCTCCAGTTC-3′) and antisense primers (HBV22; 5′-CAATWCKYTGACANACTTTCCAATCARTWGG-3′). 5 μl of the second-round PCR product was then analyzed by electrophoresis in 3% agarose gels, stained with ethidium bromide, and visualized under ultraviolet light.

The PCR primers used in this study were designed based on the most conserved regions derived from the known sequences published in DDBJ/GenBank. These primers were confirmed to be specific for HBV through a computerized search. The primers were prepared with a DNA synthesizer (Model 394, Perkin-Elmer Cetus). Standard precautions for avoiding contamination during PCR were observed. A negative control serum was also included in each run to ensure specificity.

All the PCR products were ligated into the pGEM-T vector (Promega, Madison, WI, USA). Positive clones were identified, expanded, and sequenced bidirectionally using the dideoxynucleotide chain termination method (Dye Terminator Cycle Sequencing FS Ready Reaction Kit, Perkin-Elmer Cetus). The sequences were determined using a DNA sequencer (ABI Prism 377 DNA Sequencer, Perkin-Elmer Cetus).

2.3. HBV genotyping by RFLP analysis

To design a genotyping system using RFLP, 85 S gene sequences were aligned and analyzed to determine the conserved sequences which were genotype-specific. After identifying the restriction enzyme sites that identified genotype-specific sequences, HBV DNA extracted from the 23 serum samples were amplified by nested PCR with the first-round sense (HBMF1; 5′-YCCTGCTGGTGGCTCCAGTTC-3′) and antisense primers (HBMR2; 5′-AAGCCANACARTGGGGGAAAGC-3′) and then the second-round inner sense (HBMF2; 5′-GTCTAGACTCGTGGTGGACTTCTCTC-3′) and antisense primers (HBMR2; 5′-AAGCCANACARTGGGGGAAAGC-3′), under the same conditions described above. Restriction digestions were carried out with 5 μl of the second-round PCR product for 3 h after adjust-

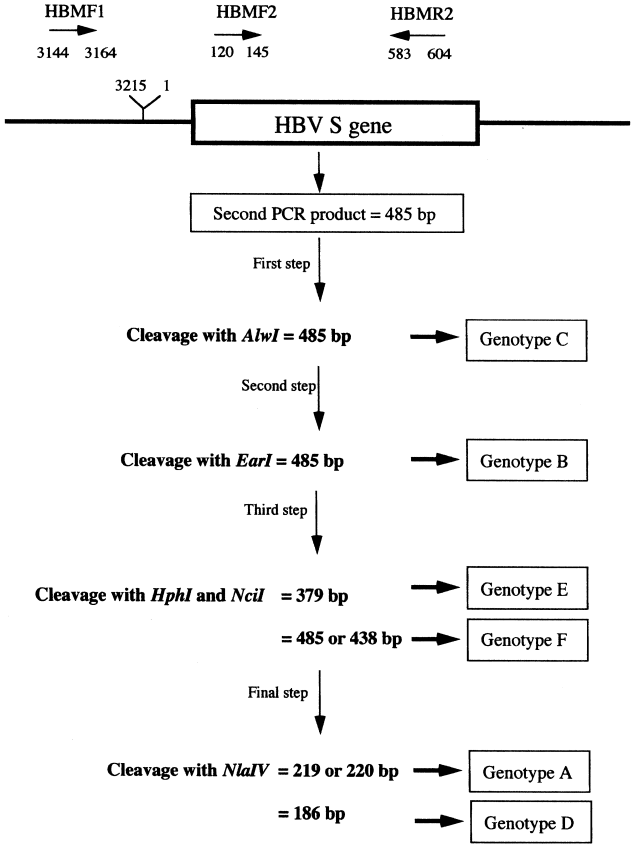


Fig. 4. Strategy of the HBV genotyping by RFLP. The second-round PCR products, which were 485 bp, were digested by *Alw*I and *Ear*I to find genotypes C and B. *Hph*I and *Nci*I digest genotypes E and F. Finally, genotypes A and D are found by *Nla*IV.

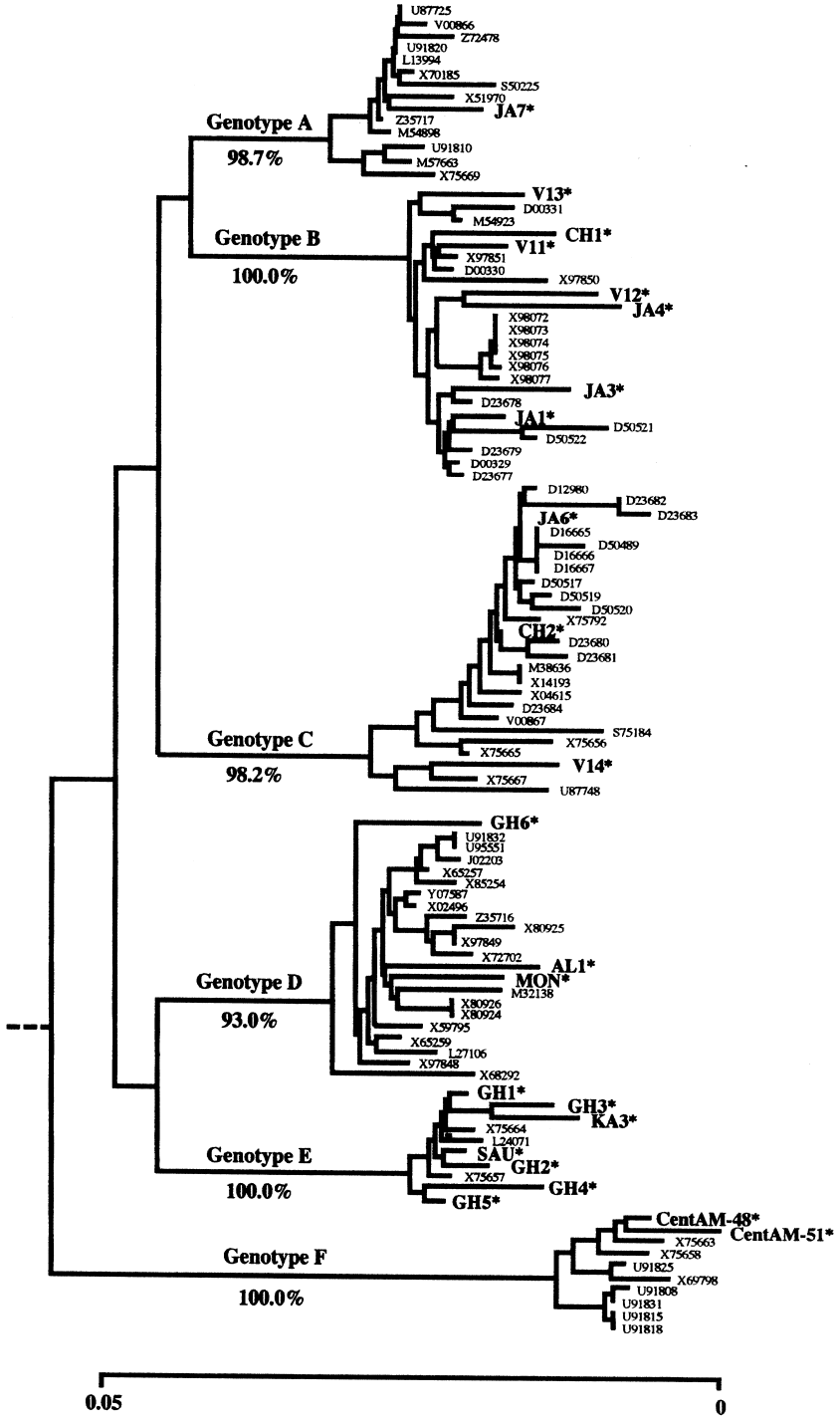


Fig. 5. Phylogenetic tree of the S gene. This phylogenetic tree was constructed with 85 sequences of the S gene from DDBJ/GenBank and 23 sequences determined in this study which are indicated by an asterisk against the name. The result of the RFLP analysis showed accordance with the classification by the phylogenetic tree. AL = Altai, CH = China, CentAM = Central America, GH = Ghana, JA = Japan, KA = Kamchatka, Mon = Mongolia, SAU = Saudi Arabia, and V = Vietnam.

ment with 10×enzyme reaction buffer according to the manufacturer's recommendations. Reactions were carried out with 10 units of *Alw*I, *Hph*I, *Nci*I, *Nla*IV or *Ear*I (New England BioLabs) at 37°C. The digested PCR products were electrophoresed on 3.0% Nusieve GTG (3:1) agarose gel in 1×TBE buffer (134 mM Tris-HCl, pH 10; 68 mM boric acid; 2.5 mM EDTA) containing 500 ng ethidium bromide per ml. The RFLP pattern was then evaluated under ultraviolet light.

3. Results

3.1. Comparison of phylogenetic analysis of complete HBV genome sequences and S gene sequences

A molecular phylogenetic tree of the 68 full-genomic HBV sequences was constructed using the N-J method is shown in Fig. 1A. In this phylogenetic tree, six major genotypes, des-

ignated A–F, are clearly identified with 100.0% bootstrap values for each of the major branches. In addition, in the 68 sequences of the S gene, the same genetic clustering of sequences as that in the full-genomic sequences was identified by the phylogenetic tree with 82.0–100.0% bootstrap values. Thus, based upon these analyses, HBV genotypes can be classified using the sequences of the S gene.

3.2. RFLP patterns

The alignment of S gene sequences was inspected to determine whether restriction sites could be identified that corresponded to genotype-specific sequences. Five restriction enzymes, *Hph*I, *Nci*I, *Alw*I, *Ear*I and *Nla*IV, were identified which should result in genotype-specific RFLP from the S gene region. Representative sequences from the S genome region of each HBV genotype were aligned to show the restriction sites by the five restriction enzymes (Fig. 2). Genotype B could be distinguished by the fact that the S gene fragment was uncut by *Ear*I, while no *Alw*I site was present in S gene sequences of genotype C. Only genotype E had a restriction site for *Nci*I at position 461, and only genotype F had a restriction site for *Hph*I at position 82. For genotype A, the specific restriction enzyme site for *Nla*IV was not found at position 299, but genotype D was digested at positions 265 and 299 by *Nla*IV. The pattern observed with these restriction enzymes is shown in panels 1–4 of Fig. 3 and the strategy for HBV genotyping using this approach is summarized in Fig. 4. Firstly, genotype C is found by *Alw*I digestion, and secondly, genotype B is identified using *Ear*I. Thirdly, genotypes E and F can be identified with *Hph*I and *Nci*I. Finally, the genotypes A and D were identified by *Nla*IV (Fig. 4).

3.3. Verification of RFLP genotype assignment by sequence analysis

Sequences of the 23 samples evaluated by RFLP were determined and compared to representative S gene sequences obtained from GenBank/DDBJ ($n = 83$) using the N-J method for phylogenetic analysis (Fig. 5). S region sequences cluster with the appropriate genotypes based upon phylogenetic analysis and confirm the validity of the RFLP technique.

4. Discussion

In this study, two important points are illustrated. Firstly, we demonstrate that phylogenetic analysis of the S gene and the full-length HBV genome by the N-J method generates trees with clearly separated genetic clusters. This means that genotypes for HBV can be assigned based upon S gene sequences. Secondly, a new simple HBV genotyping system has been established using the RFLP method.

Serologic classification of HBV into the four major subtypes, *adr*, *adw*, *ayr* and *ayw* [8], is based upon a limited number of amino acids substitutions within the '*a*' determinant of HBsAg and serotypes do not strictly correlate with genotypic classification [3,4,9]. Full-genomic sequences are the ideal for genotype assignment, but sequence determination is costly and time-consuming. To simplify classification of HBV genotypes for new isolates, a genotyping method has been developed using RFLP. HBV genotyping by RFLP is less complicated compared to direct sequencing and sequence analysis. In future it may be important to evaluate the etiological or clinical relevance of HBV genotypes to predict the progression of liver disease or to investigate routes of infection, and this simple and exact genotyping system will be useful for these investigations.

In the past, other HBV genotyping systems using RFLP analysis of a PCR product of the S gene or pre-S sequence have been reported [10–12]. However, the HBV genotype of 11 isolates deduced from DDBJ/GenBank are not appropriately classified by the RFLP system using the pre-S region (data not shown). The sequence of the S gene is more conserved than the pre-S region because the S gene overlaps the reverse transcriptase active site in the P gene which is encoded in a different frame [13]. Therefore, the S gene is more suitable for genotyping than the pre-S region. Lindh et al. [12] also reported HBV genotyping by RFLP using the S region, but they did not analyze the topologies and the reliabilities of trees generated using the S gene and full genome sequences. In conclusion, the RFLP system that we have described in relatively simple to interpret and is expected to be widely used for clinical and etiological research of HBV.

Acknowledgements: The authors would like to thank Prof. M. Hayami, Institute for Virus Research, Kyoto University, Prof. S. Sonoda, Department of Virology, Faculty of Medicine, Kagoshima University, and Dr. K. Tajima, Division of Epidemiology, Aichi Cancer Research Institute for collecting the samples from Altai, Ghana, Kamchatka and Vietnam and also thank Prof. R. Williams, and Dr. C.J. Tibbs, Institute of Liver Studies, King's College Hospital, for collecting samples from Saudi Arabia.

References

[1] Tiollais, P., Pourcel, C. and Dejan, A. (1985) Nature 317, 489–495.
[2] Okamoto, H., Tsuda, F., Sakugawa, H., Sastrosoewignjo, R.I., Imai, M., Miiyakawa, Y. and Mayumi, M. (1988) J. Gen. Virol. 69, 2575–2583.
[3] Norder, H., Courouce, A.M. and Magnius, L.O. (1994) Virology 198, 489–503.
[4] Orito, E., Mizokami, M., Ina, Y., Moriyama, E.N., Kameshima, N., Yamamoto, M. and Gojobori, T. (1989) Proc. Natl. Acad. Sci. USA 86, 7059–7062.
[5] Gojobori, T., Ishii, K. and Nei, M. (1982) J. Mol. Evol. 18, 414–423.
[6] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[7] Felsenstein, J. (1985) Evolution 39, 783–791.
[8] Bancroft, W.H., Mundon, F.K. and Russell, P.K. (1972) J. Immunol. 109, 842–848.
[9] Ohba, K., Mizokami, M., Ohno, T., Suzuki, K., Orito, E., Lau, J.Y.N., Ina, Y., Ikeo, K. and Gojobori, T. (1995) Virus Res. 39, 25–34.
[10] Lindh, M., Gonzalez, J.E., Norkrans, G. and Horal, P. (1998) J. Virol. Methods 72, 163–174.
[11] Shih, J.W., Cheung, L.C., Alter, H.J., Lee, L.M. and Gu, R.J. (1991) J. Clin. Microbiol. 29, 1640–1644.
[12] Lindh, M., Anderson, A.-S. and Gusdal, A. (1997) J. Infect. Dis. 175, 1285–1295.
[13] Mizokami, M., Orito, E., Ohba, K., Ikeo, K., Lau, J.Y.N. and Gojobori, T. (1997) J. Mol. Evol. 44, (Suppl. 1) S83–S90.