Hepatitis B virus: predominance of genotype D in primitive tribes of the Andaman and Nicobar islands, India (1989–1999)


To understand the possible origin of hepatitis B virus (HBV), three of the four hyperendemic, primitive accessible tribes of the Andaman and Nicobar islands, India, were investigated. The Nicobarese tribe was investigated in 1989 and 1999. The S gene from 65 HBV isolates was amplified by PCR and sequenced. Genotyping and serotyping were carried out on the basis of phylogenetic and amino acid analyses of S gene. All 20 Nicobarese-89 isolates, nine Onges-99 isolates and the single Andamanese-99 HBV isolate were classified as genotype D. Of the Nicobarese-99 isolates, 32 (91.4 %) and three (8.6 %) were genotypes D and A, respectively. Percent nucleotide identity between the S sequences representing different tribes varied from 98.06 to 98.59 % and varied from mainland isolates by 1.6–2.0 %. Although southeast Asian origin is postulated for the Nicobarese tribe, the presence of different genotypes suggests introduction of HBV after migration to these islands, probably from mainland India, 200 years back, when these islands became inhabited as a part of penal settlement during the British regimen.

Hepatitis B virus (HBV) infection is an important public health problem, with in excess of 350 million individuals estimated to be chronically infected worldwide. HBV is an aetiological agent of acute and chronic liver disease, including fatal fulminant hepatitis, cirrhosis and hepatocellular carcinoma, which is one of the most common human cancers and causes of death worldwide (Lee, 1997). HBV isolates have been classified into eight genotypes, A–H, and four major serotypes, ayw, ayr, adw and adr, on the basis of complete genome and S gene sequence analysis (Okamoto et al., 1988; Orito et al., 1989; Norder et al., 1992, 1994; Stuyver et al., 2000; Arauz-Ruiz et al., 2002) and reactivity of the HBV small antigen, HBsAg (Le Bouvier, 1971; Bancroft et al., 1972). Genotypes and serotypes are useful tools in understanding the epidemiology of HBV infection. Although both genotypes and serotypes segregate geographically, the same serotype can be represented by several genotypes. Recent documentation of the presence of a unique genotype C variant expressing amino acid residues characteristic of serotype ayw3 in Australian aborigines presents an important example of HBV evolution in such populations (Sugauchi et al., 2001).

The Andaman and Nicobar islands, situated in the Bay of Bengal, Union Territory of India, are the home of six primitive tribes, namely the Great Andamanese, Onges, Jarawas and Sentinelese, belonging to the Negrito race, and the Nicobarese and Shompens, belonging to the Mongolid race. The Jarawas and Sentinelese are still hostile to the outside world. The population of these tribes, except that of the Nicobarese, has declined continuously over the past several years. The total population of these tribes is 26 852; the Nicobarese tribe constitutes over 98 % of the tribal population. The population of the other tribes ranges from 37 to 275. Serological studies carried out earlier among the four accessible tribes of these islands showed that hepatitis B infection is hyperendemic, with HBsAg carrier rates ranging from 23.3 (Nicobarese) to 37.8 % (Shompens) (Murhekar et al., 2000), probably the highest reported rates in India.

The tribes of two different races in these islands indicate their different origin. No information is available regarding HBV genotype distribution among the tribes of these islands. The present study examines S gene sequences from three different tribes. Nicobarese HBV isolates recovered 10 years earlier (1989 and 1999) were also analysed.

Study subjects included individuals identified incidentally as being HBsAg-positive during the serological surveys conducted (1) by the Regional Medical Research Centre, Port Blair, in 1999, as part of tribal healthcare activities (Murhekar et al., 2000) and (2) by the National Institute of Virology, Pune, in 1989, as part of the virological survey of these islands. HBsAg-positive individuals were tested further for the presence of HBV DNA, according to the protocol and primers described previously (Gandhe et al., 2003). HBV DNA-positive samples from the Nicobarese
tribe bled in 1999 \( (n=35) \) and 1989 \( (n=20) \), as well as from the Onges \( (n=9) \) and Andamanese \( (n=1) \) tribes collected in 1999, were analysed during the present study. DNA isolation was carried out using DNAZOL (Gibco-BRL), according to the manufacturer’s instructions.

Part of the S gene was amplified by nested PCR. Of the DNA extracted, 10 \( \mu l \) was mixed with 90 \( \mu l \) PCR buffer \( (10 \text{ mM Tris}/\text{HCl} \, l-1, \text{pH} \, 8.3, 1.5 \text{ mM MgCl}_2 \, l^{-1}, 50 \text{ mM KCl} \, l^{-1}, 0.01 \% \text{ (wt/vol)} \text{ gelatin}, 1 \mu \text{M each primer} \, l^{-1} \) and 200 \mu M each of the four dNTPs \( l^{-1} \). The reaction mixture was then overlaid with 50 \( \mu l \) mineral oil and subjected to 30 cycles of amplification: denaturation at 94 \degree C for 1 min, annealing at 55 \degree C for 1 min and extension at 72 \degree C for 1 min. Primers used for first-round PCR were 5’-ACCCCTGCTGTTACAGGC-3’ (sense, nt 184–204) and 5’-AAAGCCAGACAGTGGGGAAA-3’ (antisense, nt 731–711). For second-round PCR, 1 \( \mu l \) of the first-round PCR product with primers 5’-GACTCGTGGTGGACTTCTC-3’ (sense, nt 251–271) and 5’-TAAACTGAGCCAGGAAACG-3’ (antisense, nt 679–659) (concentration of primers and PCR reactants was identical to those used for first-round PCR) was subjected to 25 cycles of amplification using the first-round PCR protocol to obtain a 429 bp product. PCR products were purified using the Wizard PCR Preps DNA Purification kit (Promega) and sequenced using the Big Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) and an automatic Sequencer (ABI PRISM 310 Genetic Analyser, Applied Biosystems).

Phylogenetic analysis of a 363 nt fragment of the S gene employing MEGA, version 2.0 (Kumar et al., 1993), and PHYLIP, version 3.54c (Felsenstein, 1993), was the basis for HBV genotyping. For analysis in MEGA and PHYLIP, Jukes–Cantor and Kimura’s 2-parameter algorithms were utilized with the neighbour-joining method. The reliability of different phylogenetic groupings was evaluated using the bootstrap test (1000 bootstrap replications). As almost identical groupings were observed with these tree-building programs, only the MEGA-based (Jukes–Cantor) tree is presented. For assessing the utility of the 363 nt fragment for accurate classification of HBV genotypes, the complete S gene (680 nt) was amplified for four samples (two each from mainland India and islands).

Classification of HBV isolates into different serotypes and subdeterminants on the basis of S gene sequences was performed as described by Magnius & Norder (1995). Three strains suspected to be mixtures of different serotypes on the basis of electrophoretograms were cloned using a TA cloning kit (pGEM-T Easy Vector System, Promega) and sequenced.

During this study, partial sequence data for the S gene, representing HBV isolates from the Nicobarese tribe collected in 1999 \( (n=35) \) and 1989 \( (n=20) \), as well as from the Onges \( (n=9) \) and Andamanese \( (n=1) \) tribes, were generated. These sequences represent the first genetic analysis of HBV from these tribes (accession nos AY155369–AY155437).

The dendrogram in Fig. 1 shows that based on phylogenetic analysis of the 363 nt fragment of the S gene, all eight genotypes were differentiated distinctly. To ascertain the suitability of this region in the genotypic analysis, the complete S gene (680 nt) was analysed in a similar manner for four HBV isolates (data not shown). Both 363 and 680 nt-based analyses yielded identical results, confirming the utility of the smaller fragment for genotyping.

**Fig. 1.** Phylogenetic analysis of the partial S gene (363 nt fragment) sequences of 65 HBV isolates from the Andaman and Nicobar islands, India. Figures in parentheses represent the total number of isolates at the same position. Accession numbers for the prototype strains used for the analysis are genotypes: A, AB014370; B, AB014366; C, AB014371; D, AF121242; E, X75664; F, X75663; G, AF160501; and H, AY090454. Accession numbers for the HBV isolates used for the comparison from the mainland are ASC (asymptomatic carrier) 1027, AF492213; ASC 1062, AF492243; ASC 1274, AF492233; CLD (chronic liver disease) 206, AF492255; and CLD 1110, AF492269. Accession numbers for isolates from the Andaman and Nicobar islands are from AY155369 to AY155437.
As seen in Fig. 1, all nine HBV isolates representing the Onges tribe and the single HBV isolate from the Andamanese tribe were classified as genotype D. Of the 35 HBV isolates from the Nicobarese tribe collected in 1999, 32 (91.4%) and 3 (8.6%) were genotypes D and A, respectively. All 20 Nicobarese HBV isolates from 1989 belonged to genotype D. Thus, genotype D is the predominant genotype circulating in the tribal population of these islands. Within genotype D, no definite clusters were noted for HBV isolates from different tribes. HBV isolates representative of both tribal and mainland populations grouped together.

Per cent nucleotide identity (PNI) between the S gene sequences of HBV isolates (genotype D) from the Nicobarese tribe bled in 1989 and 1999 was 98.29%. These isolates differed from mainland Indian HBV genotype D isolates (Gandhe et al., 2003) by 2-0 and 1-6%, respectively. The PNI for Onges isolates with Nicobarese-99, Nicobarese-99 and mainland isolates were 98.45, 98.59 and 98.28%, respectively. The only isolate from the Andamanese tribe was 98.06-98.97% similar to isolates from other tribes. Within the tribes, PNI varied from 99 (Nicobarese-99) to 98.7% (Onges). For genotype A isolates, the PNI between the S sequences of HBV isolates from Nicobarese individuals bled in 1999 and mainland India was 98.16%. Within genotype A HBV isolates (Nicobarese-99), the PNI was 98.53%.

Of the 20 isolates from the Nicobarese tribe (1989), 12 and 8 were classified as serotypes ayw<sup>2</sup> and ayw<sup>3</sup>, respectively. In 1999, the distribution was 29 serotype ayw<sup>3</sup>, two ayw<sup>2</sup>, one mixture of ayw<sup>2</sup> and ayw<sup>3</sup>, two adw<sup>2</sup> and one adw<sup>3</sup>. The only isolate from the Andamanese tribe was serotype ayw<sup>2</sup>. Among HBV isolates from the Onges tribe, five and two were identified as serotype ayw<sup>2</sup> and ayw<sup>3</sup>, respectively, whereas two isolates were a mixture of serotypes ayw<sup>2</sup> and ayw<sup>3</sup>. Thus, ayw<sup>2</sup>/ayw<sup>3</sup> represents the predominant serotype among the tribes of these islands.

Fig. 2 presents comparisons of amino acid sequences of part of the S protein (aa 4–164) for the HBV isolates from the tribes under investigation and prototype sequences for different genotypes. It is important to note that the A→V change at position 128, present in few isolates from the mainland, was not observed in any of the isolates from these islands.

Earlier studies among the primitive tribes of the Andaman and Nicobar islands have shown high endemicity of HBV infection (Murhekar et al., 2000). Horizontal transmission through close contact with carriers and perinatal routes was identified as an important mode of transmission of HBV in these tribal communities. Besides, use of unsafe injections represents an independent risk factor for acquiring HBV infection in this population (Murhekar et al., 2002). The present study provides partial sequence data with genotypic analysis of 65 HBV isolates from three of the four accessible tribes of these islands. Except for three genotype A isolates recovered from the Nicobarese tribe in 1999, the other 62 isolates representing all three tribes under investigation belonged to genotype D. As evident from the dendrogram, all tribal, as well as mainland Indian, HBV isolates grouped closely together (PNI 98–99%).

To understand the origin of HBV infection in these tribal populations, it is important to know the origin of the tribes of these islands. The tribes in the Andaman islands (Great Andamanese, Onges, Jarawas and Sentinelese) are of the Negrito race, whereas the tribes in the Nicobar group of islands (Nicobarese and Shompens) are Mongoloid, signifying a quite different origin. The Andaman tribes are primarily hunter–gatherers, while the Nicobar tribes are farmers and herdsmen. We have sequenced HBV isolates representing the Negrito (Onges and Andamanese) as well as Mongoloid races (Nicobarese).

The Negrito race represents an ancient component in the prehistoric peopling of Asia by anatomically modern humans. As such, they could go back 60,000 years or more. However, clear archaeological evidence available so far suggests that Negrito settlements in the Andaman islands go back to about 2,000 years (Weber, 2002). As regards the origin of the Negrito tribes, there are two opposing schools of thought: (1) the Negritos were one group living in a large area of tropical Asia tens of thousands of years ago when new and more aggressive immigrants arrived and pushed the ancestral Negritos into the remoter jungle areas. In such a scenario, the Negrito groups in mainland areas would have lost contact with each other gradually, leaving only a few widely separated surviving populations; or (2) all these groups share a common ancestral origin somewhere in southeast Asia or southern China. (Weber, 2002)

It is not known how and when the Nicobar islands were peopled. Nicobarese people are thought to be related to the Malays and Burmese and speak dialects related to Mon–Khmer and the languages spoken in Vietnam, Malaysia and parts of northeast India (Bellwood, 1997). This view is supported further by the result of mitochondrial DNA hypervariable region 1 sequence data, which indicated affinities to populations of mainland southeast Asian Mon–Kmer-speaking populations. (Prasad et al., 2001)

In the light of the possible origin of these tribes, it is interesting to view the genotypic distribution of HBV. The predominant genotypes reported from southeast Asian countries are genotype C from Thailand (Sugauchi et al., 2002; Theamboonlers et al., 1998), genotypes C and B from Indonesia (Sastrosoewignjo et al., 1991) and genotype C from Myanmar (Nakai et al., 2001). In China, C and B were the predominant genotypes (Xia et al., 2001). The fact that genotype D was found to be predominant among the Negrito and Mongoloid tribes of these islands suggests that HBV infection might have been introduced to these tribes after their migration to these islands.

Predominance of genotype D was observed among HBV
Fig. 2. Alignment of aa 44–164 of the partial S protein of HBV genotype D and genotype A isolates belonging to different serotypes from the Andaman and Nicobar islands, representative isolates from mainland India and prototype isolates of all genotypes. Accession numbers for the HBV isolates used for the comparison from the mainland are described in Fig. 1. NC89-11 is representative of 28 other isolates (NC99-2, NC89-16, NC89-18, NC99-10, NC99-11, NC99-12, NC99-13, NC99-14, NC99-15, NC99-17, NC99-20, NC99-21, NC99-25, NC99-26, NC99-3, NC99-30, NC99-31B, NC99-33, NC99-35, NC99-4, NC99-5, NC99-6, NC99-7, NC99-8, NC99-9, ONG1B and ONG3) showing identical amino acid sequences. NC89-12 represents four more isolates (NC99-19, NC99-34, ONG7 and ONG8B) with identical amino acid sequences. NC89-13 represents 13 more isolates (NC89-14, NC89-20, NC89-3, NC99-6, NC99-8, NC99-1, NC99-22, NC99-31A, ONG1A, ONG2, ONG5, ONG8A and ONG9) with identical amino acid sequences. NC89-17 represents one more isolate (NC89-4) with an identical amino acid sequence, as do NC99-27 (NC99-29), NC89-1 (NC99-9) and NC89-7 (NC99-10). Genotype A isolates are indicated with an asterisk.
isolates from mainland India (Gandhe et al., 2003). Thus, populations from mainland western India as well as primitive tribes from the Andaman and Nicobar islands situated in the Bay of Bengal, about 1200 km east of the Indian subcontinent, predominantly circulate genotype D. Therefore, it is logical to consider the introduction of virus from mainland India. The fact that the PNI between mainland and tribal HBV isolates varied from 1·6 to 2·0% shows recent introduction of the virus, probably about 200 years back when these islands became inhabited as part of a penal settlement during the British regime. This might have been followed by extensive spread of the virus on account of the factors described earlier (Murhekar et al., 2002). It is widely accepted that, in contrast to the non-aboriginal population, aborigines have higher rates of HBV infection worldwide (Hart, 1993; Lin et al., 2000).

Detection of a minor proportion of genotype A isolates is also similar to mainland India (Gandhe et al., 2003). It is pertinent to add here that, although individuals from neighbouring countries predominantly circulating genotype C (Xia et al., 2001; Usuda et al., 1999) visit these islands for trade, none of the HBV DNA-positive individuals investigated was infected with genotype C. Thus, transmission from neighbouring countries seems unlikely.

Nucleotide and amino acid sequence analysis of the S region revealed ayw to be the predominant serotype in the tribes of these islands, with 62/65 (95·6%) of the isolates belonging to this category. In western India, serotype ayw is predominant, at least since 1979 (Thyagarajan et al., 1979). In China, a sequence-based analysis was carried out among 280 chronic HBV carriers from 25 counties of four provinces. Serotypes adr and adw2 were the leading serotypes, representing 64·3 and 31·4%, respectively. Serotype adr was encoded completely by genotype C, while the majority of serotype adw2 was encoded by genotype B (Xia et al., 2001). In our series, all genotype D isolates (n = 62) belonged to the ayw serotype, whereas serotype adw was encoded solely by genotype A (n = 3).

In conclusion, the present study reports for the first time the partial genome sequences of HBV strains from three primitive tribes of the Andaman and Nicobar Islands. The virus might have been introduced to these communities after their migration to these islands. The genotype prevalent in mainland India and very minimal difference in the PNI between isolates from tribal populations suggests the possibility of the introduction of HBV from mainland India.

ACKNOWLEDGEMENTS

This study was carried out with financial assistance from the Ministry of Health and Family Welfare, Government of India, New Delhi. The authors gratefully acknowledge the help and co-operation extended by the Director of Health Services, Medical Superintendent, Civil Hospital, Car Nicobar, and the Director of Tribal Welfare, Andaman and Nicobar Administration, Port Blair.

REFERENCES


