

Variability of the preC/C region of hepatitis B virus genotype A from a South African cohort predominantly infected with HIV

Simnikiwe H. Mayaphi,^a Desmond J. Martin,^a M. Jeffrey Mphahlele,^b Jason T. Blackard,^c Sheila M. Bowyer^a

^aDepartment of Medical Virology, University of Pretoria / Tshwane Academic Division of NHLS, Pretoria, South Africa. ^bDepartment of Virology, University of Limpopo, MEDUNSA campus / National Health Laboratory Service, Pretoria, South Africa. ^cDivision of Digestive Diseases, Department of Internal Medicine, University of Cincinnati College of Medicine.

Institution where this study was performed: University of Pretoria / Tshwane Academic Division NHLS.

Correspondence author:

Dr S.H. Mayaphi

Clinical Virologist

Department of Medical Virology

University of Pretoria

P.O. Box 2034, Pretoria, 0001

Tel: +27 12 319 2305

Fax: +27 12 325 5550

Email: sim.mayaphi@up.ac.za; smayaphi@absamail.co.za

Funding: National Health Laboratory Service (NHLS) and Poliomyelitis Research Foundation (PRF).

Conflict of interest: The authors have declared that no competing interests exist.

ABSTRACT

Hepatitis B virus (HBV) is a serious global health problem, and HBV genotype is an important determinant of disease progression and treatment outcome. The aim of this study was to assess variations of the precore/core (preC/C) region in HBV genotype A. Sequencing of the preC/C and surface (S) genes of HBV was performed on plasma samples from 20 HBV/HIV co-infected and 5 HBV mono-infected individuals. All preC/C study sequences clustered with subgenotype A1, except for 2 which clustered with subgenotype D4 reference strains. The nucleotide and amino acid variability was far higher in the preC/C region than in the S region. Mutations associated with reduction or failure of HBV e-antigen (HBeAg) production were observed, with a preC start codon mutation being common (24%). Other mutations (e.g. P5H/L and I97L) associated with severe liver disease were also noticed, some of which were located in the major histocompatibility restricted sites. PreC/C intergenotype nucleotide divergence was >7%, while subgenotypes differed by 2.5 – 7%. Several study sequences were highly divergent from other African subgenotype A1 strains. This study showed that HBeAg-negative chronic hepatitis B is underestimated in subgenotype A1, and also highlighted the variant South African A1 strains. The major advantage of preC/C sequencing is that it informs patient management as HBeAg-negative chronic hepatitis B responds poorly to conventional interferon- α therapy, and some guidelines treat HBeAg-negative chronic hepatitis B differently from HBeAg-positive chronic hepatitis B. These data suggest that subgenotype A1 may be more involved in severe HBV-related diseases.

INTRODUCTION

The global prevalence of chronic hepatitis B virus (HBV) infection is estimated at 240 million infections, and these infections contribute to ~600 000 deaths annually due to complications such as liver failure, cirrhosis and hepatocellular carcinoma [Ott et al., 2012]. HBV is an enveloped, 3.2 kb partially double-stranded DNA virus, which belongs to the *Hepadnaviridae* family [Faure, 2006]. Like retroviruses, HBV also uses a reverse transcriptase enzyme during replication, resulting in high HBV diversity [Olinger et al., 2006].

There are at least 8 genotypes of HBV designated A – H with an inter-genotype divergence of >8% on full length genome sequencing and >4% on S gene sequencing [Kramvis et al., 2005]. There is still controversy regarding two newly proposed genotypes – I and J [Kurbanov et al., 2008; Araujo et al., 2011]. Subgenotypes have been described in HBV genotypes A, B, C, D and F [Kramvis et al., 2005; Stanojević et al., 2011] with a genetic variation of 4 – 8%, while strains within a subgenotype exhibit <4% variation on full length genome sequencing [Kramvis et al., 2005]. Three HBV genotypes – A (subgenotypes A1 and A2), D (subgenotypes D3 and D4) and E – currently circulate in South Africa, although A1 is predominant [Kimbi et al., 2004; Kew et al., 2005; Andernach et al., 2009].

The HBV genome has four open reading frames (ORFs): preS/S ORF coding for the 3 species of surface antigen; P ORF coding for polymerase; preC/C ORF coding for e and core antigens; and X ORF coding for X protein [Ganem and Prince, 2004]. Mutations within these ORFs could have clinical, diagnostic or therapeutic implications [Liang et al., 1991; Kay and Zoulim, 2007; Alavian et al., 2012]. For instance, the preC/C region contains both B and T cell epitopes; hence,

mutations within this region often lead to escape of immune responses [Sendi et al., 2009]. Additionally, mutations that affect HBeAg expression lead to phenotypic changes such as faster progression of clinical disease [Ganem and Prince, 2004; Faure, 2006].

HBeAg confers protection to hepatocytes against the cytotoxic T lymphocyte (CTL) responses; thus, HBV strains that do not express HBeAg are usually associated with active hepatitis at relatively lower HBV viral loads [Araujo et al., 2011]. The emergence of HBeAg-negative chronic hepatitis B often occurs during the natural course of infection with a wild type virus that expresses HBeAg, and usually appears during the immune clearance phase (HBeAg loss or seroconversion) of chronic HBV infection [Hadziyannis and Vassilopoulos, 2001]. The spontaneous HBeAg seroconversion is known to occur earlier in subgenotype A1 than in A2 during the natural history of chronic hepatitis B [Hadziyannis, 2011]. However, the prevalence of HBeAg-negative chronic hepatitis B due to precore mutations is reported as common in non-A genotypes only (B, C, D, E and G) [Funk et al., 2002]. This study aimed to assess the variations of preC/C region in an area where HBV subgenotype A1 infections are predominant.

METHODS

Study samples

This study included 25 HBV PCR-positive individuals with known serology and viral load results (Table I) from a cohort that was studied for the prevalence of HBV in HIV co-infected patients with AIDS and HIV-negative controls [Mayaphi et al., 2012]. None of these individuals had received HBV therapy at the time of enrollment into the above mentioned study; even those

co-infected with HIV were enrolled before anti-retroviral drug initiation. All the study samples were stored at -70°C within 24 – 72 hours after collection.

TABLE I. HBV Serology and Viral Load Results of All Study Participants

PTID	Gender	HBsAg	Anti-HBs	Anti-HBc	HBeAg	Anti-HBe	ALT	HIV status	HBV VL ^a
3354	F	+	-	+	+	-	31	+	3×10^7
3658	M	+	-	+	-	+	29	+	2×10^7
3261	M	+	-	+	+	-	42	+	$>1 \times 10^8$
2950	M	+	-	-	+	-	18	+	2×10^6
3274	F	+	-	+	+	-	60	+	1×10^7
4070	M	+	-	+	+	-	46	+	$>1 \times 10^8$
3791	M	+	-	+	+	-	58	+	$>1 \times 10^8$
3464	M	+	-	+	-	+	26	+	278
3321	M	+	-	+	-	+	56	+	6×10^4
3269	F	+	-	+	-	-	21	+	95
4312	F	+	-	+	+	-	61	+	7×10^6
3869	M	+	-	-	+	-	50	+	1×10^5
3975	F	+	-	+	+	-	19	+	2×10^5
3678	F	-	+	+	-	-	16	+	290
3319	F	-	-	+	-	-	35	+	127
T33	M	-	-	+	-	-	57	+	LDL
3734	M	-	-	+	-	-	49	+	36
3832	M	-	-	+	-	-	41	+	126
3671	F	-	-	+	-	-	51	+	44
3812	M	-	-	+	-	-	30	+	LDL
N005	M	+	-	+	-	+	21	-	2×10^3
N011	M	+	-	+	+	-	28	-	$>1 \times 10^8$
N060	F	+	-	+	-	+	44	-	157
N012	M	-	-	+	-	-	25	-	LDL
N199	M	-	-	+	-	-	30	-	99

^aAll participants had a positive nested PCR for the preC/C region of hepatitis B virus (HBV). PTID, participant identity; F, female; M, male; HBsAg, hepatitis B surface antigen; anti-HBc, hepatitis B core antibodies; anti-HBs, hepatitis B surface antibodies; HBeAg, Hepatitis B e-antigen; anti-Hbe, hepatitis B e-antibodies; ALT, alanine aminotransferase (IU/L); +, positive; -, negative; LDL, less than detectable limit; VL, viral load.

DNA extraction and amplification

Viral nucleic acid was extracted from 100 μl of plasma using the MagNa Pure LC Total Nucleic Acid Isolation kit in the MagNa Pure LC instrument (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions. HBV preC/C primers were used for nested HBV-PCR as previously described [Mphahlele et al., 2006]. Gel electrophoresis was used for visualization of 700 bp PCR products. The nested S gene PCR was performed at Inqaba Biotechnical company using published primers [Mphahlele et al., 2006].

DNA Sequencing PCR

Sequencing was performed at Inqaba Biotechnical company. The PCR amplicons were purified enzymatically using Exonuclease I and Shrimp Alkaline Phosphatase (Fermentas Life Sciences, GE Healthcare, USA) as previously described [Werle et al., 1994]. Labelling was performed using BigDye Direct Cycle Sequencing kit V3.1 from ABI (Life Technologies, Carlsbad, USA) according to the manufacturer's specifications. The labelled reactions were purified with ZR-96 DNA sequencing clean-up kit (Zymo Research, Irvine, USA) and analysed on the ABI 3500 XL genetic analyser (Life Technologies, Foster city, USA).

The following sense 5'GCTGTAGGCATAAATTGGT3' (nucleotides 1780-1798) and anti-sense 5'GAGGCGAGGGAGTTC3' (nucleotides 2382-2396) primers were used for preC/C sequencing with DreamTaq DNA polymerase (Thermo Scientific, Foster city, USA) for 40 cycles at 50°C annealing temperature. Genotypes were confirmed on S gene sequencing using the following sense and anti-sense primers, respectively: 5'CCTGCTGGTGGCTCCAGTTC3' (nucleotides 56-75) and 5'CGAACCACTGAACAAATGGC3' (nucleotides 685-704).

Phylogenetic analysis

Non-recombinant reference strains representing all HBV genotypes were obtained from National Centre for Biotechnology Information (NCBI) website for comparison with study sequences. Similar references were used for both preC/C and S gene data sets, and 3 more references, which were only S gene sequences were included in the S gene data set. MAFFT version 6 programme was used for sequence alignment [Kato and Toh, 2008], and results were confirmed on MEGA version 5.05 programme [Tamura et al., 2011]. Phylogenetic tree construction was done on

MEGA using Neighbor-Joining statistical method (with a 1000 bootstrap replicates). Phylogenetic inference was also performed using a Bayesian Markov chain Monte Carlo (MCMC) approach as implemented in the BEAST v1.5.0 program as described previously [Martin et al., 2011]. The dN/dS ratios and pairwise comparison data were obtained from MEGA. Relatedness of sequences was further characterized on a Histogram programme using pairwise comparison data.

RESULTS

Of the 25 samples analysed in this study, 20 were from HBV/HIV co-infected participants with AIDS {average CD4 count = 93,5 cells/ μ l (range: 2 – 185)}, and 5 were from HBV mono-infected participants (Table I). The average age of all participants was 35 years (range: 20 – 52). PreC/C gene was successfully amplified in all 25 samples, while S gene was amplified from 11 (44%) samples with no improvement in recovery after a second PCR. Investigation on poor S gene primers' performance revealed little variation in 3' primer binding sites but significant differences in primer pair melting points. Further testing could not be done due to insufficient samples.

PreC/C gene sequencing

542 nucleotides of the preC/C gene were analysed, starting from position 1799 (i.e. 15 bases before the start of the preC/C) to position 2340. Phylogenetic analysis showed that 23 sequences belong to genotype A, while two sequences clustered with genotype D references (Fig. 1A). A significant proportion (39.1% [n = 9]) of genotype A study sequences (N199, 3274, N60, 3269,

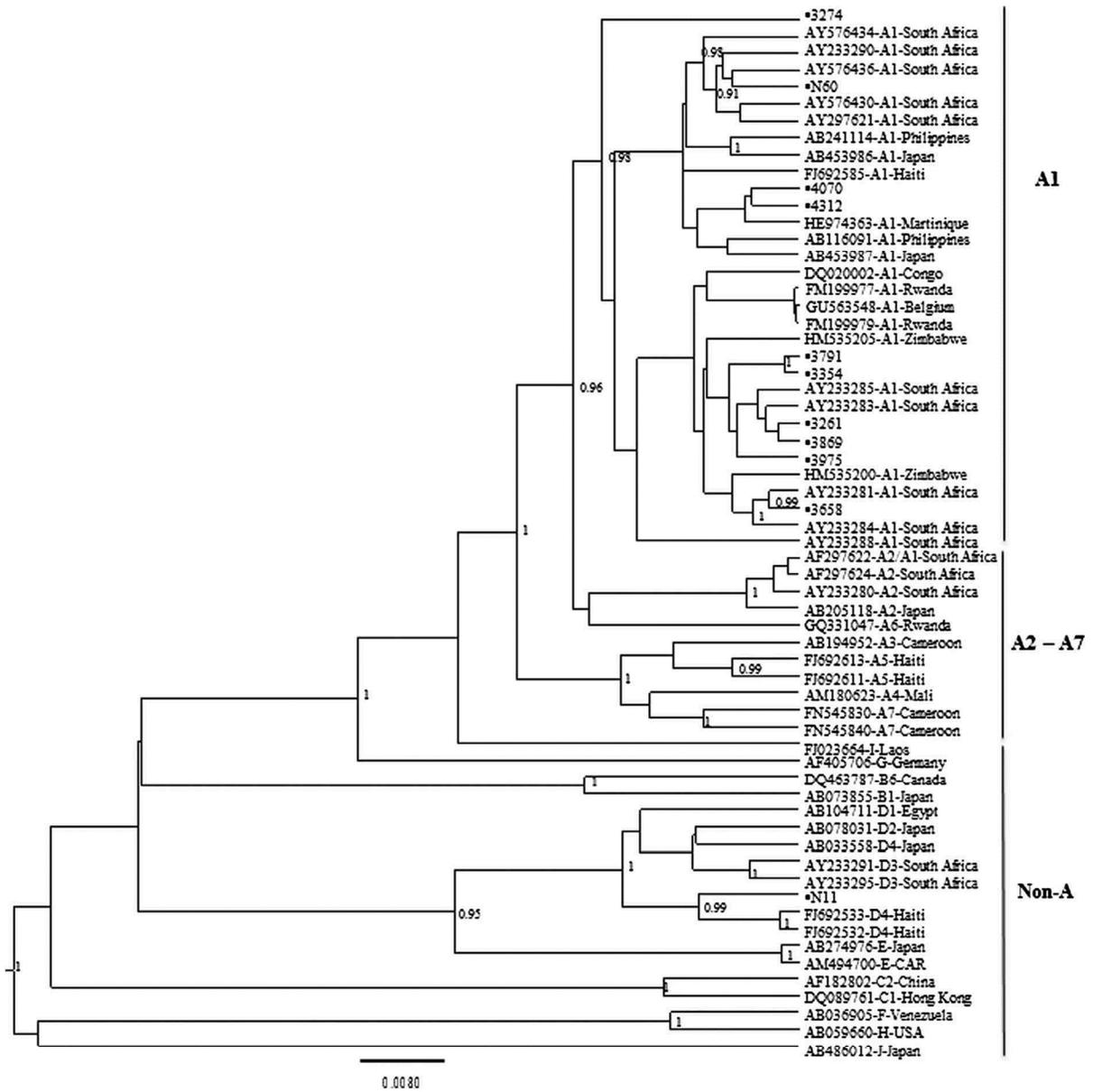


Fig. 1. (Continued).

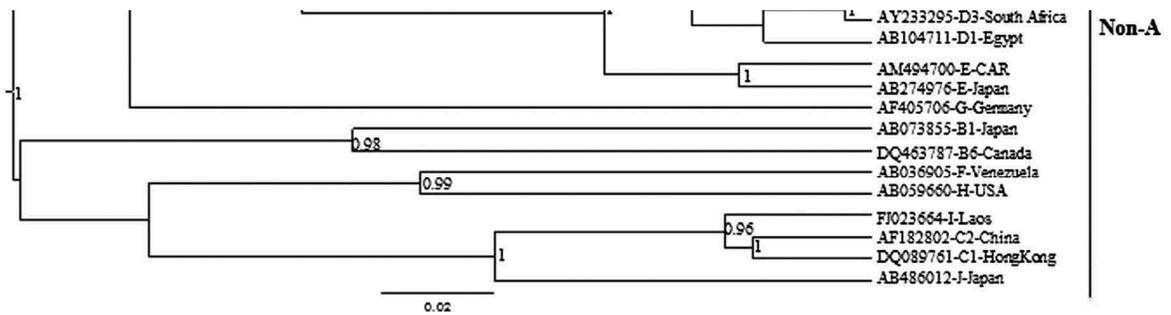


Fig. 1. Phylogenetic analysis of: (A) the preC/C gene and (B) S gene sequences using a Bayesian inference approach. Relevant posterior probabilities >0.9 are shown. CAR, Central African Republic; USA, United States of America. Study sequences are highlighted with a square shaped bullet. S gene dataset had three more reference strains, which were only S gene sequences, hence they were not used in the preC/C data set.

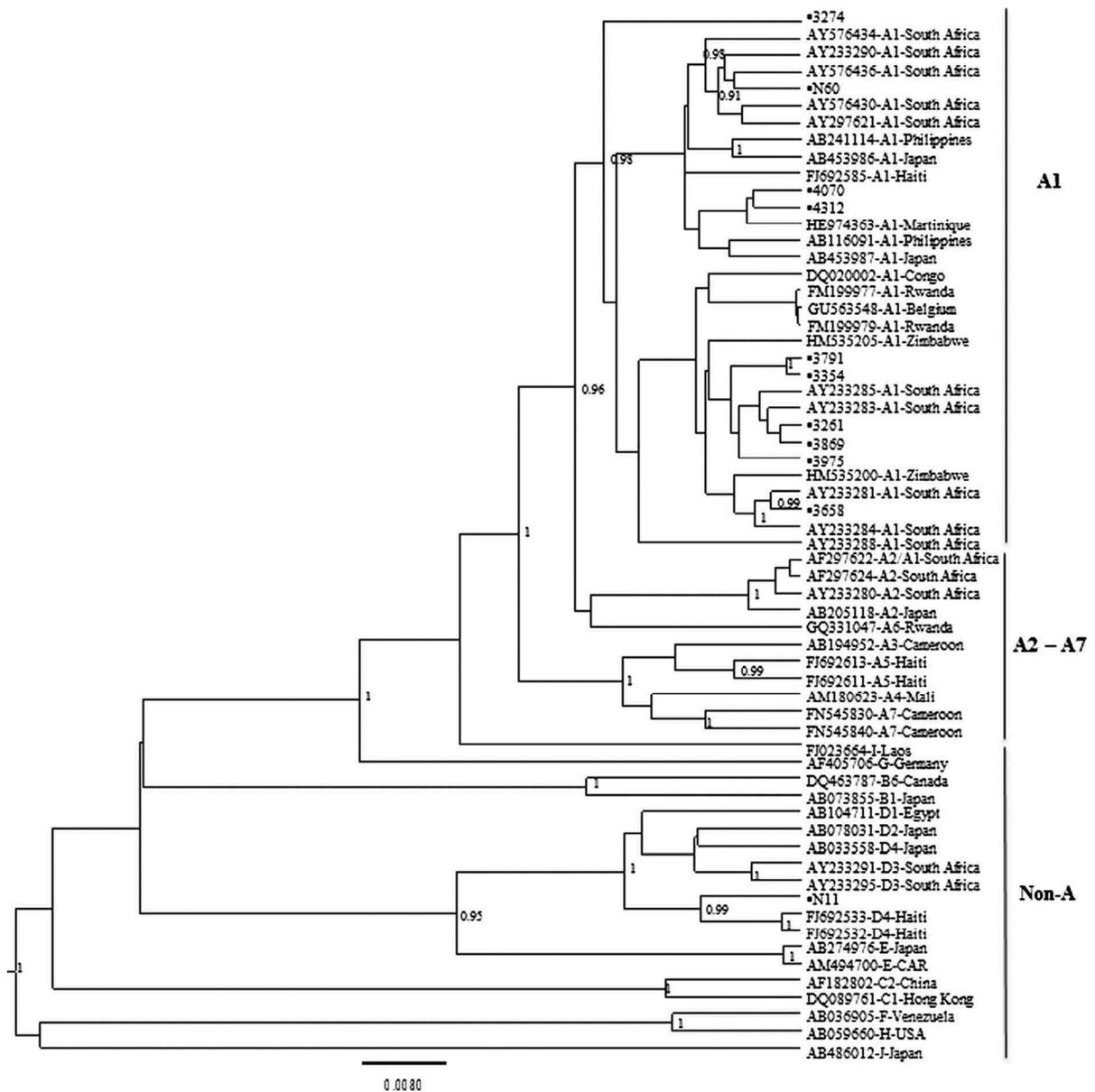


Fig. 1. (Continued).

3319, N005, 4070, 4312 and 3658) partitioned away from typical African A1 references and had longer branch lengths (more visible on Neighbour-Joining phylogenetic tree – data not shown). Sequences from participants N60, 3269, 3319 and N005 clustered with South African A1

reference strains; 4070 and 4312 clustered with Asian and Caribbean A1 reference strains; 3658 also clustered with South African A1 reference strains; while N199 and 3274 formed a separate cluster (Fig. 1A). Interestingly, one South African subgenotype A2 reference strain (AF297622) clustered with unrelated strains in the preC/C region indicating that it is a recombinant (Fig. 1A & B).

A summary of nucleotide and amino acid sequence alignment data was prepared by excluding the areas of similarity amongst all sequences so as to highlight the differences between study sequences and references. As expected, nucleotide data showed that all the genotype A study sequences had a C at 1858, while the two genotype D samples had a T at this position. None of the study sequences had a G1896A stop codon mutation.

Six (24%) of the study sequences had preC start codon mutations: three (3321, T033 and N60) had A1814C, two (3354 and 3658) had A1814T, and one (3269) had G1816T. This led to translation of a different amino acid instead of a start codon amino acid (methionine). All but one (N60) of these participants had HIV/AIDS. Two of the participants with preC start codon mutation from the AIDS group had abnormal alanine transferase (ALT), and N60 was the only participant from the HIV uninfected group with abnormal ALT (Table I).

Several random nucleotide and amino acid mutations were noticed, and trends of similarity were found amongst some sequences. More striking were nucleotides which defined the long branched clusters notably the AY233290 cluster containing N60, 3269, 3319 and N005; and the cluster containing N199 and 3274 (Table II). Some of these nucleotides were signature nucleotides while some were commonly found in non-A genotypes. Some mutations noted in this study were

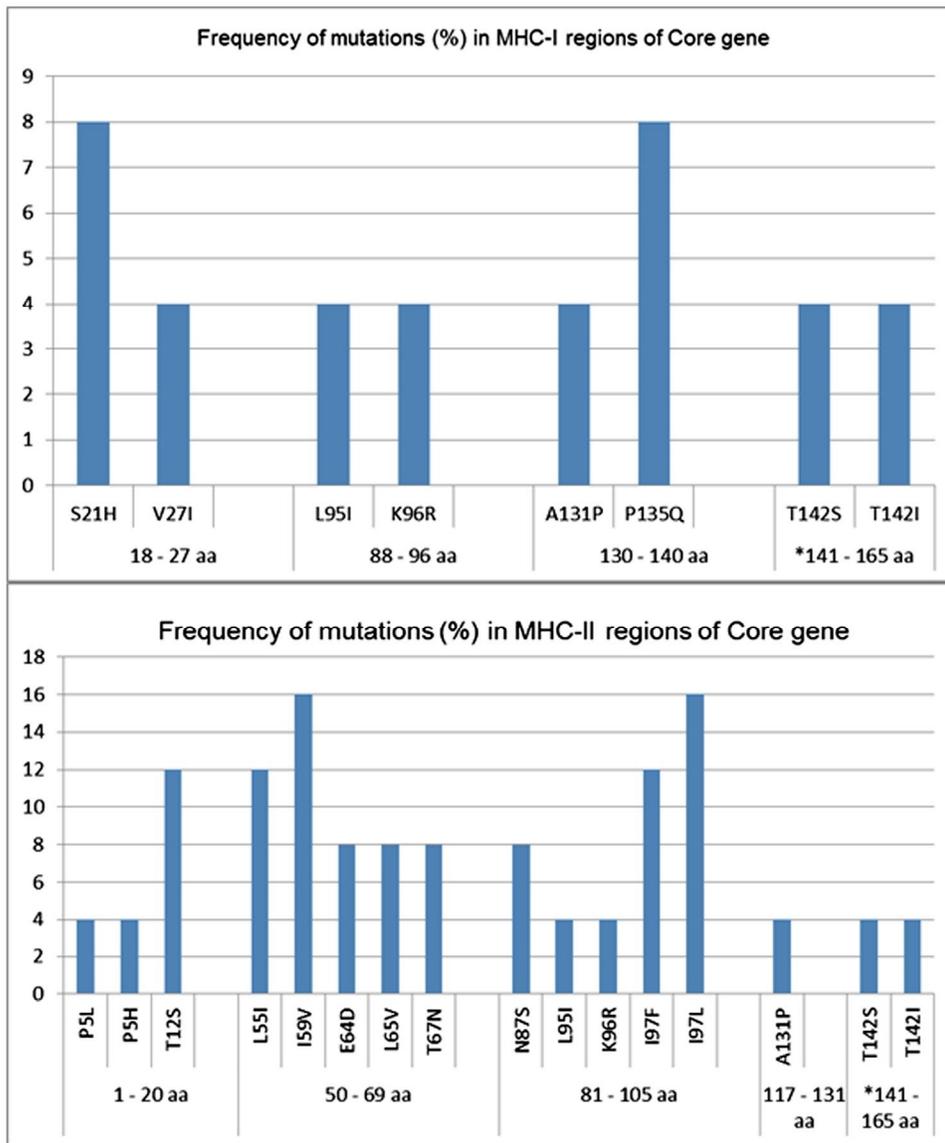


Fig. 2. Frequency of mutations in the major histocompatibility complex class I and II (MHC-I and -II) restricted sites of the Core gene. aa, amino acids, *study sequences ended at 146th amino acid (39 aa short from the end of core gene).

located in the major histocompatibility complex classes I and II (MHC-I and II) sites within the core region. Most of these mutations were commonly found in the MHC-II regions (Fig. 2).

The dN/dS ratio analysis for different codons amongst sequences did not show significant non-synonymous changes, except for the 12th codon of preC region, which was marginally significant ($p = 0.058$) (data not shown).

Pairwise sequence data analysis showed a preC/C nucleotide variation of >7% for inter-genotype strains, 2.5% to 7% for inter-subgenotype strains, and <2.5% for intra-subgenotype strains (Fig. 3). The pairwise distance points at which the nine divergent (atypical) study strains (N199, 3274, N60, 3269, 3319, N005, 4070, 4312 and 3658) meet others were plotted separately on the graph to further analyse their distances from other sequences. The majority of these points fell within the inter-subgenotype area, suggesting that most of them are different from other A1 strains, with the exception of 4070 and 4312 that appeared in the intrasubgenotype area (Fig. 3).

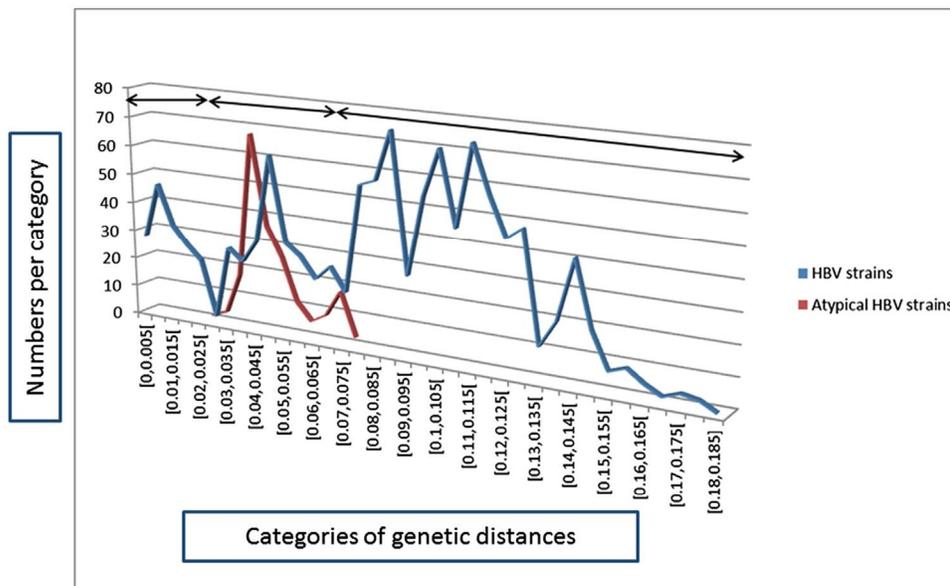


Fig. 3. Histogram data showing the distribution of pairwise genetic distances among study and reference sequences. The arrows demarcate different ranges for the intra-subgenotype (0.0 to <0.25), inter-subgenotype (0.25 to <0.07), and intergenotype isolates (≥ 0.07). The genetic distances for the atypical study sequences (N199, 3274, N60, 3269, 3319, N005, and 3658) were plotted separately (in red), and can be seen to fall in the inter-subgenotype section.

S gene sequencing

566 nucleotides of the S gene were analysed, from 76 – 641 nucleotide positions. Of the 11 study sequences, 10 clustered with genotype A while 1 clustered with genotype D reference strains. S gene sequencing revealed the same trends noticed on preC/C data. For instance, N60 still clustered with the same South African A1 references as in the preC/C phylogenetic tree together with three additional references in the S gene dataset (Fig. 1B). S gene sequences showed much less variability compared to preC/C sequences (Table II).

DISCUSSION

This is the first study to report on the criteria for assigning HBV strains into genotypes and subgenotypes based on preC/C sequencing. These criteria were validated against reference strains of HBV, and sequencing of the S gene, which is the only sub-genomic region commonly used for HBV genotyping [Kramvis et al., 2005]; and are similar to criteria described by Simmonds *et al* to define HCV genotypes [Simmonds et al., 1993]. This study also reports on 2 clusters of HBV within subgenotype A strains, which had intersubgenotype genetic variation indicating variants (or putative subgenotypes) uncharacterized previously (Table II, Fig. 1 and Fig. 3).

Nucleotide and amino acid sequence variability

There was high nucleotide variability in the preC/C region, with signature nucleotide and amino acid mutations identified for some study sequences (Table II). Variability did not seem to differ in those infected with HIV compared to the uninfected group; however, further studies would have to be conducted to elucidate the role of HIV in HBV mutations. The higher variability of the preC/C region has been reported in other studies that also genotyped HBV using both preC/C and S genes [Hur et al., 1996; Olinger et al., 2006; Malik et al., 2012].

Mutations affecting HBeAg expression

The finding that 24% (6) of the study participants had preC start codon mutations (A1814C/T and G1816T) suggests that mutations that lead to failure of HBeAg production are common in subgenotype A1. This phenomenon is well documented in participant 3269 who did not express HBeAg at all (Table I). Other participants (except 3354) with preC start codon mutations seem to have developed these mutations during the course of infection with a wild type HBV strain, because they were anti-HBe-positive. This is consistent with what is described to be the typical course of HBeAg-negative chronic hepatitis B, where HBeAg mutants develop as a result of immune pressure during the immune clearance phase of chronic HBV [Hadziyannis and Vassilopoulos, 2001]. I97L and A131P amino acid mutations have been associated with failure to express HBeAg [Jazayeri et al., 2004; Kim et al. 2012]; both these mutations have been found in this study as well mostly in patients who did not express HBeAg (3812, 3464, 3734 and 3319) (Fig. 2 and Table I). A G1862T nucleotide mutation, which was identified in two study sequences (4070 and 4312) has also been reported to negatively affect HBeAg expression at

post-translational level, however, both these participants were HBeAg-positive. Only N199 and 3274 sequences had a Kozak sequence mutation (T1812C), which is also known to interfere with HBeAg expression [Kew et al., 2005]. Co-infection with wild type and HBeAg-mutant HBV viruses or viral quasispecies could explain the presence of HBeAg in participants with mutations that are known to reduce or abolish HBeAg [Hadziyannis and Vassilopoulos, 2001].

Mutant HBV strains that do not express HBeAg are usually associated with active hepatitis at relatively lower HBV viral load [Araujo et al., 2011]. It is, however, difficult to assess the impact of these mutants in this study as most participants had AIDS and, therefore, had weak CTL responses. Amongst the participants lacking a preC start codon, ALT levels were abnormal in two of the 5 participants with AIDS (3321 and T33), while participant N60 was the only one with abnormal ALT in the HIV uninfected group (Table I). Although opportunistic infections could also account for ALT abnormalities in patients with AIDS, participant N60 exemplifies the expected course of HBeAg mutants in the setting of robust CTL immune responses [Araujo et al., 2011].

Most published data state that HBeAg-negative chronic hepatitis B due to preC mutations is common in non-A genotypes only, and that mutations that often lead to reduced production of HBeAg in genotype A are confined to the core promoter region (A1762T and G1764A) [Hadziyannis and Vassilopoulos, 2001; Funk et al., 2002]. However, preC start codon mutations noticed in this study have also been reported in recently published data from South Africa (A1814C/T, T1815C/A) [Makondo et al., 2012], and in patients from Cameroon who were infected with HBV genotype A3 (A1814C and A1814T) [Olinger et al., 2006]. This shows that HBeAg-negative chronic hepatitis B prevalence in genotype A is underestimated, particularly in

sub-Saharan Africa [Funk et al., 2002]. In this study, the preC start codon mutations were only noticed in subgenotype A1 study strains and one A5 reference strain (FJ692613). This could imply that preC start codon mutations are rare in non-A HBV genotypes.

Clinical implications of other preC/C mutations

The random mutations noticed in this study could be due to the gradual accumulation of mutations during the course of chronic HBV infection driven by immune selection pressure as the preC/C region harbours both B and T cell epitopes [Hur et al., 1996]. The higher frequency of mutations in the MHC-II restricted sites as opposed to MHC-I sites in this study (Fig. 2) corroborates with a recent report [Kim et al., 2012]. The preC/C mutations have been shown to have a negative impact on the disease progression and response to treatment [Buti et al., 2005]. G1899A found in participant 3319 has been reported in patients with fulminant hepatitis failure and has also been found in high frequency in hepatocellular carcinoma (HCC), liver cirrhosis and chronic hepatitis [Malik et al., 2012].

P5H/L and I97L amino acid mutations in genotype C of HBV (all of which were noticed in this study) were recently found to be associated with HCC [Kim et al., 2012], and I97L was found at a higher frequency (16%) among study participants (Fig. 2). These mutations could partially explain the high prevalence of HCC in South Africa [Kew et al., 2005], however, this needs to be confirmed as none of the participants in this study had symptoms of or had been diagnosed with HCC.

Genotyping with preC/C region

Pairwise data analysis showed that HBV genotypes differ by >7% on preC/C nucleotide sequencing. This variation is 2.5 – 7% for inter-subgenotype isolates and <2.5% for isolates of the same subgenotype (Fig. 3). These data provide a good alternative for genotyping using a different sub-region of HBV genome than the S gene. PreC/C genotyping provides excellent opportunities to identify mutations that are able to guide clinicians on patient management. For instance, the American Association for the Study of Liver Diseases guidelines have separate provisions for the treatment of HBeAg-negative chronic hepatitis B and HBeAg-positive chronic hepatitis B [Lok and McMahon 2009]. Furthermore, the HBeAg-negative chronic hepatitis B is associated with poor response to conventional interferon- α therapy, thus preC/C genotyping could inform the choice of drugs for treatment and predict treatment response where interferon- α therapy will be used [Paik et al., 2010]. PreC/C genotyping could also aid researchers to better understand the prevalence of HBeAg-negative chronic hepatitis B in all HBV genotypes, particularly in genotype A, where it is not well described [Hadziyannis and Vassilopoulos, 2001; Funk et al., 2002]. The preC/C region is a hot spot for recombination in HBV [Simmonds and Midgley, 2005; Shi et al., 2012], therefore, genotyping in more than one region (including preC/C) is recommended. It was interesting that a South African subgenotype A2 reference strain was found to be a recombinant in this study (AF297622 – A2/A1 recombinant) (Fig. 1).

Variant South African A1 strains

Highly divergent subgenotype A1 strains were noticed in this study in a cluster containing N60, 3269, 3319 and N005; and the other one with N199 and 3274 strains (Fig. 1 and Table II). All

these strains differed from other subgenotype A strains (outside their clusters) by >2.5% in the preC/C region (Fig. 3). Interestingly, the South African A1 references (AY233290 and AF297621) that are in the cluster containing N60, 3269, 3319 and N005 are 4% different from other genotype A strains according to full length genome BLAST search on NCBI. Genetic variation of 4 – 8% on full length sequencing describes subgenotype differences [Kramvis et al., 2005]. Data from other South African studies have also shown HBV strains that formed a distinct cluster with AY233290 reference strain [Kew et al., 2005; Makondo et al., 2012]. The variant A1 strains observed in this study could represent recombinant or new subgenotype A strains. Ultra-deep full length genome sequencing is currently being done for full characterisation of all these variant strains.

Conclusion

This study has for the first time defined preC/C genotyping criteria and highlighted the circulating sub-genotype A1 variants in South Africa. It has also highlighted the prevalence of HBeAg-negative chronic hepatitis B in subgenotype A1 strains, which is underestimated in global studies. Other mutations that are associated with severe hepatitis B disease were observed. These data suggest that subgenotype A1 might actually be more involved in severe HBV related diseases as opposed to the data that generally associates genotype A with milder disease [Raimondi et al., 2010].

The limitations of this study include small sample size and failure to amplify some sequences in the S region, and that most participants were co-infected with HIV. Also the basal core promoter region of HBV genome was not sequenced in this study, therefore, the prevalence of A1762T

and G1764A mutations could not be determined. Sequencing in this study did not cover the full subgenomic regions of HBV but ultra-deep complete genome sequencing of all study sequences is currently underway and will be used to fully characterize the variant South African A1 strains.

Table II footnotes:

X denotes the presence of a nucleotide or amino acid. The top part of the figure shows NTs and AA variation in the preC/C region and the bottom part with shaded boxes shows variation in the S region.

^a = A2/A1 recombinant found in this study, ^b = typical African A1, ^c = typical Asian A1, ¹ = amino acid located on preS2 region, ² = amino acid(s) located on small S region. All the numbering was based on GenBank sequence accession number X02763.

* = signature nucleotides or amino acids, shaded areas - denote the samples that could not be amplified on the S region.

REFERENCES

Alavian SM, Carman WF, Jazayeri SM. 2012. HBsAg variants: Diagnostic-escape and diagnostic dilemma. *J Clin Virol*: <http://dx.doi.org/10.1016/j.jcv.2012.04.027>

Andernach IE, Nolte C, Pape JW, Muller CP. 2009. Slave Trade and Hepatitis B Virus Genotypes and Subgenotypes in Haiti and Africa. *Emerg Infect Dis* 15:1222-1228.

- Araujo NM, Waizbort R, Kay A. 2011. Hepatitis B virus infection from an evolutionary point of view: How viral, host, and environmental factors shape genotypes and subgenotypes. *Infect Genet Evol* 11:1199-1207.
- Buti M, Rodriguez-Frias F, Jardi R, Esteban R. 2005. Hepatitis B virus genome variability and disease progression: the impact of pre-core mutants and HBV genotypes. *J Clin Virol* 34: S79-82.
- Faure E. 2006. Alternative peptide-fusion proteins generated by out-of-frame mutations, just upstream ORFs or elongations in mutants of Human Hepatitis B Viruses. *Virus Res* 117:185-201.
- Funk ML, Rosenberg DM, Lok AS. 2002. World-wide epidemiology of HBeAg-negative chronic hepatitis B and associated precore and core promoter variants. *J Viral Hepat* 9:52-61.
- Ganem D, Prince AM. 2004. Hepatitis B virus infection--natural history and clinical consequences. *N Engl J Med* 350:1118-1129.
- Hadziyannis SJ, Vassilopoulos D. 2001. Hepatitis B e antigen—negative chronic hepatitis B. *Hepatology* 34:617-624.
- Hadziyannis SJ. 2011. Natural history of chronic hepatitis B in Euro-Mediterranean and African Countries. *J Hepatol* 55:183-191.

- Hur GM, Lee YI, Suh DJ, Lee JH, Lee YI. 1996. Gradual accumulation of mutations in precore core region of HBV in patients with chronic active hepatitis: implications of clustering changes in a small region of the HBV core region. *J Med Virol* 48:38-46.
- Jazayeri SM, Basuni AA, Sran N, Gish R, Cooksley G, Locarnini S, Carman WF. 2004. HBV core sequence: definition of genotype-specific variability and correlation with geographical origin. *J Viral Hepat* 11:488-501.
- Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9:286-298.
- Kay A, Zoulim F. 2007. Hepatitis B virus genetic variability and evolution. *Virus Res* 127:164-176.
- Kew MC, Kramvis A, Yu MC, Arakawa K, Hodgkinson J. 2005. Increased hepatocarcinogenic potential of hepatitis B virus genotype A in Bantu-speaking sub-saharan Africans. *J Med Virol* 75:513-521.
- Kim DW, Lee SA, Hwang ES, Kook YH, Kim BJ. 2012. Naturally Occurring Precore/Core Region Mutations of Hepatitis B Virus Genotype C Related to Hepatocellular Carcinoma. *PLoS One* 7(e47372):1 - 8.
- Kimbi GC, Kramvis A, Kew MC. 2004. Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa. *J Gen Virol* 85:1211-1220.
- Kramvis A, Kew M, François G. 2005. Hepatitis B virus genotypes. *Vaccine* 23:2409-2423.

- Kurbanov F, Tanaka Y, Kramvis A, Simmonds P, Mizokami M. 2008. When should "I" consider a new hepatitis B virus genotype? *J Virol* 82:8241-8242.
- Liang TJ, Hasegawa K, Rimon N, Wands JR, Ben-Porath E. 1991. A hepatitis B virus mutant associated with an epidemic of fulminant hepatitis. *N Engl J Med* 324:1705-1709.
- Lok AS, McMahon BJ. 2009. Chronic hepatitis B: update 2009. *Hepatology* 50:661-662.
- Makondo E, Bell TG, Kramvis A. 2012. Genotyping and molecular characterization of hepatitis B virus from human immunodeficiency virus-infected individuals in southern Africa. *PLoS One* 7(e46345):1-12.
- Malik A, Singhal DK, Albanyan A, Husain SA, Kar P. 2012. Hepatitis B Virus Gene Mutations in Liver Diseases: A Report from New Delhi. *PLoS One* 7(e39028):1-7.
- Martin CM, Welge JA, Blackard JT. 2011. Hepatitis B virus (HBV) X gene diversity and evidence of recombination in HBV/HIV co-infected persons. *J Med Virol* 83:1142-1150.
- Mayaphi SH, Roussow TM, Masemola DP, Olorunju SA, Mphahlele MJ, Martin DJ. 2012. HBV/HIV co-infection: the dynamics of HBV in South African patients with AIDS. *S Afr Med J* 102:157-162.
- Mphahlele MJ, Lukhwareni A, Burnett RJ, Moropeng LM, Ngobeni JM. 2006. High risk of occult hepatitis B virus infection in HIV-positive patients from South Africa. *J Clin Virol* 35:14-20.

- Olinger CM, Venard V, Njayou M, Oyefolu AO, Maiga I, Kemp AJ, Omilabu SA, le Faou A, Muller CP. 2006. Phylogenetic analysis of the precore/core gene of hepatitis B virus genotypes E and A in West Africa: new subtypes, mixed infections and recombinations. *J Gen Virol* 87:1163-1173.
- Ott JJ, Stevens GA, Groeger J, Wiersma ST. 2012. Global epidemiology of hepatitis B virus infection: New estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine* 30:2212-2219.
- Paik YH, Kim JK, Kim DY, Park JY, Ahn SH, Han KH, Chon CY, Lee KS. 2010. Clinical Efficacy of a 24-months Course of Lamivudine Therapy in Patients with HBeAg Negative Chronic Hepatitis B: A Long-term Prospective Study. *J Korean Med Sci* 25:882-887.
- Raimondi S, Maisonneuve P, Bruno S, Mondelli MU. 2010. Is response to antiviral treatment influenced by hepatitis B virus genotype? *J Hepatol* 52:441-449.
- Sendi H, Mehrab-Mohseni M, Shahraz S, Norder H, Alavian SM, Noorinayer B, Zali MR, Pumpens P, Bonkovsky HL, Magnus LO. 2009. CTL escape mutations of core protein are more frequent in strains of HBeAg negative patients with low levels of HBV DNA. *J Clin Virol* 46:259-264.
- Shi W, Carr MJ, Dunford L, Zhu C, Hall WW, Higgins DG. 2012. Identification of novel inter-genotypic recombinants of human hepatitis B viruses by large-scale phylogenetic analysis. *Virology* 427:51-59.

- Simmonds P, Holmes EC, Cha TA, Chan SW, McOmish F, Irvine B, Beall E, Yap PL, Kolberg J, Urdea MS. 1993. Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *J Gen Virol* 74:2391-2399.
- Simmonds P, Midgley S. 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. *J Virol* 79:15467-15476.
- Stanojević B, Osiowy C, Schaefer S, Bojović K, Blagojević J, Nešić M, Yamashita S, Stamenković G. 2011. Molecular characterization and phylogenetic analysis of full-genome HBV subgenotype D3 sequences from Serbia. *Infect Genet Evol* 11:1475-1480.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731-2739.
- Werle E, Schneider C, Renner M, Volker M, Fiehn W. 1994. Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acids Res* 22:4354-4355.