A novel hepatitis B virus subgenotype, D7, in Tunisian blood donors

Birgit H. M. Meldal,1 Najet Mojaat Moula,2 If H. A. Barnes,1† Kamel Boukef2 and Jean-Pierre Allain1

1Division of Transfusion Medicine, Department of Haematology, University of Cambridge, Cambridge Blood Centre, Long Road, Cambridge CB2 0PT, UK
2Tunisian National Blood Centre, Tunis, Tunisia

Tunisia is a medium-level epidemic country for hepatitis B virus (HBV). This study characterizes, for the first time, full genome HBV strains from Tunisia. Viral load quantification and phylogenetic analyses of full genome or pre-S/S sequences were performed on 196 hepatitis B surface antigen (HBsAg)-positive plasma samples from Tunisian blood donors. The median viral load was 64.65 IU ml⁻¹ (range <5–7.7×10⁹ IU ml⁻¹) and 89 % of samples had viral loads below 10,000 IU ml⁻¹. Fifty-nine strains formed a novel subgenotype D7, 41 strains clustered in subgenotype D1, seven strains in subgenotype A2 and one strain in genotype C. The novel subgenotype D7 was defined by maximum Bayesian posterior probability, a genetic divergence from other HBV/D subgenotypes by >4 % and a stronger HBV/E signal in the X to core genes than subgenotype D1. In conclusion, HBV/D is dominant in asymptomatic Tunisian HBsAg carriers and a novel subgenotype, D7, was the most common subgenotype found in this population.

In Tunisia, the seroprevalence for the hepatitis B virus (HBV) surface antigen (HBsAg) ranges between 3 and 13 %, given regional variations, and over 50 % of adults have been exposed to the virus (Coursaget et al., 1994; Hannachi et al., 2009; Triki et al., 1997). HBV has evolved into eight recognized genotypes, A–H (Arauz-Ruiz et al., 2002; Bartholomeusz & Schaefer, 2004; Norder et al., 1992, 1994; Stuyver et al., 2000). Genotypes are defined by a genomic divergence of more than 7.5–15 % depending on genotype, while subgenotypes diverge from each other by 2.5–7.5 % (Kramvis et al., 2008).

In Tunisia, the most common HBV genotype is D and there is evidence of occasional identification of genotypes A, B, C and E (Ayed et al., 2007; Bahri et al., 2006; Borchani-Chabchoub et al., 2000; Triki et al., 2000). Little is known about the molecular diversity of HBV in Tunisia except in the basal core promoter/precore region (BCP/PC) (Ayed et al., 2007; Bahri et al., 2006; Triki et al., 2000) and five partial preS2/S sequences (Borchani-Chabchoub et al., 2000). In this study, we analysed the molecular character-

1Present address: The Wellcome Trust Sanger Institute, Morgan Building N306, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK.

The GenBank/EMBL/DDBJ accession numbers of sequences derived in this paper are FJ904328–FJ904393 (pre-SS) and FJ904394–FJ904447 (full genome).

A supplementary figure is available with the online version of this paper.
HBV subgenotype D7 in Tunisian blood donors

Phylogenetic analyses for genotyping and genomic distance calculations were performed using MrBayes version 3.1 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2001; Ronquist & Huelsenbeck, 2001), PAUP* version 4.0 beta 10 (Swofford, 2002), and SIMPLOT version 3.5.1 (Lole et al., 1999) using the same reference sequences and default parameters as in Fig. 1. Bootstrap values of <70% were considered to be non-significant. Genomic distances between HBV/D subgenotypes and HBV/E were calculated by using Kimura-2-Parameter distances (Kimura, 1980) for nucleotides and mean distances for amino acids using as many sequences from HBsAg+ asymptomatic carriers as were available in GenBank and in our laboratory. Genetic divergence was defined as the mean of all pairwise genomic distances within or between a given (sub)genotype. Statistical analyses were performed in PRISM version 4.0b (GraphPad Software).

HBV DNA load was successfully quantified in 175 of 196 (89%) Tunisian blood donor samples. Six RT-PCR-negative samples were positive by nested PCR and sequencing, while six samples were low positive by RT-PCR (<10 IU ml\(^{-1}\)) but could not be confirmed by nested PCR. The median viral load was 64.65 IU ml\(^{-1}\) (range <5-7.7 \times 10^4 IU ml\(^{-1}\)) and the majority of samples (89%) had a viral load below 10,000 IU ml\(^{-1}\). The viral load distributions of subgenotypes D1 and D7 were not significantly different [D1\(_{\text{median}}\) = 126 IU ml\(^{-1}\) (range: 1.5 \times 10^2-9.6 \times 10^4 IU ml\(^{-1}\)] vs D7\(_{\text{median}}\) = 178.5 IU ml\(^{-1}\) (range: 7.0 \times 10^2-9.3 \times 10^3 IU ml\(^{-1}\)], \(P=0.97\].

A total of 173 plasma samples were successfully sequenced for BCP/PC, 59 for the pre-S/S region and 51 for the full genome. Genotyping of all 110 pre-S/S and full genome sequences identified 102 strains (93%) in HBV/D, seven strains (6%) in HBV/A and one strain (1%) in HBV/C with maximum Bayesian posterior probability (PP) of 1.0 at the genotype level (Fig. 1). Genotype A samples were most closely related to subgenotype A2 (PP=1.0) but formed a separate clade. Amongst genotype D strains, 41 clustered in subgenotype D1 and 59 strains formed a new clade (PP=1.0 for both), proposed to be subgenotype D7; two strains could not be assigned to a subgenotype. No regional distribution of genotypes was observed because the majority of samples came from the Tunis Blood Centre.

A BLAST search and subsequent genotyping analysis of sequences from proposed subgenotype D7 found only two references falling within this clade; these were AM494716 from the Central African Republic and DQ991753 from an Irish patient of West African origin. The divergence of subgenotype D7 from other HBV/D and HBV/E subgenotypes was 4.17–5.74% and 7.89%, respectively, over the whole genome (Table 1). In the polymerase, precore/core and X proteins, the divergence was >4% from other HBV/D subgenotypes and in the large surface antigen the divergence ranged from 2.96 to 4.67%. The intra-subgenotypic divergence was below 4% for all proteins except precore/core in HBV/D1 and HBV/D7, where the intra- and inter-subgenotypic divergence was similar and larger than those for other HBV/D subgenotypes. A Bayesian inference analysis of the genes of the four open reading frames (ORFs) of HBV separated subgenotype D7 in three out of four ORFs, with genotype D being unresolved in the pre-S/S gene. The polymerase gene had maximum PP, while the PP was low (0.81) for the X gene and below 0.5 for precore/core (Supplementary Fig. S1, available in JGV Online). However, a cluster of subgenotype D7 and genotype E had high PP in the precore/core gene, matching the SIMPLOT analysis (see below). Therefore, the above data provide ample evidence for the definition of subgenotype D7.

All HBV/D sequences from Tunisia were evaluated for recombination events using SIMPLOT. All subgenotype D7 strains showed a higher signal for HBV/E than HBV/D in the X to core region (intervals approx. between bp 1700 and 2400) but only eight strains (A7, A13, A29, A42, A45, B1, B45 and B65) had a consistent recombination signal of over 70% bootstrap support with HBV/E (Fig. 2a, b). These strains did not belong to a particular subclade of D7. The same recombination signal was observed for the two reference sequences that clustered within D7 (Fig. 2c, d). D1 strains showed no recombination signal (Fig. 2e). Bayesian and SIMPLOT analysis of the four ORFs showed that A76 was a complex recombinant of subgenotypes D1 and D7 (Fig. 2f).

The basal core promoter mutations A1762T and G1764A were found in six and seven of the seven HBV/A strains, respectively, and in 29 (26.6%) and 38 (34.9%) of all HBV/D strains, respectively. Subgenotype D7 strains had significantly lower frequencies of G1764T and C1766G double mutation (2/59 vs 7/41, \(P=0.003\)) and of the G1764T and C1766G double mutation (2/59 vs 7/41, \(P=0.031\)) than subgenotype D1 strains. Precore start codon mutants were equally common in all subtypes (average 14%) but the premature precore stop codon at G1896A was twice as common in HBV/D (87/102, 85%) than in HBV/A (3/7, 43%). cP130Q was found in 6 of 31 D7 strains but not in any D1 strains.

Two strains had in-frame deletions in the pre-S1 region (5 and 33 residues) and seven strains had in-frame deletions in the pre-S2 region (4–13 residues). Six strains had premature stop codons in the S gene, truncating the surface protein by 11 residues in five strains (L216stop) and by 14 residues in one strain (L213stop). Most strains (92%) had wild-type major hydrophilic regions.
Fig. 1. Bayesian phylogeny of 51 Tunisian full genome HBV strains. Samples from Tunisia are given in bold type and indicated by an asterisk (*). Subgenotypes present in Tunisia are marked by brackets and the novel subgenotype, D7 is indicated by bold lines. Bayesian posterior probabilities are given for major nodes. Reference strains from GenBank are identified by accession number, country of origin and subgenotype (for genotypes A and D).
The study of HBV in Tunisian HBsAg+ blood donors presented here was performed in asymptomatic individuals presenting at different blood centres across the country. Although the molecular data presented are representative of the majority of HBsAg carriers in the country, some differences might exist compared with those with clinical liver disease. The vast majority of this population of HBsAg asymptomatic carriers had low or undetectable HBV DNA levels, an observation that is consistent with the low rate of HBV vertical transmission previously reported in Tunisia (Coursaget et al., 1994) and matching data from other studies on Tunisian asymptomatic HBV carriers (Bahri et al., 2006; Hannachi et al., 2009). These findings are in contrast with other genotype D-infected Mediterranean, Middle Eastern or Eastern European individuals (unpublished data) but are similar to findings in genotype E-prevalent areas (Candotti et al., 2006, 2007). These data suggest that viral replication of Tunisian HBV strains in chronically infected, asymptomatic carriers may be different from that in other genotype D strains but that it is not restricted to subgenotype D7 strains.

This is the first investigation into the molecular characteristics of HBV in Tunisia using pre-S/S and full genome sequencing data. HBV/D (93 %) was the most prevalent HBV genotype, followed by HBV/A (6 %) and one HBV/C strain. Among the genotype D strains, 41 % were subgenotype D1 and 59 % belonged to a new subgenotype D7 (Fig. 1). Only two other subgenotype D7 strains were found in GenBank, one from the Central African Republic (GenBank accession no. AM494716; Bekondi et al., 2008) and one from an Irish patient of West African origin (GenBank accession no. DQ991753; Laoi & Crowley, 2008). Few studies on HBV molecular diversity in North Africa have been published. Borchani-Chabchoub et al. (2000) published three preS2/S and two S sequences from Tunisia, and Norder et al. (2004) published five S gene sequences from Somalia. Although these sequences were too short for a robust phylogenetic analysis, four of the five Tunisian strains and all five Somaliarn strains clustered with Tunisian subgenotype D7 strains in a neighbour-joining tree of the S gene (data not shown). The remaining Tunisian strain clustered with subgenotype D1 strains. Khelifa & Thibault (2008) found a new clade of HBV genotype D from East Algerian patients based on partial HBV genome sequences. Since the sequences are not yet available in GenBank, we can only speculate that these strains may also belong to the new subgenotype D7. These authors describe three residues in the reverse transcriptase that have a higher prevalence in the Algerian strains compared with other genotype D reference strains (rtQ149K, rtR153W and rtP237T). We found the same three residues at high frequencies in subgenotype D7. However, compared with HBsAg+ strains from asymptomatic carriers of all HBV genotypes, there are no residues that are unique to D7, although other than D7, rtP237T is only found in HBV/F and HBV/H. The viral load distribution for their HBV e antigen-negative samples is also comparable to that found in Tunisia. In contrast, Saudy et al. (2003) described strains from Egypt that all belong to subgenotype D1 (Kramvis & Kew, 2007 and strain with GenBank accession no. AB104710 in Fig. 1). Therefore, it is possible that subgenotype D7 is most prevalent in the Maghreb (Tunisia, Algeria and Morocco).

Subgenotype D7 is characterized by clear phylogenetic separation (Fig. 1), a genomic divergence of over 4 % from all other HBV/D subgenotypes (Table 1) and a relatedness to HBV/E in the precore/core gene (Fig. 2). Although subgenotype D4 is the sister clade closest to D7, only a few subgenotype D4 strains have so far been isolated and these were predominantly found in Australasian carriers. It is therefore difficult to speculate on a common evolutionary history of these two clades. On the other hand, the Simplot analysis showed that all D7 strains had a higher signal for HBV/E than HBV/D in the X to core region and the phylogenetic analysis of the separate ORFs clustered D7 closer to HBV/E than to other HBV/D subtypes in both genes (except D4 in X gene). The bootstrap support for this suspected region is below 70 % for most strains and >70 % in a very short stretch (Fig. 2), making it unlikely that D7 has undergone recombination with HBV/E. In addition, very few genotype E strains have been reported in Tunisia (Ayed et al., 2007; Bahri et al., 2006).

### Table 1. Identification of HBV clade D7 as a novel genotype D subgenotype based on percentage divergence and range between HBsAg+ strains from asymptomatic carriers of HBV/D subgenotypes and HBV/E, calculated for the full genome.

<table>
<thead>
<tr>
<th>Genotype/subgenotype</th>
<th>$n$</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D7</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>44</td>
<td>2.01 ± 1.79</td>
<td>1.58 ± 2.16</td>
<td>3.48 ± 1.26</td>
<td>4.56 ± 1.18</td>
<td>4.34 ± 1.13</td>
<td>2.01 ± 1.42</td>
<td>2.79 ± 1.50</td>
</tr>
<tr>
<td>D2</td>
<td>37</td>
<td>2.29 ± 1.50</td>
<td>3.41 ± 1.25</td>
<td>5.12 ± 0.86</td>
<td>4.57 ± 1.26</td>
<td>5.13 ± 0.51</td>
<td>2.40 ± 0.08</td>
<td>1.50 ± 1.79</td>
</tr>
<tr>
<td>D3</td>
<td>15</td>
<td>3.41 ± 1.25</td>
<td>5.23 ± 0.85</td>
<td>5.12 ± 0.86</td>
<td>4.57 ± 1.26</td>
<td>5.13 ± 0.51</td>
<td>2.40 ± 0.08</td>
<td>1.50 ± 1.79</td>
</tr>
<tr>
<td>D4</td>
<td>5</td>
<td>4.57 ± 1.26</td>
<td>4.56 ± 1.18</td>
<td>5.12 ± 0.86</td>
<td>5.13 ± 0.51</td>
<td>5.27 ± 0.70</td>
<td>2.40 ± 0.08</td>
<td>1.28 ± 1.50</td>
</tr>
<tr>
<td>D5</td>
<td>2</td>
<td>5.12 ± 0.86</td>
<td>5.23 ± 0.85</td>
<td>5.13 ± 0.51</td>
<td>5.13 ± 0.51</td>
<td>5.27 ± 0.70</td>
<td>2.40 ± 0.08</td>
<td>1.28 ± 1.50</td>
</tr>
<tr>
<td>D7</td>
<td>34</td>
<td>4.64 ± 1.29</td>
<td>4.82 ± 1.61</td>
<td>6.61 ± 1.32</td>
<td>4.17 ± 1.28</td>
<td>5.74 ± 0.90</td>
<td>2.88 ± 2.42</td>
<td>7.89 ± 1.31</td>
</tr>
<tr>
<td>E</td>
<td>108</td>
<td>8.07 ± 1.42</td>
<td>8.08 ± 1.50</td>
<td>7.77 ± 0.29</td>
<td>8.27 ± 1.34</td>
<td>8.79 ± 0.77</td>
<td>7.89 ± 1.31</td>
<td>1.59 ± 1.97</td>
</tr>
</tbody>
</table>

$n$ = Number of sequences included in analysis. Divergence was calculated as percentage of the mean pairwise Kimura-2-Parameter distances over the full genome. The divergences of subgenotype D7 from other (sub)genotypes are highlighted in bold. For subtypes D4 and D5, only limited numbers of full genome sequences are available and these data include strains from symptomatic patients.
HBV subgenotype D7 in Tunisian blood donors

Because the phylogenetic analyses suggested that subgenotype D7 is the HBV/D subgenotype most closely related to HBV/E, a more plausible explanation for the SIMPLOT findings is that HBV/E and D7 share a more recent common ancestor than HBV/E and other HBV/D strains. Simmonds & Midgley (2005) also found that the region between bp 1000 and 2100 does not significantly distinguish HBV/D and HBV/E, suggesting the presence of a common evolutionary ancestor. The putative geographical area harbouring subgenotype D7 across North Africa and as far south as the Central African Republic (Bekondi et al., 2008) just borders the sub-Saharan area known to be endemic for HBV/E. Therefore, it is possible that HBV/E emerged in an area endemic for subgenotype D7 and has retained some characteristics of this subgenotype. Furthermore, Bekondi et al. (2007) also found an E/D recombinant that had a D7-typical precore/core signal in SIMPLOT (data not shown).

This study has identified that HBV has intermediate prevalence amongst Tunisian blood donors. Genotype D is most common and can be divided into two subgenotypes, D1 and a novel subgenotype, D7. This is defined by a clear phylogenetic and geographical separation from its closest sister clade, subgenotype D4, by a genomic divergence of >4% from all other HBV/D subgenotypes and by a similarity to HBV/E in the X to core gene region. This suggests that subgenotype D7 may have contributed to the emergence of HBV/E. Bayesian inference analysis was more sensitive than neighbour-joining at reconstructing phylogenetic relationships in HBV genotype D due to the overall low diversity of genotype D strains. The median viral load of HBV in Tunisia was lower than in European and Mediterranean genotype D strains but no difference was observed in the viral load distribution between Tunisian subgenotype D1 and D7 strains.

Acknowledgements

The authors would like to thank the Tunisian Blood Centre staff who participated in the collection and screening of blood donor samples and Dr D Candotti for fruitful discussion on the molecular aspects of the study. This study was supported by grant BS05/2 from the National Health Service Blood and Transplant.

References


