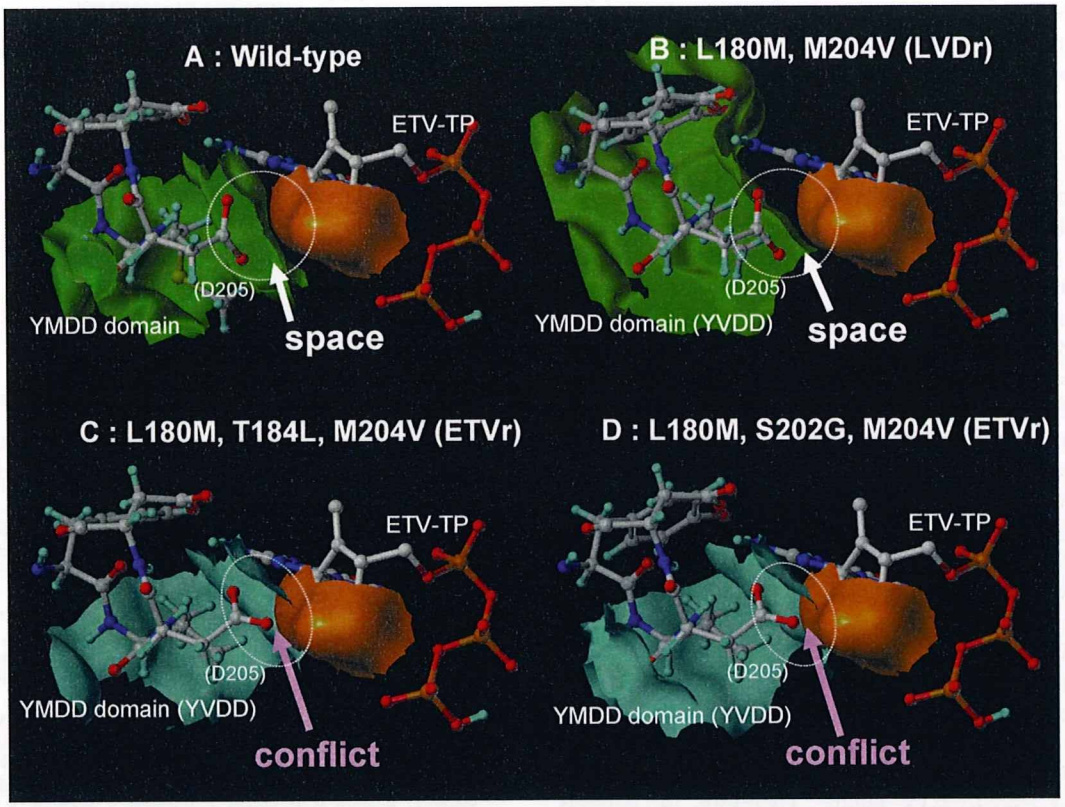

FIG. 3. 3D structures of the dNTP-binding domains of HBV RT of the wild type (A), an LVDr substitution (B), and ETVr substitutions (C and D). The molecular surfaces of the wild type and the LVDr mutant are drawn in green, those of LVDr plus ETVr mutants are drawn in blue, and that of ETV-TP is drawn in orange.

lower binding affinity for ETV-TP than the wild type. Molecular docking simulation in the present study showed that the L180M, M204V, S202G, and T184L substitutions can lessen the affinity of ETV-TP for HBV RT by heightening the potential energy between them, suggesting that S202G and T184L substitutions, in addition to M204V in the YMDD motif and L180M in domain C, could affect the initial polymerase binding of dNTP analog inhibitors.

DISCUSSION

Based on the combination of clinical observations and 3D docking simulation, this is the first report to suggest the mechanism by which ETVr substitutions (T184SCGA/ILFM and S202G, but not S202C), in addition to LVDr (L180M and M204V, but not M204I), can induce BT during ETV therapy.

TABLE 6. Minimal distances and binding potentials between ETV and the HBV RT domain in the wild type, one LVDr mutant, and 2 ETVr mutants

Strain	distance (Å)	Binding potential (GOLD score [Kcal/mol])	Reference (Fig. 3)
Wild type	1.3	−178.4	A
L180M, 204V	1.5	−141.3	B
L180M, T184L, 204V	2.2	−117.9	C
L180M, S202G, 204V	2.1	−99.8	D

First, an assessment of virological and biochemical events during a 3-year ETV treatment course showed that ETVr substitutions were absent among treatment-naïve patients but were detected in 44.8% of patients who were refractory to LVD during the preceding treatment period. Evidence of BT during ETV therapy was observed in 26.8% of LVD-refractory patients between weeks 60 and 144 of treatment. All 11 of the BT cases had both L180M and M204V/I substitutions at baseline (LVD refractory), as well as additional substitutions, such as T184 and/or S202G (not S202I/C), during the 3-year ETV treatment period.

Statistically significant risk factors for BT were the presence of LVDr (L180M and M204V) at baseline, detection of ETVr (S202G and T184SCGA/ILFM substitutions) during ETV treatment, and undetectable HBV DNA (<2.6) or more than a 2-$\log_{10}$-unit reduction in HBV DNA levels during the first year of ETV treatment. Detection of T184SCGA/ILFM and S202G was significantly associated with BT independent of age, gender, and LVDr (M204V and/or L180M) at baseline or nondetection or reduction in HBV DNA at the first year of treatment, indicating that these substitutions could be used as predictive markers for BT.

The mechanism by which combinations of ETVr (S202G and T184 SCGA/ILFM) and LVDr (L180M and M204V) can induce BT during ETV therapy is largely unknown. Note that T184L and S202G residues are located within domain B and domain C of the RT/polymerase, respectively, as well as

L180M and M204V. The modeling of HBV RT indicated that the combination changed the direction of the D205 residue (YMDD domain) and narrowed the dNTP-binding pocket in comparison with the wild type and LVDr substitutions (M204V and L180M) (Fig. 3). The results of docking simulation of HBV RT and ETV-TP showed that the ETVr (184L and S202G) plus LVDr (L180M and M204V) substitutions had significantly longer minimal distances for ETV-TP and steric conflict with the D205 residue (Fig. 3 and Table 6). These docking simulation results suggest that nucleotide analogs that have the exocyclic alkene moiety of ETV-TP replaced by a smaller atom may retain activity against ETV-resistant mutants. Differences in the mode of binding of nucleotide inhibitors to the dNTP-binding pocket of HBV polymerase, as predicted from the current modeling studies, may account for the complementary drug resistance profiles seen for different nucleotide analogs. Interestingly, a previous *in vitro* study showed that ETVr substitutions (S202I and T184G), in addition to LVDr (L180M and M204V), were associated with a >1,100-fold decrease in susceptibility to ET (20). Collectively, these data indicate that nucleoside-naïve patients treated with ETV were less likely to become resistant to ETV.

In an *in vitro* assay, the rtA181T/V clinical-isolate genome from patients refractory to LVD/ADV induced a decrease in susceptibility to LVD, ADV, and, to a lesser extent, TDF, but sensitivity to ETV remained (22). LVDr selected by LVD exposure may lead to ETV failure. Therefore, for patients refractory to LVD/ADV, a combination of emtricitabine/TDF (10) might be an effective option. Furthermore, since sequential antiviral therapy leads to the selection of multidrug-resistant HBV and fitness or maximal viral resistance (25), combination therapy using a nucleoside together with a nucleotide analog, such as emtricitabine/TDF (10), ADV/LVD, ADV/ETV, ADV/telbivudine, or TDF, would be a more appropriate treatment strategy for patients with the LVDr substitution.

Based on HBV DR v3, T184SCGA/ILMF and S202G substitutions were present at baseline in 4.8% of patients and were detected in 14.6%, 24.4%, and 44.8% during 48, 96, and 144 weeks, respectively, of ETV therapy (Fig. 2). The prevalence of ETVr in our cohort seems to be higher than that reported in previous studies, based on assessment of ETV treatment at weeks 48, 96, 144, 192, and 240 using direct sequencing, where ETVr emerged in 6%, 15%, 36%, 47%, and 51% of LVD-refractory patients, respectively (21). The differences might be attributable to the tools used to detect HBV DNA substitutions associated with drug resistance, which differed between the studies. HBV DR v.3 and v.2 performed better than direct sequencing, and monitoring of the nucleoside mutations by HBV DR v.3 and v.2 in patients before and during ETV therapy was good for selecting effective therapeutic strategies and new combination therapies.

In conclusion, the combination of clinical observations and 3D docking simulation in the present study indicated that the low binding affinity of ETV-TP for the dNTP-binding domains of HBV RT by the ETVr plus LVDr substitutions could induce BT and provides the mechanistic foundations for a mechanism of inhibition of ETV against HBV. This modeling would be useful for designing new antiviral drugs.

ACKNOWLEDGMENTS

This work was supported in part by a grant-in-aid from the Ministry of Health, Labor, and Welfare of Japan and a grant-in-aid from the Ministry of Education, Culture, Sports, and Science.

We thank Kenichi Fukai, Graduate School of Medicine, Chiba University, Chiba, Japan; Tatsuya Ide, Department of Internal Medicine, Kurume University School of Medicine, Kurume, Japan; Debbie Hana Yi, Department of Emergency Medicine, New York-Presbyterian Hospital Columbia/Cornell, New York, NY; and Robert G. Gish, California Pacific Medical Center, San Francisco, CA, for their help throughout this work.

We have no conflicts of interest to disclose, except for M. F. Yuen and C. L. Lai, who received research support from BMS.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Baldick, C. J., B. J. Eggers, J. Fang, S. M. Levine, K. A. Pokornowski, R. E. Rose, C. F. Yu, D. J. Tenney, and R. J. Colonno.** 2008. Hepatitis B virus quasispecies susceptibility to entecavir confirms the relationship between genotypic resistance and patient virologic response. J. Hepatol. **48:**895–902.
3. **Baldick, C. J., D. J. Tenney, C. E. Mazzucco, B. J. Eggers, R. E. Rose, K. A. Pokornowski, C. F. Yu, and R. J. Colonno.** 2008. Comprehensive evaluation of hepatitis B virus reverse transcriptase substitutions associated with entecavir resistance. Hepatology **47:**1473–1482.
4. **Brautigan, D. L., M. Brown, S. Grindrod, G. Chinigo, A. Kruszewski, S. M. Lukasik, J. H. Bushweller, M. Horal, S. Keller, S. Tamura, D. B. Heimark, J. Price, A. N. Larner, and J. Larner.** 2005. Allosteric activation of protein phosphatase 2C by D-chiro-inositol-galactosamine, a putative mediator mimetic of insulin action. Biochemistry **44:**11067–73.
5. **Chan, H. L., A. Y. Hui, M. L. Wong, A. M. Tse, L. C. Hung, V. W. Wong, and J. J. Sung.** 2004. Genotype C hepatitis B virus infection is associated with an increased risk of hepatocellular carcinoma. Gut **53:**1494–1498.
6. **Chang, T. T., R. G. Gish, S. J. Hadziyannis, J. Cianciara, M. Rizzetto, E. R. Schiff, G. Pastore, B. R. Bacon, T. Poynard, S. Joshi, K. S. Klesczewski, A. Thiry, R. E. Rose, R. J. Colonno, R. G. Hindes, and the BEHoLD Study Group.** 2005. A dose-ranging study of the efficacy and tolerability of entecavir in lamivudine-refractory chronic hepatitis B patients. Gastroenterology **129:**1198–1209.
7. **Degertekin, B., M. Hussain, J. Tan, K. Oberhelman, and A. S. Lok.** 2009. Sensitivity and accuracy of an updated line probe assay (HBV DR v. 3) in detecting mutations associated with hepatitis B antiviral resistance. J. Hepatol. **50:**42–48.
8. **Huang, H., R. Chopra, G. L. Verdine, and S. C. Harrison.** 1998. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science **282:**1669–1675.
9. **Jain, A. N.** 2003. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J. Med. Chem. **46:**499–511.
10. **Keeffe, E. B., D. T. Dieterich, S. H. Han, I. M. Jacobson, P. Martin, E. R. Schiff, and H. Tobias.** 2008. A treatment algorithm for the management of chronic hepatitis B virus infection in the United States: 2008 update. Clin. Gastroenterol. Hepatol. **6:**1315–1341.
11. **Lai, C. L., D. Shouval, A. S. Lok, T. T. Chang, H. Cheinquer, Z. Goodman, D. DeHertogh, R. Wilber, R. C. Zink, A. Cross, R. Colonno, L. Fernandes, and the BEHoLD AI463027 Study Group.** 2006. Entecavir versus lamivudine for patients with HBeAg-negative chronic hepatitis B. N. Engl. J. Med. **354:**1011–1020.
12. **Lee, W. M.** 1997. Hepatitis B virus infection. N. Engl. J. Med. **337:**1733–1745.
13. **Marcellin, P., T. Asselah, and N. Boyer.** 2005. Treatment of chronic hepatitis B. Rev. Prat. **55:**624–632.
14. **Miyakawa, Y., and M. Mizokami.** 2003. Classifying hepatitis B virus genotypes. Intervirology **46:**329–338.
15. **Ono, S. K., N. Kato, Y. Shiratori, J. Kato, T. Goto, R. F. Schinazi, F. J. Carrilho, and M. Omata.** 2001. The polymerase L528M mutation cooperates with nucleotide binding-site mutations, increasing hepatitis B virus replication and drug resistance. J. Clin. Invest. **107:**449–455.
16. **Orito, E., M. Mizokami, H. Sakugawa, K. Michitaka, K. Ishikawa, T. Ichida, T. Okanoue, H. Yotsuyanagi, and S. Iino.** 2001. 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 **33:**218–223.
17. **Petrey, D., X. Xiang, C. L. Tang, L. Xie, M. Gimpelev, T. Mitors, C. S. Soto, S. Goldsmith-Fischman, A. Kernytsky, A. Schlessinger, I. Y. Y. Koh, E. Alexov, and B. Honig.** 2003. Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. Proteins **53:**430–435.
18. **Sherman, M., C. Yurdaydin, J. Sollano, M. Silva, Y. F. Liaw, J. Cianciara, A. Boron-Kaczmarska, P. Martin, Z. Goodman, R. Colonno, A. Cross, G. Denisky, B. Kreter, R. Hindes, and the AI463026 BEHoLD Study Group.** 2006.

Entecavir for treatment of lamivudine-refractory, HBeAg-positive chronic hepatitis B. Gastroenterology **130:**2039–2049.

19. **Szczesny, P., G. Wieczorek, and P. Zielenkiewicz.** 2005. MOFOID—not only the protein modeling server. Acta Biochim. Pol. **52:**267–269.
20. **Tenney, D. J., R. E. Rose, C. J. Baldick, S. M. Levine, K. A. Pokornowski, A. W. Walsh, J. Fang, C. F. Yu, S. Zhang, C. E. Mazzucco, B. Eggers, M. Hsu, M. J. Plym, P. Poundstone, J. Yang, and R. J. Colonno.** 2007. Two-year assessment of entecavir resistance in lamivudine-refractory hepatitis B virus patients reveals different clinical outcomes depending on the resistance substitutions present. Antimicrob. Agents Chemother. **51:**902–911.
21. **Tenney, D. J., R. E. Rose, C. J. Baldick, K. A. Pokornowski, B. J. Eggers, J. Fang, M. J. Wichroski, D. Xu, J. Yang, R. B. Wilber, and R. J. Colonno.** 2009. Long-term monitoring shows hepatitis B virus resistance to entecavir in nucleoside-naive patients is rare through 5 years of therapy. Hepatology **49:**1503–1514.
22. **Villet, S., C. Pichoud, G. Billioud, L. Barraud, S. Durantel, C. Trepo, and F. Zoulim.** 2008. Impact of hepatitis B virus rtA181V/T mutants on hepatitis B treatment failure. J. Hepatol. **48:**747–755.
23. **Xiang, Z., and B. Honig.** 2001. Extending the accuracy limits of prediction for side chain conformations. J. Mol. Biol. **311:**421–430.
24. **Xiang, Z., C. Soto, and B. Honig.** 2002. Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. Proc. Natl. Acad. Sci. U. S. A. **99:**7432–7437.
25. **Yim, H. J., M. Hussain, Y. Liu, S. N. Wong, S. K. Fung, and A. S. Lok.** 2006. Evolution of multi-drug resistant hepatitis B virus during sequential therapy. Hepatology **44:**703–712.
26. **Yuen, M. F., W. K. Seto, D. H. Chow, K. Tsui, D. K. Wong, V. W. Ngai, B. C. Wong, J. Fung, J. C. Yuen, and C. L. Lai.** 2007. Long-term lamivudine therapy reduces the risk of long-term complications of chronic hepatitis B infection even in patients without advanced disease. Antivir. Ther. **12:**1295–1303.

JOURNAL OF VIROLOGY, Oct. 2009, p. 10538–10547 Vol. 83, No. 20
0022-538X/09/$08.00+0 doi:10.1128/JVI.00462-09

A Genetic Variant of Hepatitis B Virus Divergent from Known Human and Ape Genotypes Isolated from a Japanese Patient and Provisionally Assigned to New Genotype J▽†

Kanako Tatematsu,[1] Yasuhito Tanaka,[1]* Fuat Kurbanov,[1] Fuminaka Sugauchi,[2] Shuhei Mano,[3] Tatsuji Maeshiro,[4] Tomokuni Nakayoshi,[5] Moriaki Wakuta,[6] Yuzo Miyakawa,[7] and Masashi Mizokami[1,8]

Department of Clinical Molecular Informative Medicine[1] and Department of Gastroenterology and Metabolism,[2] Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; Nagoya City University Graduate School of Natural Sciences, Nagoya, Japan[3]; Control and Prevention of Infectious Diseases, Department of Medicine and Therapeutics, Faculty of Medicine, University of the Ryukyus, Okinawa, Japan[4]; Heart Life Hospital, Okinawa, Japan[5]; Wakusan Clinic. Okinawa, Japan[6]; Miyakawa Memorial Research Foundation, Tokyo, Japan[7]; and Research Center for Hepatitis and Immunology, International Medical Center of Japan Kohnodai Hospital, Chiba, Japan[8]

Received 5 March 2009/Accepted 24 July 2009

Hepatitis B virus (HBV) of a novel genotype (J) was recovered from an 88-year-old Japanese patient with hepatocellular carcinoma who had a history of residing in Borneo during the World War II. It was divergent from eight human (A to H) and four ape (chimpanzee, gorilla, gibbon, and orangutan) HBV genotypes, as well as from a recently proposed ninth human genotype I, by 9.9 to 16.5% of the entire genomic sequence and did not have evidence of recombination with any of the nine human genotypes and four nonhuman genotypes. Based on a comparison of the entire nucleotide sequence against 1,440 HBV isolates reported, HBV/J was nearest to the gibbon and orangutan genotypes (mean divergences of 10.9 and 10.7%, respectively). Based on a comparison of four open reading frames, HBV/J was closer to gibbon/orangutan genotypes than to human genotypes in the P and large S genes and closest to Australian aboriginal strains (HBV/C4) and orangutan-derived strains in the S gene, whereas it was closer to human than ape genotypes in the C gene. HBV/J shared a deletion of 33 nucleotides at the start of preS1 region with C4 and gibbon genotypes, had an S-gene sequence similar to that of C4, and expressed the *ayw* subtype. Efficient infection, replication, and antigen expression by HBV/J were experimentally established in two chimeric mice with the liver repopulated for human hepatocytes. The HBV DNA sequence recovered from infected mice was identical to that in the inoculum. Since HBV/J is positioned phylogenetically in between human and ape genotypes, it may help to trace the origin of HBV and merits further epidemiological surveys.

Worldwide, an estimated 400 million people are infected with hepatitis B virus (HBV) persistently, of whom three quarters live in the Southeast and Far East Asia, and one million die of decompensated cirrhosis and/or hepatocellular carcinoma (HCC) annually (8, 15). HBV is the smallest animal DNA virus and has a genome made of approximately 3,200 nucleotides (nt) that contains four open reading frames for P, C, S, and X genes; they code for DNA polymerase/reverse-transcriptase, core protein, surface protein, and X protein, respectively (49). The S gene is divided into preS1 and preS2 regions and the small S gene, and the C gene splits into PreC and C.

Eight genotypes of HBV have been recognized by a sequence divergence of >8% in the entire genome and named by capital alphabet letters (A to H) in the order of discovery (3, 26, 29, 42). HBV genotypes are further classified into subgenotypes, such as B1/Bj and B2-5/Ba (44), as well as C1/Cs, C2/Ce, and C3-5 (36). A systematic nomenclature is proposed for designating HBV subgenotypes using Arabic numbers, such as A1, A2, and A3 (25). HBV genotypes have distinct geographical distribution (16, 23). Genotype A is prevalent in Africa, Europe and India, genotypes B and C are common in Asia, and genotype E is common in sub-Saharan Africa. Genotypes F and H are restricted to Central and South American continents, whereas genotype D is distributed all over the world. HBV genotypes have clinical application, and they influence severity and progression of liver disease and the response to antiviral therapies. Previous reports indicate that HCC is more frequent in the patients infected with genotype C than B (7, 47), and interferon is more effective in those infected with genotype B than C in Asia and more effective in those infected with genotype A than D in Europe (18, 34, 51).

Recently, a ninth genotype (I) was tentatively proposed for HBV strains detected in Laos (31). These strains are phylogenetically similar to aberrant Vietnamese strains that display complex recombination over the genome (10). In the present study, an HBV isolate was recovered from a Japanese patient with HCC, who was involved in military actions in Borneo during the World War II. The isolated strain was compared against eight human (A to H) and four ape (chimpanzee, gorilla, gibbon, and orangutan) genotypes and was provisionally designated genotype J. The new genotype was assigned based on a sequence diver-

* Corresponding author. Mailing address: Department of Clinical Molecular Informative Medicine, Nagoya, City University Graduate School of Medical Sciences, Kawasumi, Mizuho, Nagoya 467-8601, Japan. Phone: (81) 52-853-8292. Fax: (81) 52-842-0021. E-mail: ytanaka@med.nagoya-cu.ac.jp.

† Supplemental material for this article may be found at http://jvi.asm.org/.

▽ Published ahead of print on 29 July 2009.

TABLE 1. Nucleotide divergence in the full-genome sequence estimated from pairwise comparison between the Ryukyu 34 strain of a provisional genotype J and 1,440 HBV strains from the database entered by September 2008

Genotype	No. of strains	Divergence (%)		
		Range	Mean	SD
A	202	12.1–15.9	13.0	0.4
B	309	11.1–13.6	11.9	0.5
C	396	11.2–13.1	11.9	0.5
D	264	12.6–15.0	13.4	0.2
E	90	12.3–13.4	12.7	0.3
F	56	15.2–16.5	15.6	0.2
G	23	12.8–14.6	13.7	0.3
H	21	15.4–16.3	15.7	0.3
I	16	11.4–12.0	11.7	0.2
Chimpanzee	14	11.6–12.7	12.1	0.3
Gorilla	1	12.2		
Gibbon	34	9.9–11.7	10.9	0.5
Orangutan	12	10.4–11.2	10.7	0.4
Woolly monkey	2	27.2–27.4	27.3	0.1

gence of 10.7 to 15.7% from other genotypes, a unique phylogenetic position between human and ape genotypes, and the absence of strong evidence of recombination.

MATERIALS AND METHODS

Patient. A Japanese man, 88 years old, developed HCC in 2006. He had a history of residing in Borneo during the World War II. No HBV infections were recorded in his family members. In October 1996, he was diagnosed with chronic hepatitis B. Hepatitis B surface antigen (HBsAg) was detected in serum, and the aspartate aminotransaminase and alanine aminotransferase levels were elevated to 83 and 73 U/liter, respectively (normal levels, <30 U/liter for both). Thereafter, the transaminase levels were normalized, and he had been monitored as an asymptomatic HBV carrier. In August 2000, the level of a tumor marker (des-γ-carboxy prothrombin) was elevated to 52 mAU/ml (normal, <40 mAU/ml), while another tumor marker (alpha-fetoprotein) remained within normal range (<10 ng/ml) as alanine aminotransferases. In October 2006, a tumor (4.3 by 4.1 cm) was detected in the liver by ultrasonography, and he received treatment with transarterial embolization. Des-γ-carboxy prothrombin was elevated to 419 mAU/ml, while the aminotransferase levels remained within normal limits. Hepatitis B e antigen (HBeAg) was negative, and the corresponding antibody (anti-HBe) was detected in his serum. The subtype of HBsAg in this serum was *ayw*.

HBV DNA was extracted from his serum specimen obtained in 2006, and the full-length genome sequence was determined for phylogenetic and biological analyses. An informed consent had been obtained from the patient, and the study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki as reflected in a priori approval by the institution's human research committee.

Markers of HBV infection. HBeAg and anti-HBe were determined by enzyme-linked immunosorbent assay (ELISA) with commercial kits (HBeAg EIA; Institute of Immunology, Tokyo, Japan), and subtypes of HBsAg by ELISA with commercial kits (HBsAg Subtype EIA; Institute of Immunology). Hepatitis B core-related antigen (HBcrAg) was determined by chemiluminescence enzyme immunoassay (13). The method allows more sensitive detection of core protein and, as was shown in previous studies, HBcrAg levels reflect HBV DNA loads and well correlate with intrahepatic covalently closed circular DNA (cccDNA) levels. The measurement of serum HBcrAg is a useful noninvasive tool for monitoring intrahepatic HBV viral status (52). HBV DNA was quantified by the S gene-targeted real-time detection PCR with a sensitivity of 100 copies/ml (equivalent to 20 IU/ml) (1). However, due to small volumes of sera available from the challenged mice, HBV DNA was extracted from 10-fold-diluted specimens, resulting in reduced assay sensitivity in the present study (1,000 copies/ml [200 IU/ml]).

Determination of the complete nucleotide sequence of HBV/J isolate. HBV DNA was extracted by using the QIAamp DNA blood kit (Qiagen, GmbH, Hilden, Germany) from 100 μl of serum that had been stored at −80°C. The complete genome sequence of an HBV/J isolate recovered from the patient was determined by the strategy previously reported (43). In brief, two sets of primers were designed to amplify overlapping fragments (A and B) covering the entire HBV genome (stat not shown). Nested PCR was carried out for 35 cycles (95°C, 30 s; 57°C, 30 s; and 72°C, 2 min) using TaKaRa LA *Taq* polymerase (Takara Biochemicals, Kyoto, Japan). Amplified fragments were inserted into the pGEM-T Easy vector (Promega, Madison, WI), and cloned in DH5a cells (Toyobo, Osaka, Japan). Obtained HBV DNA clones were confirmed to have the sequence identical to the major-clone consensus sequence determined directly on PCR products by Prism BigDye (Applied Biosystems, Foster City, CA) in the ABI 3100 automated sequencer.

Phylogenetic analysis. Full-length sequences of HBV isolates were aligned with use of the CLUSTAL W software program (48) (available at www.ebi.ac.uk), and the alignment was confirmed by visual inspection. Genetic distances were estimated by the six-parameter method, and phylogenetic trees were constructed with the neighbor-joining method (35). To confirm the reliability of phylogenetic trees, bootstrap resampling and reconstruction were carried out 1,000 times using the program

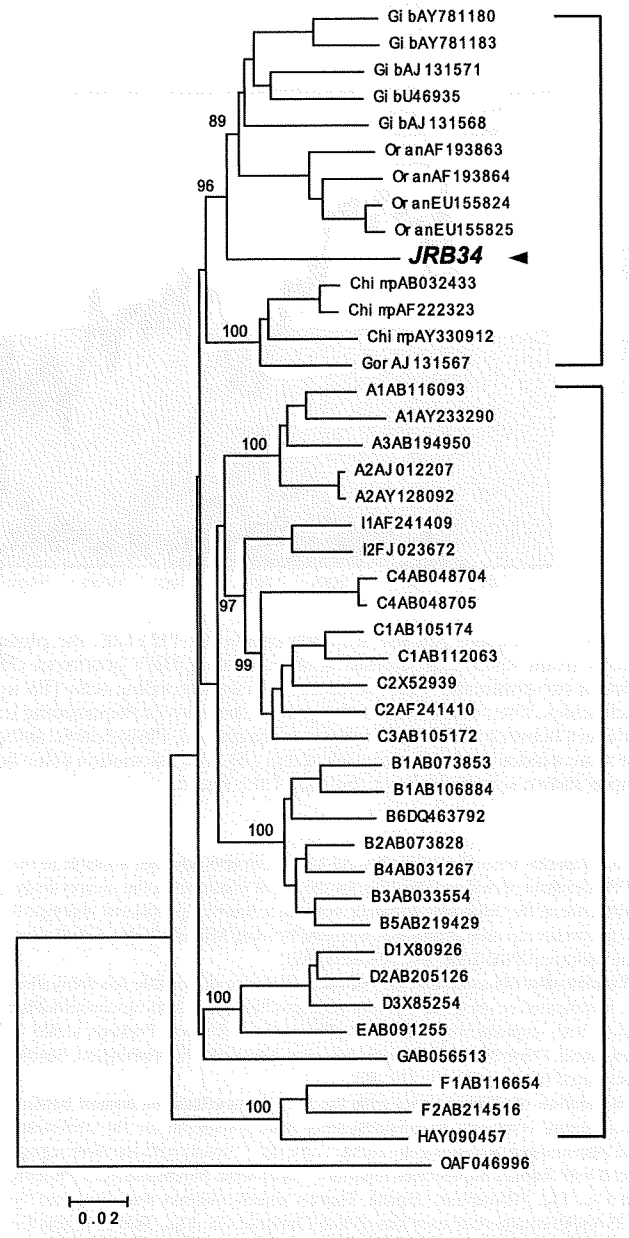

FIG. 1. Phylogenetic tree constructed on the entire genome sequences of 44 HBV isolates representing four ape and eight human genotypes. A woolly monkey HBV isolate serves as an outgroup. The HBV/J isolate (JRB34) is indicated by an arrowhead, and the genetic distance is indicated by a bar below.

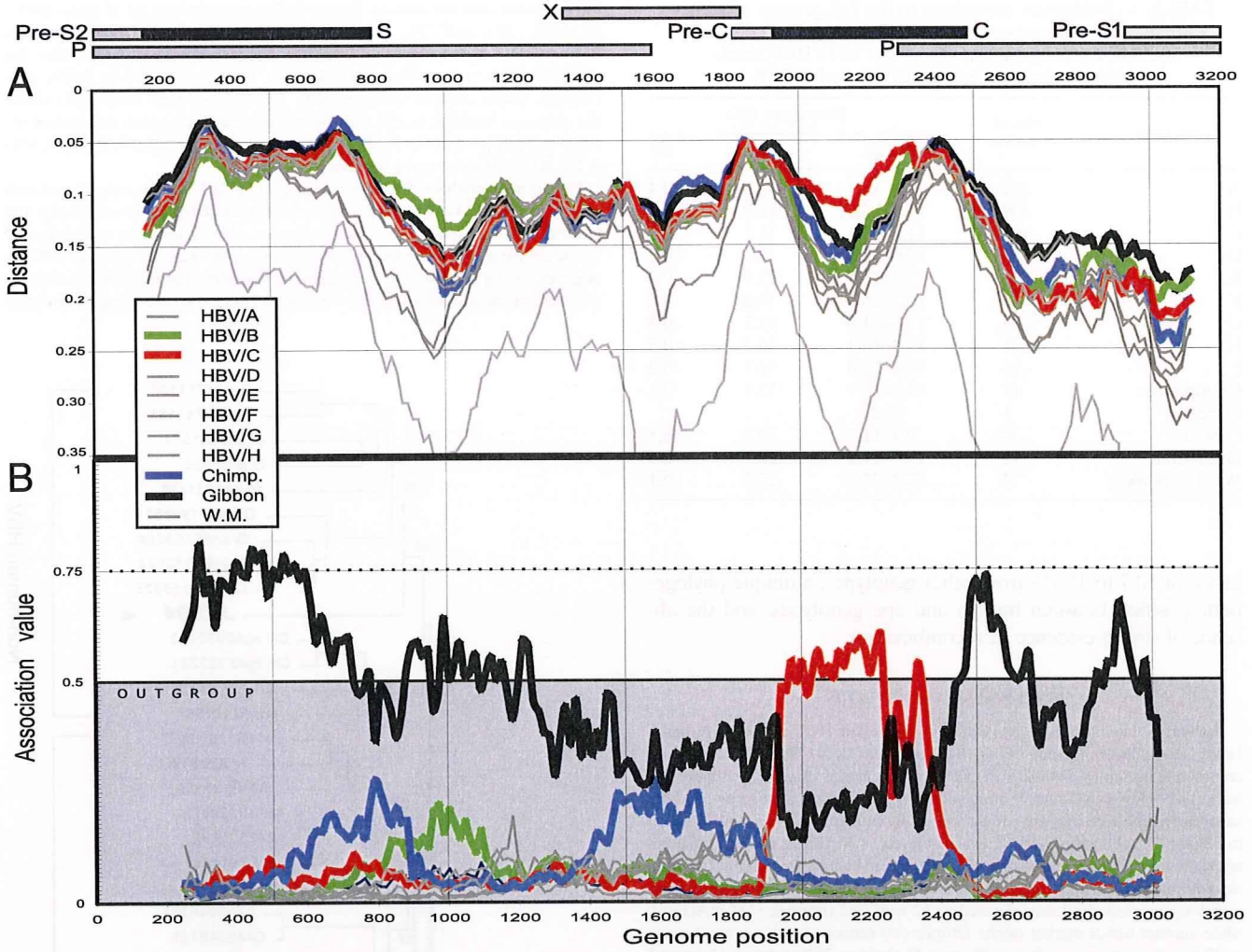

FIG. 2. Complete genome scanning carried by PHYLIP, the phylogeny inference package implemented in the Simmonic software, for the JRB34 strain versus 228 selected nonrecombinant HBV genotypes (HBV/Ba and HBV/I not included) reference strains grouped by genotype. Kimura two-parameter distance model (A) and grouping scan (B) were determined with a 300-nt size window sliding by an increment of 15 nucleotides. The *x* axis indicates the genome position (corresponding to the midpoint of the scanning fragment), and the *y* axis indicates the mean distances between JRB34 and reference groups (A). Phylogenetic association (*y* axis) was evaluated throughout entire HBV genome (*x* axis) with the same window and step size parameters (B). The association value below 0.5 was considered to represent an outgroup. The open reading frame map is shown schematically at the top of the figure.

of the Hepatitis Virus Database (39). All 1,440 complete genomes available in the DDBJ/GenBank served as references for the initial alignment in the present study. Divergence in the nucleotide sequence between a strain of provisional genotype J and previously reported strains was estimated by using MEGALIGN v.6.00 (Lazergene package; DNASTAR, Inc., Madison, WI).

Examination of recombination evidence. Evidence of possible recombination was investigated by using the software packages Simmonic 2005 v1.6 and SimPlot v3.5.1, both implementing PHYLIP (Phylogeny Inference Package v3.68; J. Felsenstein, Department of Genome Sciences, University of Washington, Seattle [distributed by the authors]) (19, 40).

Inoculation of chimeric mice with the liver repopulated for human hepatocytes. Severe combined immunodeficiency mice transgenic for the urokinase-type plasminogen activator gene ($uPA^{+/+}/SCID^{+/+}$ mice) with the liver repopulated with human hepatocytes (chimeric mice) were purchased from Phoenix Bio Co., Ltd. (Hiroshima, Japan). Human serum albumin was measured by ELISA with commercial assay kits (Eiken Chemical Co., Ltd., Tokyo, Japan) for estimating the extent of repopulation. The research complied with all relevant federal guidelines and institutional policies.

Immunofluorescence. Freshly prepared liver tissues were snap-frozen in isopentane precooled in liquid nitrogen. Frozen specimens were cut at 5 to 6 μm by cryostat, mounted on glass slides, air dried, and fixed in 100% acetone at room temperature for 10 min. Sections were blocked with antibody diluent (Dako, Tokyo, Japan) and stained for hepatitis B core antigen (HBcAg). They were incubated with rabbit anti-HBc (Dako) at room temperature for 1 h, washed in phosphate-buffered saline, and then incubated with goat anti-rabbit immunoglobulin G conjugated with Cy3 (Chemicon International, Inc., Temecula, CA) or goat anti-human albumin antibody labeled with fluorescein isothiocyanate (Bethyl Laboratories, Inc., Montgomery, TX). Sections were washed with phosphate-buffered saline and observed in a fluorescence microscope (Eclipse E800M; Nikon, Tokyo, Japan).

Nucleotide sequence accession numbers. The nucleotide sequence data reported in the present study will appear in the DDBJ/EMBL/GenBank databases under accession no. AB486012.

RESULTS

Composition of the HBV genome of genotype J. HBV DNA was extracted from serum of a patient with HCC. It was named JRB34 ("J" for Japanese; "R" after the southernmost island [Ryukyu] where the patient has spent most of his life now

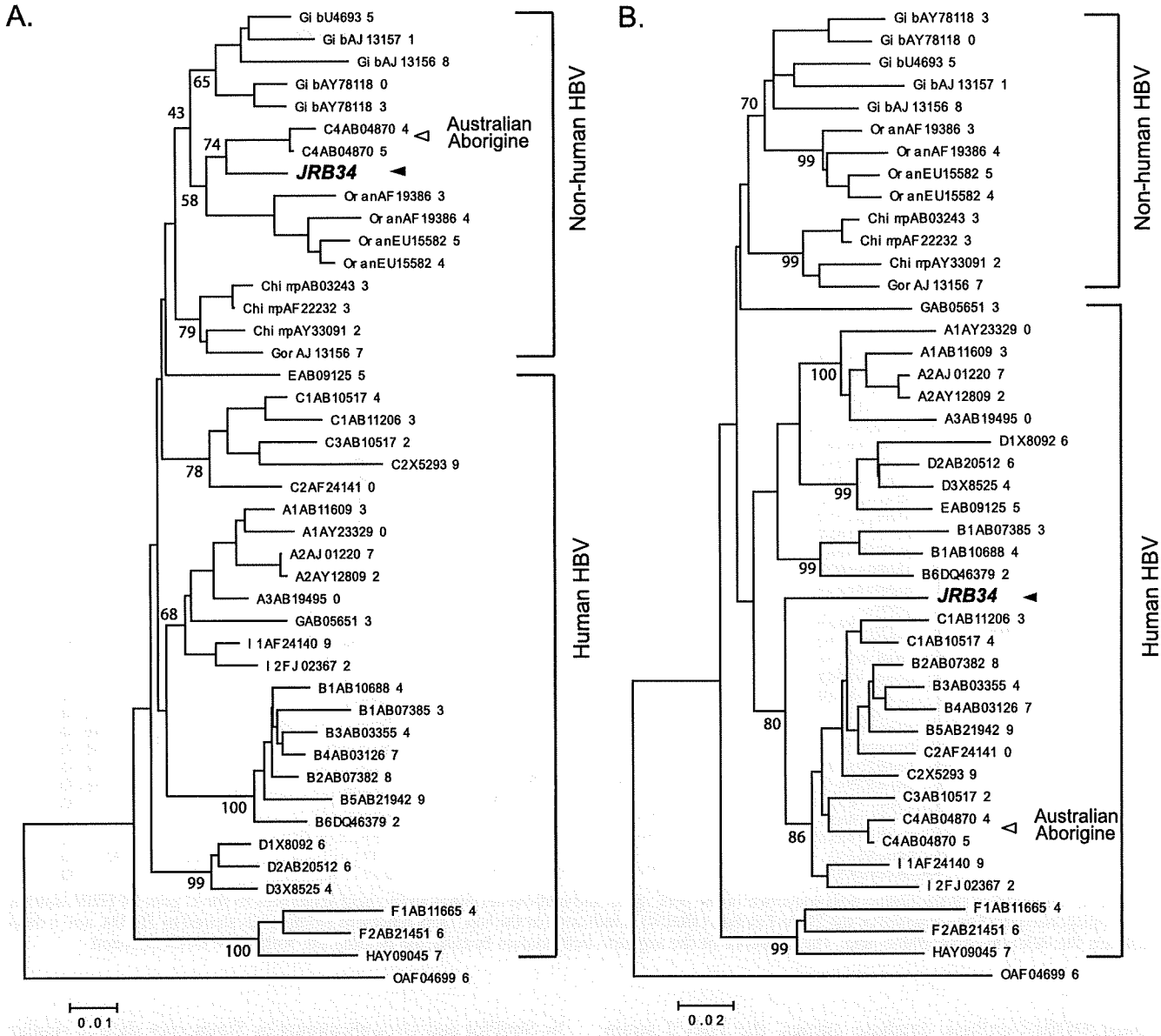

FIG. 3. Phylogenetic tree constructed on the preS/S gene (A) and C gene (B) sequences of 44 HBV isolates representing four ape and eight human genotypes. A woolly monkey HBV isolate serves as an outgroup. The HBV/J isolate (JRB34) is indicated by an arrowhead, and an HBVC4 isolate from Australian aborigine is indicated by an open triangle. The genetic distance is indicated by a bar below.

exceeding 90 years; and "B" for Borneo where he is suspected to have contracted the HBV infection). The entire nucleotide sequence was determined for the JRB34 isolate of genotype J (HBV/J). It had a genomic length of 3,182 nt, which consisted of envelope gene containing preS1 region (nt 2848 to 3171, coding for 108 amino acids [aa]), preS2 region (nt 3172 to 154 [55 aa]), and the small S gene (nt 155 to 835 [226 aa]), X gene (nt 1374 to 1838 [154 aa]), preC region (nt 1814 to 1897 [27 aa]), C gene (nt 1901 to 2452 [183 aa]), and P gene (nt 2307 to 1623 [832 aa]).

Sequence divergence of the JRB34 strain from other genotypes. The complete genome sequence of the JRB34 strain obtained in the present study was compared against those of 1,440 HBV genomes registered in the Viral Hepatitis Database (39). Estimated nucleotide sequence divergence of the JRB34 strain from four ape and nine human genotypes is summarized in the Table 1. The mean divergence by genotypes ranged from 10.7 and 10.9% (from orangutan and gibbon, respectively) to 15.6 and 15.7% (from genotypes F and H, respectively). Surprisingly, the minimum divergence of 9.9% was observed in comparison with a nonhuman HBV isolate from *Hilobates agilis* gibbon confiscated in Taiwan in 1993 (AY330917) (41). Since the sequence divergence from any documented genotypes, including recently proposed genotype I, exceeded 8%, the JRB34 strain was tentatively classified into a novel genotype J of HBV.

Phylogenetic analysis of the entire genomic sequence. In the phylogenetic tree constructed on 1,440 complete genome

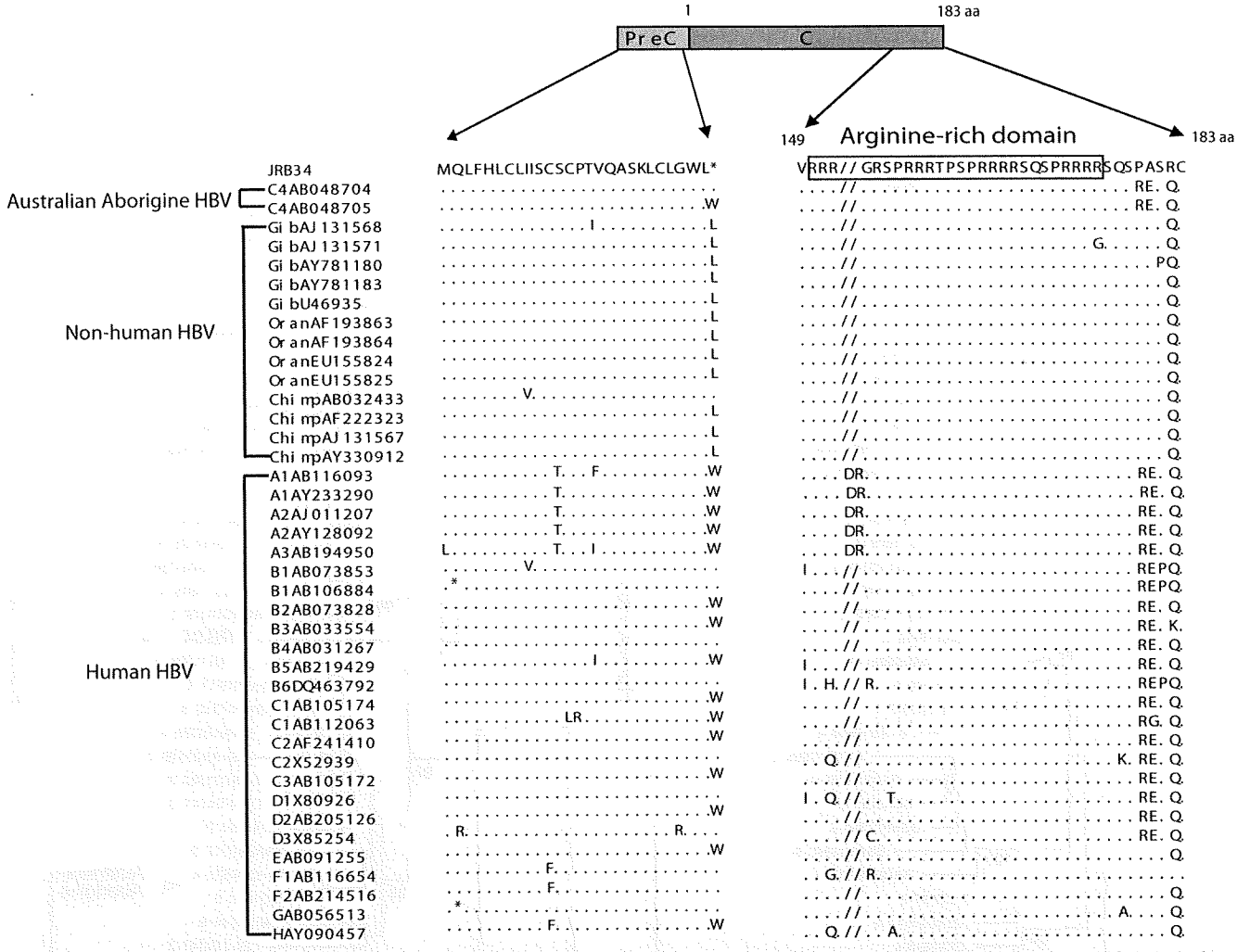

FIG. 4. Comparison of the amino acid sequence in the preC gene and carboxy-terminal amino acid sequences in the C gene of HBV isolates of various genotypes. The sequence of the HBV/J isolate (JRB34) is indicated at the top. Dots represent amino acids shared by JRB34, and a dash indicates the deletion of an amino acid. The sequence of the arginine-rich domain bearing the binding site with HBV DNA is boxed.

EMBL/DDBJ/GenBank database entries, the HBV/J strain was positioned distinctively from all known human genotypes (data not shown). It was closest to the cluster formed by gibbon- and orangutan-derived strains. However, including recombinant strains in such analyses may significantly affect the overall phylogenetic topology. This possibility was ruled out by reconstruction of the phylogeny using nonrecombinant HBV strains that further confirmed the phylogenetic peculiarity of the studied JRB34 strain (see Fig. S1 in the supplemental material). A total of 44 representative reference strains were further selected for establishing the consistency. Thus, phylogenetic topology indicating genotype-specific clustering is shown in the Fig. 1. Hence, using various sets of references, we confirmed that genotype J undoubtedly differed phylogenetically from all other known genotypes.

Lack of significant evidence of recombination with other human or ape genotypes in genotype J. To investigate possible recombination in the JRB34 genome, a window scanning analysis of aligned HBV genomes was performed by means of Simplot and Simmonics software packages. Both Bootscanning by SimPlot and GroupScanning by Simmonics showed similar output results. However, the methodological approach is different between these two software packages; GroupScanning provides more robust analysis of the phylogenetic relation between the examined strain and clusters of reference strains, whereas SimPlot does this comparison between the examined strain and parametrically generated consensus of the reference strains. The results obtained by SimPlot therefore can be significantly affected by selected parameters for the generation of consensus. This is especially undesirable when a new genotype strain (for which no references are available among known genotypes) is being analyzed (40). Figure 2 shows genome-wide distance scanning and GroupScanning plots for the JRB34 strain in comparison with a reference set consisting of 228 nonrecombinant HBV isolates retrieved from the public database (the phylogenetic tree is shown in Fig. S1 in the supplemental material). It is evident that the JRB34 strain was divergent from all known genotypes, and the closest genetic neighbors were estimated by distance and phylogenetic association scanning were the gibbon genotype (in preS, S, and P

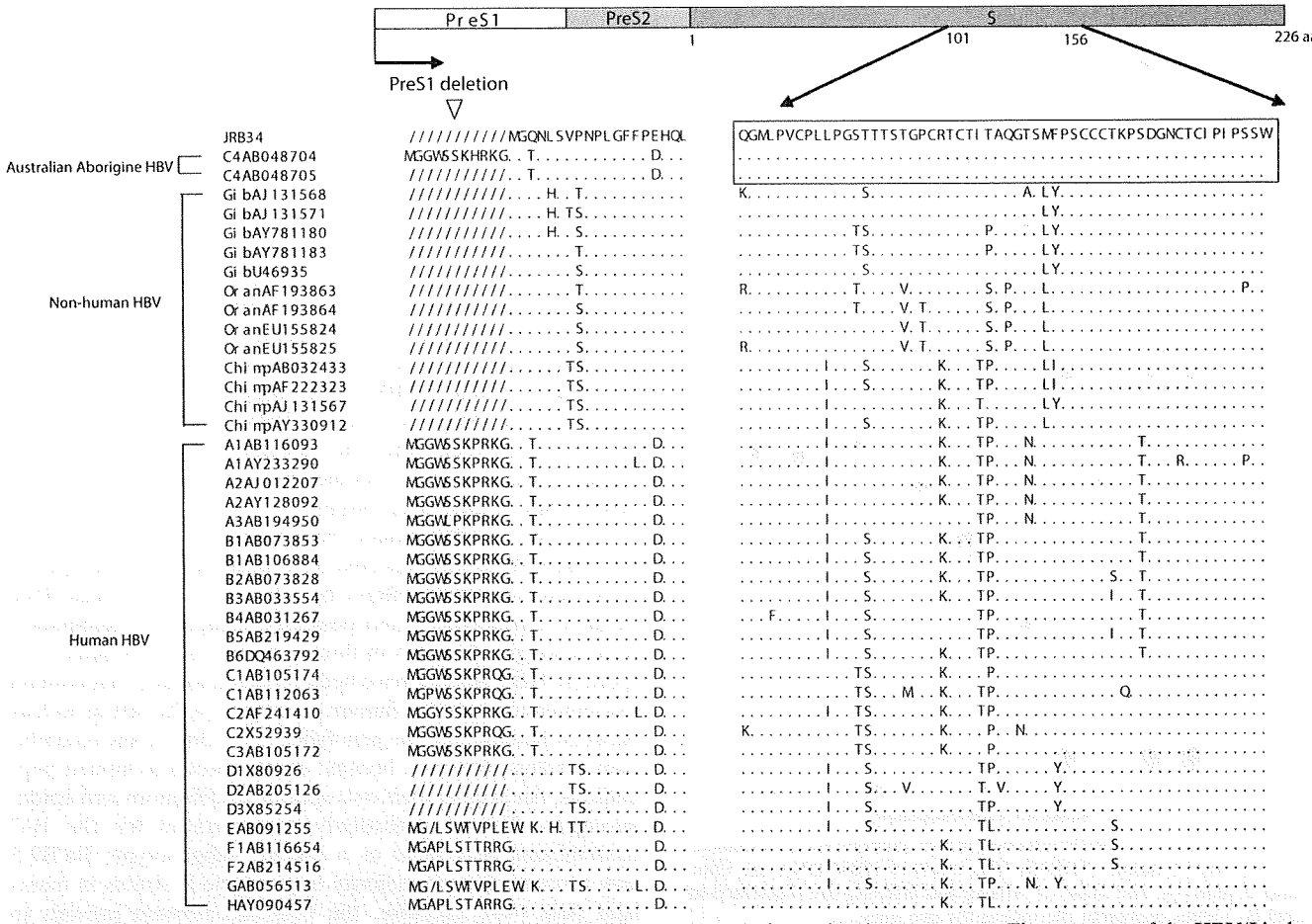

FIG. 5. Comparison of amino acid sequences of the preS/S gene among HBV isolates of various genotypes. The sequence of the HBV/J isolate (JRB34) is indicated at the top. Dots represent amino acids shared by JRB34, and a dash indicates the deletion of an amino acid. The sequence from positions 101 to 156 forming loops, bearing the common antigenic determinants of HBsAg, is boxed.

genes) and genotype C (in the core gene). However, no significant evidence of recombination between these two ape and human genotypes was revealed by the used methods. Homology scan carried out by SimPlot using the same set of reference sequences gave concordant results.

Phylogenetic analyses of the four open reading frames. Phylogenetic relationship between the JRB34 strain and other genotypes was further analyzed in four open reading frames. In the small S gene, subgenotype C4 recovered from Australian aborigines (43) changed its phylogenetic topology from the branch of human genotypes to a branch intermediate between orangutan and gibbon strains (Fig. 3A). Remarkably, genotype J and C4 strains joined together to create a clade between orangutan and gibbon strains. In contrast, genotype J clustered with human genotypes in the phylogenetic analysis of the C gene and was closely related to genotype C; it took a position outside genotype I strains, however (Fig. 3B). Genotype J was closer to gibbon and orangutan genotypes in the phylogenetic trees constructed on P and large S genes (data not shown), demonstrating its topology similar to that in the analysis of the entire genome (Fig. 1).

Amino acid sequence of the HBV/J isolate. The amino acid sequence of HBV/J was compared against those of other genotypes over three different areas of the genome. The amino acid sequence in the preC gene and arginine-rich domain in the carboxy-terminal sequence in the C gene were well conserved by genotype J (Fig. 4). In the preS1 region, genotype J had a deletion of 11 aa as gibbon and chimpanzee genotypes (Fig. 5). This deletion was shared by one of the two HBV/C4 isolates from Australian aborigines, as well as all HBV/D isolates. Amino acid sequence in the S gene of genotype J was the same as those of aborigine isolates of subgenotype C4; they would share antigenic epitopes of HBsAg. Amino acids at codons 122 and 160 were arginine (with G as nt 365) and lysine (with G as nt 479), respectively, which was consistent with subtype *ayw* of HBsAg from this patient (27).

Five domains (A to E) of DNA polymerase/reverse transcriptase in the P gene were preserved well in HBV/J, and it did not have mutations in the Tyr-Met-Asp-Asp motif in the domain C that determines the sensitivity to lamivudine (data not shown). HBV/J possessed A1762T/G1764A double mutations in the core promoter and G1896A stop codon mutation in the preC region, which was compatible with an HBeAg-minus phenotype of HBV recovered from the patient positive for anti-HBe.

Infection with HBV/J in chimeric mice with the liver repopulated for human hepatocytes. Two chimeric mice that had been transplanted with human hepatocytes were inoculated with 10^4 HBV DNA copies of genotype J. In both mice, HBV

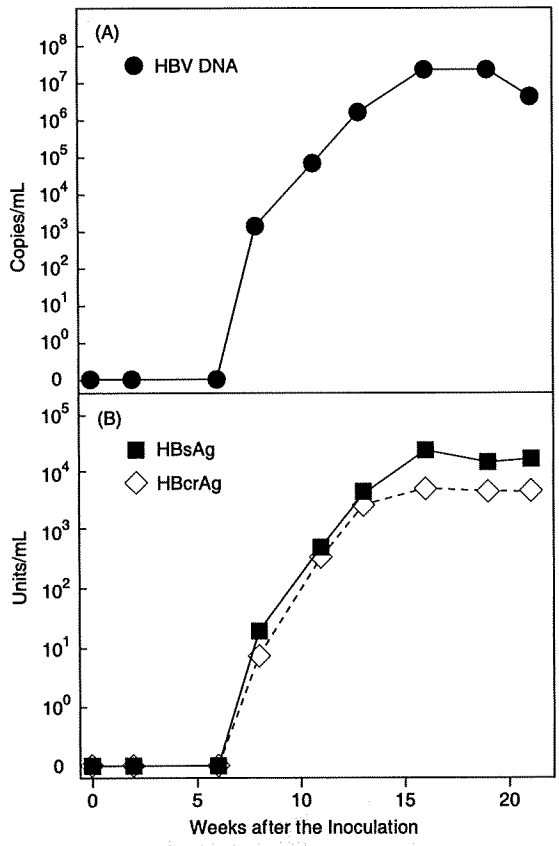

FIG. 6. Markers of HBV infection in two chimeric mice inoculated with the HBV/J isolate (JRB34). The levels of HBV DNA are illustrated in panel A, and those of HBsAg and HBcrAg are illustrated in panel B. Values represent the means for two mice.

DNA in a high titer (10^5 copies/ml) appeared in the circulation at week 7, plateaued at high levels (10^6 to 10^8 copies/ml), and stayed detectable until 22 weeks of observation after the inoculation (Fig. 6A). HBsAg and HBcrAg became detectable at week 7 and kept increasing in concentrations until week 15 when they reached a plateau at high levels (Fig. 6B). HBV strains recovered from mice at the last day of follow-up were identical in the complete genome sequence to the JRB34 strain used for inoculation.

The liver from chimeric mice infected with HBV/J was stained for HBcAg by immunofluorescence (Fig. 7A). The staining for HBcAg was confined to areas where mouse liver had been replaced for human hepatocytes, and the same areas were stained for human albumin (Fig. 7B). Colocalization of HBcAg and human hepatocytes was demonstrated by double staining for HBcAg and human albumin (Fig. 7C). Finally, expression and replication of the JRB34 strain were confirmed by successful detection of cccDNA and HBV RNA in the liver tissue from both sacrificed mice (see Fig. S2A and B in the supplemental material).

DISCUSSION

An HBV isolate (JRB34) was recovered from a male, 88-year-old Japanese patient with HCC and sequenced over the entire genome. In the full-genome sequence, the JRB34 strain had 10.9 to 15.7% divergence from 1,440 HBV strains retrieved from the DDBJ/EMBL/GenBank. The divergence exceeds 8% that has been defined originally for distinguishing between four genotypes (A to D) (29) and later for an additional four genotypes (E to H) (3, 26, 42). Phylogenetically, the sequence of JRB34 was closer to ape than human HBV genotypes. No significant evidence of recombination with eight known human and four ape genotypes was revealed by the GroupScanning analysis (40) and phylogenetic analyses. These lines of evidence have qualified the JRB34 strain to represent a possible new HBV genotype. To further confirm the epidemiological significance of this strain, capable of establishing new infections, two chimeric mice were each inoculated with 10^4 copies of JRB34 HBV DNA. They both were successfully infected with sharp increases in HBV DNA and HBsAg in serum several weeks after the inoculation. Replication in the chimeric mice was also confirmed by detection of cccDNA and HBV RNA in their liver tissues.

Recently, an HBV isolate from Vietnam (VH24 [accession no. AB231908]) was reported as a ninth human genotype (I) (12). However, VH24 differed by only 7.0% ± 0.4% from HBV isolates of genotype C and possessed complex recombination with genotypes A and G in three genomic areas. A number of sporadic HBV isolates have been reported to date that contain recombination between human genotypes (4, 24, 40), as well as between human and ape genotypes (21). Only a few recombinant variants, however, became widely spread in human populations, developing their own specific distributions and epidemiologies. This is particularly demonstrated for the B/C recombinant designated as a distinct subgenotype; Ba/B2-5 now accounts for the majority of genotype B strains in mainland Asia (44). Likewise, the C/D recombinant prevails in Tibet and northern China (50). To avoid assigning a new genotype for every newly discovered sporadic recombinant HBV variant, evidence of intergenotypic recombination should be carefully eliminated (14). However, in some cases, designation of a new genotype is proposed by a potential epidemiological significance of a novel genetic variant. Recently, a study carried out in Laos described a number of strains closely related phylogenetically with the Vietnamese genotype I strains, thereby suggesting their epidemiological significance (31). The JRB34 strain documented in the present study was genetically and phylogenetically distinct from any previously published strains, including those of genotype I from Vietnam and Laos. To avoid possible misconceptions in the future, the strain is provisionally designated genotype J.

HBV of distinct genotypes can infect great apes in the wild, including chimpanzee, gorilla, orangutan and gibbons (9, 20, 37, 51). HBV genotypes of chimpanzee and gorilla, as well as those of orangutan and gibbon, cocluster in agreement with their geographical distribution in Africa and Southeast Asia, respectively (41). Genotype J represented by the JRB34 strain clustered with gibbon/orangutan genotypes. In a phylogenetic analysis of the S region/gene sequence, JRB34 belonged to a nonhuman HBV group but was closely related to an HBV isolate of subgenotype C4 (AB048704) recovered from an Australian aborigine; C4 is most divergent from other subgenotypes of genotype C (43). In the phylogenetic analysis of the C gene, however, JRB34 clustered with human genotypes and closely related to genotype C, including C4, and was positioned

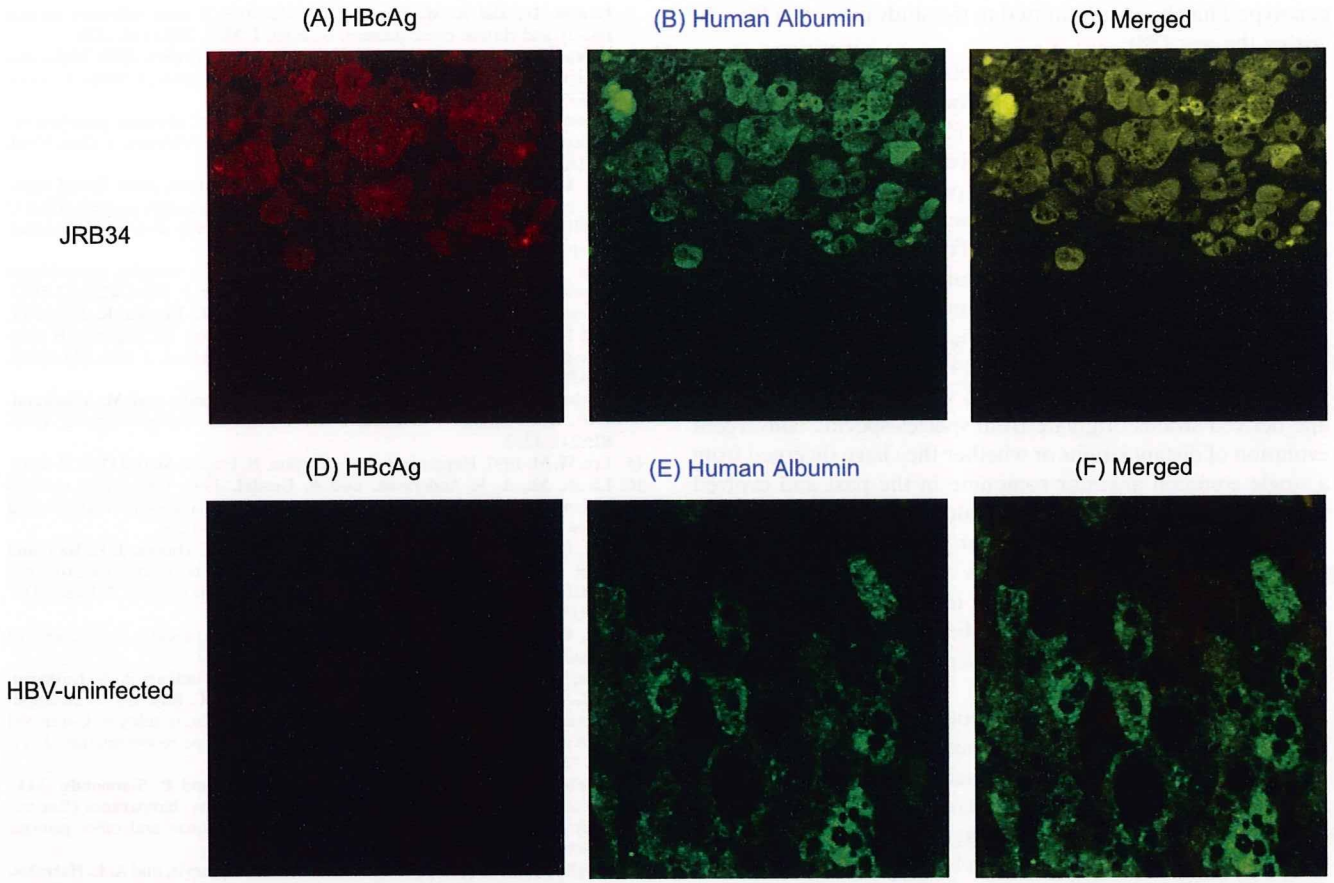

FIG. 7. (A and B) Immunofluorescent staining of a frozen liver section of a chimera mouse inoculated with the HBV/J isolate (JRB34). HBcAg is stained in panel A, and human albumin is stained in panel B. (C) Colocalization of HBcAg and human albumin is revealed by double staining. (D to F) HBV-uninfected mouse liver shows that only human albumin is stained.

outside genotype I strains (Fig. 4). Taken together, genotype J is phylogenetically close to gibbon/orangutan genotypes in the entire genome and to genotype C (C4 in particular) in the S and C genes. However, despite observed interchangeable relatedness with gibbon and genotype C/I strains, no strong evidence of recombination was confirmed in the JRB34.

In the sequence of C gene, carboxyl-terminal arginine-rich region, required for binding with HBV DNA, was preserved in JRB34. It had the G1896A stop codon in the precore region that aborts the translation of HBeAg (5, 30) and A1762T/G1764A double mutations in the core promoter that interfere with the transcription of HBeAg by downregulating preC mRNA (28, 45); they are compatible with the $HBeAg^-$ anti-HBe^+ phenotype of the patient from whom JRB34 was isolated. Since the double mutations are detected frequently in HBV DNA sequences from patients with HCC (17, 33), it could be implicated in hepatocarcinogenesis of the patient from whom JRB34 was isolated. It is not certain, however, if precore and core-promoter mutations had existed in HBV transmitted to the patient who is presumed to have been infected 60 years ago. Since amino acid sequences constituting antigenic loops of HBsAg (6) were the same as those of Australian aborigine isolates of C4, they would share antigenic epitopes of HBsAg. The amino acids at codons 122 and 160 were arginine (with G at nt 365) and lysine (with G at nt 479), respectively (27), in agreement with subtype *ayw* of HBsAg from this patient. Five domains (A to E) of DNA polymerase/reverse transcriptase in the P gene were preserved well in HBV/J, and it did not have mutations in the Tyr-Met-Asp-Asp motif in the domain C that determines the sensitivity to lamivudine (2).

How and when the patient contracted infection with HBV/J is not certain. It is very unlikely, however, that he acquired infection in Japan via perinatal or horizontal transmission. There are no wild primates in Okinawa, where the patient was originally from, and the prevalent human HBV genotypes are limited to B (60%), C (39%), and sporadic cases of A (1%) (32). Furthermore, HBV/J was not found among patient's family members who are currently alive (data not shown). The phylogenetic position within open reading frames of JRB34 in between gibbon/orangutan genotypes and human genotype C gives a clue where and when the patient had contracted HBV infection. He was drafted to Borneo during World War II (1939 to 1945); the island in the Southeast Asia is inhabited by gibbons and orangutans and has a local population mainly infected with genotypes B or C. Zoonotic infection of HBV has been previously reported (11, 46), and HBV of genotype E was recovered from a chimpanzee captured in West Africa where this genotype is common. There is a possibility that JRB34 of

genotype J had been transmitted to the study patient in Borneo during the war (38).

The origin of genotype J in gibbon/orangutan or human inhabitants in Borneo is not certain but very likely. HBV DNA and/or HBsAg was detected in 26% (55/213) and 20% (58/297) of gibbons and orangutans, respectively, captured in Southeast Asia (38). HBV is also endemic in people living there, with a prevalence of HBsAg at 2 to 8%. There would be high chances for cross-species transmission of HBV where it prevails both in human beings and nonhuman primates. Phylogenetic analysis for close relationship between human and nonhuman HBV genotypes has indicated geographical influence rather than association with particular species (41).

It is remains to be determined whether genotype J and ape-derived strains originate from species-specific convergent evolution of distant strains or whether they have diverged from a single common ancestor sometime in the past and evolved independently thereafter. The validity of cross-species infection or species-specific evolution for genotype J would be verified by sequence analysis of HBV DNA from gibbons and humans living in Borneo. If they turn out to be the same, cross-species infection will be justified. Should genotype J be restricted to human beings, in converse, species-specific infection will be confirmed.

In conclusion, a novel HBV genotype was identified in the Ryukyu isolate and provisionally named genotype J. Phylogenetic analyses over the full-length sequence and open reading frames indicate a close relationship of genotype J with gibbon/orangutan genotypes and human genotype C. The index patient would have been infected with HBV/J while he resided in Borneo inhabited by gibbons and orangutans. Although only one HBV isolate of genotype J (JRB34) has been identified, this may be only the tip of an iceberg. It would be worthwhile to examine the genotype of HBV infecting people and gibbons, as well as orangutans, living in Borneo and neighboring countries for mapping the epidemiology of genotype J and finding any clinical relevance.

ACKNOWLEDGMENTS

This study was supported in part by a grant-in-aid from the Ministry of Health, Labor and Welfare of Japan and a grant-in-aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

REFERENCES

1. **Abe, A., K. Inoue, T. Tanaka, J. Kato, N. Kajiyama, R. Kawaguchi, S. Tanaka, M. Yoshiba, and M. Kohara.** 1999. Quantitation of hepatitis B virus genomic DNA by real-time detection PCR. J. Clin. Microbiol. **37:**2899–2903.
2. **Allen, M. I., M. Deslauriers, C. W. Andrews, G. A. Tipples, K. A. Walters, D. L. Tyrrell, N. Brown, L. D. Condreay, et al.** 1998. Identification and characterization of mutations in hepatitis B virus resistant to lamivudine. Hepatology **27:**1670–1677.
3. **Arauz-Ruiz, P., H. Norder, B. H. Robertson, and L. O. Magnius.** 2002. Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. J. Gen. Virol. **83:**2059–2073.
4. **Bollyky, P. L., and E. C. Holmes.** 1999. Reconstructing the complex evolutionary history of hepatitis B virus. J. Mol. Evol. **49:**130–141.
5. **Carman, W. F., M. R. Jacyna, S. Hadziyannis, P. Karayiannis, M. J. McGarvey, A. Makris, and H. C. Thomas.** 1989. Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. Lancet **ii:**588–591.
6. **Carman, W. F., A. R. Zanetti, P. Karayiannis, J. Waters, G. Manzillo, E. Tanzi, A. J. Zuckerman, and H. C. Thomas.** 1990. Vaccine-induced escape mutant of hepatitis B virus. Lancet **336:**325–329.
7. **Fung, S. K., and A. S. Lok.** 2004. Hepatitis B virus genotypes: do they play a role in the outcome of HBV infection? Hepatology **40:**790–792.
8. **Ganem, D., and A. M. Prince.** 2004. Hepatitis B virus infection–natural history and clinical consequences. N. Engl. J. Med. **350:**1118–1129.
9. **Grethe, S., J. O. Heckel, W. Rietschel, and F. T. Hufert.** 2000. Molecular epidemiology of hepatitis B virus variants in nonhuman primates. J. Virol. **74:**5377–5381.
10. **Hannoun, C., H. Norder, and M. Lindh.** 2000. An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam. J. Gen. Virol. **81:**2267–2272.
11. **Hu, X., A. Javadian, P. Gagneux, and B. H. Robertson.** 2001. Paired chimpanzee hepatitis B virus (ChHBV) and mtDNA sequences suggest different ChHBV genetic variants are found in geographically distinct chimpanzee subspecies. Virus Res. **79:**103–108.
12. **Huy, T. T. T., T. N. Trinh, and K. Abe.** 2008. New complex recombinant genotype of hepatitis B virus identified in Vietnam. J. Virol. **82:**5657–5663.
13. **Kimura, T., A. Rokuhara, Y. Sakamoto, S. Yagi, E. Tanaka, K. Kiyosawa, and N. Maki.** 2002. Sensitive enzyme immunoassay for hepatitis B virus core-related antigens and their correlation to virus load. J. Clin. Microbiol. **40:**439–445.
14. **Kurbanov, F., Y. Tanaka, A. Kramvis, P. Simmonds, and M. Mizokami.** 2008. When should "I" consider a new hepatitis B virus genotype? J. Virol. **82:**8241–8242.
15. **Lee, W. M.** 1997. Hepatitis B virus infection. N. Engl. J. Med. **337:**1733–1745.
16. **Lindh, M., A. S. Andersson, and A. Gusdal.** 1997. Genotypes, nt 1858 variants, and geographic origin of hepatitis B virus: large-scale analysis using a new genotyping method. J. Infect. Dis. **175:**1285–1293.
17. **Liu, C. J., B. F. Chen, P. J. Chen, M. Y. Lai, W. L. Huang, J. H. Kao, and D. S. Chen.** 2006. Role of hepatitis B viral load and basal core promoter mutation in hepatocellular carcinoma in hepatitis B carriers. J. Infect. Dis. **193:**1258–1265.
18. **Liu, C. J., J. H. Kao, and D. S. Chen.** 2005. Therapeutic implications of hepatitis B virus genotypes. Liver Int. **25:**1097–1107.
19. **Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard, and S. C. Ray.** 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. **73:**152–160.
20. **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.
21. **Magiorkinis, E. N., G. N. Magiorkinis, D. N. Paraskevis, and A. E. Hatzakis.** 2005. Re-analysis of a human hepatitis B virus (HBV) isolate from an East African wild born *Pan troglodytes schweinfurthii*: evidence for interspecies recombination between HBV infecting chimpanzee and human. Gene **349:** 165–171.
22. Reference deleted.
23. **Miyakawa, Y., and M. Mizokami.** 2003. Classifying hepatitis B virus genotypes. Intervirology **46:**329–338.
24. **Morozov, V., M. Pisareva, and M. Groudinin.** 2000. Homologous recombination between different genotypes of hepatitis B virus. Gene **260:**55–65.
25. **Norder, H., A. M. Courouce, P. Coursaget, J. M. Echevarria, S. D. Lee, I. K. Mushahwar, B. H. Robertson, S. Locarnini, and L. O. Magnius.** 2004. Genetic diversity of hepatitis B virus strains derived worldwide: genotypes, subgenotypes, and HBsAg subtypes. Intervirology **47:**289–309.
26. **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.
27. **Okamoto, H., M. Imai, F. Tsuda, T. Tanaka, Y. Miyakawa, and M. Mayumi.** 1987. Point mutation in the S gene of hepatitis B virus for a *d/y* or *w/r* subtypic change in two blood donors carrying a surface antigen of compound subtype *adyr* or *adwr*. J. Virol. **61:**3030–3034.
28. **Okamoto, H., F. Tsuda, Y. Akahane, Y. Sugai, M. Yoshiba, K. Moriyama, T. Tanaka, Y. Miyakawa, and M. Mayumi.** 1994. Hepatitis B virus with mutations in the core promoter for an e antigen-negative phenotype in carriers with antibody to e antigen. J. Virol. **68:**8102–8110.
29. **Okamoto, H., F. Tsuda, H. Sakugawa, R. I. Sastrosoewignjo, M. Imai, Y. Miyakawa, and M. Mayumi.** 1988. Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. J. Gen. Virol. **69**(Pt. 10):2575–2583.
30. **Okamoto, H., S. Yotsumoto, Y. Akahane, T. Yamanaka, Y. Miyazaki, Y. Sugai, F. Tsuda, T. Tanaka, Y. Miyakawa, and M. Mayumi.** 1990. Hepatitis B viruses with precore region defects prevail in persistently infected hosts along with seroconversion to the antibody against e antigen. J. Virol. **64:** 1298–1303.
31. **Olinger, C. M., P. Jutavijittum, J. M. Hubschen, A. Yousukh, B. Samountry, T. Thammavong, K. Toriyama, and C. P. Muller.** 2008. Possible new hepatitis B virus genotype, southeast Asia. Emerg. Infect. Dis. **14:**1777–1780.
32. **Orito, E., T. Ichida, H. Sakugawa, M. Sata, N. Horiike, K. Hino, K. Okita, T. Okanoue, S. Iino, E. Tanaka, K. Suzuki, H. Watanabe, S. Hige, and M. Mizokami.** 2001. Geographic distribution of hepatitis B virus (HBV) genotype in patients with chronic HBV infection in Japan. Hepatology **34:**590–594.

33. **Orito, E., M. Mizokami, H. Sakugawa, K. Michitaka, K. Ishikawa, T. Ichida, T. Okanoue, H. Yotsuyanagi, and S. Iino.** 2001. A case-control study for clinical and molecular biological differences between hepatitis B viruses of genotypes B and C. Hepatology **33:**218–223.
34. **Palumbo, E.** 2007. Hepatitis B genotypes and response to antiviral therapy: a review. Am. J. Ther. **14:**306–309.
35. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
36. **Sakamoto, T., Y. Tanaka, E. Orito, J. Co, J. Clavio, F. Sugauchi, K. Ito, A. Ozasa, A. Quino, R. Ueda, J. Sollano, and M. Mizokami.** 2006. Novel subtypes (subgenotypes) of hepatitis B virus genotypes B and C among chronic liver disease patients in the Philippines. J. Gen. Virol. **87:**1873–1882.
37. **Sall, A. A., S. Starkman, J. M. Reynes, S. Lay, T. Nhim, M. Hunt, N. Marx, and P. Simmonds.** 2005. Frequent infection of *Hylobates pileatus* (pileated gibbon) with species-associated variants of hepatitis B virus in Cambodia. J. Gen. Virol. **86:**333–337.
38. **Sa-nguanmoo, P., C. Thongmee, P. Ratanakorn, R. Pattanarangsan, R. Boonyarittichaikij, S. Chodapisitkul, A. Theamboonlers, P. Tangkijvanich, and Y. Poovorawan.** 2008. Prevalence, whole genome characterization and phylogenetic analysis of hepatitis B virus in captive orangutan and gibbon. J. Med. Primatol. **37:**277–289.
39. **Shin-I, T., Y. Tanaka, Y. Tateno, and M. Mizokami.** 2008. Development and public release of a comprehensive hepatitis virus database. Hepatol. Res. **38:**234–243.
40. **Simmonds, P., and S. Midgley.** 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. J. Virol. **79:**15467–15476.
41. **Starkman, S. E., D. M. MacDonald, J. C. Lewis, E. C. Holmes, and P. Simmonds.** 2003. Geographic and species association of hepatitis B virus genotypes in non-human primates. Virology **314:**381–393.
42. **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.
43. **Sugauchi, F., M. Mizokami, E. Orito, T. Ohno, H. Kato, S. Suzuki, Y. Kimura, R. Ueda, L. A. Butterworth, and W. G. Cooksley.** 2001. A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. J. Gen. Virol. **82:**883–892.
44. **Sugauchi, F., E. Orito, T. Ichida, H. Kato, H. Sakugawa, S. Kakumu, T. Ishida, A. Chutaputti, C. L. Lai, R. Ueda, Y. Miyakawa, and M. Mizokami.** 2002. Hepatitis B virus of genotype B with or without recombination with genotype C over the precore region plus the core gene. J. Virol. **76:**5985–5992.
45. **Takahashi, K., K. Aoyama, N. Ohno, K. Iwata, Y. Akahane, K. Baba, H. Yoshizawa, and S. Mishiro.** 1995. The precore/core promoter mutant (T1762A1764) of hepatitis B virus: clinical significance and an easy method for detection. J. Gen. Virol. **76**(Pt. 12)**:**3159–3164.
46. **Takahashi, K., B. Brotman, S. Usuda, S. Mishiro, and A. M. Prince.** 2000. Full-genome sequence analyses of hepatitis B virus (HBV) strains recovered from chimpanzees infected in the wild: implications for an origin of HBV. Virology **267:**58–64.
47. **Tanaka, Y., and M. Mizokami.** 2007. Genetic diversity of hepatitis B virus as an important factor associated with differences in clinical outcomes. J. Infect. Dis. **195:**1–4.
48. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
49. **Tiollais, P., P. Charnay, and G. N. Vyas.** 1981. Biology of hepatitis B virus. Science **213:**406–411.
50. **Wang, Z., Z. Liu, G. Zeng, S. Wen, Y. Qi, S. Ma, N. V. Naoumov, and J. Hou.** 2005. A new intertype recombinant between genotypes C and D of hepatitis B virus identified in China. J. Gen. Virol. **86:**985–990.
51. **Wiegand, J., D. Hasenclever, and H. L. Tillmann.** 2008. Should treatment of hepatitis B depend on hepatitis B virus genotypes? A hypothesis generated from an explorative analysis of published evidence. Antivir. Ther. **13:**211–220.
52. **Wong, D. K., Y. Tanaka, C. L. Lai, M. Mizokami, J. Fung, and M. F. Yuen.** 2007. Hepatitis B virus core-related antigens as markers for monitoring chronic hepatitis B infection. J. Clin. Microbiol. **45:**3942–3947.

Journal of Human Genetics (2009) 54, 543–546

www.nature.com/jhg

SHORT COMMUNICATION

Genome-wide association database developed in the Japanese Integrated Database Project

Asako Koike[1], Nao Nishida[2], Ituro Inoue[3], Shoji Tsuji[4] and Katsushi Tokunaga[2]

The establishment of high-throughput single-nucleotide polymorphism (SNP)-typing technologies has enabled astonishing progress to be made in genome-wide association studies (GWAS), and various novel genetic factors associated with complex diseases have been discovered. Our organization has created a public repository database (DB) to achieve a continuous and intensive management of GWAS data and to facilitate data sharing among researchers. In the GWAS DB, information on study design, quality control protocols, allele frequencies, genotype frequencies and statistical genetic analysis results are stored as publicly available data and can be accessed freely, whereas individual genotyping data and raw data are stored as restricted data and can only be accessed with authorization. All data are presented by a graphic viewer, which is designed to be user friendly for researchers who are not familiar with GWAS to accelerate disease-related studies. Furthermore, the DB allows users to compare various study results obtained by different institutions and on different platforms. The same data are also managed as a distributed annotation system to call up useful data from other DBs and to superimpose them on the GWAS data for help in interpretation. The DB is accessible at https://gwas.lifesciencedb.jp/.
Journal of Human Genetics (2009) **54,** 543–546; doi:10.1038/jhg.2009.68; published online 24 July 2009

Keywords: database; genome-wide association; SNP

INTRODUCTION

The accomplishment of sequencing of the entire human genome[1,2] and the HapMap project,[3] coupled with the development of cost-effective high-throughput dense single-nucleotide polymorphism (SNP)-typing techniques, have enabled a genome-wide exploration of various complex disease-associated variants. Currently, the high-throughput SNP-typing methods are expected to cover about 80% of the human genome in linkage disequilibrium.[4] A number of large-scale genome-wide cohort studies and case–control studies, such as seven common disease GWAS by the Wellcome Trust Case Control Consortium (WTCCC, 2007), have been planned, and some of them are underway. So far, more than 100 loci of disease-related/causing candidates for about 40 common diseases and traits have been identified,[5] and some loci have led to new insights into pathophysiology and etiological pathways. Because GWAS yields large amounts of raw data and analysis results, the management of GWAS data has become a matter of serious concern. Furthermore, more and more grant-funding agencies, journal editors and research communities are beginning to require the disclosure of GWAS data. Disclosure and data sharing of GWAS data will primarily lead to the following three possibilities: (1) meta-analysis using data sets produced in multiple studies to find novel disease-related SNP candidates; (2) re-use of GWAS data combined with other experimental data, including pathway data and expression data, to deepen the exploration of each disease; and (3) development of methods to analyze and compute genetic statistics. In the case of meta-analysis in particular, the use of raw data is indispensable for quality control and for consideration of population structures. Some studies have successfully found additional disease-related SNP candidates on the basis of meta-analysis.[6,7]

The National Center for Biotechnology Information launched the database (DB) of Genotype and Phenotype in the fall of 2006 as a centralized GWAS system to archive and distribute GWAS data. Currently, results funded by the Genetic Association Information Network and voluntarily submitted data have been accumulated. The European Genotype Archive was created in the spring of 2008 as a repository system for phenotype–genotype relationships, and results primarily from WTCCC have been accumulated and redistributed. To achieve a continuous and intensive management of GWAS data and data sharing among researchers, we established a new DB that is publicly available. This DB is expected to have an essential role in providing easily accessible GWAS data to researchers in various biomedical fields. Some disease-related SNPs are assumed to be buried because of their insufficient *P*-values caused by an insufficient number of case–control samples. It is possible that these SNPs will be revealed by combining the GWAS analysis results with other data possessed by users.

In this paper, we introduce the GWAS DB.

[1]Central Research Laboratory, Hitachi Ltd, Tokyo, Japan; [2]Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; [3]Department of Molecular Life Science and Molecular Medicine, Tokai University School of Medicine, Tokyo, Japan and [4]Department of Neurology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan
Correspondence: Dr A Koike, Central Research Laboratory, Hitachi Ltd, 1-280 Higashi-koigakubo Kokubunji, Tokyo, Japan.
E-mail: asako.koike.ea@hitachi.com
Received 3 June 2009; accepted 27 June 2009; published online 24 July 2009

MATERIALS AND METHODS

Database structure

The DB system consists of an internal GWAS DB and a public GWAS DB. For a maximum of 1 year, or until the acceptance of publication, submitted data are stored in the internal GWAS DB and can be accessed only by the research team that submitted the data for greater convenience in data sharing among research team members living in various locations. Currently, the DB systems are implemented using mysql version 5.0 (http://dev.mysql.com/downloads/mysql/5.0.html), and some of the statistical analysis results are also accumulated in a distributed annotation system (DAS) server. A schematic drawing of the GWAS DB is shown in Figure 1.

In this DB, three types of data access, namely, (1) public access, (2) authorized access accompanied by a data use application, and (3) authorized access accompanied by a data use application and its review by a data access committee, are possible. Principally, frequency data of genotypes and alleles and statistical analysis results can be accessed freely. However, automatic access and frequent access are restricted to prevent the release of frequency data of genome-wide genotypes and alleles, as such a large volume of genotype/allele data leads to the specification of whether the given genome is contained in the case or in the control group, as reported previously.[8] These genome-wide frequency data can be obtained by submitting a data use application to the data access committee. For the use of genotype or raw data, an application that

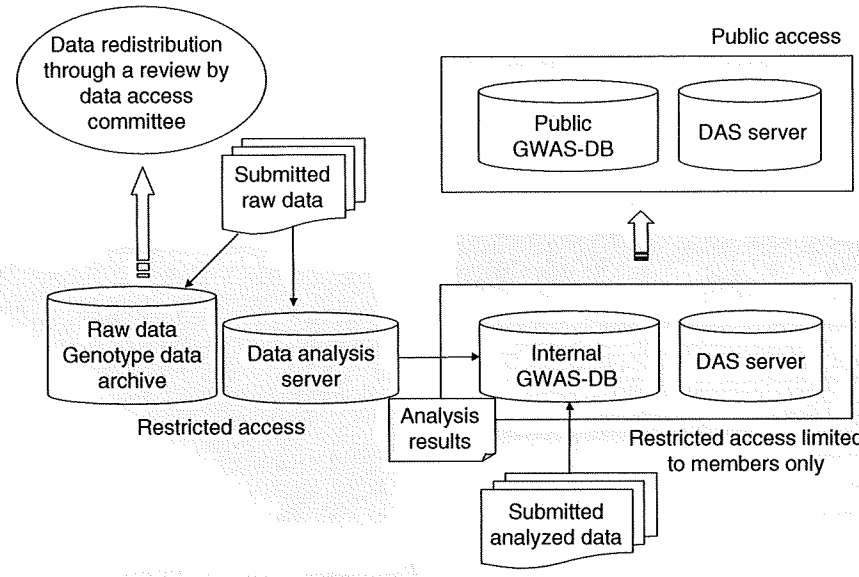

Figure 1 Schematic drawing of genome-wide association study (GWAS) database (DB) systems.

Table 1 Summary of database contents

Contents	*Data sources*
Statistics	
Frequencies of genotypes, alleles and haplotypes	
Statistical genetic analysis	
P-values and odds ratios on genotypic model and allelic model	
P-values and odds ratios on trend model, additive model and recessive model	
Permutation test results	
Bonferroni's corrections and false discovery rate for multiple testing using	
Akaike information criterion	
Hardy–Weinberg equilibrium test	
Haplotype-based χ^2-test	
Epistasis	
Linkage disequilibrium parameters (r^2, D', Lod)	
Other data	
mRNA, amino-acid sequence of each gene	NCBI (http://www.ncbi.nlm.nih.gov/)
mRNA, genome-mapped position	UCSC Hg. 18 (http://hgdownload.cse.ucsc.edu/)
SNP position and SNP kind (cSNP, sSNP, rSNP and so on)	NCBI (http://www.ncbi.nlm.nih.gov/)
OMIM	NCBI (http://www.ncbi.nlm.nih.gov/)
Copy number variation	DGV (http://projects.tcag.ca/variation/)
Gene function	Gene ontology (http://www.geneontology.org/)
Microsatellite polymorphism	UCSC (http://hgdownload.cse.ucsc.edu/)
Manually curated disease-related mutation information	

describes the research purpose and lists the research team members must be submitted to the data access committee. The data access committee deliberates on whether the applicant's research purpose meets the content of the consent form. Only applicants approved by the review committee can use individual genotype data and raw data in accordance with the data handling security rules required by the data access committee and following data use restrictions on the basis of informed consent.

Individual data and raw data are accumulated in the server in a secured computer environment that is different from the public DB server. Only authorized persons can access this server.

Data submission

In principal, both analysis results and unanalyzed data can be submitted. When data have already been analyzed, the analyzed data are accumulated in this DB, along with a detailed description of the analysis protocols. When data have not been analyzed yet, they are analyzed in our site, and the results are accumulated in this DB. When raw data are redistributable under certain conditions, they are also submitted with the contents of the consent form. All data must be submitted with documents explaining the design of the study, as well as ethical consideration.

Data cleaning for quality control

When data are submitted as individual data without analysis results, they are analyzed as follows: (1) SNPs with a call rate <95% and samples with a call rate <95% are removed. (2) SNPs, the Hardy–Weinberg equilibrium test result of which in a control group is less than 0.001 or the minor allele frequency of which is less than 0.05, are removed. (3) The principle component analysis (PCA) of these case–control data, along with HapMap data, is carried out using EIGENSTRAT[9] or other programs so that sample outliers and samples with a possible ethnic mixture or a different ethnicity are removed on the basis of the PCA result. Sample outliers in the plot of heterozygosity versus call rate are also removed. The quantile–quantile plot based on the allelic model is calculated and checked. When only genotype frequency data are submitted, PCA and heterozygosity checks are skipped,

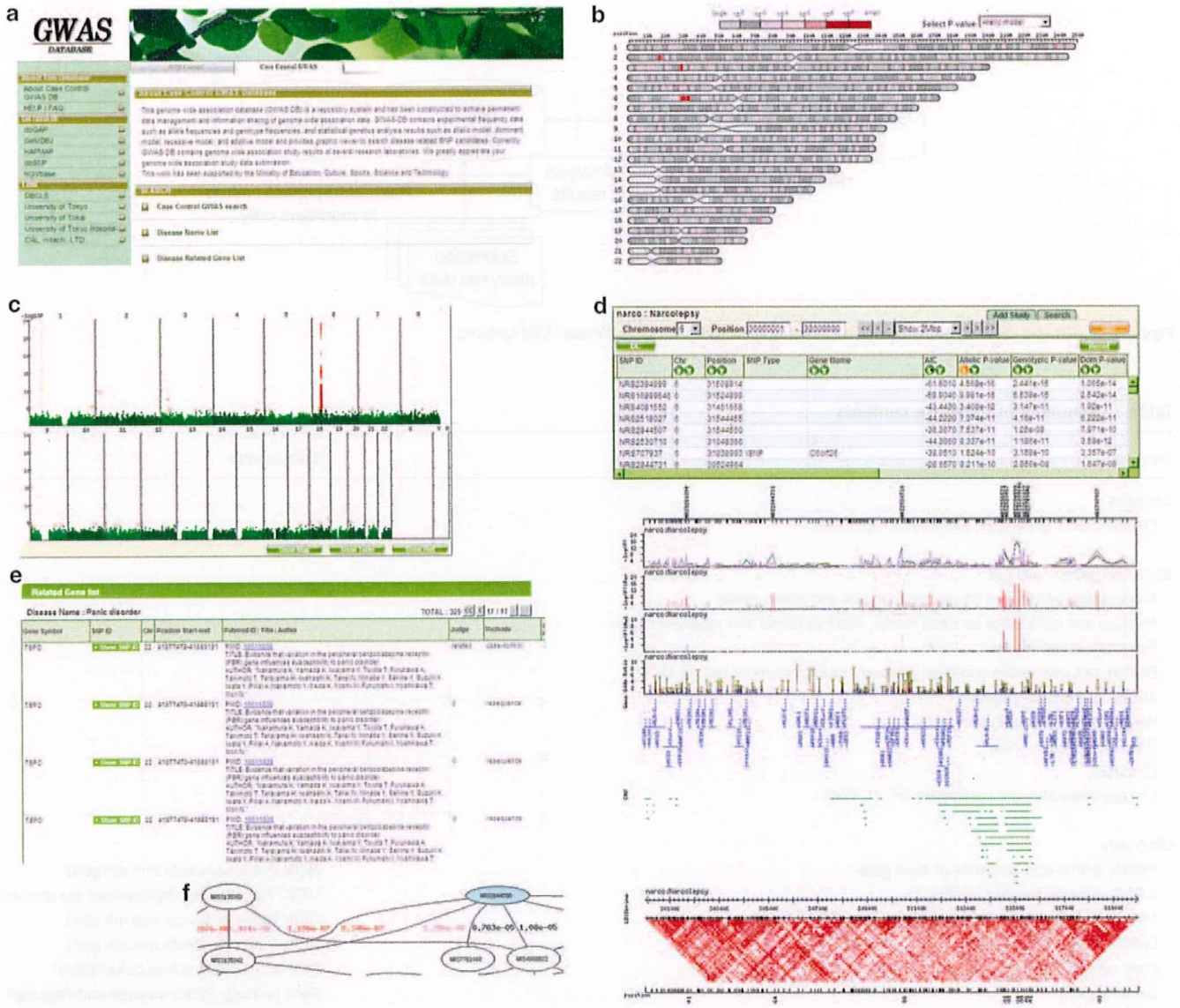

Figure 2 Snapshots of the genome-wide association study (GWAS) database. **(a)** Top page, **(b)** bird's-eye view, **(c)** Manhattan plot, **(d)** region table and graph, **(e)** disease-related gene/single-nucleotide polymorphism (SNP) lists (public data) and **(f)** SNP network based on epistasis.

as they require individual data. The cleaning results are linked from 'study details' on the web.

Data analysis

Standard statistical genetic analyses are performed by plink[10] and Haploview.[11] Additional analyses such as the Akaike information criterion, epistasis and more complicated ones (for example, genetic analysis considering potential case samples existing in the control samples, which sometimes becomes a concern for diseases that develop in old age) are calculated by internally developed programs. The major statistics include *P*-values based on an allelic model, genotypic model, trend model, dominant model, recessive model and permutation test results of these models, and Bonferroni's correction and false discovery rate for multiple testing. These methods are also shown in 'study details.' When submitted data consist of only genotype frequency data, the genome-wide permutation test is skipped.

Database contents and utility

The DB contents (as of April 2009) are summarized in Table 1.

User data other than GWAS data, such as expression data and epigenetic data, are also accumulated and can be displayed on the graph. Although clinical data are not currently accumulated in the DB, they can be added if submitted. Major tables are summarized in Supplementary Table 1.

A snapshot of the GWAS DB is shown in Figure 2. Figure 2a shows the top page of the GWAS DB. When the 'SNP control' tab is selected, the interface jumps to the SNP control DB, which is affiliated to the GWAS DB and contains allelic frequencies, genotypic frequencies, Hardy–Weinberg equilibrium tests and estimated haplotype frequencies of Japanese control samples. Bird's-eye view (Figure 2b) and Manhattan plot (Figure 2c) are provided to draw *P*-values of each model. A genome region can be selected from both (Figures 2b and c), and the results of statistical genetic analysis along with other information such as exon–intron information and copy number variations (CNVs) can be displayed in tables and graphs to facilitate the identification of disease-related SNPs, as shown in Figure 2d. Furthermore, comparisons among various study results obtained by different institutions and/or different platforms can be carried out easily by plotting their graphs on the web (using the 'add study' function in Figure 2d). When the published disease-related gene or SNP is registered as shown in Figure 2e, data are plotted as a known disease-related gene/SNP in the graph (Figure 2d). Epistasis data are also accumulated and drawn as a network graph using Graphviz (http://www.graphviz.org/), as shown in (Figure 2f). Data can be searched by SNP ID (dbSNP ID #rs, affymetrix SNP ID and so on), gene name, disease name and so on. The study design and analysis protocols can also be browsed.

Statistical results are also accumulated on a DAS server, and they can be browsed using the Gmod Gbrowse (http://gmod.org/wiki/Main_Page)-based browser (http://gwas.lifesciencedb.jp/cgi-bin/gbrowse/snpdb/). Furthermore, as a function of the DAS server, data on other DAS servers such as Ensemble can be called up. This function is useful to superimpose data from other DBs onto GWAS data. The GWAS DB is designed to be user friendly for researchers unfamiliar with GWAS to promote disease-related studies.

Further development

A recent topic of interest is genome-wide association analysis coupled with other data such as pathway data[12] to compensate for the low statistical power in disease-associated candidate SNPs. The function to browse or calculate SNP/SNP pair *P*-values on the basis of the GWAS result, along with other data, will be added to this DB to facilitate the generation and understanding of user hypotheses.

The relationships between CNVs and diseases have begun to emerge in recent studies.[13] Although concerns remain about the quality of detected CNVs, genomic locations and frequencies of CNV regions and their case–control association study results will be incorporated into this DB. Furthermore, in the near feature, new high-throughput techniques such as short-read sequencing will be applied for GWAS, and this DB will be improved to suit the new experimental techniques.

ACKNOWLEDGEMENTS

This work was supported by the contract research fund 'Integrated Database Project' from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

1 Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
2 Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
3 The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437,** 1299–1320 (2005).
4 Barrett, J. C. & Cardon, L. R. Evaluating coverage of genome-wide association studies. *Nat. Genet.* **38,** 659–662 (2006).
5 Manolio, T. A., Brooks, L. D. & Collins, F. S. A HapMap harvest of insights into the genetics of common disease. *J. Clin. Invest.* **118,** 1590–1605 (2008).
6 Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40,** 638–645 (2008).
7 Houlston, R. S., Webb, E., Broderick, P., Pittman, A. M., Di Bernardo, M. C., Lubbe, S. *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat. Genet.* **40,** 1426–1435 (2008).
8 Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J. *et al.* Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet.* **4,** e000167 (2008).
9 Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).
10 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
11 Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21,** 263–265 (2005).
12 Baranzini, S. E., Galwey, N. W., Wang, J., Khankhanian, P., Lindberg, R., Pelletier, D. *et al.* Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum. Mol. Genet.* **18,** 2078–2090 (2009).
13 McCarroll, S. A. Extending genome-wide association studies to copy-number variation. *Hum. Mol. Genet.* **17** (R2), R135–R142 (2008).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)

基礎的観点から 2
疾患感受性遺伝子とゲノムワイド関連解析

西田奈央　　徳永勝士

にしだ なお，とくなが かつし：東京大学大学院医学系研究科 人類遺伝学分野

● はじめに

単一塩基多型（single nucleotide polymorphism：SNP）を検出する技術の進展に伴って，ヒトのさまざまな多因子疾患に関わる遺伝子を探索する戦略として，ゲノムワイド関連研究（genome-wide association study：GWAS）が近年大きな注目を浴びている。2007年5月には，90万種を超えるSNP解析用プローブ，およびCNV（copy number variation）解析用の94万種を超えるプローブを搭載したキットが市販された（Affymetrix® Genome-Wide Human SNP Array 6.0：SNP Array 6.0）[1]。われわれの教室に設置したヒトSNPタイピングセンターでは，いくつかの多因子疾患についてSNP Array 6.0によるゲノムワイド関連分析を実施している。

SNP Array 6.0に搭載されたSNPは，公共のSNPデータベースおよびPerlegen社に登録された約220万種のSNPから遺伝学的情報量が最大化されるように，また連鎖不平衡やHapMapプロジェクトからの情報も考慮して選択された約44万種のSNPに，Tag SNP，X染色体およびY染色体に存在するSNP，ミトコンドリアSNPなどを加えた全909,622種のSNPである。これらのSNPについて，HapMapの3集団（Caucasian，African，Asian）におけるマイナーアリル頻度（minor allele frequency：MAF）の平均は，それぞれ19.6%，20.6%，18.2%である。

しかしながら，HapMapプロジェクトでは45検体の日本人しか解析していないため，MAFの低いSNPについては正確な頻度推定ができない。そこで，われわれはSNP Array 6.0を用いて日本人健常者200検体を解析し，日本人を対象としたGWASにおいて統計解析に用いることのできるSNP数を算出することを試みた。また，SNP Array 6.0は遺伝子型を決定するためにBirdseedアルゴリズムを用いるが，Birdseedアルゴリズムを用いた遺伝子型決定の精度を上げることが，ゲノムワイド関連分析における偽陽性関連を効果的に排除することにつながる。そこで，日本人健常者200検体のタイピング結果を用いて，Birdseedアルゴリズムによる正確な遺伝子型決定方法を検討した。

● ゲノムワイド関連研究の動向

ゲノムワイド関連研究は日本の研究者によって先駆的に行われ，これまでにいくつかのヒト多因子疾患の感受性遺伝子を特定することに成功している[2,3]。また，日本では2003年からオーダーメイド医療実現基盤を構築することを目標とした「オーダーメイド医療実現化プロジェクト」が開始され，30万人の日本人を対象とした遺伝情報解析が行われている[4]。2008年には，日本における2大プロジェクトである「オー

ダーメイド医療実現化プロジェクト」と「ミレニアムゲノムプロジェクト」から，それぞれ独立に2型糖尿病に関連する遺伝子である*KCNQ1*を発見したという報告がなされた[5,6]。また，われわれの研究室においても，*CPT1B*遺伝子と*CHKB*遺伝子の間に存在するSNPが睡眠障害のひとつであるナルコレプシーと関連していることを発見し，2008年に報告をした[7]。

SNP解析技術の著しい進展によって，近年，大規模なゲノムワイド関連研究が計画・実施されてきている。2007年には，WTCCC（The Wellcome Trust Case Control Consortium）は7種類のcommon diseases〔双極性感情障害（BD），冠動脈疾患（CAD），クローン病（CD），高血圧（HT），関節リウマチ（RA），1型糖尿病（T1D），2型糖尿病（T2D）〕について，それぞれ2,000人の患者とコントロールとして健常者3,000人の計17,000人を対象とした大規模なゲノムワイド関連研究を行った[8]。また，大規模な疫学研究として知られるFramingham Heart Studyで収集された試料のうち9,000検体について，心，肺，血液，睡眠疾患に関与する遺伝子変異を探索する計画が発表され，ゲノムワイド関連分析およびゲノムワイド連鎖解析などを行った結果が2007年にまとめて報告された[9]。

● 日本人健常者200検体を用いたゲノムワイドSNPタイピング

SNP Array 6.0によるゲノムワイドSNPタイピングでは，1検体につき500 ngのゲノムDNAを使用する（図1）。制限酵素（StyI, NspI）を用いたゲノムDNA断片化反応において，ゲノムDNA量を250 ngとなるように調整することが，タイピングの精度に大きな影響を与える

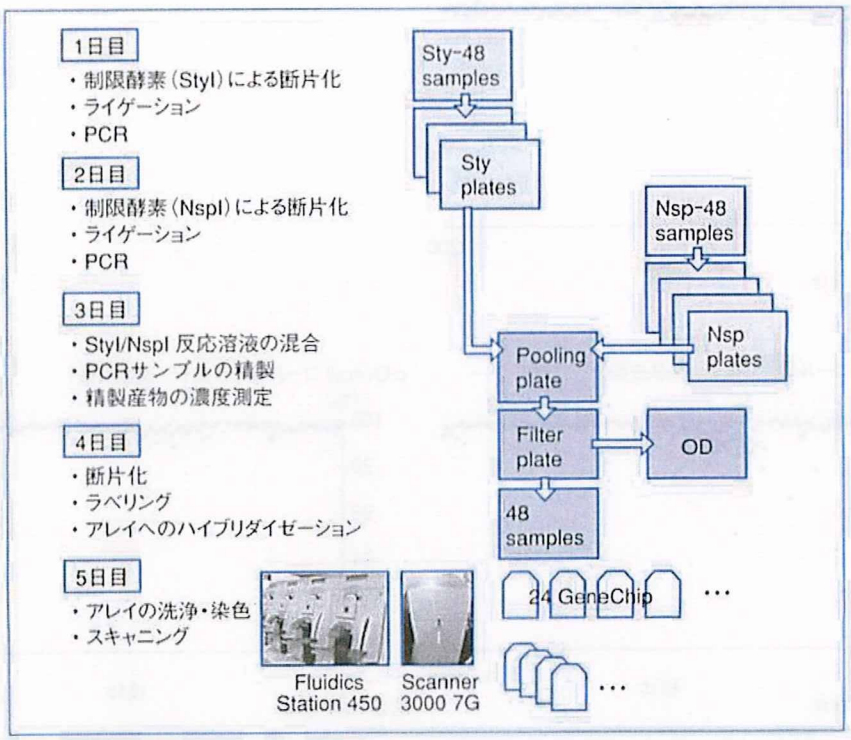

図1　Affymetrix® Genome-Wide Human SNP Array 6.0によるSNP（単一塩基多型）タイピングの流れ

制限酵素（StyI, NspI）による断片化反応からGeneChipのスキャニングまで，全5日の工程でSNPタイピングが行われる。1検体につき500 ngのゲノムDNAを用いて，全909,622種のSNPをタイピングすることができる。

ことが明らかとなっている[10]。

日本人健常者200検体のうち195検体のゲノムDNA濃度は規定濃度（50 ng/μL）を満たしており，平均54.8 ng/μL（45.0～57.8）であったが，残る5検体は規定濃度を下回り平均41.1 ng/μL（38.2～44.5）であった。そこで，規定濃度を下回った5検体は制限酵素によるゲノムDNA断片化反応に6μLを持ち込み，ゲノムDNAの総量が250 ngとなるように調整してタイピングを行った。日本人健常者200検体のSNPタイピングを行った結果，クオリティーコントロール（QC）コール率は平均97.37%となった（図2a）。ここでQCコール率とは，タイピングデータの質を評価するために用いられる指標で，SNP Array 6.0に搭載された3,022 SNPsについてのコール率を示す。これら3,022 SNPsの遺伝子型はDMアルゴリズム（confidence score＝0.17）で決定され，コール率が86%以下となった検体は解析対象から除外する。日本人健常者200検体のタイピングデータのうち，QCコール率が86%を下回った2検体を除外し，86%を上回った198検体についてBirdseedアルゴリズムを用いて全909,622 SNPsの遺伝子型を決定したところ，Overallコール率は平均99.58%（96.42～99.90）となった（図2b）。

● Birdseedアルゴリズムによる正確な遺伝子型決定方法

膨大なSNPデータを取り扱うゲノムワイド関連分析において，タイピングエラーが原因で

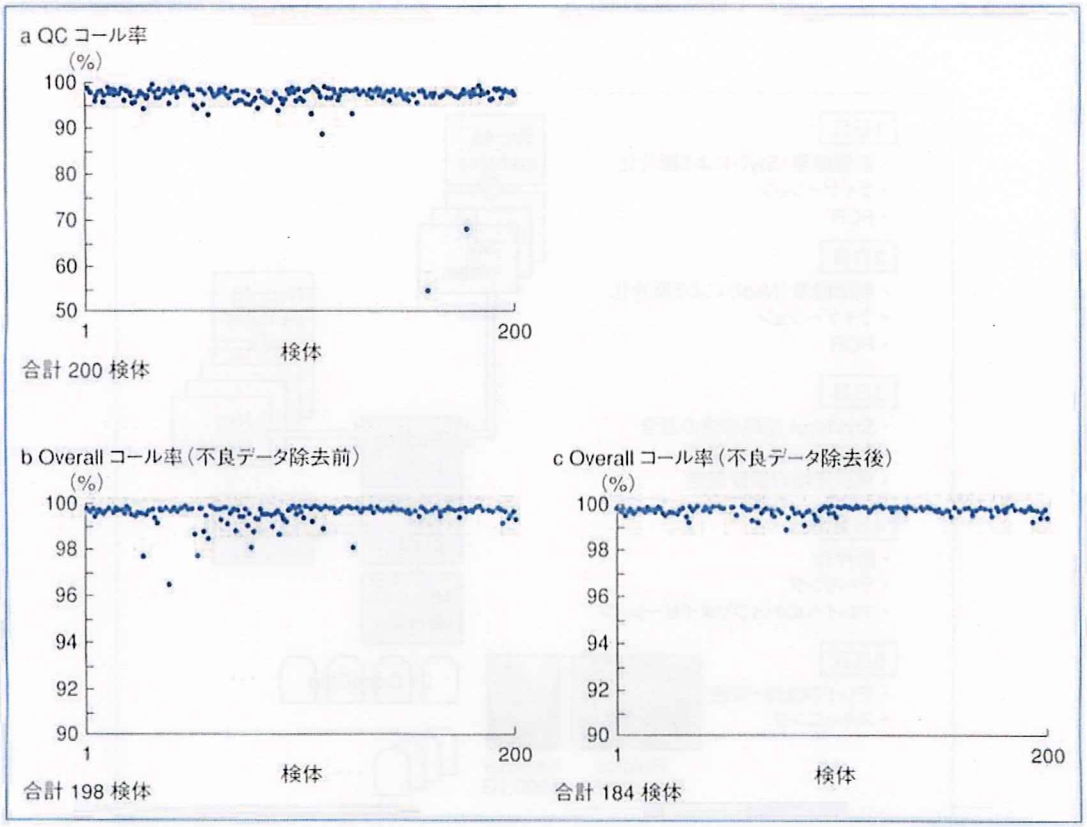

図2　Affymetrix® Genome-Wide Human SNP Array 6.0による日本人健常者200名のタイピング結果
a：クオリティーコントロール（QC）としてタイピングされた3,022 SNPs（単一塩基多型）のコール率を示す。
b：QCコール率が86%を上回った198検体を用いて決定された全909,622 SNPsのコール率を示す。
c：QCコール率を指標として不良データを除去した後の184検体を用いて遺伝子型を決定した際の全909,622 SNPsのコール率を示す。