JOURNAL OF CLINICAL MICROBIOLOGY, Nov. 2010, p. 4063–4071 Vol. 48, No. 11
0095-1137/10/$12.00 doi:10.1128/JCM.01174-10

Unique Hepatitis B Virus Subgenotype in a Primitive Tribal Community in Eastern India▽

Sumantra Ghosh,[1] Priyanka Banerjee,[1] Arindam RoyChoudhury,[2] Sumanta Sarkar,[1] Alip Ghosh,[1] Amal Santra,[1] Soma Banerjee,[1] Kausik Das,[1] Bhagirathi Dwibedi,[3] Shantanu K. Kar,[3] Vikas Gangadhar Rao,[4] Jyothi T. Bhat,[4] Neeru Singh,[4] Abhijit Chowdhury,[1] and Simanti Datta[1]*

Centre for Liver Research, School of Digestive and Liver Diseases, Institute of Post Graduate Medical Education and Research, Kolkata, India[1]; Department of Biostatistics, Columbia University, New York, New York 10032[2]; Regional Medical Research Centre, Bhubaneswar, India[3]; and Regional Medical Research Centre for Tribals, Jabalpur, India[4]

Received 10 June 2010/Returned for modification 3 August 2010/Accepted 1 September 2010

Hepatitis B virus (HBV) strains isolated from members of the primitive Paharia ethnic community of Eastern India were studied to gain insight into the genetic diversity and evolution of the virus. The Paharia tribe has remained quite separate from the rest of the Indians and differs culturally, genetically, and linguistically from the mainstream East Indian population, whose HBV strains were previously characterized. Full-length HBV DNA was PCR amplified, cloned, and sequenced. Phylogenetic relationships between the tribal sequences and reference sequences from the mainstream population were assessed, and divergence times of subgenotypes of HBV genotype D were estimated. HBV was found in 2% of the Paharias participating in the study. A predominance of hepatitis B e antigen-negative infection (73%) was observed among the Paharias, and the genome sequences of the HBV strains exhibited relative homogeneity, with a very low prevalence of mutations. The novel feature of Paharia HBV was the exclusive presence of the D5 subgenotype, which was recently identified in Eastern India. Analysis of the four open reading frames (ORFs) of these tribal HBV D5 sequences and comparison with previously reported D1 to D7 sequences enabled the identification of 27 specific amino acid residues, including 6 unique ones, that could be considered D5 signatures. The estimated divergence times among subgenotypes D1 to D5 suggest that D5 was the first to diverge and hence is the most ancient of the D subgenotypes. The presence of a specific, ancient subgenotype of HBV within an ethnically primitive, endogamous population highlights the importance of studies of HBV genetics in well-separated human populations to understand viral transmission between communities and genome evolution.

The extraordinary level of diversity in outcomes of hepatitis B virus (HBV) infection arises primarily from host-virus interactions (31). It has emerged that HBV genetic components exist in a dynamic equilibrium with host innate and acquired immune response determinants and that the steady state or deviation thereof determines the nature of clinical manifestation (7, 30). Immune selective pressure during long periods of persistence coupled with the lack of proofreading activity by the HBV polymerase drives the complexity and diversification of HBV quasispecies and led to the emergence of distinct genotypes and their subgenotypes in specific geo-ethnic populations (9, 12). Viral genotypes have been shown to correlate well with human population migration (22, 23, 29), which further supplements the fact that the HBV genome carries its own signature, preserved in defined population groups, that may reflect selection based on differences among human hosts in traits important to pathophysiology/transmission of the virus in that population. India is remarkable for its rich population genomic diversity, and it has been proposed that various founder events shaped each Indian population, whereby each group bears unique genetic imprints that have been maintained for thousands of years by social practices such as endogamy (24). The diverse Indian populations thus provide an excellent opportunity to study the features of HBV genomes harbored by the variegated ethnic groups living there, in particular the ancient tribal population, to understand the genetic underpinnings of HBV evolution and disease biology.

The Paharias are a primitive tribe settled in the hilly terrains of the Rajmahal Hills in the state of Jharkhand, in the eastern part of India. They are ethnically Dravidian, speak the Malto dialect, similar to the Khurukh dialect of the Dravidian linguistic family, and rarely intermarry with people of other ethnicities (13, 25). In contrast, the people of Eastern India have an Indo-European ancestry, and their languages (Hindi, Bengali, and Oriya) were derived from Sanskrit. The Paharias remained sufficiently isolated from mainstream Indians during much of human history, which is reflected in differences in the human gene pool (4, 26), and by extrapolation, this seems to have fostered divergence in the HBV population they carry. Here we report a population-based study of HBV infection among members of the Paharia tribe to test whether the viral gene pool is distinct from that of previously studied HBV strains from Eastern India for a better understanding of the genetic geography and evolution of this virus.

* Corresponding author. Mailing address: Centre for Liver Research, School of Digestive and Liver Diseases, Institute of Post Graduate Medical Education and Research, 244 A. J. C. Bose Road, Kolkata 700020, India. Phone and fax: (91)-033-2223 5435. E-mail: seemdatt@gmail.com.

▽ Published ahead of print on 15 September 2010.

TABLE 1. Primers used in this study

Primer	Sequence (5′–3′)	Location (nt)
F1	CACAAGAGGACTCTTGGACT	1653–1672[a]
F2	CTGCACTCAGGCAAGCAATT	2058–2077[a]
F3	CGCCTCATTTTGTGGGTCAC	2801–2820[a]
F4	CTCAGGCCATGCAGTGGAA	3164–3182[a]
F6	TGCCGATCCATACTGCGGAA	1259–1278[a]
F9	TTTACCTCTATTACCAATTTTC	789–810[b]
F10	GACCACCAAATGCCCCTATC	2298–2317[b]
R1	CCACCTTATGAGTCCAAGG	2457–2475[a]
R2	AAATTACCACCCACCCAGG	2109–2127[a]
R3	AACTGGAGCCACCAGCAG	57–74[a]
R4	AGAGGACAAACGGGCAACA	462–480[a]
R5	AAAGCCCAAAAGACCCACAAT	997–1017[a]

[a] Positions are given according to the HBV sequence under GenBank accession no. AF121242.
[b] Positions are given according to the HBV sequence under GenBank accession no. AB014371.

MATERIALS AND METHODS

Collection of samples and storage. In a community-based epidemiological approach, four Paharia hamlets (Murgathali, Chakalta, Korigha, and Mazladi; the primary settlements) were identified in the Dumka District of Jharkhand, based on the cooperation of the tribal village elders and group leaders. A 1:4 sampling of the households was done in each hamlet, and all individuals above 10 years of age in each of these households were invited to participate in the study. A total of 754 healthy individuals of both sexes belonging to the Paharia tribal community were included in the study. Informed consent was obtained from the participating subjects and, in the case of minors, from the parents. From each participant, 6 to 7 ml of blood was collected, and the serum was isolated at the site, preserved in dry ice, transported to the laboratory within 24 h of collection, and stored in a −80°C refrigerator until use. The study procedures were reviewed and approved by the ethical committee of the Institute of Post Graduate Medical Education and Research, Kolkata, India, as well as by the Tribal Welfare Department, Government of Jharkhand.

Testing of serological markers and liver enzymes. Each serum sample was tested for the levels of the liver enzymes alanine aminotransferase (ALT) and aspartate aminotransferase (AST), using commercially available kits from Bayer Diagnostics (India) in a semiautomated biochemistry analyzer (RA-50; Bayer). The normal ranges for AST and ALT are 2 to 45 IU/liter and 2 to 40 IU/liter, respectively. Serological markers for hepatitis B virus surface antigen (HBsAg) and hepatitis B virus e antigen (HBeAg) and antibodies to hepatitis C virus (anti-HCV) and HBeAg (anti-HBeAg) were also checked in each sample, using commercially available enzyme-linked immunosorbent assay (ELISA) kits from General Biologicals, Taiwan, and bioMérieux, Boxtel, Netherlands.

HBV DNA isolation and quantification. HBV DNA was extracted from 200 μl serum for HBsAg-positive samples by use of a QIAamp DNA Mini kit (Qiagen Inc., Valencia, CA), and DNA was reconstituted in 50 μl of sterile water. Extracted DNA samples were stored at −20°C for PCR amplification, viral quantification, and sequencing analysis.

Viral loads were determined by real-time PCR, using a commercial SYBR green PCR master mix (AB Applied Biosystems, Foster City, CA) as described by Garson et al. (10). The lower limit of detection was 250 copies/ml.

Amplification of full-length HBV genome, cloning, and sequencing. Full-length HBV DNA (~3.2 kb) was amplified by a one-step amplification method described by Günther et al. (11). PCR products were purified by use of a QIAquick gel extraction kit (Qiagen, CA) and were cloned into the pJET1.2 vector by blunt end ligation, using a CloneJET PCR cloning kit (Fermentas, Canada). Recombinant plasmid DNAs were used to transform *Escherichia coli* DH5α. Five colonies with an HBV insert were selected for each sample. Plasmid DNA was purified from the colonies by use of a Qiagen Plasmid Mini kit and was checked for the presence of HBV genomes by performing 2 different single-step PCR amplification assays with primers F3 and R3 and primers F1 and R2 (Table 1), specific for the surface and core genes, respectively. The nucleotide sequence of full-length HBV was determined by sequencing representative cloned plasmids (usually 3 clones per sample) or the 3.2-kb PCR-amplified genome directly, using a BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems) on an automated DNA sequencer (ABI Prism 3130). Six appropriate internal primers (Table 1) designed to cover the complete genome were used for sequencing. DNA sequence editing and analysis were performed using Seqscape V2.5 (Applied Biosystems) software. For sequence alignment as well as phylogenetic analysis, reference sequences were selected from GenBank, and ClustalW was used to perform the alignment.

Estimation of phylogenetic trees and divergence times. We computed phylogenetic trees among genotypes of HBV (A to J) and among the subgenotypes of genotype D, using the Kimura two-parameter matrix and neighbor-joining (NJ) method in Mega software, version 4, and the likelihood method implemented in the DNAMLK component of PHYLIP. To confirm the reliability of the phylogenetic tree analysis, bootstrap resampling and reconstruction were carried out 1,000 times.

We especially examined the evolutionary relationships between the subgenotypes of D that are prevalent in India (D1 to D5). To do this with a high degree of accuracy, we used a likelihood method with the software DNAMLK. First, we identified the insertion-deletion (indel)-free regions of the DNA. A total of 684 base substitutions were identified in the indel-free regions. Next, the program DNAMLK from the PHYLIP package was used to build a tree for the 64 strains from the indel-free regions of the DNA. This software uses the base substitutions as well as the matching parts of the DNA to estimate a rooted tree, assuming a "molecular clock."

Nucleotide sequence accession numbers. The nucleotide sequence data reported in this paper are available in the GenBank database (http://www.ncbi.nlm.nih.gov/GenBank/index.html) under accession no. GQ205377 to GQ205389.

RESULTS

Demographic profile. Among 753 persons who participated in the study, 16 tested positive for anti-HCV, and hence they were excluded from the study. Thus, the remaining 737 individuals constituted the final study group. Among them, 15 (~2%) were found to be positive for HBsAg. Details of the demographic data are given in Table 2. Among the 15 HBsAg-positive individuals, 4 (26.6%) were also positive for HBeAg, while the rest were HBeAg negative. Thirteen had ALT levels in the normal range, with a mean value of 29.86 (±13.15) IU/liter, while in all of them AST levels were a little high, with a mean value of 48.86 (±17.73) IU/liter. HBeAg-to-anti-HBeAg seroconversion was observed in 5 of 15 (33.3%) individuals.

HBV DNA could be detected in 11 of 15 samples by real-time PCR, with the median serum HBV DNA load being 2.5×10^6 (range, 1,967 to 4.59×10^7) copies/ml. In the remaining 4 samples, the HBV DNA level was below the detection limit (<250 copies/ml). Full-length HBV DNAs were amplified from all 11 samples that were positive during viral load quantification and were cloned into the pJET1.2 vector. The complete nucleotide sequence of the HBV genome was determined for 3 representative clones per sample or directly from the amplified full-length PCR product. While for 9 of 11 samples there were uniformities in the HBV sequences of the representative clones derived from each sample, for 2 samples two different clones of HBV, varying in sequence from each other, were identified. Thus, a total of 13 full-length HBV sequences were obtained from 11 individuals.

HBV genotype phylogenetic analysis. Thirteen full-length HBV genome sequences obtained in this study, along with a total of 47 representative sequences of HBV strains of all known genotypes retrieved from GenBank, were used to calculate the phylogenetic relationships between the sequences (Fig. 1). All 13 tribal HBV sequences were found to be closely related to each other and clustered together with previously reported genotype D strains. Genotype D has been segregated further into 7 subgenotypes (D1 to D7), and to determine the subgenotype of the Paharia HBV strains, a second phyloge-

TABLE 2. Baseline characteristics of HBsAg-positive Paharia individuals participating in the study

ID	Age (yr)	Sex[a]	ALT level (IU/liter)	AST level (IU/liter)	HBV DNA concn (copies/ml)	Presence of HBeAg	Presence of anti-HBeAg
T1029	35	F	26	38	<250	Negative	Positive
T1159	38	M	38	54	<250	Negative	Positive
T1178	61	F	24	41	14,100	Negative	Negative
T1181	38	F	26	44	9,190	Negative	Positive
T1195	19	M	54	80	<250	Negative	Positive
T1234	40	M	62	52	1.42×10^7	Positive	Negative
T1245	21	M	31	55	1,967	Negative	Negative
T1258	35	M	16	27	<250	Negative	Positive
T1360	60	M	31	93	5.05×10^6	Positive	Negative
T1434	20	M	31	50	2.25×10^7	Positive	Negative
T1469	10	M	14	28	2.31×10^6	Negative	Negative
T1503	64	M	23	52	2.40×10^7	Negative	Negative
T1545	70	F	22	37	4.59×10^7	Positive	Negative
T1562	14	M	19	38	1.90×10^4	Negative	Negative
T1667	35	M	31	44	1.60×10^4	Negative	Negative

[a] F, female; M, male.

netic tree was built, using 52 different HBV sequences representing the 7 subgenotypes of genotype D (Fig. 2). Interestingly, all 13 tribal HBV sequences formed a cluster within subgenotype D5, which was described recently from Eastern India (3). The intragenotypic divergence for the tribal HBV strains was 1.49%, and the mean intergenotypic divergence data among the nucleotide sequences of the 7 subgenotypes of D are indicated in Table 3. No HBV recombinants were observed in this study. The sequences of the two different HBV clones obtained from each of two samples (T1360 and T1434) clustered together in the same clade (Fig. 2).

In our likelihood estimation of the tree for D1 to D5 (Fig. 3), all of the representative strains (except one) were grouped according to their subgenotypes. The only exception was the sequence AY057948, which was placed outside the D4 group. In fact, this strain, initially referred to as D4, was later identified as a D/C recombinant (27), and consequently it is estimated to have diverged much earlier than the rest of the strains. If we disregard strain AY057948, then the D5 line comprising the tribal strains was the first to diverge from the other strains. This suggests that D5 evolved in isolation for a period, during which all other subgenotypes (D1 to D4) evolved together. Thus, the branching event between D5 and the other strains is the most ancient among the HBV D genotypes.

Analysis of HBV open reading frames (ORFs). (i) pre-S/S ORF. The amino acid sequences for the surface antigen, as deduced from DNA sequence analysis, proved to be highly conserved in almost all of the tribal HBV isolates. A 61-amino-acid (61-aa) deletion in the pre-S1 region was observed in only one of the clones (clone b) of sample T1434, which also exhibited a 29-amino-acid truncation at the C terminus of S due to the introduction of a premature stop codon at position 199. The major hydrophilic region encompassing amino acids 101 to 160 of HBsAg and the "a" determinant region within it (aa 122 to 148) were also found to be very conserved in Paharia HBV; only two changes (M125T and Q129H) were noted in this region, in the case of sample T1245. All tribal HBV strains belonged to subtype *ayw3*, except for T1245, which was subtyped as *ayw2*.

(ii) Pre-C/C ORF. The basal core promoter (BCP) region (nucleotides [nt] 1744 to 1804) was found to be highly conserved in 12 of 13 tribal HBV strains, with no mutations at nucleotide positions 1753, 1762, and 1764, which are often responsible for decreased precore mRNA synthesis (2). The G1896A mutation associated with HBeAg-negative hepatitis was not detected in any of the 13 sequences. In one of the clones (clone a) of sample T1360, a deletion was observed between nt 1753 and 1771 that included the TA repeat region of BCP along with a 26-nt deletion in the core gene that disrupted the reading frame and introduced a premature stop codon at position 24. In a clone (clone b) of T1434, two consecutive point mutations, T1972A and T1973A, generated a stop codon, resulting in translational termination of the core protein. However, in both T1360 and T1434, apart from the above mutant clones, HBV strains with wild-type core sequences were also identified. In contrast, for sample T1503, all of the HBV clones examined in the study as well as direct sequencing of the PCR-amplified full-length HBV genome showed the presence of a core internal deletion (CID), with a small 11-amino-acid (aa 83 to 93) region of the core protein deleted in frame.

(iii) Pol ORF. In all of the tribal HBV strains, the reverse transcriptase domain of the polymerase (Pol) ORF was found to be highly conserved, and no mutation that could affect the susceptibility of HBV to antiviral nucleotide/nucleoside analogs was detected. However, a single amino acid change in the RNase H domain, namely, Leu^{791} to Ile, was found in 7 of 13 sequences. Since the pre-S/S ORF completely overlaps the Pol gene, the 61-amino-acid deletion observed in the pre-S1 region of clone b of T1434 theoretically corresponds to a deletion in the nonessential spacer region of the Pol ORF. However, the polymerase gene of this particular clone had a premature stop codon in the region encoding the amino terminus of the terminal protein (TP) domain that would terminate the translation of Pol after only 95 amino acids.

(iv) X ORF. Among 13 tribal HBV sequences, 12 had an intact X ORF; only HBX of clone a of sample T1360 was found to be truncated at the C terminus, by 27 amino acids. The Ser^{31} and Ser^{38} residues, implicated previously (32) as risk factors

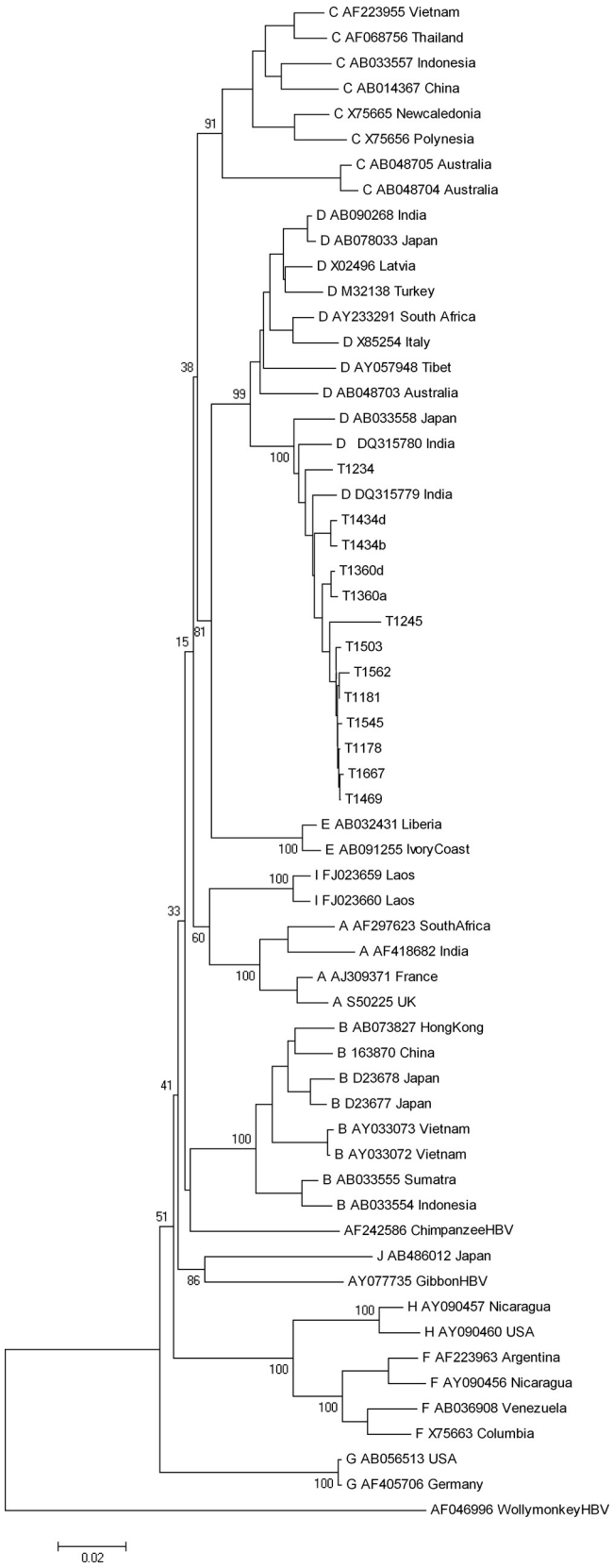

FIG. 1. Phylogenetic tree analysis of full-length sequences of 13 Paharia HBV isolates, along with reference sequences of HBV strains belonging to different genotypes (A to J) derived from GenBank, including 3 sequences of HBV strains from nonhuman primates. HBV sequences from GenBank are indicated by their genotypes, followed by accession numbers and countries of origin, while the sequences determined in this study are indicated by isolate numbers starting with "T." Bootstrap values are given at the internal nodes.

for the development of hepatocellular carcinoma, were found in all HBX sequences, along with the His^{94} and Ser^{101} residues, which are associated with growth-suppressive and apoptotic effects (14).

Signature residues in HBV genotype D5. Since only three sequences of the HBV D5 subgenotype were available in GenBank (GenBank accession no. DQ315779, DQ315780, and AB033558) prior to this study, we compared the deduced amino acid sequences of all 13 tribal D5 isolates with the existing D5 sequences as well as with residues at corresponding positions for the other D subgenotypes (D1 to D7) in order to identify the signature motif residues of subgenotype D5. As shown in Table 4, 27 specific amino acid residues were identified that were present in all 16 HBV D5 strains analyzed in the study and hence could be considered the signature motif residues of the HBV D5 subgenotype. Twenty-one of these residues are found in the Pol ORF, and three each are found in the pre-S/S and X gene products. Of these 27 residues, 6 residues (V^{30} and S^{36} in the X ORF, R^{28} and A^{85} in the pre-S region, and L^{269} and M^{336} in the reverse transcriptase domain of the HBV polymerase) are unique to D5, i.e., are not shared or exhibited by the other D subgenotypes (D1 to D7).

DISCUSSION

Since the evolution of our pathogens is intimately intertwined with human history, it is expected that both microbe and human subpopulations exhibit distinctive genetic features, likely caused by population isolation and genetic drift, and the study of microbial genetic markers might represent one of the most promising alternatives to classical human genetic markers for analyzing ancestry, migratory episodes, genetic fluxes, and disease dissemination. The tribal people of India are considered the original inhabitants of the land, and a recent genetic study suggested that the present caste system grew directly out of tribe-like organizations during the formation of Indian society (24).

Two separate studies have been carried out to assess the prevalence and genotypes of HBV among the primitive tribes of the Andaman and Nicobar islands at the southern tip of the country (1) and among the Idu Mishmi tribe of northeastern India (5). Providing important insights into HBV genome divergence in India, these initial studies motivated further investigations of the molecular epidemiology of HBV among tribal communities in other parts of the country. The Paharia tribes of Eastern India have remained quite separate from the mainstream population in terms of language, culture, ethnicity, and ancestry.

The prevalence of HBV infection was found to be low (2%) among the Paharia tribal community, in sharp contrast to the alarming rates of HBV infection among other tribes of India, particularly the Jarwas (60%), Shompens (37.8%), Onges

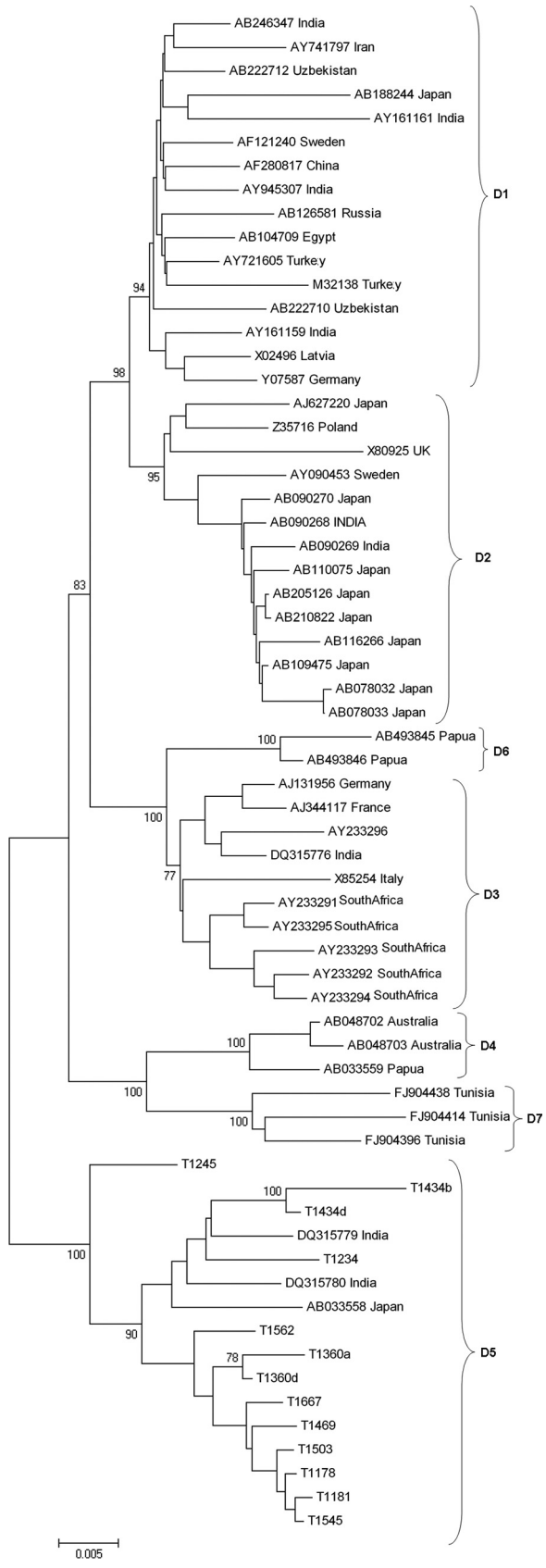

FIG. 2. Phylogenetic tree showing the subgenotypic distribution of HBV strains belonging to genotype D. The tree was formed by 52 full-length reference sequences of HBV strains belonging to different subgenotypes of D, namely, D1 to D7, as indicated by their accession numbers followed by countries of origin. Thirteen Paharia HBV sequences determined in the study are indicated by respective isolate numbers beginning with "T."

TABLE 3. Mean intersubgenotypic divergence among the subgenotypes of HBV genotype D

Subgenotype	Intersubgenotype divergence (%)[a]					
	D1	D2	D3	D4	D5	D6
D1						
D2	2.53					
D3	3.19	3.39				
D4	3.88	4.05	3.89			
D5	**4.68**	**4.82**	**4.85**	**4.81**		
D6	3.55	3.73	2.5	4.19	**5.16**	
D7	4.25	4.61	4.52	3.51	**5.38**	4.94

[a] The genetic distances of D5 from the other subgenotypes of D are indicated in bold. There was no divergence for the D7 column.

(31.0%), and Nicobarese (23.3%) of the Andaman and Nicobar islands (5, 19, 20) and the Idu Mishmi tribe (21.2%) in northeastern India (5). However, this prevalence rate is comparable to the average carrier rate (2.97%) reported in our previous community-based epidemiological study of HBV infection in Eastern India (8). Most of the Paharia individuals with HBV infection had ALT levels that were close to normal, despite high viral loads in some cases (Table 2), invoking host or environmental factors specific to the tribal people that might be more important as determinants of disease outcome. However, the AST values among Paharias (including those negative for HBsAg) were always higher than ALT values, and this may be correlated with the high level of consumption of alcohol (homemade country liquor) as part of their daily food habit. The presence of other significant medical/surgical illness in the HBsAg-positive Paharias participating in the study was adequately excluded by clinical examination and ultrasound screening.

The present study documented that overall, the level of sequence heterogeneity of HBV strains was much lower in the Paharia community, which can be attributed to the relative segregation and expected uniformity of ethnicity of this particular tribe. Although HBV clones with deletions/mutations in the pre-S1/S and core genes and having defective polymerase were identified in different serum samples, in most cases they were found to coexist with wild-type HBV clones that might help in efficient propagation of the defective genomes through *trans*-complementation.

The predominance of HBeAg-negative infection (73%) and the complete absence of mutations, particularly those in the BCP regions and PC stop codon, characterize this clinically asymptomatic HBV-infected tribal community. Since we analyzed only three clones per sample, we cannot exclude the possibility that viral subpopulations with PC and BCP mutations may also be present in the HBeAg-negative sera studied. But we also directly sequenced the PCR products containing the full-length HBV genomes from these samples, and the

FIG. 3. Estimation of divergence times of subgenotypes D1 to D5 of HBV. The tree was constructed by the DNAMLK program of the PHYLIP package, using 64 indel-free HBV sequences of different subgenotypes of D, derived from GenBank, along with 13 Paharia HBV sequences. Reference sequences are indicated by their specific subgenotypes followed by accession numbers, and the tribal sequences are indicated by respective isolate numbers beginning with "T."

resulting data, which certainly represent the highly predominant species of HBV in the respective samples, indicate the presence of wild-type sequences, with no well-known PC/BCP mutations. It has been shown that the genetic diversity of HBV is highest during periods of high viral replication and a mounting immune response, while the mutation rate appears to be low, with little variability of sequences over time, in occult HBV infection, in which the host immune response is relatively quiescent. The majority of the HBeAg-negative subjects in the present study had a biochemically inactive infection. It is possible that the relatively quiescent and inactive virological status and stable immunological status of HBV in the majority of the HBV-infected Paharias we studied negatively influenced the emergence and evolution of PC/BCP mutants as the predominant species over the wild-type sequences. Naturally occurring CID mutants of HBV, with an in-frame deletion of 11 amino

TABLE 4. Conserved amino acid residues in four ORFs of Paharia HBV isolates compared with the previously reported residues at the same positions in different subgenotypes of D (D1 to D7)

Amino acid position	Amino acid in reported subgenotype[a]							Amino acid in sequences of tribal D5 isolates[a]	% of tribal isolates with major aa[c]
	D1	D2	D3	D4	D5	D6	D7		
X gene									
x30	F	F	F	F	V	F	F	**V**	100
x36	T	T	T	A	S	T	D	**S**	100
x40	P	P	P/S	P/S	S	P	S	**S**	100
x87	H/Q	H	R	Q	Q		Q	**Q**/I	81.28
x88	F	F	M	I/F	F/I	Q	S/I	**I**/F	84.61
x92	V	V	V	V	I	V	V	**I**/L	92.3
x102	V	V	T/A	A	T/A	A	A/T	**T**/A	84.61
x110	A	A	A	A	Q/A	A	A	**Q**/P	92.3
Pre-C/C gene									
c5	H	H	H	H	H	H	H	**D**/H	54.55
c11	S	S	S	S	F	S	S	**F**/S	72.73
c64	S	S	S	A	A	S	S	**A**/S	54.55
c69	E	E	E	D	D	E	D	**D**/E	63.64
c103	G	V	V	T	T	V	V/G/A	**T**/G/V/I	63.64
c145	I	I	I	L	L	I	L/V	**L**/I	81.82
c209	E	E	E	E	E	E	G	**D**/**E**	90.91
Pre-S/S gene									
preS28	A	A	A	N	R	A	N	**R**	100
preS85	S	S	S	S	A	S	S	**A**	100
preS103	N	N/D	N	T	D	N	D/N	**D**/V	91.67
preS112	N	N	N	N	T	N	N	**T**/N	91.67
preS118	T	T	T	T	A	T	T/N	**A**/T	91.67
preS147	V	A	A	A	A	A	A/V	A/V	91.67
preS151	S	S	S	S	L	S	S	**L**/S	91.67
preS156	R	R	R	R	K	R	R	**K**/R	91.67
preS162	L	L	L	L	P	L	L/V	**P**/L	91.67
s125	T	T	M	T	M	T	T	**M**/T	91.67
s127	P	T	T	P	T	P	P	**T**/P	91.67
s143	L	S	L/S	S	S	S	S	S	100
Polymerase gene[b]									
tp45	N	N	N	N	N	N	D/N	**T**/N	92.3
tp78	S	S	S	S	T	S	S	**S**/T	61.53
tp90	K	K	K	K	N	K	K	**N**/K	84.61
tp119	V	D/V	V	F	F/V	V	V/F/C	**F**/V/A	84.61
tp161	E	E	E	V	E	E	E	E	100
tp162	T	T	T	S	T	T	T	T	100
tp164	H	R	H	H	H	H	H	H	100
tp178	E	K	D	E	K	E	E	K	100
sp31	R	S/R	R	Q	Q	R	Q	Q	100
sp58	G	G	G	R	R/G	G	G/R	R	100
sp59	I	I	F	V	V	F	V	**V**/A	83.33
sp72	S	S	S	S	A	S	S	**A**/P	91.67
sp88	F/L	I	L	F/L	S/C	L	L	S	100
sp100	S	A	A	A	T	A	A/P	**T**/S	91.67
sp104	S/F	F	F	S	S	S	F/Y	S	100
sp106	K	K/R	K	N	R	K	R/K	**R**/S	91.67
sp115	E	E	E	E	E/D	E	E	**D**/E	91.67
sp122	N	N	N	N	S	N	N/K	**S**/N	91.67
sp139	Q	Q	Q	Q	Q/N	Q	Q	**N**/Q	83.33
rt7	V/A	A	A	A	T	A	A/G	T	91.67
rt54	Y/H	H/P	Y	Y	Y	Y	Y	**Y**/H/N	58.33
rt122	F	F/V	L	I	F	L	F	F	100
rt123	N	N	N	N	D	N	N	**D**/N	91.67
rt135	S	Y	S	Y	Y	S	S	**Y**/S	91.67
rt149	Q	Q	Q	K	Q	Q	K	Q	100
rt248	H	N	N	N	H	H	H	H	100
rt257	Y	Y	Y	W	W	Y	H	**W**/Y	91.67
rt263	D/E	D	E	N/D	E/D	E	D	E	100
rt267	Q	Q	Q	H	H	Q	H	H	100
rt269	I	I	I	I	L	I	I	L	100

Continued on following page

TABLE 4—*Continued*

Amino acid position	Amino acid in reported subgenotype[a]							Amino acid in sequences of tribal D5 isolates[a]	% of tribal isolates with major aa[c]
	D1	D2	D3	D4	D5	D6	D7		
rt278	V	V	I	V	V	V	V	V	100
rt279	N	N	N	H	N	N	N	N	100
rt317	S	S	S	S	A/S	S	S	**A**/S	91.67
rt336	L	L	L	L	M	L	L	M	100
rh2	P	P	P	P	S/T	P	P/S	T	100
rh30	L	L	S	L	Q	M	P/R/Q	**Q**/P	91.67
rh55	L	L	I	L	I/L	I	L/I	I	100
rh107	L	L	L	P	L	L	I	L	100
rh108	S	Y	S	C	C	S	Y/S	C	100

[a] D5 subgenotype-specific residues are shaded.
[b] The conserved domains of the HBV polymerase open reading frame are indicated as follows: tp, terminal protein domain; sp, spacer region; rt, reverse transcriptase domain; rh, RNase H region.
[c] The major amino acid is indicated in bold.

acids in the core antigen, were identified in an HBeAg-negative tribal participant with a high viral load. This internal deletion in CID also coincides with a potent major histocompatibility complex (MHC) class I- and class II-restricted T-cell epitope region of core protein (amino acids 80 to 120), which raises the possibility that these variants can escape immune surveillance. It has been reported that CID mutants are replication defective but can be rescued by wild-type helper virus. However, they interfere with the replication of wild-type HBV and also have the ability to enrich their proportion in the total viral yield in human hepatoma cell lines infected with wild-type and CID variant DNAs (33).

The surface protein was found to be highly conserved in almost all of the tribal HBV isolates. Of the two point mutations found in the "a" determinant region of HBsAg in one of the samples, the Q129H mutation is a well-known vaccine escape mutation, while the immunological significance of the M125T mutation needs to be elucidated further. Since there was no history of HBV vaccination among the Paharia tribe, the escape mutants might have arisen spontaneously in the host under selection forces other than a vaccine-induced immune response, and this finding seems to be epidemiologically relevant and should be considered in designing future prevention strategies. As expected, none of the tribal HBV strains were found to harbor mutations in the polymerase gene that conferred resistance to commonly used antiviral drugs, since the use of such drugs is unknown among the tribal people.

Although HBV genotypes A, D, and C could be found in comparable proportions in the mainstream population of Eastern India (3), the uniqueness of Paharia HBV strains is the exclusive presence of the D genotype and the D5 subgenotype. The geographical distribution of the subgenotypes within genotype D (D1 to D7) was found to be less restricted than that of other HBV genotypes, although strains from the Middle East belong mainly to D1 (18), those from Russia and the Baltic region belong to D2 (28), strains from Serbia, South Africa, and Alaska belong to D3 (15), and those from Oceania and Somalia are specific for D4 (21). HBV D5 was reported from Japan and Eastern India (3), whereas D6 is most prevalent in Papua, Indonesia (16), and D7 is most prevalent in the Maghreb (Tunisia, Algeria, and Morocco) (17). In Eastern India, among the D genotype strains, D3 is the most prevalent (34%), followed by D2 (29%), D5 (20%), and D1 (17%) (6). However, the present study revealed that the HBV epidemiological profile in the Paharia tribal community is quite distinct, with an absolute predominance of HBV D5. Interestingly, by analyzing viral sequences of the different subgenotypes of D, namely, D1 to D5, which are prevalent in India and calibrating a molecular clock based on observed nucleotide changes, it was inferred that D5 was the first to diverge from the other subgenotypes (D1 to D4); in other words, D5 could be considered the most ancient of the five D subgenotypes. To examine whether other Indian tribes carried HBV D5, we subgenotyped the HBV D strains of Andamanese, Nicobarese, Onges, and Idu Mishmi peoples, based on 65 partial sequences of the surface genes available in GenBank (data not shown). However, the sequences belonged to D1 (36%), D2 (38%), or D3 (17%), and none grouped with D5.

Prior to this study, only three full-length HBV D5 sequences, two from Eastern India (GenBank accession no. DQ315779 and DQ315780) and one from Japan (GenBank accession no. AB033558) could be found in GenBank, and that posed difficulties in ascertaining the unique signature motifs of D5. By compiling the 13 tribal HBV D5 amino acid sequences with the previous 3 and comparing them with representative sequences of the D1 to D4, D6, and D7 subgenotypes, we were able to identify the signature substitutions for all the subgenotypes (D1 to D7), particularly D5. This will mostly enable the classification of HBV strains into subgenotypes by limited sequencing.

Given the exclusive presence of HBV subgenotype D5 among this tribal population and that D5 diverged much earlier than other D subgenotypes prevalent in India, it is tempting to speculate that D5 may be endogenous to the Paharia tribe, reflecting an old association and intimate evolution of this viral subgenotype within this relatively isolated primitive community and the consequent spread of D5 to the mainstream population. However, there is a possibility that the presence of D5 among the Paharias may correspond to a recent introduction of the virus from the mainstream population and its transmission within this close-knit community. Independent of how HBV D5 became associated with the tribal people, the results of this study highlight the importance of analyzing HBV strains from relatively understudied geo-

graphic regions and human ethnic groups, as such geographic genomics will increase our understanding of host-virus interactions in colonization and disease and provide new insights into the evolution of this diverse and globally distributed human pathogen as well as the history and demography of its host.

ACKNOWLEDGMENTS

We thank the Secretary DHR and Director General, Indian Council of Medical Research, for funding this study through project Tribal/22/2005-ECD-II.

We have no conflicts of interest in connection with the submitted report.

This research work is part of a multicenter project on the epidemiology of viral hepatitis in tribal populations in India, coordinated by Deepak A. Gadkari, Vidya Arankalle from the National Institute of Virology Pune, and Rashmi Arora from the Division of Epidemiology and Communicable Diseases, ICMR.

REFERENCES

1. **Arankalle, V. A., K. M. Murhekar, S. S. Gandhe, M. V. Murhekar, A. Y. Ramdasi, V. S. Padbidri, and S. C. Sehgal.** 2003. Hepatitis B virus: predominance of genotype D in primitive tribes of the Andaman and Nicobar islands, India (1989–1999). J. Gen. Virol. **84:**1915–1920.
2. **Banerjee, A., S. Banerjee, A. Chowdhury, A. Santra, S. Chowdhury, S. Roychowdhury, C. K. Panda, S. K. Bhattacharya, and R. Chakravarty.** 2005. Nucleic acid sequence analysis of basal core promoter/precore/core region of hepatitis B virus isolated from chronic carriers of the virus from Kolkata, Eastern India: low frequency of mutation in the precore region. Intervirology **48:**389–399.
3. **Banerjee, A., F. Kurbanov, S. Datta, P. K. Chandra, Y. Tanaka, M. Mizokami, and R. Chakravarty.** 2006. Phylogenetic relatedness and genetic diversity of hepatitis B virus isolates in Eastern India. J. Med. Virol. **78:**1164–1174.
4. **Basu, A., N. Mukherjee, S. Roy, S. Sengupta, S. Banerjee, M. Chakraborty, B. Dey, M. Roy, B. Roy, N. P. Bhattacharyya, S. Roychoudhury, and P. P. Majumder.** 2003. Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res. **13:**2277–2290.
5. **Biswas, D., B. J. Borkakoty, J. Mahanta, L. Jampa, and L. C. Deouri.** 2007. Hyperendemic foci of hepatitis B infection in Arunachal Pradesh, India. J. Assoc. Physicians India **55:**701–704.
6. **Chandra, P. K., A. Biswas, S. Datta, A. Banerjee, R. Panigrahi, S. Chakrabarti, B. K. De, and R. Chakravarty.** 2009. Subgenotypes of hepatitis B virus genotype D (D1, D2, D3 and D5) in India: differential pattern of mutations, liver injury and occult HBV infection. J. Viral Hepat. **16:**749–756.
7. **Chisari, F. V., and C. Ferrari.** 1995. Hepatitis B virus pathogenesis. Annu. Rev. Immunol. **13:**29–60.
8. **Chowdhury, A., A. Santra, R. Chakravorty, A. Banerji, S. Pal, G. K. Dhali, S. Datta, S. Banerji, B. Manna, S. R. Chowdhury, S. K. Bhattacharya, and D. G. Mazumder.** 2005. Community-based epidemiology of hepatitis B virus infection in West Bengal, India: prevalence of hepatitis B e antigen-negative infection and associated viral variants. J. Gastroenterol. Hepatol. **20:**1712–1720.
9. **Datta, S.** 2008. An overview of molecular epidemiology of hepatitis B virus (HBV) in India. Virol. J. **5:**156.
10. **Garson, J. A., P. R. Grant, U. Ayliffe, R. B. Ferns, and R. S. Tedder.** 2005. Real-time PCR quantitation of hepatitis B virus DNA using automated sample preparation and murine cytomegalovirus internal control. J. Virol. Methods **126:**207–213.
11. **Günther, S., B. C. Li, S. Miska, D. H. Krüger, H. Meisel, and H. Will.** 1995. A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. J. Virol. **69:**5437–5444.
12. **Hollinger, F. B.** 2007. Hepatitis B virus genetic diversity and its impact on diagnostic assays. J. Viral Hepat. **14**(Suppl. 1)**:**11–15.
13. **Kumari, S.** 2005. Sauria Paharia economy in time perspective—a case of Sahebganj district, p. 69–84. *In* S. K. Chaudhari and S. Sen Chaudhari (ed.), Primitive tribes in contemporary India—concept, ethnography and demography. Mittal Publications, New Delhi, India.
14. **Kwun, H. J., and K. L. Jang.** 2004. Natural variants of hepatitis B virus X protein have differential effects on the expression of cyclin-dependent kinase inhibitor p21 gene. Nucleic Acids Res. **32:**2202–2213.
15. **Lazarevic, I., M. Cupic, D. Delic, N. S. Svirtlih, J. Simonovic, and T. Jovanovic.** 2007. Distribution of HBV genotypes, subgenotypes and HBsAg subtypes among chronically infected patients in Serbia. Arch. Virol. **152:**2017–2025.
16. **Lusida, M. I., V. E. Nugrahaputra, Soetjipto, R. Handajani, M. Naganno-Fujii, M. Sasayama, T. Utsumi, and H. Hotta.** 2008. Novel subgenotypes of hepatitis B virus genotypes C and D in Papua, Indonesia. J. Clin. Microbiol. **46:**2160–2166.
17. **Meldal, B. H., N. M. Moula, I. H. Barnes, K. Boukef, and J. P. Allain.** 2009. A novel hepatitis B virus subgenotype, D7, in Tunisian blood donors. J. Gen. Virol. **90:**1622–1628.
18. **Mohebbi, S. R., S. Amini-Bavil-Olyaee, N. Zali, B. Noorinayer, F. Derakhshan, M. Chiani, M. Rostami Nejad, M. H. Antikchi, F. Sabahi, and M. R. Zali.** 2008. Molecular epidemiology of hepatitis B virus in Iran. Clin. Microbiol. Infect. **14:**858–866.
19. **Murhekar, M. V., K. M. Murhekar, D. Das, V. A. Arankalle, and S. C. Sehgal.** 2000. Prevalence of hepatitis B infection among the primitive tribes of Andaman and Nicobar Islands. Indian J. Med. Res. **111:**199–203.
20. **Murhekar, M. V., K. M. Murhekar, and S. C. Sehgal.** 2003. Alarming prevalence of hepatitis-B infection among the Jarawas—a primitive Negrito tribe of Andaman and Nicobar Islands, India. J. Viral Hepat. **10:**232–233.
21. **Norder, H., A. M. Couroucé, 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.
22. **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.
23. **Palumbo, E., G. Scotto, D. C. Cibelli, G. Faleo, A. Saracin, and G. Angarano.** 2008. Immigration and hepatitis B virus: epidemiological, clinical and therapeutic aspects. East. Mediterr. Health J. **14:**784–790.
24. **Reich, D., K. Thangaraj, N. Patterson, A. L. Price, and L. Singh.** 2009. Reconstructing Indian population history. Nature **461:**489–494.
25. **Risley, H. H.** 1891. The tribes and castes of Bengal, vol. 2. Bengal Secretariat Press, Calcutta, India.
26. **Roychoudhury, S., S. Roy, A. Basu, R. Banerjee, H. Vishwanathan, M. V. Usha Rani, S. K. Sil, M. Mitra, and P. P. Majumder.** 2001. Genomic structures and population histories of linguistically distinct tribal groups of India. Hum. Genet. **109:**339–350.
27. **Simmonds, P., and S. Midgley.** 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. J. Virol. **79:**15467–15476.
28. **Tallo, T., V. Tefanova, L. Priimägi, J. Schmidt, O. Katargina, M. Michailov, S. Mukomolov, L. Magnius, and H. Norder.** 2008. D2: major subgenotype of hepatitis B virus in Russia and the Baltic region. J. Gen. Virol. **89:**1829–1839.
29. **Toro, C., V. Jiménez, C. Rodríguez, J. Del Romero, B. Rodés, A. Holguín, P. Alvarez, M. García-Campello, C. Gómez-Hernando, A. Guelar, J. Sheldon, C. de Mendoza, A. Simón, and V. Soriano.** 2006. Molecular and epidemiological characteristics of blood-borne virus infections among recent immigrants in Spain. J. Med. Virol. **78:**1599–1608.
30. **Wang, F. S., and Z. Zhang.** 2009. Host immunity influences disease progression and antiviral efficacy in humans infected with hepatitis B virus. Expert Rev. Gastroenterol. Hepatol. **3:**499–512.
31. **Webster, G. J., and A. Bertoletti.** 2002. Control or persistence of hepatitis B virus: the critical role of initial host-virus interactions. Immunol. Cell Biol. **80:**101–105.
32. **Yeh, C. T., C. H. Shen, D. I. Tai, C. M. Chu, and Y. F. Liaw.** 2000. Identification and characterization of a prevalent hepatitis B virus X protein mutant in Taiwanese patients with hepatocellular carcinoma. Oncogene **19:**5213–5220.
33. **Yuan, T. T., M. H. Lin, D. S. Chen, and C. Shih.** 1998. A defective interference-like phenomenon of human hepatitis B virus in chronic carriers. J. Virol. **72:**578–584.