Hepatitis B virus genome analysis in patients of hepatocellular carcinoma and asymptomatic carriers from northern, southern and North East India

Manash Pratim Sarma[1,2], Giasuddin Ahmed[1,2], Subhash Medhi[1,3] and Premashis Kar[1,*]

[1]Department of Medicine, Maulana Azad Medical College, University of Delhi, New Delhi 110 002, India
[2]Department of Biotechnology, Gauhati University, Guwahati 781 014, India
[3]Present address: Department of Bioengineering and Technology, Gauhati University, Guwahati 781 014, India

The present study was designed to analyse the whole genome and mutational profile of hepatitis B virus (HBV) isolates in hepatocellular carcinoma (HCC) and asymptomatic carriers from three regions of India. Seventy-five HBV-related HCC and 15 HBV-related asymptomatic carriers were included in the study. HBV DNA was amplified by six sets of walking primers. Amplicons were sequenced commercially, submitted to GenBank translated into amino acid and aligned using BioEdit v7.0.9. Mutations membering 60, 15, 23 and 1 were observed in PC/C, X, P and S genes respectively. Mutations like 10I → L were significantly associated with HCC cases from North East India (NEI) [(P = 0.01; OR = 5.63) versus South India (SI)] and [(P < 0.01; OR=16.63) versus North India (NI)]. Mutations like 41S → T (P < 0.001; OR = 19.01), 92V → G (P < 0.001; OR = 19.01), 96N → T (P < 0.001; OR = 19.01) and 164Q → P (P = 0.0279; OR = 3.085) were significantly associated with HCC cases from NI [vs SI]. Widely reported 28 W → stop mutation was found in a few HCC cases. Also, 132 → stop [(P = 0.004486; OR = 5.479 versus SI) and [(P = 0.004486; OR = 5.479) versus NEI] was interesting. 267I → N and 268D → T were exclusive to HCC from NEI, while 270S → F was exclusive to HCC from NI. Reported drug mutants (80L → I, 236N → T, 169I → T and 181A → V) were observed. The PC/C region was most prone to mutation followed by P, X and S regions. Maximum variation in HBV genome was observed in HCC cases from NI and least in asymptomatic HBV carriers. Novel mutations in surface (132 stop), polymerase (frameshift mutation at 178), core (10I → L, 41S → T, 92V → G, 96N → T and 164Q → P) and X (33P → S) genes need further studies.

Keywords: Genome analysis, hepatitis B virus, hepatocellular carcinoma, mutation.

HEPATOCELLULAR CARCINOMA (HCC) is a major cause of morbidity and mortality in the world[1]. It is the third leading cause of cancer-related death in males and fourth in females, with more than 600,000 deaths attributed per year[2]. In countries like Japan and Egypt, where hepatitis C virus (HCV) infection is endemic, high prevalence of infection is reported among people with HCC. On the other hand, hepatitis B virus (HBV) infection is the major risk factor associated with the development of HCC in regions like China and southern Asia because of the high endemicity of the virus in these places[1]. Many published works are available on HBV mutation from different parts of India, but data from North East India (NEI) are scanty.

HBV gene mutation has been widely studied now. Mutations in the precore (pre-C) gene are more frequent in patients with persistent viremia and end-stage disease[3]. Mutations in the basal core promoter (BCP) have also been associated with fulminant hepatitis and severe disease[4–6]. The HBV X gene shares sequences with both the polymerase and precore genes, and carries several regulatory signals critical to the replicative cycle; its product has a trans-activating function. In the polymerase gene, mutations are more commonly associated with the drug-resistant mutants. Antiviral resistance to lamivudine has been mapped to the YMDD locus in the catalytic or C domain of HBV Pol (ref. 7), whereas resistance to adefovir dipivoxil is found to be associated with mutations in the D and B domains of the enzyme[7,8].

In the past researchers had used overlapping primer sets to amplify the entire HBV genome[9,10]. Few groups carried out the whole-genome analysis in HCC cases infected with HBV genotype C by the walking primer[11,12]. Research groups all over the world mostly follow the protocol proposed by Günther *et al.*[13] which is based on amplification of complete virion-encapsidated HBV genomes followed by cloning and analysis of cloned HBV DNA by PCR and sequencing. This has been the most widely used whole-genome analysis followed by many researchers across the globe[14–16]. Although cloning increases the authenticity of the results, it involves pain, labour and most importantly, it limits the number of cases to be studied to a small number. On the other hand,

*For correspondence. (e-mail: premashishkar@gmail.com)

primer-based approach followed by direct sequencing is a cheap method and can process large number of samples, though it is laborious. Only a few studies on whole genome of HBV have been carried out in India. Recent studies involve genomic analysis of HBV.

The objectives of the present study are: (1) Whole-genome analysis of HBV isolates in HCC and asymptomatic carriers from North India, South and NEI; (2) Analyse the mutational profile of HBV genome and correlate it with the final outcome of the disease.

Materials and methods

Enrollment of cases

The study was designed to include 25 cases of HCC each from the three different regions, namely North India, South India and NEI. With reference to HBV carriers, it was decided to study five cases each of asymptomatic HBV carriers from North India, South India and NEI. The study was designed so as to include cases from three locations, namely Lok Nayak Hospital, New Delhi (for cases from North India); Rajaji Government Hospital, Madurai (for cases from South India) and Gauhati Medical College (for cases from NEI) with different cultural and linguistic ethnicity.

Ethics committee approval

The study was approved by the respective ethical committees of the three hospitals and the study conformed to the ethical committee guidelines of EASL Helsinki 1975. All the study subjects provided written informed consent prior to enrollment.

Diagnostic criteria

The EASL diagnostic criteria for diagnosis of HCC were followed[17]. The asymptomatic HBV carriers were diagnosed based on the AASLD practice guidelines[18].

Initial workout

All the patients were evaluated on the basis of history, clinical examination, liver function profile and various serological markers of hepatitis B (HBsAg, HBeAg, AntiHBcIgG and AntiHBe) using commercially available third-generation ELISA kits. Serological tests for anti-HCV were carried out in all the cases to exclude those infected with HCV. Upper GI endoscopy and ultrasound examination of the patients were done, whenever indicated. Prior information about different risk factors such as intramuscular injections, intravenous medication, blood transfusions, tattooing and any past history of jaundice in the family was collected on a pre-designed questionnaire with respect to the HBV-related HCC cases.

Confirmed cases of HCC and asymptomatic HBV carriers with HBsAg positivity, HBV DNA positivity (DNA extraction by phenol chloroform method as described by Saiki *et al.*[19]) and with high HBV viral load (real-time PCR based on Taqman principle) were included in the study for whole-genome analysis.

PCR standardization

The whole genome of HBV was amplified by six sets of primers which were designed using the BLAST software. The approach adopted to amplify the whole genome was based on walking primer sets that cover the overlapping regions of the HBV genome. The primers were designed to cover the entire genome of HBV and so were overlapping in nature.

Nucleotide sequencing

PCR amplicons for HCC and asymptomatic cases were sequenced commercially from Macrogen, South Korea.

Sequencing results and whole genome analysis

The gene sequences of HBV obtained in the study were submitted to GenBank (http://www.ncbi.nlm.nih.gov). BLAST search was carried out to confirm the identity of the strains. For comparison, other HBV sequences from India and geographical locations across the world from the global database were retrieved. Lasergene 5 software package (DNASTAR Inc, USA) was used to examine the percentage of identity and diversity among sequences. The sequences were then translated into amino acid and aligned using BioEdit v7.0.9. Phylogenetic analysis was carried out using MEGA version 3.1 (ref. 20). Phylogenetic tree was constructed employing neighbour joining method[21]. Fifty-five sequences of precore/core, 61 sequences of polymerase, 21 sequences of X and 22 sequences of surface genes (totally 159 sequences) were submitted under the following BankIt numbers of NCBI GeneBank. BankIt1600217: (55); BankIt1600355: (61); BankIt1600378: (21) and BankIt1600408: (22). Accession numbers for these sequences are awaited.

Also, 40 sequences of precore/core and surface genes of HBV were submitted earlier and the following accession numbers were assigned: BankIt1494013 (JQ038375-JQ038414).

Recombination studies and phylogenetic analysis

Recombination events for sequences with conflicting phylogenetic positions were searched using the Simplot

program and boot scanning analysis[22]. The sequences obtained were compared with those of the eight HBV genotypes (A–H) retrieved from GenBank. Alignments were carried out using CLUSTAL-X. The Kimura two-parameter model integrated into PAUP* v. 4.0b6 software was used to calculate genetic distance and pairwise distance comparisons. Phylogenetic trees were constructed by the neighbour joining method. GenBank accession numbers for the reference sequences used in the phylogenetic analyses were specified.

Clinical outcome

To assess the clinical outcome in HCC cases, staging of HCC was done with the help of radiological findings, wherever available. The Kaplan–Meier curves were constructed for analysing the survival of patients included in the study based on different parameters. Survival curves were also constructed for HCC cases with respect to the mutants in different genes of the HBV genome in all the three regions of the country.

Results

Whole-genome analysis

The 132 stop-codon mutation was observed most frequently in HCC cases, which falls in the Pre S2 (aa119–aa174) region of the HBV surface protein and is involved in HBV attachment and entry into the liver. However, no other mutation was observed in the surface region from any of the HBV isolates in Pre S1 (aa1–aa119), major–surface (aa174–aa400) and inside or outside and MHR. The mutation was more often associated with HCC cases from all the three geographical regions than in the asymptomatic HBV carriers. A majority of the mutations were observed in the Primase (aa1–aa178) and Spacer (aa178–aa304) regions of the gene, while the other two regions of catalytic determinant (A–H; aa336–aa680) and RNase (aa680–aa838) of the gene were not studied in detail. A total of four different drug-resistant mutations were observed in three patients. Mutations were observed in all the regions of the X protein such as regulatory domain (aa1–aa48) which is a strongly conserved region, XAP-binding region (aa61–aa91) and P 53-binding region (aa101–aa153). In the core/precore gene of HBV, mutations were detected in both parts of the gene, namely amino terminal (aa1–aa150) and carboxyl terminal (aa150–aa179) (Table 1).

Mutations documented

Polymerase gene: The mutation patterns observed in the polymerase gene were either substitution or frameshift mutation due to deletion. In the present study, a majority of the mutations in polymerase gene of HBV were documented in HCC cases from North India. A frameshift mutation was seen at codon 178 that continued till 184 was due to deletion (179Q, 180H, 181G, 182R, 183L, 184V). Again mutations like 237S → R, 239T → Q, 240S → G, 242G → W, 247K → G, 251S → T, 255Y → P, 267I → T, 237S → C, 371V → C, 324S → F, 298S → A and 307C → D were associated with few cases of HCC from North India. On the other hand, 267I → N and 268D → T were exclusively found in three HCC cases each from NEI. Also, mutations like 178L → deletion and 270S → F were found in cases from all the three regions, irrespective of their geographical locations. The deletion mutation at codon 178 was observed in all the 25 cases of HCC from North India, while in five cases from South India and one case from NEI. This deletion mutation was significantly associated with the development of HCC North Indian cases when compared to HCC cases from South India (P = <0.0000001; OR = 20.86) and HCC cases from NEI (P = <0.0000001; OR = 62.28). The same mutation was found to increase the risk of HCC sixfold in South Indian cases only when compared against HCC cases from NEI [OR = 5.811]. Already reported drug-resistant mutations were observed in three HCC cases. The drug-resistant mutations observed were 80L → I, 236N → T, 169 I → T and 181A → V. All the drug-related mutants were associated with genotype D. These were expected, as few of the HCC cases were under treatment. The most widely found polymerase gene mutation in HCC cases was a deletion mutation at position 178 that has been previously described (Table 2).

Precore/core gene: A total of 60 different mutations were observed at 38 different amino acid positions in the precore/core region. A majority of the mutations were substitutions, while a few were deletion mutations. Two of the mutations observed were stop-codon mutations. Few mutations were confined with respect to the disease, while few others were specific to geographical locations. Mutation like 10I → L, 28W → stop, 2Q → deletion, 103S → V, 148L → S, 171T → H, 172L → F, 173P → R, 176T → N, 177V → X, 179R → deletion, 1M → P, 41S → T, 92V → G, 96N → T, 103S → G, 109S → I, 164Q → P, 1M → Q, 2Q → X, 109A → I, 164Q → P, 1M → Q, 2Q → X, 109A → I, 159P → T, 164Q → P, 10I → X, 13S → T, 103S → N, 116S → N, 109A → T, 29G → D, 148L → S, 155I → T, 161Y → N, 100W → G, 120V → T, 103S → V, 170S → H, 172L → S, 173P → G, 176V → S, 173P → D and 174E → H were exclusively found in HCC cases and entirely absent in the asymptomatic carriers. Similarly, few mutations of precore/core gene found only in asymptomatic cases such as 93E → D, 109A → V and 159P → Q. Again, some of the mutations were specific to geographical distribution of the HBV genome. Mutations like 120V → T,

Table 1. Whole genome analysis of HBV genome in three different geographical locations from HCC and asymptomatic HBV carriers (significant mutations are tabulated)

Geographical regions/types of patients →	North India (25)					South India (25)				North East India (25)						North India (5)				South India (5)	North East India (5)	
↓Genes and parameters (total number of mutation)	HCC (Numbers inside bracket indicate percentage of cases)															Asymptomatic HBV carriers (Numbers inside bracket indicate percentage of cases)						
Surface gene (1)	132 stop codon (52)					132 stop codon (16)				132 stop codon (16)						132 stop codon (20)				ND	ND	
Pre S1 [aa1–aa119]	ND					ND				ND						ND				ND	ND	
Pre S2 [aa119–aa174]	√ (52)					√ (16)				√ (16)						√ (20)				ND	ND	
Major surface [aa174–aa400]	ND					ND				ND						ND				ND	ND	
Inside a and MHR																						
First loop (aa124–aa137)	ND					ND				ND						ND				ND	ND	
Second loop (aa139–aa147)	√ (52)					√ (16)				√ (16)						√ (20)				ND	ND	
Outside a and MHR	ND					ND				ND						ND				ND	ND	
X gene (23)	31S → A(12)	76A → S(20)	78S → C(12)	89L → V(12)	135 G→ S(12)	31S → A(12)	33P →S (20)	78S → C(20)	89L → V(20)	31S → A(12)	33P → S(12)	78S → C(12)	89L → V(12)	98L → deletion (12)	135G → S(12)	31S → A(20)	135G → S(20)	113K → stop (20)	139 H→E (20)	ND	ND	
Regulatory domain (aa1–aa 48) [strongly conserved region]	√	√	ND	ND	ND	√	√	ND	ND	√	√	ND	ND	ND	ND	√	ND	ND	ND	ND	ND	
XAP binding region (aa61–aa91)	ND	ND	√	√		ND	ND	√	√	ND	ND	√	√		ND	ND	ND	ND	ND	ND	ND	
P 53 binding region (aa101–aa153)	ND	ND	ND	ND	√	ND	ND	ND	ND	ND	ND	ND	ND	ND	√		√	√	√	ND	ND	
Polymerase gene (15)	178→ deletion (100)		270S→F (24)			178→ deletion (20)				178→ deletion (4)	267I→ N (12)		268D→ T (12)			178→ deletion (20)		270S→ F(20)		ND	ND	
Primase (aa1–aa178)	√		ND			√				√	ND		ND			√		ND		ND	ND	
Spacer (aa178–aa 304)	ND		√			ND				ND	√		√			ND		√		ND	ND	
Catalytic determinant (A–H) (aa336–aa680)	NS		NS			NS				NS	NS		NS			NS		NS		NS	NS	
RNase (aa680–aa838)	NS		NS			NS				NS	NS		NS			NS		NS		NS	NS	
Precore/core (60)	10I → L(76)	28W → Stop (20)	92V → G(88)	120V → T(88)	96N → T(92)	10I → L(36)	41S → T(36)	92V → G(36)	109A → I(36)	10I → L(16)	28W → stop (16)	148L → S(28)	155I → T(40)	161Y → N(28)		93E → D(100)	109A →V (60)	120V → T(60)		ND	10I → L(20)	30M → R(20)
Amino terminal (aa1– aa150)	√	√	√	√	√	√	√	√	√	√	√	ND	ND	ND		√	√	√		ND	√	√
Carboxyl terminal (aa150–aa179)	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	√	√	√		ND	ND	ND		ND	ND	ND
Reliability of phylogenetic approach to detect the HBV genotype																						
Based on X gene	√					√				√						√				√	√	
Based on P gene	√					√				√						√				√	√	
Based on C gene	√					√				√						√				√	√	
Based on S gene	√					√				√						√				√	√	

Novel mutations documented from the HCC cases in the present study are tabulated. These are as follows: surface (132 stop), polymerase (frameshift at codon 178), core (10I→L, 41S→T, 92V→G, 96N→T and 164 Q→P) and X (33P→S) gene. These mutations have been reported for the first time and were found significantly associated in HCC cases from the three regions.
NI: North India, SI: South India, NEI: North East India, * ND: Not detected, aa: Amino acid and NS: Not studied.

Table 2. Mutations detected in the polymerase gene across three geographical regions of India

Locations	Mutation pattern	HCC			Asymptomatic HBV carriers		
		NI (25) [A]	SI (25) [B]	NEI (25) [C]	NI (5) [D]	SI (5) [E]	NEI (5) [F]
178*	L→deletion	25 (100.0%)	5 (20.0%)	1 (4.0%)	1 (20.0%)	–	–
179	Q→deletion	2 (8.0%)	–	–	–	–	–
180	H→deletion	2 (8.0%)	–	–	–	–	–
181	G→deletion	2 (8.0%)	–	–	–	–	–
182	R→deletion	2 (8.0%)	–	–	–	–	–
183	L→deletion	2 (8.0%)	–	–	–	–	–
184	V→deletion	2 (8.0%)	–	–	–	–	–
237	S→R	2 (8.0%)	–	–	–	–	–
239	T→Q	2 (8.0%)	–	–	–	–	–
240	S→G	4 (16.0%)	–	–	–	–	–
242	G→W	2 (8.0%)	–	–	–	–	–
247	K→G	2 (8.0%)	–	–	–	–	–
251	S→T	2 (8.0%)	–	–	–	–	–
255	Y→P	2 (8.0%)	–	–	–	–	–
267	I→T	2 (8.0%)	–	–	–	–	–
237	S→C	5 (20.0%)	–	–	–	–	–
371	V→C	5 (20.0%)	–	–	–	–	–
324	S→F	5 (20.0%)	–	–	–	–	–
267	I→N	–	–	3 (12.0%)	–	–	–
268	D→T	–	–	3 (12.0%)	–	–	–
298	S→A	4 (16.0%)	–	–	–	–	–
307	C→D	4 (16.0%)	–	–	–	–	–
270**	S→F	6 (24.0%)			1 (20.0%)	–	–
80	L→I	1 (4%)	1 (4%)	–	–	—	–
236	N→T	1 (4%)					
169	I→T	1 (4%)					
181	A→V	1 (4%)					

A deletion mutation at codon 178 and a substitution mutation at codon 270S → F of the polymerase gene were found to be significantly associated with HCC cases from North India. Also 80L→I, 236N→T, 169I→T and 181A→V were the four drug resistant mutation observed in the present study. Values in bold indicates statistical significance. NI: North India, SI: South India and NEI: North East India.

*A vs B : (P = <0.0000001; OR = 20.86 [undefined])/A vs C : (P = <0.0000001; OR = 62.28 (Undefined))/B vs C : (P = 0.05294; (OR = 5.811 {0.7312, 148.2}])/A vs D : (P = 0.01190; [OR = UD{1.488 – UD}]). **A vs D : (P = 0.2444 [OR = 2.549{0.2444, 76.29}]).

2Q → deletion, 103S → N, 171T → H, 172L → F, 173P → R, 176T → N, 177V → X, 179R → deletion, 1M → P, 109S → I, 164Q → P, 10I → X, 13S → T, 103S → N, 116S → N, 109A → T and 29G → D were documented only in cases from North India. A mutation at amino acid position 159P → T was exclusively found in cases from South India. Similarly, 16T → S, 30M → R, 41S → D, 46S → stop, 148L → S, 155I → T, 161Y → N, 100W → G, 120V → T, 103S → V, 170S → H172 L → S, 173P → G, 176V → S, 173P → D, 174E → H, 50S → C, 53F → L, 59L → V, 60L → I, 66L → Q, 67Y → F, 82T → I and 16T → S were exclusively found in cases from NEI. Epi Info analysis using two-by-two tables showed that mutations like 10I → L were risk factors for HCC cases of North India when compared to those from South India (P = 0.01 [OR = 5.63{1.64–19.23}]) and NEI (P < 0.01 [OR = 16.63 {4.04–68.03}]). Another mutation 28W → stop was associated with HCC cases from NEI and North India, and was completely absent in HCC cases from South India.

Mutations like V → T at position 120 were found to increase the risk of HCC by around 4.5-fold when compared to asymptomatic cases in North India (P = 0.102 [OR = 4.54 {0.402–45.63}]). Again, association of mutations like 41S → T (P < 0.001; [OR = 19.01{4.03–143.6}]), 92V → GA versus B (P < 0.001; [OR = 19.01 {4.03–143.6}]), 96N → TA versus B (P < 0.001; [OR = 19.01{4.03–143.6}]), 164Q → P (P = 0.0279; [OR = 3.085 {0.9735–10.29}]) and 103S → G (P = 0.05; [OR = 2.613 {0.8312–8.578}]) were significantly associated with HCC from North Indian cases when compared to those from South India. These mutations were found to be risk factors associated with HCC development in cases from North India when compared to those from South India. An increased risk (4.5 times) was observed for the mutation 11S → F in HCC cases when compared to asymptomatic cases from the same region. Overall mutations were more frequently observed in HCC cases than in asymptomatic cases for the precore/core gene of HBV isolates in this study. Overall, the frequency of precore/core

Table 3. Mutations detected in the precore/core gene across three geographical regions of India

Locations	Mutation pattern	HCC NI (25) [A]	HCC SI (25) [B]	HCC NEI (25) [C]	Asymptomatic HBV carriers NI (5) [D]	Asymptomatic HBV carriers SI (5) [E]	Asymptomatic HBV carriers NEI (5) [F]	Statistical analysis
10	I→L	19 (76.0%)	9 (36.0%)	4 (16.0%)	–	–	1 (20.0%)	A vs B(P = 0.01; [OR = 5.63{1.64–19.23}]) A vs C(P < 0.01; [OR = 16.63{4.04–68.03}]) B vs C(P = 0.06; [OR = 2.95{0.769–11.34–19.23}])
28	W→stop	5 (20.0%)	–	4 (16.0%)	–	–	–	
93	E→D	–	–	–	5 (100.0%)	–	–	
109	A→V	–	–	–	3 (60.0%)	–	–	
120	V→T	22 (88.0%)	–	–	3 (60.0%)	–	–	A vs D(P = 0.102; [OR = 4.54{0.402–45.63}])
159	P→Q	–	–	–	5 (100.0%)	–	–	
2	Q→deletion	1 (4.0%)	–	–	–	–	–	
103	S→V	1 (4.0%)	–	–	–	–	–	
148	L→S	1 (4.0%)	–	–	–	–	–	
171	T→H	1 (4.0%)	–	–	–	–	–	
172	L→F	1 (4.0%)	–	–	–	–	–	
173	P→R	1 (4.0%)	–	–	–	–	–	
176	T→N	1 (4.0%)	–	–	–	–	–	
177	V→X	1 (4.0%)	–	–	–	–	–	
179	R→deletion	1 (4.0%)	–	–	–	–	–	
16	T→S	–	–	1 (4.0%)	–	–	–	
30	M→R	–	–	1 (4.0%)	–	–	1 (20.0%)	C vs F(P = 0.161; [OR = 0.175{0.004–7.7}])
41	S→D	–	–	1 (4.0%)	–	–	1 (20.0%)	C vs F(P = 0.161; [OR = 0.175{0.004–7.7}])
46	S→stop	–	–	1 (4.0%)	–	–	1 (20.0%)	C vs F(P = 0.161; [OR = 0.175{0.004–7.7}])
1	M→P	6 (24.0%)	–	–	–	–	–	
41	S→T	23 (92.0%)	9 (36.0%)	–	–	–	–	A vs B(P < 0.001; [OR = 19.01{4.03–143.6}])
92	V→G	23 (92.0%)	9 (36.0%)	–	–	–	–	A vs B(P < 0.001; [OR = 19.01{4.03–143.6}])
96	N→T	23 (92.0%)	9 (36.0%)	–	–	–	–	A vs B(P < 0.001; [OR = 19.01{4.03–143.6}])
103	S→G	15 (60.0%)	9 (36.0%)	–	–	–	–	A vs B(P = 0.05; [OR = 2.613{0.8312–8.578}])
109	S→I	6 (24.0%)	–	–	–	–	–	
1	M→deletion	1 (4.0%)	6 (24.0%)	1 (4.0%)	1 (20.0%)	–	–	A vs B(P = 0.049; [OR = 0.14{0.002756–1.029}])
1	M→Q	16 (64.0%)	3 (12.0%)	–	–	–	–	A vs B(P = 0.432; [OR = 1.17{0.1208–9.371}])
2	Q→X	3 (12.0%)	6 (24.0%)	–	–	–	–	A vs B(P = 0.1505; [OR = 0.439{0.08–2.013}])
109	A→I	11 (44.0%)	9 (36.0%)	–	–	–	–	A vs B(P = 0.290; [OR = 1.387{0.438–4.47}])
159	P→T	–	6 (24.0%)	–	–	–	–	
164	Q→P	16 (64.0%)	9 (36.0%)	–	–	–	–	A vs B(P = 0.0279; [OR = 3.085{0.9735–10.29}])
10	I→X	3 (12.0%)	–	–	–	–	–	
13	S→T	4 (16.0%)	–	–	–	–	–	
103	S→N	4 (16.0%)	–	–	–	–	–	
116	S→N	4 (16.0%)	–	–	–	–	–	
109	A→T	2 (8.0%)	–	–	–	–	–	
29	G→D	2 (8.0%)	–	–	–	–	–	
11	S→F	4 (16.0%)	–	1 (4.0%)	–	–	1 (20%)	A vs C(P = 0.099; [OR = 4.47{0.51–117.4}]) C vs F(P = 0.166; [OR = 0.182{0.004–8.078}])
148	L→S	–	–	7 (28.0%)	–	–	–	
155	I→T	–	–	10 (40.0%)	–	–	–	
161	Y→N	–	–	7 (28.0%)	–	–	–	
100	W→G	–	–	3 (12.0%)	–	–	–	
120	V→T	–	–	3 (12.0%)	–	–	–	
103	S→V	–	–	3 (12.0%)	–	–	–	
170	S→H	–	–	3 (12.0%)	–	–	–	
172	L→S	–	–	3 (12.0%)	–	–	–	
173	P→G	–	–	3 (12.0%)	–	–	–	
176	V→S	–	–	3 (12.0%)	–	–	–	
173	P→D	–	–	2 (8.0%)	–	–	–	
174	E→H	–	–	2 (8.0%)	–	–	–	
50	S→C	–	–	1 (4.0%)	–	–	1 (20.0%)	C vs F(P = 0.166; [OR = 0.182{0.004}])
53	F→L	–	–	1 (4.0%)	–	–	1 (20.0%)	C vs F(P = 0.166; [OR = 0.182{0.004}])
59	L→V	–	–	1 (4.0%)	–	–	1 (20.0%)	C vs F(P = 0.166; [OR = 0.182{0.004}])
60	L→I	–	–	1 (4.0%)	–	–	1 (20.0%)	C vs F(P = 0.166; [OR = 0.182{0.004}])
66	L→Q	–	–	1 (4.0%)	–	–	1 (20.0%)	C vs F(P = 0.166; [OR = 0.182{0.004}])
67	Y→F	–	–	1 (4.0%)	–	–	1 (20.0%)	C vs F(P = 0.166; [OR = 0.182{0.004}])
82	T→I	–	–	1 (4.0%)	–	–	1 (20.0%)	C vs F(P = 0.166; [OR = 0.182{0.004}])
16	T→S	–	–	1 (4.0%)	–	–	1 (20.0%)	C vs F(P = 0.166; [OR = 0.182{0.004}])

41S→T (P < 0.001; OR = 19.01), 92V → G (P < 0.001; OR = 19.01), 96N→T (P < 0.001; OR = 19.01) and 164Q→P (P = 0.0279; OR = 3.085) were found significantly associated with HCC cases from North India when compared to the HCC cases from South India. Again 10I→L was significantly associated in HCC cases from North East India when compared to the cases from South India (P = 0.01; OR = 5.63) and North East India (P < 0.01; OR = 16.63). Values in bold indicates statistical significance. NI, North India; SI, South India and NEI: North East India.

Table 4. Mutations detected in the X gene across three geographical regions of India

Locations	Mutation pattern	HCC NI (25) [A]	HCC SI (25) [B]	HCC NEI (25) [C]	Asymptomatic HBV carriers NI (5) [D]	Asymptomatic HBV carriers SI (5) [E]	Asymptomatic HBV carriers NEI (5) [F]	Statistical analysis
76	A→S	5 (20.0%)	–	–	1 (20.0%)	–	–	A vs D (P = 0.7017; [OR = 1.0 {0.09668–29.45}])
76	A→T	3 (12.0%)	–	–	1 (20.0%)	–	–	A vs D (P = 0.5384; [OR = 0.5583 {0.04539–17.79}])
113	K→I	3 (12.0%)	–	–	1 (20.0%)	–	–	A vs D (P = 0.5384; [OR = 0.5583 {0.04539}])
133	V→R	3 (12.0%)	–	–	1 (20.0%)	–	–	A vs D (P = 0.5384; [OR = 0.5583 {0.04539}])
135	G→stop	3 (12.0%)	–	–	1 (20.0%)	–	–	A vs D (P = 0.5384; [OR = 0.5583 {0.04539}])
137	C→stop	3 (12.0%)	–	–	1 (20.0%)	–	–	A vs D (P = 0.5384; [OR = 0.5583 {0.04539}])
31	S→A	3 (12.0%)	5 (20.0%)	3 (12.0%)	1 (20.0%)	–	–	A vs D (P = 0.5384; [OR = 0.5583 {0.04539–17.79}]) A vs B (P = 0.2372; [OR= 0.5521 {0.09748–2.698}]) A vs C (P = 0.5000; [OR = 1.0 {0.1564–6.396}]) B vs C (P = 0.2372; [OR = **1.811** {0.3706–10.26}])
33	P→S	–	5 (20.0%)	3 (12.0%)	–	–	–	B vs C (P = 0.2372; [OR = 1.811{0.3706–10.26}])
78	S→C	3 (12.0%)	5 (20.0%)	3 (12.0%)	1 (20.0%)	–	–	A vs D (P = 0.5384; [OR = 0.5583 {0.04539–17.79}]) A vs B (P = 0.2372; [OR = 0.5521{0.09748–2.698}]) A vs C (P = 0.5000; [OR = 1.0 {0.1564–6.396}]) B vs C (P = 0.2372; [OR = **1.811** {0.3706–10.26}])
89	L→V	3 (12.0%)	5 (20.0%)	3 (12.0%)	1 (20.0%)	–	–	A vs D (P = 0.5384; [OR = 0.5583 {0.04539–17.79}]) A vs B (P = 0.2372; [OR = 0.5521{0.09748–2.698}]) A vs C (P = 0.5000; [OR = 1.0 {0.1564–6.396}]) B vs C (P = 0.2372; [O = **1.811** {0.3706–10.26}])
98	L→deletion	–	–	3 (12.0%)	–	–	–	
113	K→stop	3 (12.0%)	–	3 (12.0%)	1 (20.0%)	–	–	A vs C (P = 0.5000; [OR= 1.0 {0.1564– 6.396}]) A vs D (P = 0.5384; [OR = 0.5583 {0.04539–17.79}])
134	L→R	3 (12.0%)	–	3 (12.0%)	1 (20.0%)	–	–	A vs C (P = 0.5000; [OR = 1.0 {0.1564}])
135	G→S	3 (12.0%)	–	3 (12.0%)	1 (20.0%)	–	–	A vs D (P = 0.5384; [OR = 0.5583 {0.04539}])
139	H→E	3 (12.0%)	–	3 (12.0%)	1 (20.0%)	–	–	A vs C (P = 0.5000; [OR = 1.0 {0.1564}])

33P→S was entirely absent in HCC cases from North India but were observed in HCC cases from North East and South India while 98L→deletion was strictly observed in HCC cases from North East India. NI, North India; SI, South India and NEI, North East India.

mutants was maximum in HCC cases, of which North Indian cases were found to harbour maximum mutants followed by NEI and South India. Few of the documented mutations were nonsignificantly associated with HCC cases (Table 3).

X gene: A total of 15 mutations were documented in the X region of the HBV sequences analysed at 15 different codon positions. The mutation patterns commonly observed were of deletion, substitution or stop codon. Mutations like 76A → S, 76A → T, 113K → I, 133V → R, 135G → stop and 137C → stop were exclusively found in cases from North India in both HCC and asymptomatic HBV carriers; thus these can be categorized as exclusive mutations observed from North India. However, mutations at 31S → A, 78S → C and 89L → V were found in HCC cases from all the three regions and can be termed as X gene mutations associated with advanced stage of the disease. Again, mutation at codon 33P → S was entirely absent in HCC cases from North India, but was observed in those from the other two regions. Again, a deletion mutation at codon 98 was strictly observed in three HCC cases from NEI. Many mutations were absent in HCC cases from South India, but were found in those from North India and NEI; for example 113K → stop, 134L → R, 135G → S and 139H → E. However, significantly raised values of odds ratios were not found associated upon statistical analysis for a majority of the mutations of X gene documented in the present study. An increased risk of around twofold was observed for some of the mutations observed in HCC cases from South India when compared to those from NEI. It is of interest whether the 89L → V and 98L → deletion mutation affects protein–protein interactions between the X protein and XAP3 (Table 4).

Surface gene: The only mutation observed in the surface gene of HBV genome was the stop codon mutation at position 132. Around 52% of the North Indian HCC cases were mutants for this, while 16.0% were from South India and another 16.0% from NEI. A single case of asymptomatic carrier was found to harbour this mutation which belonged to North India. Also, 132 stop-codon mutation was a independent risk factor for development of HCC in North Indian cases when compared to those from South India (P = 0.004486 [OR = 5.479{1.496–23.47}]) and NEI (P = 0.004486 [OR = 5.479{1.496–23.47}]). Among the North Indian HCC cases and asymptomatic HBV carriers, this particular mutation was found to increase the risk for HCC development by

Table 5. Mutations detected in the surface gene across three geographical regions of India

Locations	Mutation pattern	HCC NI (25) [A]	HCC SI (25) [B]	HCC NEI (25) [C]	Asymptomatic HBV carriers NI (5) [D]	Asymptomatic HBV carriers SI (5) [E]	Asymptomatic HBV carriers NEI (5) [F]	Statistical analysis
132	Stop codon mutation	13 (52.0%)	4 (16.0%)	4(16.0%)	1(20.0%)	–	–	A vs B (P = **0.004486**; [OR = **5.479** {1.496–23.47}]) A vs C P = **0.004486**; [OR = **5.479** {1.496–23.47}]) B vs C (P = 0.5000; [OR = 1.0 {0.2006–4.984}]) A vs D (P = 0.1201; [OR = **4.143** {0.4476–114.8}])

The significant mutation observed in the surface region of HBV was the 132 stop codon mutation. This mutation was a independent risk factor for development of HCC in North Indian HCC cases when compared against the HCC cases from South India (P = 0.004486; OR = 5.479) and also when compared with HCC cases from North East India (P = 0.004486; OR = 5.479). NI, North India; SI, South India and NEI, North East India.

almost five fold (P = 0.1201 [OR = 4.143{0.4476–114.8}]) (Table 5).

Overall mutation in whole-genome of HBV

HCC cases were analysed with respect to the four genes, mutation documented and genotype. It was observed that mutation was observed for all the cases of HCC from North India in the polymerase and core gene of the HBV genome. Again, 21 out of the 25 cases from North India were mutants for the X gene. About 64% of the North Indian HCC cases were mutated for the surface gene. Similarly, 20% of the South Indian HCC cases were mutants for S gene, 60% for P gene, 35% for X gene and 80% were mutants for the C gene. The mutation percentages were 16, 35, 35 and 80 for the S, P, X and C genes respectively in the HCC cases from North East India (see Supplementary Information online).

Distribution of HBV mutants with respect to their genotypes

A majority of the cases from all the genotypes were mutants for one or more regions of the HBV genome. Mutation was more commonly associated with genotype D compared to genotype A in North (P = 0.004) and South (P = 0.002) Indian HCC cases. However, all the HCC cases from NEI haboured mutant genome of HBV irrespective of their genotype. A majority of the mutants were significantly associated with the development of HCC irrespective of their genotype and geographical location. The following curves of Kaplan–Meier were constructed for HCC cases from the three different regions with reference to mutants for each gene of the virus. Survival was found to be least for patients harbouring core gene mutation followed by X, P and surface gene mutants for all the HCC cases included in the study, irrespective of their geographical location. Again, the overall survival period was least for HCC cases from NEI when compared to the survival of HCC cases from the other two regions.

Staging of HCC cases

HCC cases were staged according to the Okuda staging system, Cancer of the Liver Italian Program (CLIP) scoring system and Barcelona Clinic Liver Cancer (BCLC) scoring system. Okuda stage II was observed in a majority of the HCC cases from all the three regions. When the cases were staged with respect to the CLIP scoring system, patients were found to be almost evenly distributed among the five stages of the system. The mean survival of patients was on the lower side if found associated with advanced stage of the disease with either of the staging systems. However, no HCC cases from South India or NEI were found with the final stage (4–6) of the disease with CLIP scoring system. With respect to BCLC staging system, a majority of the cases from all the three regions were in stage A1.

Survival curve with respect to different mutants

The curves of Kaplan–Meier were constructed for HCC cases from the three different regions with reference to mutants for each gene of the virus. Survival was found to be least for patients harbouring core gene mutation followed by X, P and surface gene mutants for all the HCC cases included in the study, irrespective of their geographical location. Again, the overall survival period was least for the HCC cases from NEI when compared to the survival of HCC cases from the other two parts of the country (Figures 1–3). It is clearly evident that higher Okuda stage, MELD score (<10), low albumin levels (<3.5 mg/dl), high AFP levels (>400 ng/ml) and larger nodule (>5 cm) decrease the survival of the HCC patients in the present study.

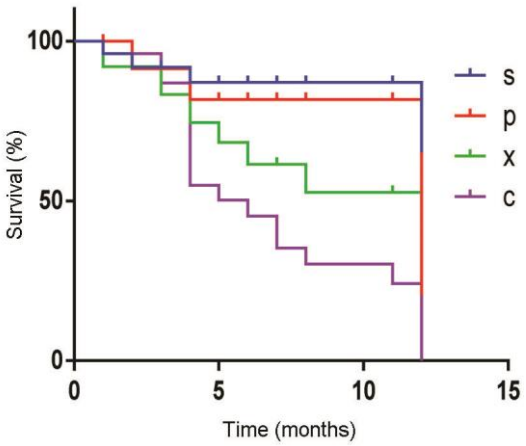
Discussion

There were many significant findings in the HBV genome infecting the study population across the three geographical regions.

The novel 132 stop-codon mutation observed in the surface gene is of special interest. As several lines of

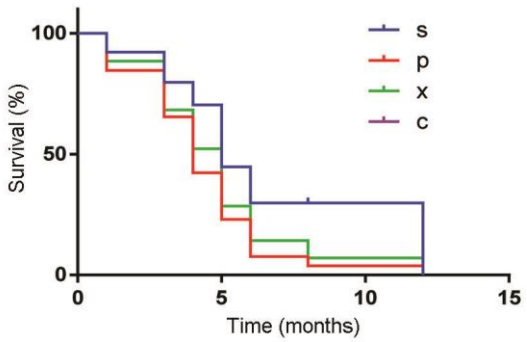
evidence suggest, naturally occurring mutants in the pre-S region are correlated with a more progressive form of liver disease and have been documented by a few studies[23,24]. Also, integration of the truncated large or middle envelope protein into the host DNA is reported to enhance the possibility of HCC development by increasing a transactivating capacity[25]. Although the 132 stop-codon mutation was not documented earlier, mutation in the PreS2 region of the HBV gnome has been reported in HCC cases[26], where the group discovered a novel pre-S2 substitution (F141L). Also, 132 is of particular interest because it falls in the epitomic region of the virus and leads to a truncated protein which is of special interest to a group of scientists who have been arguing against the common belief that mutation in the regulatory segment of the virus is of greater importance. Due to overlapping of S and P genes, variation in S gene may affect the catalytic activity of reverse transcriptase (rt) domain of the P gene, and vice versa[27]. The stop-codon mutation in the present study is in the corresponding rt domain of the polymerase gene. Evidence of few reported drug-resistant mutants like rtF166L and rtP177L in the polymerase gene may be attributed to stop-codon mutation in the surface gene and its catalytic effect[7]. Thus, there is a possibility that the stop-codon mutation observed in the present study may affect the drug resistance regions of the polymerase gene in HCC patients. The 132 stop codon mutation may be a similar to that which falls well inside the codons 25–150. The stop codon at amino acid 132 was observed in a majority of HCC cases from all the three regions, similar to a study by Ito *et al.*[28], where a premature stop-codon mutation was observed at aa156 (TGG → TAG), aa191 (TGG → TGA), aa196 (TGG → TGA), aa206 (TAC → TAG) and aa223 (TGG → TGA) respectively. But mutations like 125 (M → T) and 127 (T → P) amino acid substitutions observed in earlier studies were not found in the present study[29].

The four drug-related mutations documented in HCC cases in the present study have been previously



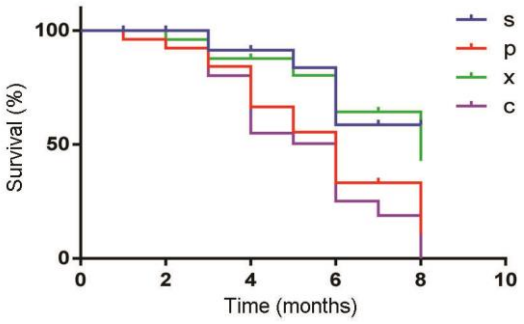
Statistical analysis	
Comparison of Survival curves	
Log-rank (Mantel–Cox) test (recommended)	
Chi square	17.09
d*f*	3
P-value	0.0007
P-value summary	***
Are the survival curves sig different?	Yes
Logrank test for trend (recommended)	
Chi square	15.35
d*f*	1
P-value	<0.0001
P-value summary	****
Sig. trend?	Yes
Gehan-Breslow-Wilcoxon test	
Chi square	0.003550
d*f*	3
P-value	0.9999
P-value summary	ns
Are the survival curves sig different?	No

Figure 1. Survival curve of HCC cases with respect to surface(s), polymerase (p), x and core (c) mutant amongst HCC cases of North east India.



Statistical analysis	
Comparison of Survival Curves	
Log-rank (Mantel-Cox) test (recommended)	
Chi square	6.530
d*f*	3
P-value	0.0885
P-value summary	ns
Are the survival curves sig different?	No
Logrank test for trend (recommended)	
Chi square	3.483
d*f*	1
P-value	0.0620
P-value summary	ns
Sig. trend?	No
Gehan–Breslow–Wilcoxon test	
Chi square	0.001187
d*f*	3
P-value	1.0000
P-value summary	ns
Are the survival curves sig different?	No

Figure 2. Survival curve of HCC cases with respect to surface(s), polymerase (p), x and core (c) mutant amongst HCC cases of South India.



Statistical analysis	
Comparison of survival curves	
Log-rank (Mantel-Cox) test (recommended)	
Chi square	15.66
d*f*	3
P-value	0.0013
P-value summary	**
Are the survival curves sig different?	Yes
Logrank test for trend (recommended)	
Chi square	6.462
d*f*	1
P-value	0.0110
P-value summary	*
Sig. trend?	Yes
Gehan–Breslow–Wilcoxon test	
Chi square	0.003183
d*f*	3
P-value	1.0000
P-value summary	ns
Are the survival curves sig different?	No

Figure 3. Survival curve of HCC cases with respect to surface(s), polymerase (p), x and core (c) mutant amongst HCC cases of North India.

described. Mutations like L80I observed in this study have been documented to be associated with lamivudine resistance[30], which is a compensatory mutation for M204V/I (ref. 30). Ogata *et al.*[30] have shown that main M204V/I resistant mutation and additional compensatory mutations such as L80I/V, V173L and L180M are associated with high-level resistance to lamivudine (3TC), telbivudine (L-dT) and emtricitabine (FTC). Also, the importance of this particular mutation observed in this study needs to be further assessed as this mutation is not located in the supposed drug binding domain. Ability of rtL80I to mediate lamivudine resistance is of general interest and deserves foremost attention primarily for two reasons. First, the fact that mutations in the periphery of the nucleoside/nucleotide binding site may lead to resistance has to be taken into account in all future resistance testing approaches. Second, it may also have an association with other drugs available in the market for HBV treatment[31]. Similarly, dual drug-resistant mutations were observed in two HCC cases from North India. The first HCC case harboured L80I and N236T mutations. L80I is a lamivudine mutation, while N236T is associated with adefovir resistance[32]. The main M204V/I resistant mutation and additional compensatory mutations L80I/V, V173L and L180M are associated with high-level resistance to lamivudine, telbivudine and emtricitabine[32]. The other patient with dual drug mutation was found to harbour I169T and A181V mutations. I169T has been associated with the entecavir resistance mutations in lamivudine-resistant patients and is believed to play a secondary role in entecavir resistance[8], while A181V/T is associated with resistance to adefovir and/or lamivudine[33,34]. These mutants were expected as a few of the HCC cases were under nucleotide/nucleoside therapy in the past. Documentation of dual mutants observed in the present study can be explained as many patients undergoing treatment developed lamuvidine mutations and were either changed to adefovir or entacavir and later progressed to HCC. Again, association of rtM204V with rtL180M double mutation was not seen in the present study, although the later mutation was documented in a few HCC cases, unlike the study by Li *et al.*[34] where combinatorial mutation was observed for 43.75% cases. This mutation is often found to be associated with lamuvidine resistance and its absence may be attributed to the fact that few of the HCC patients included in the present study were under lamuvidine treatment in the past. Evaluation of anti-virus therapy by surveillance of the two site mutations is of importance[35].

Mutations such as a deletion at position 180 have been reported earlier[35,36]. The present study did not show mutations at certain positions like 173, 204 and 223 as observed by Lu *et al.*[35]. This may be because the patient population in the present study consisted of HCC cases compared to the study by Lu *et al.*[35], who considered mutations in chronic hepatitis cases. Another study[37] found that mutations at amino acid positions 180, 181 and 184 are similar to the frame shift mutation observed in the present study. The uncommon point mutations reported in this study like 183L → deletion, 237S → R, 239T → Q, 240S → G, 242G → W, 247K → G, 251S → T, 255Y → P, 267I → T, 237S → C, 371V → C, 324S → F, 267I → N, 268D → T, 298S → A, 307C → D and 270S → F have not been documented previously and might be the cause of cross-drug resistance and multiple resistance as described by Scott *et al.*[38]. Geographical isolation of the HBV strain along with the host immune factors may also play a role in these variable mutations. Frame shift mutation in the polymerase gene of HBV has been reported earlier[39], where defective DNA was found in 90% of patients with CH-B, much like the findings of the present study where a frame shift mutation was observed at codon 178 and continued till codon 184.

Two of the mutations in the present study were located in the regulatory domain of the HBx protein (positions 20–50), i.e. 31S → A and 33P → S. The association of S → A was commonly found in HCC cases from all the three regions studied. However, the later mutation was

strictly observed in HCC cases from South India and NEI. The hot spot mutation in position 31 (serine to proline) as reported earlier[32], has been documented in the present study where alanine replaces serine at codon 31. Many other mutations reported in the present study fall in between the codon positions 111–135, much like previous findings[39–41]. These mutations are of special interest as the dominant mutations of HBV variants and are accumulated in the T-cell epitope of the HBx protein. A few mutations were observed in the p53 binding regions not only in HCC cases but also in the asymptomatic HBV carriers. The association between HBx protein and p53 has been long debated and the presence of these X gene mutants in this study needs further research and explanation for their clinical significance. Detection of stop codon mutation at positions 135 and 137 in the present study is in accordance with the earlier findings[15] and is attributed to a frame shift mutation. HBV X gene shares sequences both with polymerase and core genes and carries several regulatory cycles critical to the replicative cycles; its product has transactivation function. Detection of stop-codon mutation at positions 135 and 137 in the present study is in accordance with earlier findings[15] and is attributed to a frame shift mutation. Some other mutations observed were not significantly associated with HCC cases when compared to their presence in asymptomatic HBV carriers, indicating that these mutations cannot be used as prognostic markers for diagnosis of HCC.

Highest numbers of mutations were observed in the core/precore region of HBV genome analysed in this study compared to the other three genes, as expected. Mutations in the precore/core gene were found in both the amino and carboxyl terminals. However, the most notable point observed was that the amino terminus mutants were more common in HCC cases from North and South India, while the carboxyl terminal mutants were more frequently observed in HCC cases from NEI. These mutations in the carboxyl terminus of the core protein need to be studied in detail as this region is involved in a number of important functions in the viral replication cycle, including RNA packaging and DNA synthesis. Although few studies have reported mutations in the C-terminus of core protein in HCC cases[42,43], the mutations documented in the present study need attention as the C-terminal of the core protein is highly conserved with a repetitive structure which is essential for viral production. However, presence of precore stop-codon mutation was seen in five of the HCC cases from North India and in four cases from NEI. Importantly, a majority of the stop-codon mutations observed in the present study were infected with genotype D and less frequently with genotype A, which is in agreement with published data[32], where it was documented that the stop-codon mutation was associated in 65–75% genotype D cases, while it was around 8–18% for genotype A. The most common and expected mutation of T → V at codon 120 was observed in a majority of the HCC cases, which is in agreement with published data[44]. Mutation at 159 from Pro → Thr and a change from Pro → Gln at the same codon in this study is associated with Arg-rich region located at the C-terminus of the core. This has been identified as important for DNA and RNA binding, and has been documented by an earlier study[45]. Again, detection of 13 Ser → Thr in four HCC cases in the present study is in accordance with the findings of Lee[45]. In the core gene, the most common and expected mutation of T → V at codon 120 was observed in majority of the HCC cases, which is in agreement with published data[44]. On the other hand, A → T at position 109 was documented in two cases of HCC, and was also reported earlier[46]. High frequency for many mutations like 28W → stop and 41S → T is perhaps due to patient age and duration of infection as a consequence of chronic HBV infections. The 28W → stop-codon mutation was observed in the precore region from the HBV genome of HCC cases from North India and NEI, which was associated with HBeAg negativity status. This may be the reason for the negativity of HBeAg status on serological analysis of many HCC cases in this study. Mutation at 159 from Pro → Thr and a change from Pro → Gln at the same codon in this study are associated with Arg-rich region located at the C-terminus of the core, and have been identified as important for DNA and RNA binding and has been previously documented by an earlier study[45]. Again detection of 13Ser → Thr in four HCC cases in the present study is in accordance to the study findings of[45] and is a part of the $Cd4^+$ epitopes. Mutations like 109S → I and 109A → T are of significance as these amino acids are part of B-cell epitopes, and such mutations have been documented earlier[45]. The same study also supports our findings pertaining to the stop codon mutation at positions 177 and 46.

On the other hand, common A1762T/G1764A double mutation was not observed in this study. The possible explanation for not finding A1896 mutation in the present study population may be the fact that this precore stop-codon mutation is more common in patients with fulminant hepatitis (FH) and is mostly found in genotype-B cases as explained earlier[47], and also that the study group did not consist of either FH cases or any cases infected with HBV genotype B. Again, mutation at codon 103 (S → F, S → G and S → V) has been documented by a study from eastern India on a primitive tribal community[48], suggesting the association of the most primitive sub genotype D5 in these patients, similar to the findings of the present study.

A majority of the core mutants observed in this study were in the N-terminus of the core protein and were HCC cases from North and South India. The low prevalence of N-terminus mutants in HCC cases from NEI needs to be investigated. However, there is another school of thought based on the three-dimensional structure of HBV

nucleocapsids[49–53]. According to them, the three regions of core protein, i.e. aa78–82, 127–130 and 145–153 are exposed on the shell of nucleocapsids. Neither the N- nor the C-terminus of the core protein is on the external surface, and this rules out the significance of N- or C-terminus mutants in the core gene. According to these researchers, mutations in the three above-mentioned regions are of importance. If we view our results with respect to these three domains, we still find a few mutations at the important codons which are not reported earlier, and their association with the disease needs further investigation.

The multiple sequence alignment files for all the genes showed numerous mutations in the nucleotide and protein levels throughout the HBV genome. This is because although HBV is a DNA virus, it is prone to mutations with nucleotide substitutions estimated at a rate of 1×10^{-5}–3×10^{-5} per site per year as previously proposed[54]. This is related to the reverse transcription to an RNA intermediate during the replication cycle of HBV. Mutation at amino acid levels was considered for documentation for its widespread acceptance in place of nucleotide levels. The nucleotides were converted to amino acids and compared with the GenBank data with special reference to a recently published paper[55].

The phylogenetic tree constructed from sequences, of the polymerase gene yielded a robust form with 100% accuracy. Also, its findings were exactly the same as derived when the genotype for the same samples was determined using the precore/core or surface sequence of the respective samples, techniques usually followed to determine the HBV genotypes. This genotyping method utilizing a fragment of the HBV DNA polymerase gene can be employed as a general genotyping strategy, which has also been previously confirmed by Ma *et al.*[56]. The most widely found polymerase gene mutation in HCC cases was a deletion mutation at position 178. This particular mutation has been reported earlier[57]. Many point mutations have been documented in this study. Deletion mutations have also been reported earlier[58], much like the deletions observed in this study.

Genotype distribution with respect to mutation in the whole genome was analysed in all the HCC and asymptomatic cases from the three geographical locations. We found that HBV genotype D was significantly associated with HCC development in all the three geographical locations in the country. This is in accordance with the findings of Mumtaz *et al.*[26], who showed that precore and basal core promoter mutations are found in a significant number of patients infected with HBV genotype D infecting HCC cases. Many patients in our study were infected with dual mutations which are believed to enhance viral replication[59].

The staging of HCC in the present study reveals a high proportion of patients with multiple lesions, larger tumour size and advanced stage of the disease as documented by an earlier study[60]. Most of the lesions were hypoechoic or heterogeneous in the present study; this is in accordance with the findings of Kumar *et al.*[60]. Approximately half of the HCC cases from all the three regions had tumour size above 5 cm and high incidence of vascular invasion with very low resection rate, as documented by previous studies from India[60,61]. Higher Okuda stage, MELD score (<10), low albumin levels (<3.5 mg/dl), high AFP levels (>400), and larger nodule (>5 cm) decrease the survival of HCC patients. The survival was least for patients harbouring core gene mutation followed by X, P and surface gene mutations for all HCC cases included in the study, irrespective of their geographical location. Thus it can be concluded that mutation in the HBV genome, i.e. the four genes of HBV is directly associated with the survival of patients, especially in HCC cases.

Sequencing remains the gold standard for molecular characterization. However, its interpretation can be influenced by several biases and requires specific skills. At least three different runs are needed to get the whole 3.2 kb HBV sequence[58], similar to the technique adopted in the present study. Several molecular methods have been used for HBV genotyping like nucleotide sequencing in pre-S/S regions; but full-length genome sequencing remains the best method. Although the common genotyping method uses the surface gene amplification process, there is an exception as HBV strains can be accurately genotyped by sequencing two regions of the HBV genome as well, namely the S region and BCP/PC region[62].

Conclusion

The whole-genome analysis of HBV clearly shows that the HBV genome differs significantly among the three geographical locations and also between the two disease groups. Extensive analysis demonstrates that HBV genome associated with HCC cases harbours many significant mutations compared to carriers across the viral genome, irrespective of their geographical distribution. Among HCC cases, the HBV genome isolated from North India harbours the highest number of mutations followed by NEI and South India.

The present study reports few novel mutations in surface (132 stop), polymerase (frameshift mutation at codon 178), core ($10I \rightarrow L$, $41\ S \rightarrow T$, $92V \rightarrow G$, $96N \rightarrow T$, $164Q \rightarrow P$) and X($33\ P \rightarrow S$) genes of the genome of HCC cases. Molecular and cellular studies are required to prove the potentiality of these newly reported mutants to be used as a novel kit for the diagnosis and prognosis of HCC in future. Association of few exclusive mutants observed in HCC cases from NEI needs prior attention, as there is a changing epidemiology of HBV in that region. Also, mutations that are restricted to a particular geographical location need closer attention for developing

region-based HBV-related HCC management in India. Analysis of mutation in the conserved regions of the HBV genome in the present study is critical and needs further research to understand its effect on disease progression.

The sequences submitted to GenBank in this study will contribute to current knowledge on the genetic diversity of the virus worldwide and will provide a platform for whole-genome-based studies in future, particularly in India. Also, complete genome sequencing is more reliable for genotyping and sub-genotyping classification of HBV rather than genotyping based on a particular gene, as the use of short sequences significantly limits the evaluation of genetic relatedness among HBV strains.

1. Perz, J. F., Armstrong, G. L., Farrington, L A., Hutin, Y. J. and Bell, B. P., The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J. Hepatol.*, 2006, **45**(4), 529–538.
2. Mori, M., Hara, M., Wada, I., Hara, T., Yamamoto, K., Honda, M. and Naramoto, J., Prospective study of hepatitis B and C viral infections, cigarette smoking, alcohol consumption, and other factors associated with hepatocellular carcinoma risk in Japan. *Am. J. Epidemiol*., 2000, **151**(2), 131–139.
3. Brunetto, M. R. *et al.*, Hepatitis B virus unable to secrete e antigen and response to interferon in chronic hepatitis B. *Gastroenterology*, 1993, **105**(3), 845–850.
4. Jardi, R. *et al.*, Mutations in the basic core promoter region of hepatitis B virus and relationship with precore variants and HBV genotypes in a Spanish population of HBV carriers. *J. Hepatol.*, 2004, **40**(3), 507–514.
5. Ogata, N., Miller, R. H., Ishak, K. G. and Purcell, R. H., The complete nucleotide sequence of a pre-core mutant of hepatitis B virus implicated in fulminant hepatitis and its biological characterization in chimpanzees. *Virology*, 1993, **194**(1), 263–276.
6. Sato, S. *et al.*, Hepatitis B virus strains with mutations in the core promoter in patients with fulminant hepatitis. *Ann Intern Med*., 1995, **122**(4), 241–248.
7. Stuyver, L. J., Locarnini, S. A., Lok, A., Richman, D. D., Carman, W. F., Dienstag, J. L. and Schinazi, R. F., Nomenclature for antiviral-resistant human hepatitis B virus mutations in the polymerase region. *Hepatology*, 2001, **33**(3), 751–757.
8. Angus, P. *et al.*, Resistance to adefovir dipivoxil therapy associated with the selection of a novel mutation in the HBV polymerase. *Gastroenterology*, 2003, **125**(2), 292–297.
9. Sugauchi, F. *et al.*, A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. *J. Gen. Virol*., 2001, **82**(4), 883–892.
10. Utsumi, T. *et al*., Indones. *Clin. Microbiol.*, 2009, **47**(6), 1842–1847.
11. Tatsukawa, M. *et al.*, Hepatitis B virus core promoter mutations G1613A and C1653T are significantly associated with hepatocellular carcinoma in genotype C HBV-infected patients. *BMC Cancer*, 2011, **21**(11), 458.
12. Abdou Chekaraou, M., Brichler, S., Mansour, W., Le Gal, F., Garba, A., Dény, P. and Gordien, E., A novel hepatitis B virus (HBV) subgenotype D (D8) strain, resulting from recombination between genotypes D and E, is circulating in Niger along with HBV/E strains. *J. Gen. Virol*., 2010, **91**(6), 1609–1620.
13. Günther, S., Li, B. C., Miska, S., Krüger, D. H., Meisel, H. and Will, H., 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*., 1995, **69**(9), 5437–5444.
14. Zhang, Q. *et al.*, Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nature Genet.*, 2012, **44**(7), 765–769.
15. Cui, X. J., Cho, Y. K., Song, H. J., Choi, E. K., Kim, H. U. and Song, B. C., Molecular characteristics and functional analysis of full-length hepatitis B virus quasispecies from a patient with chronic hepatitis B virus infection. *Virus Res*., 2010, **150**(1–2), 43–48.
16. Bozdayi, G., Türkyilmaz, A. R., Idilman, R., Karatayli, E., Rota, S, Yurdaydin, C. and Bozdayi, A. M., Complete genome sequence and phylogenetic analysis of hepatitis B virus isolated from Turkish patients with chronic HBV infection. *J. Med. Virol.*, 2005, **76**(4), 476–481.
17. Bruix, J. *et al.*, EASL Panel of Experts on HCC. Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL Conference. European Association for the Study of the Liver. *J. Hepatol.*, 2001, **35**(3), 421–430.
18. Shrivastava, S., Lole, K. S., Tripathy, A. S., Shaligram, U. S. and Arankalle, V. A., Development of candidate combination vaccine for hepatitis E and hepatitis B: a liposome encapsulation approach. *Vaccine*, 2009, **27**(47), 6582–6588.
19. Saiki, R. K. *et al.*, Primer-directed enzymatic amplification of DNA with a thermo stable DNA polymerase. *Science*, 1988, **239**(4839), 487–491.
20. Kumar, S., Tamura, K. and Nei, M., Alignment. *Brief. Bioinform.*, 2004, **5**(2), 150–163. PubMed PMID: 15260895.
21. Saitou, N. and Nei, M., The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*., 1987, **4**(4), 406–425.
22. Lole, K. S. *et al.*, Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.*, 1999, **73**(1), 152–160.
23. Bock, C. T., Tillmann, H. L., Maschek, H. J., Manns, M. P. and Trautwein, C., A preS mutation isolated from a patient with chronic hepatitis B infection leads to virus retention and misassembly. *Gastroenterology*, 1997, **113**(6), 1976–1982.
24. Wang, H. C., Huang, W., Lai, M. D. and Su, I. J., Hepatitis B virus pre-S mutants, endoplasmic reticulum stress and hepatocarcinogenesis. *Cancer Sci.*, 2006, **97**(8), 683–688.
25. Caselmann, W. H., Meyer, M., Kekulé, A. S., Lauer, U., Hofschneider, P. H. and Koshy, R., A trans-activator function is generated by integration of hepatitis B virus preS/S sequences in human hepatocellular carcinoma DNA. *Proc. Natl. Acad. Sci. USA*, 1990, **87**(8), 2970–2974.
26. Mumtaz, K. *et al.*, A study of genotypes, mutants and nucleotide sequence of hepatitis B virus in Pakistan: HBV genotypes in Pakistan. *Hepat. Mon.*, 2011, **11**(1), 14–18.
27. Kerr, R., Stevens, G., Manga, P., Salm, S., John, P., Haw, T. and Ramsay, M., Identification of P gene mutations in individuals with oculocutaneous albinism in sub-Saharan Africa. *Hum. Mutat*., 2000, **15**(2), 166–172; Erratum in: *Hum. Mutat.*, 2000, **16**(1).
28. Ito, K. *et al.*, Impairment of hepatitis B virus virion secretion by single-amino-acid substitutions in the small envelope protein and rescue by a novel glycosylation site. *J. Virol.*, 2010, **84**(24), 12850–12861.
29. Ozaslan, M., Ozaslan, E., Barsgan, A. and Koruk, M., Mutations in the S gene region of hepatitis B virus genotype D in Turkish patients. *J. Genet.*, 2007, **86**(3), 195–201.
30. Ogata, N. *et al.*, Novel patterns of amino acid mutations in the hepatitis B virus polymerase in association with resistance to lamivudine therapy in Japanese patients with chronic hepatitis B. *J. Med. Virol.*, 1999, **59**, 270–276.
31. Tenney, D. J. *et al.*, Clinical emergence of entecavir-resistant hepatitis B virus requires additional substitutions in virus already

resistant to lamivudine. *Antimicrob. Agents Chemother*., 2004, **48**(9), 3498–3507.
32. Grandjacques, C. *et al.*, Rapid detection of genotypes and mutations in the pre-core promoter and the pre-core region of hepatitis B virus genome: correlation with viral persistence and disease severity. *J. Hepatol*., 2000, **33**(3), 430–439.
33. Yeh, C. T., Shen, C. H., Tai, D. I., Chu, C. M. and Liaw, Y. F., Identification and characterization of a prevalent hepatitis B virus X protein mutant in Taiwanese patients with hepatocellular carcinoma. *Oncogene*, 2000, **19**(46), 5213–5220.
34. Li, M. W., Hou, W,, Wo, J. E. and Liu, K. Z., Character of HBV (hepatitis B virus) polymerase gene rtM204V/I and rtL180M mutation in patients with lamivudine resistance. *J. Zhejiang Univ. Sci. B*, 2005, **6**(7), 664–667.
35. Lu, H. Y., Zeng, Z., Xu, X. Y., Zhang, N. L., Yu, M. and Gong, W. B., Mutations in surface and polymerase gene of chronic hepatitis B patients with coexisting HBsAg and anti-HBs. *World J. Gastroenterol*., 2006, **12**(26), 4219–4223.
36. Sayan, M., Cavdar, C. and Dogan, C., Naturally occurring polymerase and surface gene variants of hepatitis B virus in Turkish hemodialysis patients with chronic hepatitis B. *Jpn. J. Infect. Dis*., 2012, **65**(6), 495–501.
37. Ye, F., Li, X., Guo, J. C. and Wu, J., Analysis of drug-resistant multi-loci mutation of P area of HBV genome in 32 patients with chronic hepatitis B, 2011, **25**(3), 208–210.
38. Scott, P. *et al.*, Defective hepatitis B virus DNA is not associated with disease status but is reduced by polymerase mutations associated with drug resistance. *Hepatology*, 2008, **48**(3), 741–749.
39. Jung, M. C. *et al.*, Immune response of peripheral blood mononuclear cells to HBx-antigen of hepatitis B virus. *Hepatology*, 1991, **13**(4), 637–643.
40. Koh, H., Baek, S. Y. and Chung, K. S., Lamivudine therapy for Korean children with chronic hepatitis B. *Yonsei Med J*., 2007, **48**(6), 927–933.
41. Lin, S. Y. *et al.*, Association of response to hepatitis B vaccination and survival in dialysis patients. *BMC Nephrol.*, 2012, **30(**13), 97.
42. Minami, M., Poussin, K., Kew, M., Okanoue, T., Brechot, C. and Paterlini, P., Precore/core mutations of hepatitis B virus in hepatocellular carcinomas developed on noncirrhotic livers. *Gastroenterology*, 1996, **111**, 691–700.
43. Brind, A., Jiang, J, Samuel, D., Gigou, M., Feray, C., Bréchot, C. and Kremsdorf, D., Evidence for selection of hepatitis B mutants after liver transplantation through peripheral blood mononuclear cell infection. *J. Hepatol.*, 1997, **26**(2), 228–235.
44. Yuan, J. *et al.*, Implications. *Clin. Virol.*, 2007, **39**(2), 87–93.
45. Lee, C. A., Hope for haemophilic patients with hepatitis. *Gut*, 1996, **39**(6), 887–888.
46. Asim, M., Potukuchi, S. K., Arora, A., Singh, B. and Kar, P., Hepatitis-G virus infection in multi-transfused patients and intravenous drug abusers: New Delhi experience. *Digest. Dis. Sci*., 2008, **53**(5), 1383–1389.
47. Yuasa, R. *et al.*, Properties of hepatitis B virus genome recovered from Vietnamese patients with fulminant hepatitis in comparison with those of acute hepatitis. *J. Med. Virol*., 2000, **61**(1), 23–28.
48. Ghosh, S. *et al.*, Unique hepatitis B virus subgenotype in a primitive tribal community in eastern India. *J. Clin. Microbiol.*, 2010, **48**(11), 4063–4071.
49. Crowther, R. A., Kiselev, N. A., Böttcher, B., Berriman, J. A., Borisova, G. P., Ose, V. and Pumpens, P., Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell*, 1994, **77**(6), 943–950.
50. Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J. and Steven, A. C., Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature*, 1997, **386**(6620), 91–94.
51. Conway, J. F. *et al.*, Hepatitis B virus capsid: localization of the putative immunodominant loop (residues 78 to 83) on the capsid surface, and implications for the distinction between c and e-antigens. *J. Mol. Biol*., 1998, **279**, 1111–1121.
52. Zlotnick, A., Cheng, N., Conway, J. F., Booy, F. P., Steven, A. C., Stahl, S. J. and Wingfield, P. T., Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry*, 1996, **35**(23), 7412–7421.
53. Böttcher, B., Wynne, S. A. and Crowther, R. A., Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature*, 1997, **386**(6620), 88–91.
54. Okamoto, H., Imai, M., Tsuda, F., Tanaka, T., Miyakawa, Y. and Mayumi, M., 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*., 1987, **61**(10), 3030–3034.
55. Aron, M.-B. *et al.*, CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.*, 2013, **41**, D348–D352.
56. Ma, Y., Ding, Y., Juan, F. and Dou, X. G., Genotyping the hepatitis B virus with a fragment of the HBV DNA polymerase gene in Shenyang, China. *Virol. J.*, 2011, **22(**8), 315.
57. Bartenschlager, R., Junker-Niepmann, M. and Schaller, H., The P gene product of hepatitis B virus is required as a structural component for genomic RNA encapsidation. *J. Virol*., 1990, **64**(11), 5324–5332.
58. Gauthier, M. *et al.*, Microarray for hepatitis B virus genotyping and detection of 994 mutations along the genome. *J. Clin. Microbiol.*, 2010, **48**(11), 4207–4215.
59. Buckwold, V. E., Xu, Z., Yen, T. S. and Ou, J. H., Effects of a frequent double-nucleotide basal core promoter mutation and its putative single-nucleotide precursor mutations on hepatitis B virus gene expression and replication. *J. Gen. Virol.*, 1997, **78**(8), 2055–2065.
60. Kumar, M., Kumar, R., Hissar, S. S., Saraswat, M. K., Sharma, B. C., Sakhuja, P. and Sarin, S. K., Risk factors analysis for hepatocellular carcinoma in patients with and without cirrhosis: a case-control study of 213 hepatocellular carcinoma patients from India. *J. Gastroenterol. Hepatol.*, 2007, **22**(7), 1104–1011.
61. Saini, N., Bhagat, A., Sharma, S., Duseja, A. and Chawla, Y., Evaluation of clinical and biochemical parameters in hepatocellular carcinoma: experience from an Indian center. *Clin. Chim. Acta*, 2007, **371**(1–2), 183–186; Epub 2006 May 2; Erratum: *Clin. Chim. Acta*, **377**(1–2), 289.
62. Kramvis, A., Arakawa, K., Yu, M. C., Nogueira, R., Stram, D. O. and Kew, M. C., Virus. *Med. Virol.*, 2008, **80**(1), 27–46.

ACKNOWLEDGEMENT. We thank the University Grants Commission, New Delhi for funds.

Received 24 September 2015; accepted 14 October 2015

doi: 10.18520/cs/v111/i4/648-661