Virology 441 (2013) 40–48

Contents lists available at SciVerse ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Influence of overlapping genes on the evolution of human hepatitis B virus

Carolina Torres [a,b], María Dolores Blanco Fernández [a,b], Diego Martín Flichman [a,b], Rodolfo Héctor Campos [a,b], Viviana Andrea Mbayed [a,b,*]

[a] Cátedra de Virología, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina
[b] CONICET, Argentina

ARTICLE INFO

Article history:
Received 8 January 2013
Returned to author for revisions
5 February 2013
Accepted 28 February 2013
Available online 27 March 2013

Keywords:
Hepatitis B virus
Evolution
Overlapping genes
Selective pressure
Codon usage bias

ABSTRACT

The aim of this work was to analyse the influence of overlapping genes on the evolution of hepatitis B virus (HBV). A differential evolutionary behaviour among genetic regions and clinical status was found. Dissimilar levels of conservation of the different protein regions could derive from alternative mechanisms to maintain functionality. We propose that, in overlapping regions, selective constraints on one of the genes could drive the substitution process. This would allow protein conservation in one gene by synonymous substitutions while mechanisms of tolerance to the change operate in the overlapping gene (e.g. usage of amino acids with high-degeneracy codons, differential codon usage and replacement by physicochemically similar amino acids). In addition, differential selection pressure according to the HBeAg status was found in all genes, suggesting that the immune response could be one of the factors that would constrain viral replication by interacting with different HBV proteins during the HBeAg(−) stage.

Introduction

Hepatitis B virus (HBV) has a DNA genome and it replicates via a RNA intermediate. The substitution rate of HBV is thus expected to be determined by the low fidelity of its reverse transcriptase, but also, by the partial overlapping of the four open reading frames (ORFs) encoding the polymerase (P), surface antigen (S), nucleocapsid (C) and X protein (X). Particularly, gene P accounts for the third part of the genome and is overlapped with all the other genes of the genome. Besides, regulatory signals are also embedded in protein-encoding genes, and secondary structures of the HBV RNA might constrain the number and nature of substitutions occurring in the HBV genome (Baumert et al., 1998; Beck and Nassal, 1998; Kidd and Kidd-Ljunggren, 1996).

Such a complex organisation of the HBV genome allows assuming different rates of variation among sites along the genome. It is well known that substitution rates are different at the three positions of the codons due to selective constraints at the protein level, although a bias in the rate differences could be expected in the HBV genomic regions where two ORFs overlap. Besides, different types of overlapping might have different impacts on this bias (Krakauer, 2000). Previous works have studied some aspects of the HBV evolutionary behaviour (Mizokami et al., 1997; Yang et al., 1995; Zaaijer et al., 2007; Zhang et al., 2010). However, a study from an overall point of view about the evolutionary characteristics of the complete HBV genome on all major genotypes is still missing.

Chronic infections are characterised by an initial phase with high serum HBV DNA levels and detection of HBeAg (HBeAg(+) stage). Most of the carriers eventually loses HBeAg and develop antibody to HBeAg (HBeAg(−) stage) (Lok and McMahon, 2009). The transition between these stages is a process known as HBeAg seroconversion and would occur through an immune mechanism.

Thus, viral sequences obtained from HBeAg(+) or HBeAg(−) stages of the infection could retain information about the process involved. Differences in selection pressures among these stages have been proposed based mainly on the studies on gene C. Nevertheless, those studies were performed without covering all major genotypes (Abbott et al., 2010; Lim et al., 2007; Warner et al., 2011).

The aim of this work was to analyse the influence of overlapping genes on the evolution of HBV, particularly: (a) to analyse the codon-site nucleotide variation for the first, second and third codon positions of all the HBV genes, (b) to study selective pressures and synonymous and nonsynonymous substitutions, (c) to analyse amino acid and codon usages in HBV, and (d) to study the evolutionary characteristics in relation to the HBeAg status.

Results

Codon-site nucleotide variation

In order to evaluate the influence of the overlapping reading frames on the nucleotide variation pattern in the HBV genome,

* Correspondence to: Cátedra de Virología, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Junín 956, 4° piso, Ciudad Autónoma de Buenos Aires 1113, Argentina. Fax: +54 11 45083645.
E-mail address: vmbayed@ffyb.uba.ar (V.A. Mbayed).

0042-6822/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2013.02.027

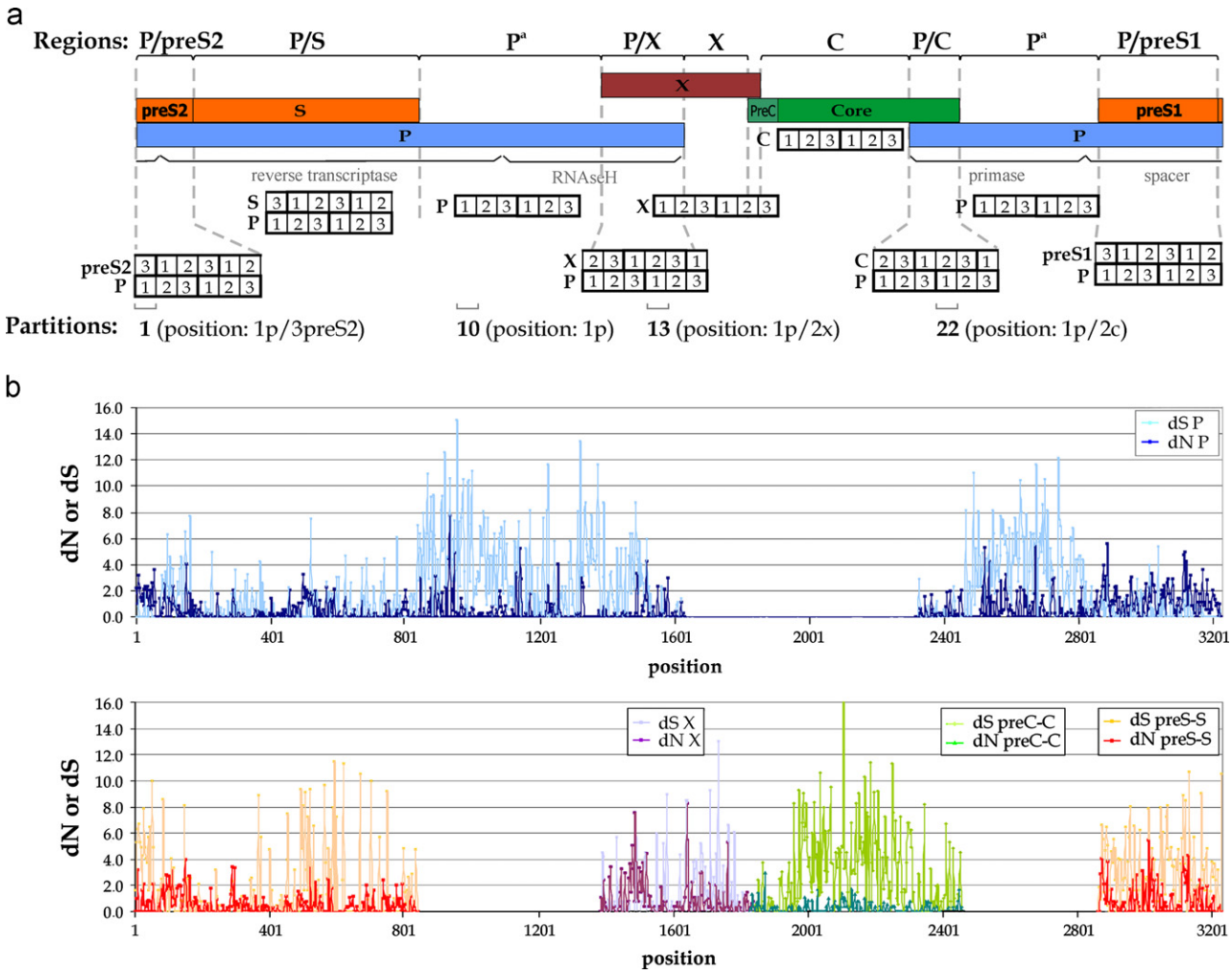

Fig. 1. (a) HBV genetic regions analysed (some examples of partitions are shown, see Table 1 for the complete list of partition), and (b) dN and dS estimates for all HBV genes. [a]P nonoverlapping regions were analysed as a whole.

relative substitutions rates for the first, second and third codon positions of overlapping and nonoverlapping genetic regions were estimated for Datasets A (Fig. 1a and Table 1).

As a result, in the nonoverlapping regions of genes P, C and X, their respective third positions (3p, 3c and 3x) were the most variable. The third position of gene P displayed the highest rate of variation.

For the P/X overlapping region, position 3p/1x was the most variable. In contrast, in other regions, the third codon positions of genes overlapped with P were the most variable. In the P/preS-S region, the most variable codon position was 1p/3preS-S, whereas in the P/C region, the most variable codon position was 2p/3c, although with a high standard error (SE) value.

In addition, the discrete gamma distribution was used to characterise the among-site rate variation for each codon position in all genetic regions. This analysis showed that three positions (3p, 3c, 1p/3preS1) displayed α values higher than or equal to 1. In addition, these three positions reached high substitution rates. It means that most of the sites involved in these codon positions displayed a uniform high substitution rate (Table 1). On the contrary, all of the other codon positions in the genome showed $\alpha<1$, which means that a high substitution rate would be displayed only for some of these sites.

Similar results were obtained on seven genotype-specific HBeAg(+) datasets (available upon request).

Analyses of selective pressure and synonymous and nonsynonymous substitutions

To characterise the pattern of selection that operates on the HBV genome, two approaches were applied.

The first method is particularly designed to analyse overlapping genetic regions and allows finding signatures of purifying selection on genes (pattern of conservation) comparing the expected and observed number of synonymous and nonsynonymous substitutions in both ORFs (Table 2). Signatures of purifying selection were found in preS1, preS2 and S regions of ORF preS-S and in ORF P in P/S and P/X regions.

The second method (SLAC) is one of the traditional approaches commonly applied to study selective pressure. Here, it was used to detect patterns of nonconservation and to compare the behaviour of overlapping and nonoverlapping regions, which is not possible using the first method since it is focused on detecting purifying selection on overlapping genes. This method accounts for an independent estimation of dN and dS, which becomes especially important when, as it occurs in HBV, a nonuniform distribution of rates is present (Kosakovsky Pond and Frost, 2005).

Different analyses were performed with the SLAC method:

i. First, to analyse overall patterns of conservation (dN<dS) or nonconservation (dN>dS), the dS and dN values were

Table 1
Codon-site nucleotide variation.

HBV genetic region[a]	Partition	Region length (nucleotides)	Codon position	Relative rate±SE	α parameter
P/preS2	1	55	1p/3preS2	1[b]	0.91
	2	55	2p/1preS2	0.243±0.072	0.33
	3	55	3p/2preS2	0.587±0.139	0.77
P/preS1	4	119	1p/3preS1	1.190±0.214	1.83
	5	119	2p/1preS1	0.497±0.106	0.53
	6	119	3p/2preS1	0.278±0.061	0.62
P/S	7	227	1p/3s	0.505±0.097	0.23
	8	227	2p/1s	0.112±0.026	0.21
	9	227	3p/2s	0.246±0.051	0.35
Nonoverlapping P	10	311	1p	0.475±0.088	0.22
	11	311	2p	0.244±0.055	0.16
	12	311	3p	2.087±0.357	1.05
P/X	13	83	1p/2x	0.240±0.109	0.18
	14	83	2p/3x	0.414±0.095	0.10
	15	84	3p/1x	0.903±0.191	0.37
Nonoverlapping X	16	63	1x	0.599±0.139	0.48
	17	64	2x	0.248±0.069	0.38
	18	63	3x	0.780±0.184	0.45
Nonoverlapping C	19	156	1c	0.268±0.060	0.30
	20	156	2c	0.094±0.025	0.21
	21	156	3c	1.400±0.258	0.97
P/C	22	49	1p/2c	0.083±0.031	0.62
	23	49	2p/3c	0.378±0.100	0.21
	24	48	3p/1c	0.218±0.071	0.34

[a] Results of the C/X overlapping region are not shown because it is composed of less than 10 codons.
[b] Estimated substitution rates are relative to this position.

statistically compared for each region. A conservative pattern was observed in all nonoverlapping regions and also in the overlapping regions on ORF S (preS1 and preS2) and on ORF P (P/X region) (Table 3). This last result is coincident with that obtained from the first method used. Besides the first method being the appropriate one to analyse selective pressures on HBV given the overlapping nature of its genome, performing the analysis using traditional methods (such as the SLAC) allowed us to obtain results that did not disagree with those obtained with the first method. However, the purifying selection on P/S overlapping region in both ORFs was only detected by the first approach but not by the SLAC method. This is consistent with the lower sensitivity of the SLAC method.

On the other hand, the nonconservative patterns were infrequent according to SLAC and occurred in overlapping regions (P/preS1 –ORF P-, P/C –ORF P-, P/X –ORF X-) (Table 3).

Some results were notorious. A dual behaviour was observed in some overlapping regions: while one of the genes showed a conservative pattern, the overlapping gene showed a nonconservative one. This was observed for the P/X region (with a dN<dS for gene P and dN>dS for gene X) and for the P/preS1 region (where the nonconservative pattern on gene P was accompanied by a conservative pattern on the ORF preS-S).

Another noteworthy result was the pattern of conservation/nonconservation in the P/S region. Given the high nucleotide substitution rate of position 1p/3s (see Table 1), a nonconservative pattern on gene P along the overlapping P/S region was expected. However, purifying selection was observed. Then, the mechanism that allows gene P to vary its first codon position without causing a high number of nonsynonymous changes was further studied (see Amino acid and codon usages section).

ii. The SLAC method was also applied to analyse the behaviour of overlapping vs nonoverlapping regions of each gene. The dS values were always higher in the nonoverlapping regions; whereas, most of the dN values did not present significant difference between overlapping and nonoverlapping regions of each gene (Table 3 and Fig. 1b). As an exception, gene P in the overlapping P/preS1 and P/preS2 region showed dN values higher than in the nonoverlapping region.

iii. Finally, site-to-site selection was analysed. In this case, 31 sites had a significant dN>dS and 27 were found in overlapping regions. Besides, 22 out of those 27 sites with dN>dS in an overlapping reading frame were associated with sites having significant dN<dS in the other ORF (Fig. S1 and Table S3), evidencing the interdependence of genes.

Similar results for the SLAC method were obtained on seven genotype-specific HBeAg(+) datasets (available upon request).

Amino acid and codon usages

To evaluate possible mechanisms of tolerance to the change in HBV, amino acid and codon usages between overlapping and nonoverlapping genetic regions were studied.

A differential amino acid usage was observed between the overlapping and the nonoverlapping regions of all genes. In particular, the overlapping regions showed higher usage of amino acids codified by six synonymous codons and lower usage of amino acids codified by

Table 2
Signature of purifying selection on HBV overlapping regions.

Overlapping region	ORF 1/*p*-value[a]	ORF 2/*p*-value[a]	Ts/Tv[b]	Possible/Observed	Nonsynonymous substitutions in ORF 2[c]		Synonymous substitutions in ORF 2[c]	
					N_2N_1	N_2S_1	S_2N_1	S_2S_1
P/preS1	**preS1**	**P**	**Ts**	**P**	100.5	98.3	99.3	2.4
				O	32.0	87.0	31.0	3.0
	0.000	0.077	**Tv**	**P**	360.8	116.4	118.1	5.7
				O	90.0	84.0	14.0	8.0
P/preS2	**preS2**	**P**	**Ts**	**P**	55.2	53.9	49.9	0.1
				O	10.0	33.0	28.0	2.0
	0.000	0.060	**Tv**	**P**	197.4	68.8	50.6	1.1
				O	18.0	26.0	5.0	1.0
P/S	**S**	**P**	**Ts**	**P**	228.4	208.4	216.6	19.1
				O	21.0	38.0	51.0	23.0
	0.000	**0.003**	**Tv**	**P**	833.1	249.1	256.5	6.4
				O	39.0	41.0	24.0	5.0
P/C	**C**	**P**	**Ts**	**P**	41.3	51.7	47.0	1.0
				O	3.0	2.0	11.0	1.0
	0.706	0.135	**Tv**	**P**	159.8	48.2	69.4	4.6
				O	10.0	3.0	9.0	4.0
P/X	**X**	**P**	**Ts**	**P**	79.2	79.5	81.8	4.6
				O	9.0	15.0	41.0	3.0
	0.159	**0.000**	**Tv**	**P**	243.6	122.5	122.1	2.0
				O	12.0	14.0	41.0	2.0

[a] Significant *p*-values ($p<0.05$) are indicative of purifying selection and are shown in bold.
[b] Ts: transitions. Tv: transversions.
[c] N_2N_1: Nonsynonymous substitution on genes 2 and 1, N_2S_1: Nonsynonymous substitution on gene 2 and synonymous substitutions on gene 1, S_2S_1: Synonymous substitution on genes 2 and 1, S_2N_1: Synonymous substitution on gene 2 and nonsynonymous substitutions on gene 1.

Table 3
dS and dN values for each HBV genetic region in the HBeAg(+) dataset.

HBV genetic region	ORF	Region length (codons)	dS			dN			dN vs dS	*p*-value (dN vs dS)[e]
			Median	25–75% Percentile	*p*-value (dS vs dS)[a]	Median	25–75% Percentile	*p*-value (dN vs dN)[a]		
P/preS1	**preS-S**	119	**2.481**	0.809–4.638	–	**0.405**	0.000–1.323	–	dN<dS	**<0.001**
P/preS2		55	**1.619**	0.000–4.918	–	**0.809**	0.000–1.829	–	dN<dS	**<0.05**
P/S		226	**0.000**	0.000–1.166	–	**0.000**	0.000–0.552	–		>0.05
P/PreS1	**P**	118	**0.000**	0.000–0.922	**<0.001**[b]	**1.157**	0.559–2.157	**<0.001**[b]	dN>dS	**<0.001**
P/PreS2		55	**0.583**	0.000–2.049	**<0.001**[b]	**0.862**	0.000–1.612	**<0.001**[b]		>0.05
P/S		227	**0.000**	0.000–0.915	**<0.001**[b]	**0.000**	0.000–0.420	>0.05[b]		>0.05
P/C		48	**0.000**	0.000–0.000	**<0.001**[b]	**0.123**	0.000–0.409	>0.05[b]	dN>dS	**<0.05**
P/X		82	**0.583**	0.000–2.333	**<0.001**[b]	**0.000**	0.000–0.292	>0.05[b]	dN<dS	**<0.001**
Nonoverlapping P		309	**3.499**	1.081–5.852	–	**0.000**	0.000–0.583	–	dN<dS	**<0.001**
P/C	**preC-C**	47	**0.000**	0.000–1.243	**<0.001**[c]	**0.000**	0.000–0.377	>0.05[c]		>0.05
Nonoverlapping C		155	**3.020**	0.755–5.284	–	**0.000**	0.000–0.377	–	dN<dS	**<0.001**
P/X	**X**	83	**0.000**	0.000–0.747	**<0.001**[d]	**0.374**	0.000–1.244	>0.05[d]	dN>dS	**<0.01**
Nonoverlapping X		62	**1.048**	0.000–3.452	–	**0.346**	0.000–1.121	–	dN<dS	**<0.05**

Significant *p*-values ($p<0.05$) are shown in bold.
[a] Overlapping vs nonoverlapping regions within each gene.
[b] Compared with nonoverlapping P.
[c] Compared with nonoverlapping C.
[d] Compared with nonoverlapping X.
[e] Intra-region comparison.

one or two synonymous codons (Fig. 2a and b). As an example, in the P/S region, ORF P showed a high frequency of the amino acids serine and leucine with respect to the nonoverlapping P region (data not shown). These amino acids are codified by six synonymous codons and for which a change on the first codon position allows synonymous changes to occur (although, in the case of serine, a change in the second codon position is also needed). The increased usage of 6-fold degenerate amino acids could explain the fact that the high variation in position 1p/3s was not accompanied by an increase in the dN along gene P (see "Codon-site nucleotide variation", in Results section).

On the other hand, if an amino acid change is committed, almost all HBV genomic regions showed a higher probability of nonsynonymous replacements with amino acids with similar physicochemical characteristics than that expected by chance (Observed vs Expected comparison, Table 4). However, this mechanism was constrained in the overlapping regions, given that, in general, they showed lower number of amino acid substitutions within a single class that the nonoverlapping regions of each ORF (Observed substitutions for overlapping vs nonoverlapping regions, Table 4).

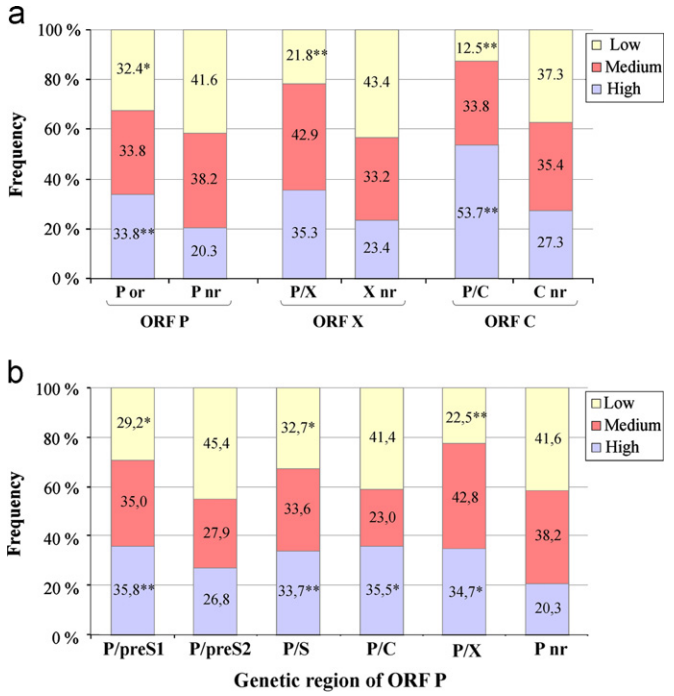

Fig. 2. Amino acid usage according to the level of degeneracy of codons for genes P, preC-C and X in overlapping and nonoverlapping regions (a) and for different regions of gene P (b). $^{*}p<0.05$ and $^{**}p<0.01$ values corresponding to the comparison between overlapping and nonoverlapping regions.

In addition, intra-region amino acid diversity was estimated to evaluate the tolerance to the change in each genetic region. Different regions showed dissimilar tolerance to nonsynonymous changes. Gene P in the P/preS1 and P/preS2 regions showed the highest amino acid diversity and thus, the highest tolerance to the change (Table S4).

When the frequency of use of alternative codons was analysed, the results showed that, in most cases, the codon usage in the nonoverlapping region of each gene did not show a significant correlation with the one used in the overlapping regions (Table 5). In contrast, the codon usage of the nonoverlapping HBV regions did show correlation with the human codon usage of the liver-specific highly expressed genes but not with brain-specific ones (as representative of a tissue that does not sustain HBV replication) (Table 5).

Differential adaptive evolution according to the HBeAg(+)/(−) status

In order to analyse the evolutionary characteristics in HBV in relation to the HBeAg status, two strategies were followed to compare HBeAg(+) and HBeAg(−) groups: the comparison of dN and dS in both groups and the detection of differential selection pressures on individual codon sites between both populations.

On one hand, the results showed that although both groups had conservative patterns in the nonoverlapping C region, the HBeAg (−) dataset was less conservative (the dN/dS ratio is higher in the HBeAg(−) dataset), owing to a higher number of nonsynonymous substitutions (dN) and a lower number of synonymous substitutions (dS) (Table S5). In addition, in the overlapping P/S region the HBeAg(−) dataset showed a nonconservative pattern in gene S along with a conservative pattern in gene P, which was not observed for the HBeAg(+) group (Tables 3 and S5). Other feature found in the HBeAg(+) group and not in the HBeAg(−) dataset include a conservative pattern in ORF S in P/preS2 and in nonoverlapping X region and the nonconservative one in ORF P in P/C region.

Table 4
Amino acid substitution pattern.

HBV genetic region	ORF	Amino acid substitutions within a single class (%)[a]		*p*-value (observed vs expected)	*p*-value (or vs nr)[b]
		Observed	Expected		
P/preS1	**preS-S**	24.5	5.2	**<0.001**	–
P/preS2		19.7	8.2	**<0.001**	–
P/S		27.1	8.2	**<0.001**	–
P/preS1	**P**	15.7	7.6	**<0.001**	**<0.001**
P/preS2		28.7	5.9	**<0.001**	>0.05
P/S		22.8	11.0	**<0.001**	**<0.05**
P/C		30.8	10.3	**<0.001**	>0.05
P/X		20.8	8.9	**<0.001**	**<0.01**
Nonoverlapping P		28.9	7.9	**<0.001**	–
P/C	**preC-C**	5.9	5.1	>0.05	**<0.001**
Nonoverlapping C		47.2	9.5	**<0.001**	–
P/X	**X**	12.4	9.1	**<0.05**	>0.05
Nonoverlapping X		14.9	7.9	**<0.001**	–

Significant *p*-values ($p<0.05$) are shown in bold.

[a] Amino acid substitutions within a single class were counted according to the classification of the IMGT® database. The classes formed by more than one amino acid were considered: Aliphatic (A, I, L, V), Basic (R, H, K), Sulphur (C, M), Hydroxyl (S, T), Acidic (D, E), Amide (N, Q).

[b] Comparison between substitutions (Observed values) of overlapping regions (or) and nonoverlapping regions (nr) of ORFs P, preC-C and X.

Table 5
Correlation analyses among relative synonymous codon usage values.

HBV genetic region	ORF	Spearman r coefficient	Confidence interval (95%)	*p*-value
P/preS1	**P**	0.072	−0.195–0.329	>0.05[a]
P/preS2		0.032	−0.234–0.292	>0.05[a]
P/S		0.407	0.161–0.606	**0.001**[a]
P/C		0.184	−0.083–0.427	>0.05[a]
P/X		−0.096	−0.351–0.171	>0.05[a]
P/C	**preC-C**	0.333	0.076–0.548	**0.01**[b]
P/X	**X**	0.159	−0.109–0.405	>0.05[c]
Nr vs liver genes	–	0.286	0.025–0.511	**<0.05**[d]
Nr vs brain genes	–	0.072	−0.195–0.329	>0.05[e]

Significant p-values ($p<0.05$) are shown in bold.

[a] Compared with nonoverlapping P.

[b] Compared with nonoverlapping C.

[c] Compared with nonoverlapping X.

[d] Nonoverlapping regions vs highly expressed human liver genes.

[e] Nonoverlapping regions vs highly expressed human brain genes.

On the other hand, sites that have a significantly different dN/dS ratios among the two populations were analysed. Five codons in ORF preS-S, eight in P, 25 in preC-C and five in X were identified as differentially selected between the HBeAg(+) and HBeAg(−) groups ($p\leq0.05$) (Table S6). Most of the sites were associated with a higher dN/dS ratio in the HBeAg(−) dataset and were related to epitopes B and T.

Discussion

The complex genetic organisation of HBV allows assuming a different evolutionary behaviour along the genome. The evolution of a genetic region with overlapping genes is under constraints, since a nucleotide substitution would have impact, simultaneously, on two genes. However, the direction and the relative strength of

these constraints are not obvious and they have not yet been comprehensively studied.

In this study, we showed that the nonoverlapping regions present nucleotide patterns of variation, synonymous substitution rates and amino acid and codon usages different from those of the overlapping regions of the HBV genome.

Concerning the pattern of variation for codon positions, we found that the nonoverlapping regions of the HBV genome displayed the typical behaviour of nonoverlapping genes, whose third positions varied with moderate or high substitution rates. On the other hand, in the overlapping regions, the highest variation was observed in the third position of one of the overlapping genes but restricted to few sites (excepting overlapping P/preS1 that showed an homogeneously high substitution rate), which means that most of the sites that belonged to this third position remained almost constant (see Table 1). Although the highest nucleotide variation was observed in nonoverlapping regions, these regions were more conserved from the point of view of amino acids. From our results, the overlapping limited the synonymous substitutions, increasing the dN/dS relation. According to the traditional methods to evaluate selective pressures, a dN>dS could be qualified as "positive selection". However, these nonsynonymous changes could better represent a higher tolerance to the change, operating on one of the overlapping genes, than a real positive selection. In line with this, our results showed that the presumptive positive selection detected on one ORF was generally accompanied by a purifying selection on the overlapping ORF. It is worth noting that none of the sites presenting dN>dS was associated with genotype-specific polymorphisms, immunotherapy escape or failure in detection by diagnostic kits (Weber, 2005; Zoulim and Locarnini, 2009; Zuckerman and Zuckerman, 2003).

Previously, it has been suggested an independent evolution of HBV genes (Zaaijer et al., 2007). However, based on our results, we propose that the substitution process in the overlapping regions of the HBV genome would be driven by selective constraints that operate on one of the genes, rather than on both, which has been proposed for other viruses (Hughes and Hughes, 2005; Hughes et al., 2001; Narechania et al., 2005; Pavesi, 2006). Thus, a protein encoded in one frame would be relatively constrained by purifying selection whilst the protein encoded in the other frame would need to be able to tolerate amino acid changes. As a noteworthy exception, the behaviour of genes S and P in the P/S region that showed simultaneous purifying selection is also described here (see later in Discussion).

As examples of the more general rule of constrained and tolerant overlapping genes, a nonconservative pattern in gene P was observed in the overlapping with the preS1, which conversely exhibited a conservative pattern of changes (see Table 3). This behaviour may be explained by the critical role of preS1 region in the biology of HBV (Seeger and Mason, 2000), which might constrain its amino acids changes. The amino acids codified into the preS1 region, as part of the Large protein of the Surface antigen, would participate as the viral anti-receptor and the ligand for C particles during the virion assembly (Bruss et al., 1994; Ostapchuk et al., 1994; Prange and Streeck, 1995). In this region, gene P codifies for the "Spacer" domain without a known function in the infectious cycle of the HBV, which was able to tolerate amino acids changes. Another example can be pointed out in the overlapping P/X, in which the P protein codifies for the C-terminal region of the "RNAseH" domain that is responsible for the degradation of the RNA template during reverse transcription (Chen and Marion, 1996), displaying a conservative pattern of substitution. Whereas, in this region, gene X codifies for the N-terminal end of X protein with an elusive role (Tang et al., 2005) being able to tolerate a nonconservative pattern. Therefore, an inverse relationship between the functional importance of regions and the level of diversity has been found. This was previously proposed for HBV based on partial sequences (Zhang et al., 2010), although in this work it is demonstrated based on complete genomes.

A dissimilar level of conservation displayed by different protein coding regions could derive from the use of alternative mechanisms to maintain functionality, involving not the whole gene but a partial region with a specific function, process that we call as a "regional" evolution of genes.

Some mechanisms are here proposed to explain the tolerance to the changes in different genetic regions:

First, an increase in the usage of amino acids with a high level of degeneracy in the overlapping regions. This has been previously shown to occur in other viruses such as coliphages, HIV-1 and some avian HBV (Pavesi et al., 1997). Here, we also demonstrated that this increase was at the expense of a decrease in the usage of amino acids with a low level of degeneracy (see Fig. 2). This mechanism might imply that nonsynonymous changes in one ORF could be diminished by increasing the use of six-fold degenerate amino acids that would allow synonymous substitutions by changes in the first codon position.

Second, a codon usage bias in most of the overlapping regions of the HBV genome (see Table 5). This strategy, combined with the previous one, could offer higher flexibility (higher number of options) to use codons that better stand the requirement of the overlapping ORF.

Third, nonsynonymous replacements by amino acids with similar physicochemical characteristics. This mechanism, common in nonoverlapping genes, would allow virus to evolve even in overlapping regions (see Table 4).

All these mechanisms can be exemplified in the analysis of the P/preS-S region, where position p1/preS-S3 was the most variable, suggesting that nonsynonymous changes in gene P and synonymous changes in the preS-S would prevail. This was confirmed in the P/preS region but failed to predict the evolutionary behaviour of gene P in the P/S region, which showed low dN and even, purifying selection. In other words, according to the analysis of nucleotide variation the purifying selection acting on gene S in the P/S region is not surprising; however, it is noteworthy to obtain this finding in gene P. This dual purifying selection is very infrequent in nature and probably reflects a long-term evolution of this virus. This was previously reported on the complete P/preS-S region (Sabath et al., 2011), although in this work it is demonstrated that this pattern is driven only for the behaviour of the P/S and not for the P/preS region (see Table 2).

Thus, the usage of amino acids with six synonymous codons in gene P is the mechanism proposed to conserve the amino acid sequence in the P/S region. Besides, the replacement of amino acids with others that share physicochemical properties could also be an evolutionary key to tolerate nonsynonymous changes in this region of gene P, which corresponds to the Retrotranscriptase (RT) domain. In this way the amino acids found in high frequency in the RT would be codified by codons that better stand the requirement of the S protein.

In addition, we tested the codon usage bias between nonoverlapping HBV regions and human genes. This analysis showed that HBV codon usage correlated with that of liver-specific but not with brain-specific human genes, suggesting a high level of viral adaptation to the main target tissue of its host, given that the relative tRNA abundance might modulate the expression of proteins (Hershberg and Petrov, 2008).

An important stage of the HBV chronic infection, the HBeAg seroconversion, was also studied under this evolutionary perspective. Different investigators associated the HBeAg seroconversion process with an increase in the number of nonsynonymous substitutions (Abbott et al., 2010; Akarca and Lok, 1995; Lim

et al., 2007; van de Klundert et al., 2011). It has been hypothesised that amino acid changes could be selected to avoid recognition by the immune system (escape mutations) but also, not mutually exclusive, that they might restore any loss of function (compensatory mutations). However, most of the studies that evaluated these possibilities focused on gene C (Abbott et al., 2010; Bertoletti et al., 1994; Tsai et al., 1996). Here, we showed that the increase of nonsynonymous substitutions was also detectable on genes P, X and S during the HBeAg(−) phase of infection. Most of these amino acid changes were located on described HBV epitopes (Khakoo et al., 2000; Liu et al., 2011; Maman et al., 2011; Mizukoshi et al., 2004; Nayersina et al., 1993; Norton et al., 2010; Rehermann et al., 1995), which is consistent with an immune selection. Nevertheless, sites unrelated with epitopes were also found to be under different selection pressures among the HBeAg(+) and HBeAg(−) populations. It is worth noting that none of the differentially selected sites were related to sites previously associated with antiviral therapy. These findings remark the importance of analysing the full-length genome to have a complete outlook of the HBV molecular evolution in relation to different clinical stages of the infection.

On the other hand, factors other than the virus *per se*, which were not considered along this work owing to the lack of information, such as age of infection, time to HBeAg seroconversion, population genetic background, etc., could be involved in HBV molecular evolution (Chatzidaki et al., 2011; Liaw et al., 2011; Locarnini and Zoulim, 2010; Singh et al., 2007; Wu et al., 2010). Thus, the relative importance of these factors should be taken into account to improve the understanding of the virus-host interaction in HBV infections.

Conclusions

A differential evolutionary behaviour among the HBV genetic regions and HBV clinical status was found. The evolutionary process in overlapping genes would allow protein conservation in one frame by synonymous substitutions while mechanisms of tolerance to the change or adaptation might occur in the other frame, including the usage of amino acids codified by high-degeneracy codons, a differential codon usage and nonsynonymous substitutions by amino acids with similar physicochemical properties. In addition, a differential selection pressure according to the HBeAg status was found in all HBV genes, increasing the evidence that the immune response could be one of the factors that constrain viral replication by interacting mainly with protein C but also with the other proteins. However, sites unrelated with epitopes were also found, suggesting that other nonimmune selective pressures could also be acting.

Material and methods

Datasets

Complete genomes belonging to almost all major HBV genotypes (A–F and H) – including 39 isolates previously sequenced in our laboratory – were obtained from GenBank. Sequences were classified as HBeAg(+)/(−) according to the absence/presence of molecular markers of HBeAg(−) status: (a) mutations G1896A or AG1762/4TA, (b) absence of the ATG preC translational initiation codon, and (c) presence of any stop codon in the preC region (Brunetto et al., 1989; Carman et al., 1989; Günther, 2006; Yim and Lok, 2006). Sequences with insertions, deletions or internal stop codons in any ORF as well as recognised or potentially proposed recombinant sequences were excluded from the analysis (Kramvis et al., 2008; Morozov et al., 2000; Simmonds and Midgley, 2005). Genotype G was excluded because it possesses two constitutive translational stop codons at positions 2 and 28 of the preCore region, whereas the recently proposed genotypes I and J were excluded because of the lack of suitable sequence representation in databases (Tatematsu et al., 2009; Tran et al., 2008).

All analyses performed on this work were based on a dataset composed of sequences classified as HBeAg(+) with an equilibrated representation of the genotypes (n sequences = 140) (Dataset A). In addition, analyses of selective pressures mentioned below were also performed on a dataset composed of sequences classified as HBeAg(−) with an equilibrated representation of the genotypes (n sequences = 119) (Dataset B)

The nucleotide sequence accession numbers used in this work can be found in the Supplementary material (Tables S1 and S2).

Codon-site nucleotide variation

The magnitude and uniformity of the codon-site variation were evaluated under a Maximum Likelihood approach implemented in the Baseml module of PAML v4.1 package (Yang, 2007).

The magnitude of the variation was studied estimating relative substitution rates for the first, second and third codon positions. The genetic regions analysed were preS1, preS2, S, preC-C, X and P, in both overlapping and nonoverlapping regions (Fig. 1a). Alignments and phylogenetic trees were used as input for the analysis. Sequences were aligned with ClustalX v2.0.5 (Thompson et al., 1997) and edited with Bioedit v7.0.9.0 (Hall, 1999). Maximum Likelihood (ML) phylogenetic trees were obtained by using the PhyML v3.0 software (Guindon and Gascuel, 2003) with the nucleotide substitution model estimated with the jModeltest v0.1.1 software (Posada, 2008) according to the Akaike Information Criterion.

For this analysis, the complete genome was divided into eight regions: overlapping P/preS1, P/preS2, P/S, P/X, P/C, and nonoverlapping P, X and C; and each codon position was analysed, resulting in 24 different partitions corresponding to each codon position in each of the eight regions (see Fig. 1a and Table 1 for an exemplification and definition of the partititions). For instance, partition 1 grouped the nucleotides at the first codon position of P in the P/preS2 overlapping region that corresponds to the third position of preS2 (1p/3preS2), while partition 10 grouped the nucleotides at the first codon position in the P nonoverlapping region (1p). The three partitions of the C/X overlapping region were not considered since they are formed by less than 10 nucleotides.

The uniformity of variation was studied estimating the α parameter value of the function of the gamma distribution for each partition. This function can be used to describe the variation in substitution rates over nucleotide sites and then, from the estimation of the α parameter (which confers the shape on the function) it would be possible to infer the level of substitution rate uniformity (Yang, 1994). Then, α values lower than 1 represent extreme rate variation (combination of some sites varying with very high substitution rates and others with low rates) and α values higher or equal than 1 represent a high uniformity among substitution rates (Yang and Kumar, 1996). Thus, for each partition, a gamma distributed function was assumed and the α parameter value was estimated.

Analyses of selective pressure and synonymous and nonsynonymous substitutions

To study synonymous and nonsynonymous substitutions and selectives pressures on HBV two approaches were used.

First, a method particularly designed to analyse multiple alignments of related sequences with overlapping genetic regions was

applied (Sabath et al., 2011). This method estimates the observed and expected number of synonymous and nonsynonymous substitutions in both ORFs and allows finding signatures of purifying selection on genes. Briefly, the method consists of the construction of an unrooted phylogenetic tree (performed as above), the reconstruction of ancestral sequences, the classification of the changes along the tree into the substitutional categories (transitions/trasversions and synonymous/ nonsynonymous), and the testing for the signature of purifying selection through comparisons between categories (Sabath et al., 2011).

Second, rates of synonymous (dS) and nonsynonymous (dN) substitutions on the P, preS-S, preC-C and X genes were estimated applying the SLAC method (Kosakovsky Pond and Frost, 2005) implemented in HYPHY v2.0 (Kosakovsky Pond et al., 2005). Alignments and phylogenetic trees were used as input for the analysis and they were constructed as previously. To enable comparisons across genes and datasets, dN and dS values were scaled by the total length of the codon tree and thus, the scaled values represent the expected number of nucleotide substitutions per codon-site (Kosakovsky Pond and Frost, 2005). Scaled dN and dS values were compared with Mann-Whitney or Kruskall-Wallis tests ($p<0.05$) by using GraphPad v4.02 (GraphPad Software, San Diego California USA).

Amino acid and codon usages

Amino acid usage was evaluated by three approaches. First, amino acid composition differences between the overlapping and nonoverlapping genetic regions were estimated calculating amino acid frequencies and values were compared grouping amino acids according to the level of degeneracy of their codifying codons (high: 6, medium: 3 and 4, or low: 1 and 2 synonymous codons). Second, the amino acid substitution pattern was studied for each genetic region. This analysis was carried out by using DAMBE v5.0.7 (Xia and Xie, 2001). Standardized physicochemical classes of the 20 common amino acids were defined according to the international ImMunoGeneTics information system® (IMGT®) database (Pommie et al., 2004). This classification was used to analyse the type of nonsynonymous changes occurring in the different regions of the genome. Observed and expected amino acid substitutions within a single class were compared for each genetic region, whereas observed substitutions for overlapping and nonoverlapping regions were also compared. For these comparisons, the Proportion tests ($p<0.05$) implemented in GraphPad v4.02 was used. Finally, the amino acid diversity for each genetic region was estimated by using MEGA v5.05 (Tamura et al., 2011). The amino acid substitution matrix used for each region was selected among those available in MEGA with the ProtTest v2.4 software (Abascal et al., 2005), according to the Akaike Information Criterion.

Codon usage was evaluated calculating the Relative Synonymous Codon Usage (RSCU) index for each synonymous codon in all genetic regions by using DAMBE v5.0.7. The RSCU value of a codon is its observed frequency divided by its expected frequency in the absence of usage bias (Sharp and Li, 1986). The correlation among RSCU values from overlapping and nonoverlapping regions was analysed using the Spearman correlation test ($p<0.05$) with the GraphPad software. A significant correlation would mean a similar codon usage among the regions compared, whereas a nonsignificant correlation would imply a lack of evidence of a similar codon usage. Stop codons and codons that code only for one amino acid (ATG: methionine and TGG: tryptophan) were not considered in the analysis.

On the other hand, to comparatively evaluate the HBV and human codon usages, human tissue-specific gene expression data were obtained from previous studies (Hsiao et al., 2001; Plotkin et al., 2004) and the correlation among RSCU values was analysed using the most expressed liver ($n=33$; ~37800 codons) and brain ($n=40$; ~39100 codons) human genes.

Differential adaptive evolution according to the HBeAg(+)/(−) status

To test whether differential adaptive evolution according to the HBeAg status occurs in the HBV genome, a method for detecting differential selective pressure on individual sites between populations was applied (Pond et al., 2006), by using the HYPHY v2.0 software. Thus, sequences classified as HBeAg(+) (Dataset A) were compared with sequences classified as HBeAg(−) ($n=119$) (Dataset B). Alignments and phylogenetic trees were built as previously.

The association of the differentially selected sites with epitopes B and T was evaluated with the *Immune Epitope Database* (*IEDB*) (Vita et al., 2010).

Acknowledgments

This work was supported by grants from Universidad de Buenos Aires (SECyT-UBA 20020100100405) and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET; PIP 112–200801-01169), Argentina.

We thank the anonymous reviewers for their constructive comments.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2013.02.027.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.

Abbott, W.G., Tsai, P., Leung, E., Trevarton, A., Ofanoa, M., Hornell, J., Gane, E.J., Munn, S.R., Rodrigo, A.G., 2010. Associations between HLA class I alleles and escape mutations in the hepatitis B virus core gene in New Zealand-resident Tongans. J. Virol. 84, 621–629.

Akarca, U.S., Lok, A.S., Naturally occurring hepatitis B virus core gene mutations. Hepatology 22, 1995, 50-60.

Baumert, T.F., Marrone, A., Vergalla, J., Liang, T.J., 1998. Naturally occurring mutations define a novel function of the hepatitis B virus core promoter in core protein expression. J. Virol. 72, 6785–6795.

Beck, J., Nassal, M., 1998. Formation of a functional hepatitis B virus replication initiation complex involves a major structural alteration in the RNA template. Mol. Cell. Biol. 18, 6265–6272.

Bertoletti, A., Costanzo, A., Chisari, F.V., Levrero, M., Artini, M., Sette, A., Penna, A., Giuberti, T., Fiaccadori, F., Ferrari, C., Cytotoxic T lymphocyte response to a wild type hepatitis B virus epitope in patients chronically infected by variant viruses carrying substitutions within the epitope. J. Exp. Med. 180, 1994, 933-943.

Brunetto, M.R., Stemler, M., Schodel, F., Will, H., Ottoberlli, A., Rizzetto, M., Bonino, F., 1989. Identification of HBV variants which cannot produce precore derived HBeAg and may be responsible for severe hepatitis. Ital. J. Gastroenterol. 21, 151–154.

Bruss, V., Lu, X., Thomssen, R., Gerlich, W.H., 1994. Post-translational alterations in transmembrane topology of the hepatitis B virus large envelope protein. EMBO J. 13, 2273–2279.

Carman, W.F., Jacyna, M.R., Hadziyannis, S., Karayiannis, P., McGarvey, M.J., Makris, A., Thomas, H.C., 1989. Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. Lancet 2, 588–591.

Chatzidaki, V., Kouroumalis, E., Galanakis, E., Hepatitis B virus acquisition and pathogenesis in childhood: host genetic determinants. J. Pediatr. Gastroenterol. Nutr. 52, 2011, 3-8.

Chen, Y., Marion, P.L., 1996. Amino acids essential for RNase H activity of hepadnaviruses are also required for efficient elongation of minus-strand viral DNA. J. Virol. 70, 6151–6156.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Günther, S., 2006. Genetic variation in HBV infection: genotypes and mutants. J. Clin. Virol. 36 (Suppl 1), S3–S11.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
Hershberg, R., Petrov, D.A., Selection on codon bias. Annu. Rev. Genet. 42, 2008, 287-299.
Hughes, A.L., Hughes, M.A., Patterns of nucleotide difference in overlapping and non-overlapping reading frames of papillomavirus genomes. Virus Res. 113, 2005, 81-88.
Hughes, A.L., Westover, K., da Silva, J., O'Connor, D.H., Watkins, D.I., Simultaneous positive and purifying selection on overlapping reading frames of the tat and vpr genes of simian immunodeficiency virus. J. Virol. 75, 2001, 7966-7972.
Hsiao, L.L., Dangond, F., Yoshida, T., Hong, R., Jensen, R.V., Misra, J., Dillon, W., Lee, K. F., Clark, K.E., Haverty, P., Weng, Z., Mutter, G.L., Frosch, M.P., MacDonald, M.E., Milford, E.L., Crum, C.P., Bueno, R., Pratt, R.E., Mahadevappa, M., Warrington, J. A., Stephanopoulos, G., Stephanopoulos, G., Gullans, S.R., 2001. A compendium of gene expression in normal human tissues. Physiol. Genomics 7, 97–104.
Khakoo, S.I., Ling, R., Scott, I., Dodi, A.I., Harrison, T.J., Dusheiko, G.M., Madrigal, J.A., Cytotoxic T lymphocyte responses and CTL epitope escape mutation in HBsAg, anti-HBe positive individuals. Gut. 47, 2000, 137-143.
Kidd, A.H., Kidd-Ljunggren, K., 1996. A revised secondary structure model for the 3′-end of hepatitis B virus pregenomic RNA. Nucleic Acids Res. 24, 3295–3301.
Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22, 1208–1222.
Kosakovsky Pond, S.L., Frost, S.D., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676–679.
Krakauer, D.C., 2000. Stability and evolution of overlapping genes. Evolution 54, 731–739.
Kramvis, A., Arakawa, K., Yu, M.C., Nogueira, R., Stram, D.O., Kew, M.C., 2008. Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. J. Med. Virol. 80, 27–46.
Liaw, Y.F., Brunetto, M.R., Hadziyannis, S., The natural history of chronic HBV infection and geographical differences. Antivir. Ther. 15 (Suppl 3), 2011, 25-33.
Lim, S.G., Cheng, Y., Guindon, S., Seet, B.L., Lee, L.Y., Hu, P., Wasser, S., Peter, F.J., Tan, T., Goode, M., Rodrigo, A.G., 2007. Viral quasi-species evolution during hepatitis Be antigen seroconversion. Gastroenterology 133, 951–958.
Liu, Y., Testa, J.S., Philip, R., Block, T.M., Mehta, A.S., A ubiquitin independent degradation pathway utilized by a hepatitis B virus envelope protein to limit antigen presentation. PLoS One 6, 2011, e24477.
Lok, A.S., McMahon, B.J., 2009. Chronic hepatitis B: update 2009. Hepatology 50, 661–662.
Locarnini, S., Zoulim, F., Molecular genetics of HBV infection. Antivir. Ther. 15 (Suppl 3), 2010, 3-14.
Mizokami, M., Orito, E., Ohba, K., Ikeo, K., Lau, J.Y., Gojobori, T., 1997. Constrained evolution with respect to gene overlap of hepatitis B virus. J. Mol. Evol. 44 (Suppl 1), S83–S90.
Maman, Y., Blancher, A., Benichou, J., Yablonka, A., Efroni, S., Louzoun, Y., Immune-induced evolutionary selection focused on a single reading frame in overlapping hepatitis B virus proteins. J. Virol. 85, 2011, 4558-4566.
Mizukoshi, E., Sidney, J., Livingston, B., Ghany, M., Hoofnagle, J.H., Sette, A., Rehermann, B., Cellular immune responses to the hepatitis B virus polymerase. J. Immunol. 173, 2004, 5863-5871.
Morozov, V., Pisareva, M., Groudinin, M., 2000. Homologous recombination between different genotypes of hepatitis B virus. Gene 260, 55–65.
Narechania, A., Terai, M., Burk, R.D., Overlapping reading frames in closely related human papillomaviruses result in modular rates of selection within E2. J. Gen. Virol. 86, 2005, 1307-1313.
Nayersina, R., Fowler, P., Guilhot, S., Missale, G., Cerny, A., Schlicht, H.J., Vitiello, A., Chesnut, R., Person, J.L., Redeker, A.G., Chisari, F.V., HLA A2 restricted cytotoxic T lymphocyte responses to multiple hepatitis B surface antigen epitopes during hepatitis B virus infection. J. Immunol. 150, 1993, 4659-4671.
Norton, P.A., Menne, S., Sinnathamby, G., Betesh, L., Cote, P.J., Philip, R., Mehta, A.S., Tennant, B.C., Block, T.M., Glucosidase inhibition enhances presentation of de-N-glycosylated hepatitis B virus epitopes by major histocompatibility complex class I in vitro and in woodchucks. Hepatology 52, 2010, 1242-1250.
Ostapchuk, P., Hearing, P., Ganem, D., 1994. A dramatic shift in the transmembrane topology of a viral envelope glycoprotein accompanies hepatitis B viral morphogenesis. EMBO J. 13, 1048–1057.
Pavesi, A., Origin and evolution of overlapping genes in the family Microviridae. J. Gen. Virol. 87, 2006, 1013-1017.
Pavesi, A., De Iaco, B., Granero, M.I., Porati, A., On the informational content of overlapping genes in prokaryotic and eukaryotic viruses. J. Mol. Evol. 44, 1997, 625-631.
Plotkin, J.B., Robins, H., Levine, A.J., 2004. Tissue-specific codon usage and the expression of human genes. Proc. Natl. Acad. Sci. USA 101, 12588–12591.
Pommie, C., Levadoux, S., Sabatier, R., Lefranc, G., Lefranc, M.P., 2004. IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. J. Mol. Recognition 17, 17–32.
Pond, S.L., Frost, S.D., Grossman, Z., Gravenor, M.B., Richman, D.D., Brown, A.J., 2006. Adaptation to different human populations by HIV-1 revealed by codon-based analyses. PLoS Comput. Biol. 2, e62.
Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
Prange, R., Streeck, R.E., 1995. Novel transmembrane topology of the hepatitis B virus envelope proteins. EMBO J. 14, 247–256.
Rehermann, B., Fowler, P., Sidney, J., Person, J., Redeker, A., Brown, M., Moss, B., Sette, A., Chisari, F.V., The cytotoxic T lymphocyte response to multiple hepatitis B virus polymerase epitopes during and after acute viral hepatitis. J. Exp. Med. 181, 1995, 1047-1058.
Sabath, N., Morris, J.S., Graur, D., 2011. Is there a twelfth protein-coding gene in the genome of influenza A? A selection-based approach to the detection of overlapping genes in closely related sequences. J. Mol. Evol. 73, 305–315.
Seeger, C., Mason, W.S., Hepatitis B virus biology. Microbiol. Mol. Biol. Rev. 64, 2000, 51-68.
Sharp, P.M., Li, W.H., 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. 24, 28–38.
Simmonds, P., Midgley, S., 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. J. Virol. 79, 15467–15476.
Singh, R., Kaul, R., Kaul, A., Khan, K., A comparative review of HLA associations with hepatitis B and C viral infections across global populations. World J. Gastroenterol. 13, 2007, 1770-1787.
Tsai, S.L., Chen, M.H., Yeh, C.T., Chu, C.M., Lin, A.N., Chiou, F.H., Chang, T.H., Liaw, Y.F., Purification and characterization of a naturally processed hepatitis B virus peptide recognized by CD8+ cytotoxic T lymphocytes. J. Clin. Invest. 97, 1996, 577-584.
van de Klundert, M.A., Cremer, J., Kootstra, N.A., Boot, H.J., Zaaijer, H.L., Comparison of the hepatitis B virus core, surface and polymerase gene substitution rates in chronically infected patients. J. Viral Hepat. 19, 2011, e34-40.
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.
Tang, H., Delgermaa, L., Huang, F., Oishi, N., Liu, L., He, F., Zhao, L., Murakami, S., 2005. The transcriptional transactivation function of HBx protein is important for its augmentation role in hepatitis B virus replication. J. Virol. 79, 5548–5556.
Tatematsu, K., Tanaka, Y., Kurbanov, F., Sugauchi, F., Mano, S., Maeshiro, T., Nakayoshi, T., Wakuta, M., Miyakawa, Y., Mizokami, M., 2009. A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. J. Virol. 83, 10538–10547.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.
Tran, T.T., Trinh, T.N., Abe, K., 2008. New complex recombinant genotype of hepatitis B virus identified in Vietnam. J. Virol. 82, 5657–5663.
Vita, R., Zarebski, L., Greenbaum, J.A., Emami, H., Hoof, I., Salimi, N., Damle, R., Sette, A., Peters, B., 2010. The immune epitope database 2.0. Nucleic Acids Res. 38, 854–862.
Warner, B.G., Tsai, P., Rodrigo, A.G., ofanoa, M., Gane, E.J., Munn, S.R., Abbott, W.G., 2011. Evidence for reduced selection pressure on the hepatitis B virus core gene in hepatitis B e antigen-negative chronic hepatitis B. J. Gen. Virol. 92, 1800–1808.
Weber, B., 2005. Genetic variability of the S gene of hepatitis B virus: clinical and diagnostic impact. J. Clin. Virol. 32, 102–112.
Wu, J.F., Wu, T.C., Chen, C.H., Ni, Y.H., Chen, H.L., Hsu, H.Y., Chang, M.H., Serum levels of interleukin-10 and interleukin-12 predict early, spontaneous hepatitis B virus e antigen seroconversion. Gastroenterology 138, 2010, 165-172 e161-163.
Xia, X., Xie, Z., 2001. DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. 92, 371–373.
Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39, 306–314.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.
Yang, Z., Kumar, S., 1996. Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites. Mol. Biol. Evol. 13, 650–659.
Yang, Z., Lauder, I.J., Lin, H.J., 1995. Molecular evolution of the hepatitis B virus genome. J. Mol. Evol. 41, 587–596.
Yim, H.J., Lok, A.S., 2006. Natural history of chronic hepatitis B virus infection: what we knew in 1981 and what we know in 2005. Hepatology 43, S173–S181.
Zaaijer, H.L., van Hemert, F.J., Koppelman, M.H., Lukashov, V.V., 2007. Independent evolution of overlapping polymerase and surface protein genes of hepatitis B virus. J. Gen. Virol. 88, 2137–2143.
Zhang, D., Chen, J., Deng, L., Mao, Q., Zheng, J., Wu, J., Zeng, C., Li, Y., 2010. Evolutionary selection associated with the multi-function of overlapping genes in the hepatitis B virus. Infect. Genet. Evol. 10, 84–88.
Zoulim, F., Locarnini, S., 2009. Hepatitis B virus resistance to nucleos(t)ide analogues. Gastroenterology 137, 1593–1608.
Zuckerman, J.N., Zuckerman, A.J., 2003. Mutations of the surface protein of hepatitis B virus. Antiviral Res. 60, 75–78.