The \( T^{1858} \) variant predisposing to the precore stop mutation correlates with one of two major genotype F hepatitis B virus clades

Heléne Norder,\(^1\) Patricia Arauz-Ruiz,\(^1\) Linda Blitz,\(^2\) Flor H. Pujol,\(^3\) José M. Echevarría\(^4\) and Lars O. Magnius\(^1\)

\(^1\)Virological Department, Swedish Institute for Infectious Disease Control, S-171 82 Solna, Sweden
\(^2\)Laboratorio Regional de Referencia Virologica, Instituto de Investigaciones Clinicas, LUZ, Maracaibo, Venezuela
\(^3\)Laboratorio de Virologia Molecular, CMBC, IVIC, Caracas, Venezuela
\(^4\)Centro Nacional de Microbiologia, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain

Received \( 9 \) December 2002
Accepted \( 23 \) March 2003

The precore mutation \( G^{1896} \rightarrow A \) occurs frequently in anti-HBe-positive carriers of HBsAg with \( T^{1858} \) in the stem of the encapsidation signal. Hepatitis B virus (HBV) genotype F, considered an Amerindian genotype, subdivides into two clades and the precore mutation occurs in Central American F strains. To investigate the relationship between substitutions at position 1858 and these clades, the precore and small S genes of 48 strains of HBV genotype F were subjected to phylogenetic analyses. Isolates of one clade, formed mainly of Central American strains, all had \( T^{1858} \) and Thr\(^{45} \) in the S gene, whereas in the other clade, formed mainly of South American strains and one strain from Polynesia, all had \( C^{1858} \) and Leu\(^{45} \). The latter strain was related to strains from Venezuela and Colombia, supporting an Amerindian contribution to the Polynesian population. The position of the Polynesian strain in the phylogenetic tree indicates that the two clades have resulted from an early split, showing a high degree of genetic stability of the stem of the HBsAg encapsidation signal.

Hepatitis B virus (HBV) is a well-known agent of acute and chronic hepatitis, infecting around 300 million individuals worldwide. HBV is the prototype strain of the family Hepadnaviridae. The virus has a 3-2 kb circular, double-stranded DNA genome with four open reading frames: P, C, S and X. The P and S regions encode the polymerase and the envelope proteins, respectively. The C region encodes two gene products, the nucleocapsid or core protein and HBeAg, translated from two different transcripts of 3-5 and 3-6 kb, respectively.

HBV strains are classified into eight genotypes, designated A–H (Okamoto \textit{et al.}, 1988; Norder \textit{et al.}, 1992, 1993a; Stuyver \textit{et al.}, 2000; Arauz-Ruiz \textit{et al.}, 2002). Genotypes A and D are distributed widely in the Old World, genotypes B and C are confined to populations in East Asia and genotype E to populations in subSaharan Africa (Norder \textit{et al.}, 1993b). Genotypes F and H are considered the indigenous genotypes of Amerindians and encompass HBV strains that are the most divergent compared to strains from other parts of the world (Naumann \textit{et al.}, 1993; Norder \textit{et al.}, 1993a; Arauz-Ruiz \textit{et al.}, 2002). Genotype H has only been found in populations from Central America, Mexico and California (Arauz-Ruiz \textit{et al.}, 2002). The origin of genotype G is not known (Stuyver \textit{et al.}, 2000). Phylogenetic analyses of S genes as well as complete genomes have shown that F strains may be classified into two major clades, one composed mainly of strains from Central America and the other of strains from South America (Arauz-Ruiz \textit{et al.}, 1997b, 2002). Another subdivision of genotype F has defined three clusters of F, designated I, II and IV (Mbayed \textit{et al.}, 2001), while cluster III of genotype F corresponds to genotype H (Arauz-Ruiz \textit{et al.}, 2002).

HBeAg is present in the early phase of acute infection and during the replicative state in chronic hepatitis B. The HBeAg-positive phase subsides later in life and is replaced by the anti-HBe-positive state, associated usually with loss of infectivity. Because HBeAg and HBcAg are encoded by different transcripts, the emergence of a Trp to a translational stop mutation at codon 28 in the precore leader sequence abrogates the synthesis of HBeAg, still allowing the synthesis of nucleocapsids and infectious particles. This immune escape mutation corresponds to a single \( G \rightarrow A \) transition at nucleotide position 1896 (Brunetto \textit{et al.}, 1989;
Carman et al., 1989; Okamoto et al., 1990). The emergence of this mutant may be associated with significant liver pathology in the chronic carrier as well as fulminant hepatitis in contacts (Laskus et al., 1993). Its frequency varies for different genotypes, depending on the substitution of position 1858 opposing position 1896 in the stem of the precore encapsidation signal or epsilon motif (Tong et al., 1990; Laskus et al., 1993; Liang et al., 1991). Thus, C1858, in general present in genotypes A, F and H, and in some genotype C strains, constrains the G<sup>1896</sup>→A mutation due to its destabilizing effect on the stem of the epsilon motif (Li et al., 1993; Alestig et al., 2001). In agreement with these base-pairing requirements, the precore codon 28 stop mutant occurs frequently in strains of genotypes B–E, which, in general, have T<sup>1858</sup> (Li et al., 1993; Rodriguez-Frias et al., 1995; Chan et al., 1999). We have, however, found T<sup>1858</sup> together with the precore stop mutation in genotype F strains from Central America (Arauz-Ruiz et al., 1997a).

The aim of this study was to investigate the prevalence of T<sup>1858</sup> in the precore region of genotype F strains of different geographical origin and to assess by comparing their S gene sequences whether genotype F strains with T<sup>1858</sup> may have a common origin. The precore region was, therefore, sequenced for 34 genotype F strains, 29 with S genes characterized previously (Norder et al., 1993b; Arauz-Ruiz et al., 1997b; Blitz et al., 1998). The S gene was sequenced for three strains from Spain and one from Polynesia.

The S and the precore regions were amplified as described (Norder et al., 1993b; Arauz-Ruiz et al., 1997a, b). The S gene sequences were aligned with 54 sequences from other studies and from GenBank. A total of 26 complete HBV genotype F and H genomes from GenBank were aligned using TREEALIGN (Hein, 1990). Genetic distances were calculated with DNADIST from the PHYLIP program, version 3.53, using Kimura’s 2-parameter model (Felsenstein, 1993). Phylogenetic trees were constructed using UPGMA and the neighbour-joining method in NEIGHBOR and the maximum-likelihood method in the DNAML program. SEQBOOT was used for bootstrapping 1000 data sets.

The subdivision of genotype F into two major clades demonstrated previously was confirmed and supported by significant bootstrap analysis, when 58 S genes (Fig. 1a) as well as 26 complete genomes (Fig. 1b) were analysed. The two clades, tentatively designated 1 and 2, correlated strictly with substitutions at amino acid residue 45 of the S gene and with the nucleotide substitution at position 1858 of the precore gene. Thus, all strains in clade 1 had Thr<sup>45</sup> in the S gene and T<sup>1858</sup> in the precore region, while the strains in clade 2 had Leu<sup>45</sup> and C<sup>1858</sup>.

Clade 1 corresponds to cluster I of Mbayed et al. (2001). A subcluster within this clade was formed by nine strains from Central America and two from Spain (Fig. 1a). The latter two strains also had the precore stop mutation at codon 28. The S genes of these Central American strains diverged by up to three nucleotides and one amino acid residue from each other. Another subcluster was formed by three strains from Argentina and one from Alaska. Two divergent strains, 1980Nic from Nicaragua and Jpf1130 from Japan, formed separate branches and diverged by 11 amino acid residues from each other and by five to nine residues from the other strains in the clade.

Clade 2 comprised 31 strains (Fig. 1a). Two strains from Central America, two from Venezuela and one from Brazil, formed a cluster corresponding to cluster II of Mbayed et al. (2001). Seven strains from Argentina and three from France formed a cluster corresponding to cluster IV, while the strains from Polynesia clustered with strains from Venezuela, Colombia and Spain. The small S gene of the strains in clade 2 diverged by up to 25 nucleotides and 12 amino acid residues from each other and by up to 32 nucleotides and 15 amino acid residues from the strains of clade 1.

The presence of two major clades of genotype F, each characterized by the nucleotide substitution at position 1858 in precore and the amino acid substitution at position 45 of the S gene, is compatible with the emergence of a variant encoding Thr<sup>45</sup> and T<sup>1858</sup>. All but two of the Central American strains belonged to this variant. The lower nucleotide and amino acid divergences for S genes of Central American strains belonging to this clade may be due to the fairly recent introduction of this variant into the Hispanic population of Central America. Strains expressing T<sup>1858</sup> are also present in Amerindians from Argentina and Brazil (Lopez et al., 2002; De Castro et al., 2001). Although the S genes of these strains were not characterized, the S genes of three other strains from Argentina express Thr<sup>45</sup> and cluster in clade 1 (Mbayed et al., 2001), indicating that this clade has also spread among Amerindians in South America.

Random genetic drift may lead to the takeover of a strain due to the size of the susceptible population being small or due to the selected variant having a higher replication capacity. By phylogenetic analysis of the envelope gene, the emergence of a new variant of dengue virus was shown to be mediated by a cross-species transfer (Zanotto et al., 1996). The finding of a common ancestor for both C<sup>1858</sup> variants in genotype C (Alestig et al., 2001) and T<sup>1858</sup> variants in genotype F shows an unexpectedly high genetic stability of the nucleotides in the precore stem-loop region. The C<sup>1858</sup>→T mutation in genotype F, and the mutation in the opposite direction in genotype C, have both apparently overcome the error threshold for survival, as has been shown for quasi-species of human immunodeficiency virus, foot-and-mouth disease virus and vesicular stomatitis virus (Eigen & Riberecher, 1988; Lee et al., 1997; Martinez et al., 1997). Whether there is a positive selection for either precore variant due to differences in their replicating ability needs further investigation. Anyway, it seems to have evolved independently from the dominating variant in both genotypes C and F and possibly also in genotype A (Alestig et al., 2001; Lopez et al., 2002).

Genotypes F and H are considered to be the indigenous HBV
strains of the New World. The finding of distinct genetic clades within genotype F is not unexpected, considering the 16 000 years since early Amerindians crossed the Behring strait and the wide geographical dispersion of present Amerindian populations. Although the origin and number of colonization waves from Asia are still under discussion, linguistic, phenotypic and genetic markers have grouped the Amerindians into three or four different groups, each restricted geographically (Greenberg et al., 1986; Torroni et al., 1993; Cann, 1994; Crawford, 1998; Novick et al., 1998). The formation of three main clades of Amerindian HBV strains, genotype H and two within genotype F, may reflect the segregation of native American populations into different groups. JC virus has been used also as a marker for
human migration in the Americas and early settlers were shown to have brought JC virus type 2A from Asia (Agostini et al., 1997). This virus was shown to have co-evolved with its host, further subdividing JC virus type 2A into three groups, one modern Japanese, followed by a second division leading to two Amerindian groups.

The finding of a Polynesian HBV strain clustering with the South American genotype F clade and the occurrence of subtype adw4q— in Polynesia (Couroucé et al., 1983) support the suggestion of early Pacific voyagers in direct contact between South America and Remote Oceania (Heyerdahl, 1952; Cann, 1994). This is also supported by data on RFLP haplotypes of mitochondrial DNA. Of the four major RFLP haplotype clusters (A–D) found in Amerindians, the B lineage was not found in Siberia but was found in high frequency in Polynesia, Indonesia and South America, and also in unmixed native tribes in Argentina (Torroni et al., 1993; Cann, 1994; Cann & Lum, 1996). Additional genetic studies are, however, needed on genotype F strains, in particular those from Amerindians in North America, to relate different Amerindian groups to the clades of genotypes F and H.

REFERENCES


