HBe antigen (HBeAg)-negative carriers are a heterogeneous group with different levels of viral DNA, clinical course and prognosis. Deletion mutants have been described in immunosuppressed patients with chronic hepatitis B by amplification of whole hepatitis B virus (HBV) genomes (Gunther et al., 1995; Preikschat et al., 2002). HBV variants identified by subgenomic amplification and conventional cloning that lack regulatory elements for pregenome synthesis, a functional encapsidation signal or both of the direct repeats are not consistent with the wild-type replication strategy (Sommer et al., 1997). However, a recent report suggests that some of these non-conventional genomes can produce pregenomic-like RNA molecules and a low level of core and P proteins in the context of a wild-type genome in vitro (Chang et al., 2004).

In order to study this highly complex process of molecular evolution of the HBV genomes, we have performed a longitudinal follow-up (3 years) during e-minus chronic hepatitis B infection in an F genotype-infected patient experiencing a severe disease course. The follow-up was performed by using a complete-genome amplification strategy and sequencing, or long-fragment amplification (nearly half of the genome) and molecular cloning, to increase the probability of detecting non-conventional deletion mutants. Also, the dynamics of intrahost viral evolution were analysed, including the rate of mutation accumulation (Supplementary Fig. S1).

A 49-year-old Argentine male (patient RI), whose biopsy confirmed a chronic active hepatitis with cirrhosis, was retrospectively studied by analysing samples obtained in 1995 (RI95) and 1998 (RI98). No markers of other hepatotropic viral infections were observed when using commercially available enzyme immunoassays.

For the analysis of the entire nucleotide sequence, the HBV DNA isolated from serum samples (RI95 and RI98) was amplified by PCR with primer set P1 and P2 (Gunther et al., 1995). Two subgenomic regions were also amplified (from RI95, RI97 and RI98). The first one, named fragment 1, spanning the COOH-S, RT and RNaseH domains in the Pol, X and Core proteins (1791 bp), was amplified using primers HBV19 [nucleotide (nt) positions 688–708] and HBV2 (nt 2484–2465). The second one, named fragment 2 (1428 bp), spanning the C-terminal of the c gene, the spacer and RT domains within the s gene, and the complete s gene, was amplified using primers HBV15 (nt 2281–2300) and HBV6 (nt 893–874).
Subgenomic PCR fragments 1 and 2 were purified. Fragments 1.8 kb in length and several minor bands in the range 0.5–1.8 kb were recovered together, and were cloned into the vector pGEMT-Easy (Promega) according to the manufacturer’s instructions.

Direct PCR- and clone-derived sequences were obtained. Sequence data were analysed by using CLUSTAL_X (Thompson et al., 1997) and MEGA2 (Kumar et al., 2001), and a maximum-likelihood (ML) phylogenetic tree was estimated by using the PAUP* package (Swofford et al., 2001). The data analysed here were selected with the Akaike Information Criterion (AIC) by using modeltest version 3.06 (Posada & Crandall, 1998). A bootstrap resampling test with 1000 replicates was performed to assess the robustness of the neighbour-joining tree.

Sequence alignments were performed by comparing them with 46 AH genotype sequences (accession numbers and primers sequences can be obtained upon request).

Analyses of numerical data and statistical analyses were performed with Microsoft Excel and InfoStat (Universidad Nacional de Córdoba, Estadística y Diseño-FCA, Córdoba, Argentina). Statistical comparison of mutation distribution in multiple regions was performed by one-way analysis of variance (ANOVA), followed by comparison of individual regions by using the Student–Newman–Keul post hoc test.

Both RI95 and RI98 sequences carried the most frequently observed pre-core mutation, the G to A transversion at nt 1896, which prevents the expression of the HBeAg precursor.

PCR-derived direct sequences from RI95 and RI98 samples were ascribed to the 1b cluster of genotype F (Mbayed et al., 2001; Piñeiro y Leone et al., 2003) with a high statistical support (Fig. 1a). As observed in Fig. 1(b), a monophyletic origin was also observed for both PCR-derived direct sequences as well as for all clone-derived sequences. These results proved unequivocally that the RI95 sequence was directly ancestral to the RI98 sequence. Thus, this result convincingly supported the value obtained from the mutation rate calculation for the RI strain given below.

When the nucleotide variability between RI95 and RI98 during a 3 year period was analysed, several findings deserved to be stressed: 26 mutations emerged, which implied a mutation rate of 2.7 × 10⁻³ nt per site per year; such changes were mainly located at single coding regions (P = 0.002); 10 nucleotide ambiguities suggested that variants were coexisting within the individual.

The 26 nucleotide differences between the RI95 and RI98 consensus sequences and the 10 ambiguities were responsible for 16 synonymous and 31 non-synonymous changes (Supplementary Table S1). In addition, nucleotide substitutions were significantly associated with amino acid changes (P = 0.0036). When we compared RI95 and RI98 consensus sequences with 60 other sequences from GenBank belonging to different genotypes, two kinds of non-synonymous changes were observed. Twelve out of 31 amino acid substitutions were exclusive to the RI95 and RI98 samples under comparison, while the remaining changes involved amino acids or, at least, polymorphic positions already existing in the GenBank analysed sequences (accession number can be provided on request).

Due to the accumulation of mutations in single coding regions, and the presence of mixed populations in sequences obtained by direct PCR, it was decided to amplify and clone fragment 1 (Supplementary Fig. S1, Supplementary Table S1). This fragment includes single and double (overlapping) coding regions.

Several consistent independent results from fragment 1 amplifications were obtained. Amplicons were heterogeneous in size, since they were composed of the expected-size fragment together with a smear encompassing smaller fragments. Thirty-two clones were analysed from RI95 and RI98 samples. The insert size diversity was higher for RI95 than for RI98 clones, as shown in Fig. 2(a). A similar result was obtained from the amplification and cloning of fragment 2 from RI95, RI97 and RI98 samples, in which the diversity of cloned inserts followed the pattern RI95 > RI98 > RI97 (data not shown).

The insert size diversity described above added a new reason for the sequencing of cloned fragment 1 genomes. Accordingly, 20 clones derived from RI95 and RI98 samples were sequenced. The selection criterion for sequencing was to analyse a mixed population of full and deleted insert sizes. The sequencing analysis results revealed that a near-identical insert size pattern corresponded to almost the same deletion mapping for all the clones (Fig. 2a, b). Supplementary Table S2 shows the boundaries of the deletions on each of the 13 deletant clones as well as the genes involved. Interestingly, nine out of 13 clones had deletion boundaries at the DRI sequence. HBx was the gene most commonly deleted (11 out of 13 clones; 10 of 11 completely deleted and 1 of 11 partially deleted), while in addition, the p and c genes were deleted in some clones (Fig. 2b). Only two of the 13 studied truncated clones showed no deletions at the x gene, although both of them had double deletions (Supplementary Table S2, Fig. 2b), encompassing 458 and 42 nt, respectively.

Regarding the complexity of RI95 and RI98 HBV populations, a quasispecies structure of cloned genomes was observed. None of the cloned genomes showed an identical sequence to any other (Figs 1b and 2b), all of them being ascribed to a monophyletic clade. Nucleotide substitutions within the RI95 and RI98 consensus, as well as within RI95 and RI98 clone-derived sequences, were mainly distributed at single coding regions. The hypothesis suggested above of the co-circulation of variants was confirmed by individual clone-derived sequence analysis (positions I–VI, Fig. 2b). Therefore, a dynamic intrahost evolution process is proposed for HBV genomes.

http://vir.sgmjournals.org
In order to provide statistical support for this observation, an ANOVA one-way method was applied. The results obtained showed that the mean mutation rates were dissimilar among the A–F regions (Fig. 2b), with \( P = 0.0001 \). Subsequently, the Student–Newman–Keul post hoc test sorted the mean mutation rates of all these regions by decreasing value, showing that such mutations were mainly located in the non-overlapping coding regions E (\( c \) gene; mean mutation rate 0.01110) and B (3' end of \( p \) gene; mean 0.01068), followed by regions F (mean 0.00654), C (mean 0.00610), A (mean 0.00515) and D (mean 0.00476), with \( P < 0.05 \) (Supplementary Table S3).

Genetic diversity was assessed by using pairwise genetic distances among clone-derived sequences from RI95 and RI98 (i.e. intragroup and intergroup analyses). Genetic distances within the RI95 sequences were greater than those from RI98 (mean ± SD 0.0167 ± 0.00864 versus 0.0101 ± 0.00514; \( P < 0.05 \); Student's \( t \) test). To sum up, the higher diversity of insert size observed for RI95 clones (Fig. 2a) was associated with their higher pairwise genetic distances of nucleotide sequences.

Overall, a dual variation mechanism, brought about by point mutations and long genomic deletions, was observed.

**Fig. 1.** (a) Phylogenetic analysis of RI95 (white triangles) and RI98 (white ovals) direct PCR sequences. Both were ascribed to genotype F cluster 1b after comparison with GenBank-deposited sequences belonging to known AH genotypes. The monophyletic behaviour of RI95 and RI98 was supported by a significant bootstrap value (98). (b) Phylogenetic analysis of both RI95 (white triangles) and RI98 (white oval) direct PCR and clone-derived [RI95 (black triangles) and RI98 (black ovals)] sequences. The highest bootstrap values are shown.
Fig. 2. (a) Subgenomic fragment 1 length polymorphisms from RI95 and RI98 samples. Lanes 1–8 and 10–13 show the electrophoretic pattern of 20 sequenced clones. Numbers in parentheses represent the fraction of clones belonging to each pattern. (b) Schematic structure of RI95 and RI98 PCR-derived direct sequences and their respective clones. The RI95 sequence was considered the ancestral sequence and all of the nucleotide changes were referred to it with a lollipop symbol and a colour code. A–F regions were arbitrarily established for the analyses of mutations and diversity.
The genomic structure of all deletants described in this study reveals the generation of a linear DNA intermediate prior to viral integration. Deletion of the DR2 element in 11 out of 13 defective mutants does not allow the translocation of the RNA primer to the homologous DR2 site on the DNA-minus strand. For wild-type HBV, this translocation step is essential for production of the relaxed circular form of the HBV genome. Without the translocation of the RNA primer, which also occurs in wild-type HBV at low frequency, double-stranded linear HBV DNA molecules are formed by in situ priming of DNA-plus-strand synthesis (Staprans et al., 1991). In 10 of 13 deletant clones, the x gene was completely absent. As has been recently reported, the C-terminal region of the X protein exhibits inhibitory effects on cell proliferation and transformation; thus, deletion of this region may enhance the transformation ability of ras and c-myc (Tu et al., 2001). In addition, another recently published study on hepatocellular carcinoma (Wang et al., 2004) shows similar X-protein deletions to those observed in this report.

The truncated genomes are not competent for efficient autonomous replication. However, their presence and interaction with wild-type genomes may contribute to the progress and persistence of infection. It has been proposed, particularly for quasispecies-like viral populations such as those of HCV, that a large proportion of the virions circulating in the bloodstream contain defective genomes that may decrease the pathogenetic potential of the agent and facilitate persistence (Esteban et al., 1999).

This study has described some viral factors which shed some light upon the late, persistent interactions of HBV with the host. The results obtained demonstrate several interesting features: (a) the mutation fixation rate for the HBeAg-negative RI strain throughout the course of infection was about two orders of magnitude higher than those for HBeAg-positive HBV genomes ($10^{-3}$ versus $10^{-5}$ nt per site per year, respectively); (b) ambiguities (mixed populations) at the consensus sequences from RI95 and RI98 samples are indicative of the complexity of the genomic population; (c) the distribution of nucleotide substitutions is related to the structural constraints imposed by the highly compact genomic structure of HBV; (d) the nucleotide changes are significantly associated with non-synonymous substitutions, which denotes a driving force for positive selection; and (e), despite the minor contribution of HBV deletant genomes to the whole viral population, they may be important, both to maintain HBV persistence and to promote the integration of HBV DNA.

**Acknowledgements**

This work was partly supported by grants from Universidad de Buenos Aires (UBA; B076 and M057), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET; PID 723), and Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT; PICT 208355 and 10871). The authors are grateful to Professor Dr G. Carbállal, Department of Virology, CEMIC Institute, Buenos Aires, Argentina, for providing an aliquot of serum samples from patient RI.

### Table 1. Non-overlapping coding regions in positively and negatively selected sites of HBV Sym (partially analysed region within fragment 1), x and c genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Positively selected sites (amino acid position)</th>
<th>Bayes factor ($\geq 1/p$)</th>
<th>Negatively selected sites (amino acid position)</th>
<th>Bayes factor ($\geq 1/p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>592</td>
<td>350</td>
<td>643</td>
<td>3678</td>
</tr>
<tr>
<td></td>
<td>615</td>
<td>370</td>
<td>651</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>644</td>
<td>402</td>
<td>661</td>
<td>1101</td>
</tr>
<tr>
<td></td>
<td>698</td>
<td>60</td>
<td>709</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>716</td>
<td>96</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>748</td>
<td>68</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>x</td>
<td>122</td>
<td>1007</td>
<td>125</td>
<td>195</td>
</tr>
<tr>
<td></td>
<td>127</td>
<td>103</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>613</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>134</td>
<td>24585</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>c</td>
<td>31</td>
<td>427</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>429</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>82</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>82</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>72</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>97</td>
<td>205</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>73</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
References


