Deletions and recombinations in the core region of hepatitis B virus genotype E strains from asymptomatic blood donors in Guinea, west Africa

Penelope Garmiri, André Loua, Nyankoye Haba, Daniel Candotti and Jean-Pierre Allain

1Division of Transfusion Medicine, Department of Haematology, University of Cambridge, Cambridge, UK
2National Blood Transfusion Center, Conakry, Guinea
3National Health Service Blood and Transplant, Cambridge Blood Centre, Cambridge, UK

The prevalence of hepatitis B virus (HBV) surface antigen (HBsAg) chronic carriage in west Africa is the highest in the world, but its molecular epidemiology remains relatively poorly investigated. Plasma samples from random asymptomatic carriers of HBsAg in Conakry, Guinea, were studied and the complete genome sequences of 81 strains were obtained. Three additional samples from Kumasi, Ghana, were also included in the analysis. Phylogenetic analyses confirmed the dominance of genotype E (95.1%), including 8.6% of strains (viral load, $5 \times 10^8$ to $2.6 \times 10^9$ IU ml$^{-1}$) comprising dominant variants with large deletions in the core region and minority wild-type variants. The presence of two different patterns of deletions in two and four donors suggested targeted genome fragility between nt 1979 and 2314. The remaining sequences included one subgenotype A3 (1%) and six A/E recombinant forms (4–7%). A/E strains with identical points of recombination in three donors suggested strongly that these recombinant HBV strains are circulating and transmitted in the population. Recombination points were concentrated in the core gene. The detection of similar A/E recombinant strains in Ghana suggested a geographical extension of recombinant HBV to the region. The quasispecies of one additional Ghanaian strain sequenced in the pre-S/S region resolved into dominant clones of either the A or E genotype, but also three different patterns of A/E recombinant variants. The observation that both deletions of genotype E strains and A/E recombination points are mostly located in the core gene at specific positions indicates a region of the genome where genetic rearrangements preferentially take place.

INTRODUCTION

Hepatitis B virus (HBV) is the agent of the most common blood-borne virus infection in west Africa. In this region, 10–25% of the general population carry chronic infection, indicated by the presence of circulating hepatitis B surface antigen (HBsAg). The prevalence of antibodies against hepatitis B core antigen (anti-HBc), indicating contact with the virus, is >75% in the adult population (Allain et al., 2003). Such prevalences are the highest in the world and represent a serious threat to the population, as HBV causes cirrhosis and hepatocellular carcinoma (HCC), and to the safety of the blood supply.

The history of HBV infection in west Africa is unique, as HBV genotype E is overwhelmingly dominant in the region, but of relatively recent origin (Mulders et al., 2004; Candotti et al., 2006). Preliminary phylogenetic information suggests a spread over approximately 300 years (Mulders et al., 2004). This estimation is supported historically by the fact that HBV subgenotype A1 is prevalent in populations of African Americans studied in Brazil (Araujo et al., 2004; Motta-Castro et al., 2005), suggesting that the spread of HBV genotype E occurred after the end of the slave trade. It also suggests that the prevalent HBV genotype in west Africa prior to the mid-1800s was genotype A, subgenotype A1 (Kurbanov et al., 2005; Makuwa et al., 2006).
Previous genetic studies of HBV in west Africa found dominance of genotype E with few strains of genotype A. However, HBV subgenotype A3, initially described in Cameroon, has also been found in west African countries such as Benin, Ghana, Mali, Côte d’Ivoire and the Gambia (Suzuki et al., 2003; Fujiwara et al., 2005; Huy et al., 2006; Candotti et al., 2007).

Recombination has been reported to occur in HBV when genomes of different genotypes co-exist in the same host and is a way of introducing variability (Bollyky et al., 1996; Morozov et al., 2000). In some regions, recombinant forms are taking over and are distributed more widely than the parental genotypes, as reported with the B/C recombinant in Asia (Sugauchi et al., 2003).

This study involved HBsAg-carrier blood donors from Conakry, Guinea, where the prevalence of HBsAg in first-time blood donors is 15 % (Loua et al., 2004). Full genome sequences were obtained and provided maximum phylogenetic precision, as well as the opportunity to study all HBV genes.

**METHODS**

**Samples.** HBsAg-containing plasma samples (n=117) were obtained from asymptomatic replacement (80 %) and volunteer (20 %) blood donors from Conakry, Guinea, during 2006. Moreover, three samples identified during an earlier study (Allain et al., 2003) from deferred Ghanaian blood donors were included in the present study. Samples were stored at −20 °C or below until tested.

**HBV DNA quantification.** Virus DNA was extracted from plasma samples (500 μl) using a Total Nucleic Acid Isolation kit (Roche Diagnostics) according to the manufacturer’s instructions. Virus DNA was quantified by quantitative (Q)-PCR using a TaqMan-based methodology and the Mx3000 or Mx4000 Multiplex Quantitative PCR systems (Stratagene) as described previously (Allain et al., 2003; Zahn et al., 2008). An in-house internal control was used and calibrated regularly against the WHO International Standard for HBV DNA for nucleic acid testing (NAT) assays 97/746 (National Institute for Biological Standards and Controls, Potters Bar, UK).

**Serological testing.** Antibodies against HBcAg (anti-HBc) were detected with a Bioelisa anti-HBc kit (Biokit). Hepatitis B e antigen (HBeAg) and anti-HBe were detected with a Monolisa HBe Ag-Ab kit (Biokit). Hepatitis B e antigen (HBeAg) and anti-HBe were detected with a Bioelisa anti-HBc kit (Biokit). Hepatitis B e antigen (HBeAg) and anti-HBe were detected with a Bioelisa anti-HBc kit (Biokit).

**Amplification of the HBV genome.** In order to obtain the complete nucleotide sequence of the HBV genome, a nested PCR using the Expand High Fidelity PCR system (Roche) provided a product of approximately 3000 bp, as reported previously (Zahn et al., 2008). Another set of PCRs targeted the basic core promoter/pre-core (BCP/PC) region (276 bp), using a semi-nested PCR as described previously (Candotti et al., 2006) to include the 50 bp not amplified with the nearly full-length primers.

In addition, to study strains presenting deletions or recombination further, specific regions of the genome were amplified. In particular, core was amplified for deleted strains and most recombinant strains using primers P3WRS (Zahn et al., 2008) and IF3 (TGACCCCRCAAAATGAGGC, reverse), or pre-surface/surface (pre-S/S) only for a recombinant strain by using primers described previously (Candotti et al., 2007).

**HBV genome sequencing.** Full-length and BCP/PC amplicons were run on low-melting-temperature agarose gel. Bands of the appropriate size were excised and applied to a QIAEX II Gel Extraction kit (Qiagen) for purification. Purified DNA was sequenced using an ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit, version 1.0, and an ABI Prism 3100 Genetic Analyzer or Applied Biosystems 3730xl DNA Analyzer. A set of primers that covered the whole HBV genome was used as described previously (Zahn et al., 2008).

**Cloning of recombinant strains.** In order to provide evidence of the recombination event observed in several samples, a region predicted to include a recombination point was amplified by using previously described primers (Zahn et al., 2008). Amplicons were cloned with a TOPO TA Cloning kit (Invitrogen) according to the manufacturer’s instructions. Plasmid DNA from these clones was purified with a QIAprep spin miniprep kit (Qiagen) and six or more clones were sequenced.

**Sequence and phylogenetic analysis.** Analysis of the electropherograms was done with the SeqMan Pro program from the Lasergene package version 7.1 (DNASTAR Inc). The consensus sequences of the full-length and BCP/PC regions were assembled with the MacVector software version 7.2 and the CLUSTAL W alignment option. Translation of the four HBV genes was done with the SeqBuilder program of the Lasergene package. Phylogenetic analysis was performed by using PAUP* version 1.0 b10 (Swofford, 2003) and the neighbour-joining algorithm based on Kimura two-parameter distance estimation, ignoring all positions with gaps in pairwise comparisons. Bayesian analysis was performed using MrBayes version 3.1 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) with testing parameters according to MODELTEST version 3.7 (Posada & Crandall, 1998). Diversity between subgenotypes A3 and A4 and within the four genes of wild-type genotype E strains was calculated using PAUP* as described by Meldal et al. (2009). Potential recombinants were tested with the genotyping function of NCBI and verified with the SIMPLOT version 3.5.1 software (Lole et al., 1999).

Window sizes of 300 and 200 bp and steps of 100 and 20 bp were used with the NCBI and SIMPLOT programs, respectively. RDP3 software (Martin et al., 2005) was also used for further confirmation of recombination. The Shimodaira–Hasegawa test in PAUP* was used to compare the phylogenetic reconstructions of the different recombinant segments. All complete genome sequences were submitted to GenBank under the accession numbers GQ161753–GQ161838 and the pre-S/S clones under GQ161839–GQ161846. The accession numbers for the partially deleted and recombinant strains are specified in Figs 2 and 4, respectively.

**Statistical analysis.** Statistical analysis of the data produced in the present study was performed with Microsoft Excel 2004 version 11.5.1 and RISM version 4.0b (GraphPad Software). The non-parametric Mann–Whitney test was used to compare non-normally distributed data and Fisher’s exact test was used for the analysis of contingency tables.

**RESULTS**

**Sample quantification and stratification**

In total, 117 HBsAg+ samples from Guinea were processed. All samples were quantified for HBV DNA by Q-PCR and the median viral load was 4.33 × 10⁵ IU ml⁻¹. The donors’ ages ranged between 18 and 55 years and the median age was 23 years. The majority of donors were male (88.6 %).
Phylogenetic analysis
The complete HBV genome sequence was obtained for 81 of 117 (69.2 %) samples. The genotype distribution of the strains from which full genome sequences were obtained is shown in Table 1. The full-length sequences were used for phylogenetic analysis and genotyping. Both neighbouring analysis and Bayesian phylogeny determined that 77 strains clearly clustered with genotype E (95.1 %) (100 % bootstrap support, 1.0 maximum Bayesian posterior probability) and one (1.2 %) with genotype A (Fig. 1; Supplementary Fig. S1, available in JGV Online). This strain clustered with a minor clade of genotype A including a strain from Mali, published as subgenotype A4 (Olinger et al., 2006). Nucleotide-diversity analysis between subgenotypes A3 and A4 showed that the divergence was 3.6 %, below the 4 % divergence required to qualify as separate subgenotype. As a result, our strain was classified as subgenotype A3. As indicated in Table 1, seven of 77 strains of genotype E carried various large deletions (see below). Three strains were outliers at the root of genotype E (one strain) and genotype A (two strains) (Fig. 1). Three extra strains with incomplete sequences (GU1125, GU1127 and GU1607) were also identified as outliers.

Molecular analysis of genotype E pre-core/core (pre-C/C)
Molecular analyses were done on either 77 genotype E full genomes or 70 wild-type sequences (90.9 %), excluding the seven strains with large pre-C/C deletions (9.1 %). Overall, 20 of 77 samples (26.0 %) of genotype E had one or more mutations at position(s) 1762, 1764 and/or 1896, and four samples (5.2 %) had triple mutations at positions 1762, 1764 and 1896. The double 1762/1764 mutation was present in six samples (7.8 %). A premature stop codon related to the G1896A mutation was observed in 16 sequences (20.8 %). The pre-C initiation codon was mutated in two samples (2.6 %) and another sample had a T deletion at position 1847 that disrupted the pre-C protein. The median amino acid divergence within the pre-C/C of the 70 wild-type genotype E strains was 4 % (range 0–24 %). One sample had a cR151P substitution and cR152del, whilst two other strains each had an arginine substitution, to cysteine and glutamine at positions C164 and C165, respectively.

Molecular analysis of pre-S/S and polymerase genes of genotype E strains
In total, 77 of 81 pre-S/S genotype E sequences were obtained. In pre-S2, eight of 77 strains (10.4 %) had deletions. Two strains presented a 9 aa deletion (positions ps132–140), two others had a 6 aa deletion (positions ps134–139), two strains had a 4 aa deletion (positions ps136–139) and one had a 3 aa deletion (positions ps138–140). There were two strains truncated of the last 11 aa and another had an amino acid substitution at the pre-S2 initiation codon. The median overall amino acid divergence within the pre-S/S region of 70 wild-type genotype E strains was 4 % (range 0–20 %). In the major hydrophilic region (positions 99–169), the substitution rate was 0.1 %. One strain exhibited an sM133I substitution. The variability of the cellular epitopes of the HBV envelope proteins was very low.

The sequence of the genotype E polymerase gene was obtained for 76 of 81 strains, because in one of the strains with large deletions in the core region (see below), the initiation codon for polymerase was deleted and a clear sequence could not be obtained from the wild-type variant of the strain. The median amino acid divergence within the polymerase of 70 wild-type genotype E strains was 11 % (range 0–29 %). There was one strain with a 1 aa deletion in the terminal protein domain that corresponded to the core region and eight strains were affected by pre-S2 deletions in the spacer domain. One strain had an alanine to threonine substitution at position rt194.

Molecular analysis of the genotype E X gene
The X gene sequences of 77 of 81 genotype E strains were obtained. There were four strains (5.2 %) with the xH94Y substitution, eight strains (10.4 %) with the xK130M substitution and seven strains (9.1 %) with xV131I. In most strains, amino acid substitutions were single events.

Table 1. Distribution of 81 Guinean HBV strains with complete genome sequences

<table>
<thead>
<tr>
<th>Sample†</th>
<th>n (%)</th>
<th>Genotype</th>
<th>Viral load (IU ml⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Median</td>
</tr>
<tr>
<td>Wild-type</td>
<td>70/81 (86.4)</td>
<td>E</td>
<td>2.00 × 10⁵</td>
</tr>
<tr>
<td>Wild-type</td>
<td>1/81 (1.2)</td>
<td>A</td>
<td>–</td>
</tr>
<tr>
<td>DEL</td>
<td>7/81 (8.6)</td>
<td>E</td>
<td>7.62 × 10⁶</td>
</tr>
<tr>
<td>REC</td>
<td>3/81 (3.7)</td>
<td>A/E</td>
<td>2.58 × 10³</td>
</tr>
<tr>
<td>REC*</td>
<td>6/84 (7.1)</td>
<td>A/E</td>
<td>9.84 × 10⁵</td>
</tr>
</tbody>
</table>

†DEL, Strains with large deletions; REC, strains with suspected recombinations; REC*, including three additional incomplete sequences with suspected recombinations.
Fig. 1. Neighbour-joining phylogenetic tree of complete HBV genome sequences from Guinea and Ghana, and reference sequences from GenBank. Bootstrap-support values obtained from 1000 replicates are indicated. Accession number, origin and genotype/subgenotype of sequences obtained from GenBank are displayed. Sequences in bold are from this study. The wild-type strains are indicated by a circumflex accent (ˆ) at the end of the strain name. Suspected recombinant strains are indicated by an asterisk (*) at the end of their code. The first two letters of the strains' code indicates their geographical origin (GU, Guinea; GH, Ghana). GU1125, GU1127 and GU1607 sequences are incomplete, but are included as suspected recombinants. Bar, 0.02 substitutions per site.
However, two strains had coexisting substitutions of H94Y and K130M and seven other strains of K130M and V131I. In 70 wild-type genotype E X proteins, the median amino acid divergence was 2% (range 0–9%).

**Strains with large deletions**

Seven of 77 genotype E strains had large single and/or double deletions in and out of frame in the core gene and are referred to as DEL strains. The exact sizes and positions of the deletions are illustrated in Fig. 2(a). There were three different patterns of deletions. One strain had a single deletion of 262 bp, two strains had deletions of 49 and 147 bp separated by 111 bp, and four strains had a 37 bp deletion and another of 78 bp, separated from the first by a 100 bp fragment. For two of these samples (GU489 and GU1086), the initiation codon for polymerase was missing in the deleted genomes, but in one of them (GU489), this codon was present in the minor wild-type species. Such deletions affecting the HBV capsid and, in two cases, the polymerase are incompatible with virion formation. However, the viral load of the samples ranged between $5.0 \times 10^3$ and $2.6 \times 10^8$ IU ml$^{-1}$, with a median of $7.6 \times 10^6$ IU ml$^{-1}$ (Table 1), and their distribution was similar to that of the wild-type strains (Fig. 3) (Mann–Whitney test, $P = 0.096$), suggesting the presence of complete variants in the same sample. Different primer sets covering the core region of the genome were used for amplification and multiple amplicon bands of molecular masses consistent with complete or deleted genomes, were observed (Fig. 2b). Close examination suggested that strains GU489 and GU1086 had an identical band pattern. This difference of 196 nt was identical to the sum of the two deleted regions (49+147 nt). The band patterns of samples GU732, GU1405, GU1410 and GU1520 appeared to be identical. The dominant band had a molecular mass consistent with 862 bp, 81 bp lower than the 943 bp wild-type band. Each band, when sequenced individually, corresponded to HBV sequences. The smallest band included the deleted version of the strain and the largest band the wild-type (data not shown).

All seven DEL samples were positive for anti-HBc, none had the 1896 pre-C stop codon and five samples (GU732, GU1086, GU1410, GU1405 and GU1520) were HBeAg-negative and anti-HBc-positive. GU489 and GU1365 were HBeAg-positive and anti-HBc-negative. A summary of the HBV marker status for all DEL samples is given in Supplementary Table S1, available in JGV Online. In comparison, the 70 genotype E wild-type strains had a similar viral load distribution (Mann–Whitney test, $P = 0.096$) (Fig. 3); 23% had the 1896 stop codon and 86% carried anti-HBe, whereas the remaining 14% were anti-HBe-negative. Neither the presence of the 1896 stop codon (Fisher’s exact test, $P = 0.59$) nor the anti-HBe status (Fisher’s exact test, $P = 0.31$) was significantly different.
between the DEL and the wild-type strains. In neither group was the presence of anti-HBe correlated with low viral load.

Recombination events in west African HBV strains

During assembly of particular strain sequences, it was observed that the products of specific primers clustered preferably with either genotype E or A (GU754, GU763, GU1125, GU1127 and GU1616). Some strains were outliers of the main genotype A (GU1125, GU1127, GU1607 and GU1616) or genotype E (GU754 and GU763) groups with high bootstrap values (Fig. 1). Strains GU1125, GU1127 and GU1607 were included in the phylogenetic analysis, although their sequences were incomplete. The presence of recombination within these HBV genomes was therefore suspected.

Two more strains identified previously in Ghana that presented similar phylogenetic profiles were included in this study. Strains GH2537 and GH16 were outliers of genotypes A and D, respectively (Fig. 1). The subtyping function of the NCBI showed that GU754 and GU763 strains were mostly genotype E with a section of genotype A. GU1616, GU1125, GU1127 and GU1607 had a similar profile, with the majority of the genome classified as genotype A and a minor part as genotype E. Strains GH2537 and GH16 from Ghana were identified as recombinants of genotypes A and E and genotypes D and E, respectively. The complete HBV sequences were analysed with SIMPLOT software that provided more details, including the exact recombination points. The results obtained for representative strains are shown in Supplementary Fig. S2 (available in JGV Online) and the exact recombination points of all the strains in Fig. 4(a). Two main areas of recombination emerged, both in the pre-C/C gene. The first was located at the junction of pre-C and core at nt 1896 (four strains) and nt 1906 (one strain) (Fig. 4a). The second was in the core region, overlapping the polymerase reading frame at nt 2394 in one A/E strain, nt 2406 in one D/E strain and nt 2419 in four A/E strains (Fig. 4a). The RDP software identified the same recombination events. Most of the breaking points were within 20–50 bp of those indicated by the SIMPLOT software. For example, SIMPLOT indicated position 2419 bp as a breaking point in sample GU1616 and RDP indicated position 2439 bp. In clones of this sample (see below), position 2430 bp was indicated as a breaking point by both software types. Phylogenetic analysis of the A part of the recombinant strains indicated that they clustered with subgenotype A3 in all cases (data not shown). The phylogenetic reconstructions of the different fragments within the sequences were significantly different for all recombinants ($P<0.05$), supporting their existence.

In the complete genome examination of these recombinant sequences, diversity and amino acid substitutions appeared similar to those of the wild-type strains. Molecular analysis of the complete genome sequences showed that one strain had a triple mutation at positions 1762, 1764 and 1896 and three of four strains had the G1896A mutation. In the pre-S/S gene, the pre-S2 start codon was deleted in one of the recombinant strains (GU1616). In the X gene, four substitutions reported as being associated with HCC were identified in three strains: GU1616 had a V116L substitution, GU763 had an A146S substitution and GU1607 had an M130K and an I131V substitution. The part of the genome where the substitutions occurred in the X gene was genotype A.

An additional, potentially recombinant strain from Ghana (GH99) was identified as an A/E recombinant. The full genome sequence of this strain could not be obtained. However, cloning and sequencing of the successfully amplified pre-S/S gene revealed that the same plasma sample contained three genotype A and two genotype E variants and three different single-recombinant variants of those genotypes. A schematic representation of the clones obtained is given in Fig. 4(b) and the actual recombination points obtained from the SIMPLOT analysis are indicated.

In order to examine the variant distribution and sequences in Guinean A/E recombinants in more detail, a 1000 bp region that included one recombination point in the core region of strains GU1127, GU1607 and GU1616 was cloned. In each clone, the points of recombination observed in the consensus sequences remained present. The diversity of the clones inside a distinct phylogenetic clade is shown in Fig. 5. All clones grouped with genotype A (84 % bootstrap support) and they belong to a monophyletic group (100 % bootstrap support). The viral load distribution of the Guinean recombinant strains was very similar to that of the wild-type strains (Mann–Whitney test, $P=0.538$) (Fig. 3).

DISCUSSION

The present study analysed 81 complete HBV genome sequences from asymptomatic HBsAg$^+$ Guinean blood
donors. Studies by Hannoun et al. (2005) and Olinger et al. (2006) presented two and three complete HBV genome sequences from west African HBV carriers, respectively. In 2006, five full-length HBV sequences from Ghanaian HBsAg+ blood donors were published (Huy et al., 2006).

The phylogenetic analysis of 81 full HBV genome sequences presented here showed that genotype E is largely dominant in Guinea (95.1%), confirming previous studies in west Africa (Odemuyiwa et al., 2001; Suzuki et al., 2003; Olinger et al., 2006). A single genotype A sequence was classified as subgenotype A3. Molecular analysis of the wild-type genotype E strains showed that all HBV genes presented low variability (pre-C/C, 4%; pre-S/S, 4%; polymerase, 11%; X, 2%). These data are in accordance with the observation of Mulders et al. (2004) that although HBV genotype E is hyperendemic, its genome is essentially conserved. Most mutations were found in the X gene and have been reported to be associated with HCC.

The group of sequences obtained from asymptomatic blood donors in the capital of Guinea, Conakry, presented two unexpected features: a relatively large number of sequences with large deletions in the core region and the presence of several recombinant A/E strains.

Some genotype E strains (seven of 77) had large single or double deletions in and/or out of frame in the core region of the HBV genome (Fig. 2). The pattern of amplicon bands shown in Fig. 2(b) was obtained with a higher-resolution electrophoresis than that used for routine amplification and sequencing. When purified and sequenced, the dominant species was the deleted one (Fig. 2b). These apparent deletions did not seem to be PCR artefacts, as they appeared at the same positions whether the nearly full genome or the core region was amplified. The consensus sequences showed the deletions, although amplicons indicating full genome length were also present, but as a minority. Such deletions would affect HBV capsid assembly and presumably prevent virion formation. However, the high viral load of these samples (Fig. 3) suggested the presence of fully competent genomes forming quasispecies in the same sample and this was supported by the presence of multiple amplicon bands (Fig. 2b). Cases with large single or double deletions in the core gene were reported by Kim et al. (2007) and are

---

**Fig. 4.** Schematic representation of Guinean and Ghanaian HBV A/E recombinant strains. (a) Schematic representation of recombination points in all relevant strains (EcoRI-cuts). Numbers above the bars represent the exact genome positions of the recombination points as indicated by SIMPLOT in an alignment of 3257 bp that includes all genotype-specific gaps. The grey areas on the bars indicate sequence gaps and the numbers below the bars indicate their exact position. A schematic representation of the three open reading frames of the HBV genome is shown above the figure. (b) Schematic representation of the diversity of clones isolated from a single plasma sample from Ghana (GH99) in the pre-S/S region. Numbers indicate the exact nucleotide positions (bp) of recombination points as indicated by SIMPLOT.
common in the pre-S2 region. The presence of a relatively important wild-type molecular mass band, confirmed as wild-type by sequencing, was consistent with the high viral load observed with the DEL strains. The presence of anti-HBc demonstrated an immune response against HBV core and indicated that all samples contained some intact core protein. There was no correlation between anti-HBe status and viral load for either DEL or wild-type strains, although anti-HBe has been reported to correlate with low viral load (Allain et al., 2003; Candotti et al., 2006).

Considering that the production of full virions with a largely deleted core protein is unlikely, the infectivity of these variants is doubtful. One therefore needs to understand why deletions take place at the same locations in multiple infected individuals. It can be hypothesized that the core gene of genotype E strains is particularly prone to genome rearrangement in discrete areas. These areas could be assigned to a relatively short region spanning from nt 1979 to 2314 (335 nt). If such was the case, it might be expected that this region could be a preferential site of recombination. In our series of recombinant strains, preferential recombination sites were indeed found at nt 1896, 2120 and 2419. The first two were included in the region where deletions had taken place and the most frequent recombination site was just over 100 nt downstream. One might, therefore, conclude that the 440 nt region between nt 1979 and 2419 is prone to genome rearrangement and may lead to either deletions, when a single genotype replicates, or recombination, when strains of two different genomes infect single hepatocytes. Simmonds & Midgley (2005) reviewed HBV recombination and identified the pre-C/C start and the end of the core gene as weak points in the HBV genome, although only a few A/D and B/C recombinations have been identified in this region. Four main weak points appeared around nt 1900, 2000, 2200 and 2400 in genotype E and A3 strains, all located in the pre-C/C gene. For the only A/E recombinant recorded by Simmonds & Midgley (2005), the weak point was located at the end of the S gene. At the pre-C/C region, the presence of direct repeat 2 (DR2) (1824–1834 bp) was reported by Dejean et al. (1984) to facilitate HBV DNA integration into the host genome. These authors also reported position 2360 ± 50 bp to be a junction between virus and host DNA in clones selected from human liver tumours.

From the full-genome phylogenetic analysis (neighbour-joining tree), three of 81 (3.7 %) strains were outliers at the root of genotype A (two strains) or E (one strain). Further analysis with theSIMPLOT software identified these strains as possible recombinants of genotypes E and A (Supplementary Fig. S2). Three extra strains with incomplete sequences (GU1125, GU1127 and GU1607) were also identified as recombinants during the sequence-assembly process and by the SIMPLOT software (Supplementary Fig. S2). The same result was obtained when recombination was tested with the RDP software. There was no evidence that the recombinant strains might be artefactual, as whole genome, specific core and cloned core regions presented the same pattern. The overall frequency of HBV recombination in Guinea was 7.1 % (six of 84) (REC*, Table 1).

Makuwa et al. (2006) reported a genotype A/E recombinant in rural Gabon and Kurbanov et al. (2005) found another case in Cameroon. Two other strains from Ghana presenting the same phylogenetic outlying as the Guinean recombinants at the root of genotype A or D (Fig. 4a) were added to the six Guinean strains for further analysis of HBV recombination in west Africa. Moreover, another case was identified in Ghana where, after cloning and sequencing in the pre-S/S region, the quasispecies contained wild-type genotype A and E variants, as well as three different recombinant of A/E (Fig. 4b).

**Fig. 5.** Neighbour-joining phylogenetic tree illustrating the clone diversity of selected genotype A/E recombinant strains. Six to eight 1 kb clones were sequenced for samples GU1127 (●), GU1607 (●) and GU1616 (▲), which represented the most frequent pattern of recombination. Numbers indicate the bootstrap value at the corresponding node. The genotype and country of origin of the reference strains are indicated after the GenBank accession number. Bar, 0.01 substitutions per site.
The recombination data presented here raise several important questions. First is the issue of whether HBV recombinant strains are circulating and are an integral part of HBV epidemiology in the area. This could be supported by three arguments. (i) Multiple individual blood-donor HBV strains carried exactly the same recombination sites and, when 1 kb clones were sequenced, the individual variants constitutive of the quasispecies had one recombination point at exactly the same location. In addition, the quasispecies were each formed of closely related variants differing by a few nucleotides, as expected in a wild-type HBV strain (Fig. 5). (ii) The relatively high frequency of A/E recombinant strains (4–7%) does not match well with the distribution of genotypes found in the 2006 population of blood donors studied here. With a ratio of 77:1 between genotype E and A in the Guinean population of asymptomatic blood donors, the likelihood of dual genotype infection and recombination is clearly very low. The relatively high percentage of recombinant strains therefore suggests that these strains are infectious and represent a significant part of the HBV epidemiology in Guinea. These transmissible recombinants may have arisen at a time when genotype A strains were more frequent than they are presently. Such a situation has been reported for recombinant genotype C/D HBV in western China, where the recombinant form constitutes the dominant subgenotype in the region (Wang et al., 2005). (iii) In Guinea, the viral load distributions of the recombinant strains and of the wild-type strains identified in blood donors were similar (Fig. 3), suggesting that the recombinant strains are infectious. Further work to verify the replication capacity of these strains could include testing on hepatoma cell lines such as HuH7.

A second question is which genotype A subgenotype is involved in the recombination. Phylogenetic analysis of the genotype A fragments obtained in the recombinant strains showed preferential clustering with subgenotype A3. This information provides indirect evidence that subgenotype A3, rather than A1, was prevalent in west Africa prior to or concomitantly with the extension of genotype E. Subgenotype A3 has been identified in Pygmies in Cameroon and in the genotype A strains occasionally identified in west Africa (Kurbanov et al., 2005). These data apparently contradict the exclusive presence of subgenotype A1 among Brazilians of African origin (Araujo et al., 2004; Motta-Castro et al., 2005; Mello et al., 2007). It is possible that subgenotype A3 has evolved from subgenotype A1 over the period of time during which genotype E extended.

Lastly, the data presented in Fig. 4(b) may provide clues to the mechanisms by which recombinant forms arise and eventually survive in an infected individual. The simultaneous detection of wild-type and recombinant variants in the quasispecies of a single individual (GH99) is a strong argument linking recombination to dual infection, unless the recombinant is itself infectious and transmitted. In GH99, the wild-type sequences were dominant and the recombinant clones were minority variants. In addition, the sites of recombination appeared different in each variant, suggesting that recombinant variants are submitted to selection and only one or a few variants ultimately become dominant. As indicated before, deletions in the pre-S2 region are relatively frequent and this may constitute a second region of genome rearrangement. The sites of recombination of the three observed recombinants are relatively close to each other, but clearly different (nt 362 and 402 for two variants and nt 753 and 778 in two others) (Fig. 4b).

The present study confirms the dominance and the low level of genetic diversity of HBV genotype E with sequences from Guinean asymptomatic blood donors. The presence of only one subgenotype A3 strain and, more convincingly, of recombinant A3/E strains suggests strongly that subgenotype A3 was a circulating strain in the area and that it was dominated relatively recently by HBV genotype E. The identification of a similar A/E recombinant in Ghana suggests that the Guinean epidemiological features probably extend to a large part of west Africa. The identical pattern of A/E recombination in multiple donors suggests strongly that several A/E recombinants are infectious and circulating in the region. The observation that both deletions and recombinations occur in similar areas of the core gene suggests that this area of the HBV genome is where genetic rearrangements preferentially take place.

ACKNOWLEDGEMENTS

The authors wish to thank the staff of the National Blood Transfusion Center in Conakry for collecting and testing donor samples. Dr I. H. A. Barnes is also thanked for determining the HBV BCP/PC sequences of the strains included in this study. The authors are grateful to Dr B. Meldal for her valuable input into the phylogenetic analysis of the data. This work was supported in part by grant BS05/02 from the National Health Service Blood and Transplant, UK.

REFERENCES


HBV genotype E: deletions and recombination


