Complex Genotype Mixtures Analyzed by Deep Sequencing in Two Different Regions of Hepatitis B Virus

Andrea Caballero[1☯], Josep Gregori[2,3,4☯], Maria Homs[1,2], David Tabernero[1,2], Carolina Gonzalez[1], Josep Quer[2,3], Maria Blasi[1,2], Rosario Casillas[3], Leonardo Nieto[1], Mar Riveiro-Barciela[2,5], Rafael Esteban[2,5], Maria Buti[2,5], Francisco Rodriguez-Frias[1,2]*

1 Liver Pathology Unit, Departments of Biochemistry and Microbiology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, 2 Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III, Madrid, Spain, 3 Liver Diseases Unit, Vall d'Hebron Research Institute, Barcelona, Spain, 4 Roche Diagnostics S.L., Sant Cugat del Vallès, Spain, 5 Liver Unit, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain

☯ These authors contributed equally to this work.
* frarodri@vhebron.net

Abstract

This study assesses the presence and outcome of genotype mixtures in the polymerase/surface and X/preCore regions of the HBV genome in patients with chronic hepatitis B virus (HBV) infection. Thirty samples from ten chronic hepatitis B patients were included. The polymerase/surface and X/preCore regions were analyzed by deep sequencing (UDPS) in the first available sample at diagnosis, a pre-treatment sample, and a sample while under treatment. HBV genotype was determined by phylogenesis. Quasispecies complexity was evaluated by mutation frequency and nucleotide diversity. The polymerase/surface and X/preCore regions were validated for genotyping from 113 GenBank reference sequences. UDPS yielded a median of 10,960 sequences per sample (IQR 16,645) in the polymerase/surface region and 11,595 sequences per sample (IQR 14,682) in X/preCore. Genotype mixtures were more common in X/preCore (90%) than in polymerase/surface (30%) (p<0.001). On X/preCore genotyping, all samples were genotype A, whereas polymerase/surface yielded genotypes A (80%), D (16.7%), and F (3.3%) (p = 0.036). Genotype changes in polymerase/surface were observed in four patients during natural quasispecies dynamics and in two patients during treatment. There were no genotype changes in X/preCore. Quasispecies complexity was higher in X/preCore than in polymerase/surface (p = 0.004).

The results provide evidence of genotype mixtures and differential genotype proportions in the polymerase/surface and X/preCore regions. The genotype dynamics in HBV infection and the different patterns of quasispecies complexity in the HBV genome suggest a new paradigm for HBV genotype classification.

Introduction

Hepatitis B virus (HBV) infection is a global health problem. Currently, around 240 million people are chronically infected by HBV, which confers a high risk of developing liver cirrhosis and hepatocellular carcinoma (HCC) [1].

HBV, a member of the Hepadnaviridae family, has a small (3.2 kb) partially double-stranded DNA genome with four highly overlapping open reading frames (ORFs): polymerase, surface, Core, and X [2]. To date, HBV has been classified into eight well-characterized (A-H) and two putative (I-J) genotypes (Gen/A-I) by >8% divergence in the entire genomic sequence. Genotypes are further divided into subgenotypes when genome divergence is 4% to 8%. The geographic and ethnic distribution of HBV genotypes and subgenotypes has been well characterized [3]. In the population of Spain, approximately 90% of HBV chronic carriers are infected by genotype A or D, and 6% by genotype F [4]. Several studies have reported associations of HBV genotype and subgenotype with HBeAg seroconversion status, clinical outcome, and prognosis [5–7]. In addition, HBV genotype has an impact on the therapeutic response to interferon and nucleos(t)ide analogues (NUCs) [8].

The most conclusive method for HBV genotyping is phylogenetic analysis of the nucleotide sequences of the whole HBV genome, although this is not performed in routine diagnostics. Instead, genotyping is done by sequencing a representative part of the genome. There is no consensus as to which region of the genome should be used for this purpose. Currently, the surface and polymerase regions are the ones most widely used, mainly because commercial line probe assays (LiPA) hybridize at the polymerase region [9].

Recombination events have been described between different HBV genotypes in HIV-positive patients [10], in integrated sequences from HCC patients [11], and in 30% of the 3300 HBV genome sequences available in public databases [12]. Intergenotype recombinations tend to have the same geographical distributions as their parent genotypes [13]. Inconsistencies in the local phylogenies of different fragments of the HBV genome may be a source of HBV genotype/subgenotype misclassification [14]. LiPA genotyping identified mixed genotype infections in 8% of HBV-infected individuals from Central and Eastern Europe [15]. In another study, LiPA documented mixed genotype infections in 22% of patients [16]. Nonetheless, detection of genotype mixtures or minor strains may be limited by the technology used [17]. Therefore, we determined the presence of HBV genotype mixtures by ultra-deep pyrosequencing (UDPS), which enables analysis of thousands of clonal sequences, 400 nucleotides (nt) in length. UDPS has been extensively used for detecting and quantifying low-frequency variants, for genotyping, and for investigating the viral quasispecies in HBV and other viruses [18–24]. In this study, we aimed to quantify genotype mixtures and evaluate changes occurring in the genotypic pattern of the HBV quasispecies in the absence and presence of nucleoside analogue treatment using UDPS. Two regions of the viral genome were selected for the study: a region including the polymerase/surface genes, and another essential region for regulation of HBV replication that includes the 5' end of the X gene and the preCore region.

Patients and Methods

Patients

This is a retrospective study approved by the Ethics Committee of Vall d'Hebron Research Institute, and including sequential samples from chronic hepatitis B patients, who gave written consent for participation. All patients were born in Spain, had failed lamivudine (LMV) therapy, had complete clinical documentation, and tested negative for hepatitis delta virus (HDV), hepatitis C virus (HCV), and human immunodeficiency virus (HIV). Ten chronically infected

HBV patients (Pt) were randomly selected for the study. Three samples per patient were analyzed: one sample obtained at the diagnosis (baseline), one taken before starting LMV, and the third taken at the time of partial virological response [25]. However, in two patients (Pt 5 and 7), a sample at viral breakthrough was included, because the partial virological response sample was not available. Clinical, biochemical, virological, and serological markers in the ten patients are summarized in Table 1. Two patients were HBeAg-positive at baseline (Pts 7 and 10). HBeAg loss occurred in one patient during antiviral treatment (Pt 10), and one patient (Pt 5) was HBeAg-negative at baseline and HBeAg-positive at the second and third sample.

Serological and virological determinations

Serological markers for HBV (HBsAg, HBeAg, and anti-HBe), HCV (anti-HCV), HDV (anti-HDV), and HIV (anti-HIV) were tested by commercial enzyme immunoassays. HBV-DNA was quantified by real-time PCR with a detection limit of 20 IU/mL (COBAS TaqMan HBV V2.0, Roche Diagnostics GmbH, Mannheim, Germany). HBV genotype was determined by LiPA (INNO-LiPA, Innogenetics, Ghent, Belgium). This method analyzes a fragment covering the polymerase and surface regions (nucleotide positions: nt 456–798), detects mixtures at proportions of 20%, and is considered to have a genotyping accuracy of >99% [9].

HBV genome regions analyzed by UDPS

HBV-DNA was extracted from serum by QIAampMiniSpin columns (QIAampDNA Mini Kit, QIAGEN, Hilden, Germany), according to the manufacturer's instructions. Two HBV genomic fragments were selected for UDPS (GS FLX, 454 Life Sciences, Roche) (Fig 1). One was a 355 bp (nt 615–969) fragment including the reverse transcriptase domain and overlapping the surface gene (P/S region); P/S includes 183 bp of the region amplified by LiPA. The other was a 356 bp (nt 1580–1936) fragment including the 3'-end of the X gene and the 5'-end of the pre-Core/Core gene (X/preCore region). The X/preCore region was chosen because it is clinically relevant, as it includes the ENHII region, the main motifs involved in HBV replication and regulation, and the sequence where synthesis of the HBV DNA minus strand is initiated (DR1). In addition, it is likely to be a region in which the viral quasispecies would show high complexity. Two PCRs with high-fidelity polymerase (Pfu Ultra-II, Stratagene, La Jolla, United States) were performed for P/S and X/preCore. PCR products were isolated from 1.5% agarose gel and their quality and quantity were checked. A detailed description of the reactions, concentrations, and primer sequences used is provided in supplementary materials and methods in S1 Protocol. Amplicons were pooled to a concentration of $2x10^6$ molecules. This working solution was enriched with capture beads for forward and reverse clonal amplification (emPCR kits II and III, 454 Life Sciences). UDPS was carried out on the FLX Genome Sequencer system (454Life Sciences).

UDPS data treatment

UDPS sequences were demultiplexed and primers were trimmed. After a quality filter step, sequences covering the full amplicon and common to both strands were established as consensus haplotypes, as described elsewhere [26,27]. Details of the procedure are given in the S1 Protocol. Computations were made in the R language and platform [28] using an in-house-developed pipeline, with the help of functions provided in the Biostrings [29] and APE [30] packages.

The UDPS sequencing data from the samples analyzed in this study have been submitted to the GenBank SRA database. The BioProject accession number is PRJNA291355 and BioSample accession numbers are included in S1 Table.

Table 1. Baseline biochemical and serological markers of the ten patients included in the study.

Pt	Sex	First sample at diagnosis (baseline)					Second sample (before starting treatment)					Third sample (under treatment)				
		Age, years	ALT, IU/mL	HBV DNA, IU/mL	HBeAg	Gen	Months after 1st sample	ALT, IU/mL	HBV DNA, IU/mL	HBeAg	Gen	Months after 2nd sample	ALT, IU/mL	HBV DNA, IU/mL	HBeAg	Gen
1	M	51	105	3.13E+06	N	A	7	364	2.32E+07	N	A	25	38	1.64E+04	N	A
2	M	56	391	2.9E+07	N	DF	25	656.9	1.16E+08	N	A	13	19.5	6.67E+04	N	A
3	M	30	46.4	1.68E+08	N	A	20	101.8	9.67E+07	N	A	14	34.7	1.18E+04	N	A
4	M	29	116	1,37E+03	N	A	28	70.7	8.64E+06	N	D	5	49.5	2.93E+04	N	D
5	M	56	75	2.65E+05	N	A	7	60	7.87E+08	P	A	19	55.9	1.03E+09	P	A
6	M	45	92	7.80E+08	N	A	36	129	2.78E+09	N	A	32	38	2.38E+03	N	A
7	F	20	59	1.66E+10	P	A	4	60	2.23E+09	P	A	14	75	2.36E+09	P	A
8	M	32	78	9.64E+04	N	AF	16	242	4.42E+05	N	A	22	57	3.17E+04	N	A
9	M	48	65	6.12E+05	N	A	43	109	7.43E+06	N	D	11	50	9.73E+03	N	A
10	M	35	262	9.66E+08	P	D	17	137	8.99E+06	P	D	18	93.8	1.31E+04	N	A

Pt; Patient, M; Male, F; Female, P; Positive, N; Negative, Gen; genotype determined by LiPA

doi:10.1371/journal.pone.0144816.t001

Genotyping and quasispecies complexity indexes

The consensus haplotypes of each sample were clustered in each region using identity thresholds of 96% (P/S) and 97% (X/preCore) (Genotyping in S1 Protocol). The centroid of each cluster was taken as the most abundant haplotype in the cluster. The Clustal Omega program was used to align the cluster centroids with a set of reference sequences from GenBank. Genotyping was done by distance-based discriminant analysis (DB rule) [31,32], which takes into account the intraclass variability of each genotype. Genetic distances were computed according to the Kimura-80 model [33]. Visualization of genetic distances between sequences was provided by UPGMA trees [34] and Multidimensional Scaling (MDS) plots [35].

HBV quasispecies (QA) complexity in the two regions was quantified using the mutation frequency (Mf) and nucleotide diversity (Pi) indexes [23,26]. Possible differences between the second and first sample were considered to result from the natural dynamics of the quasispecies; that is, changes occurring in haplotype fitness without external pressure (eg, antiviral
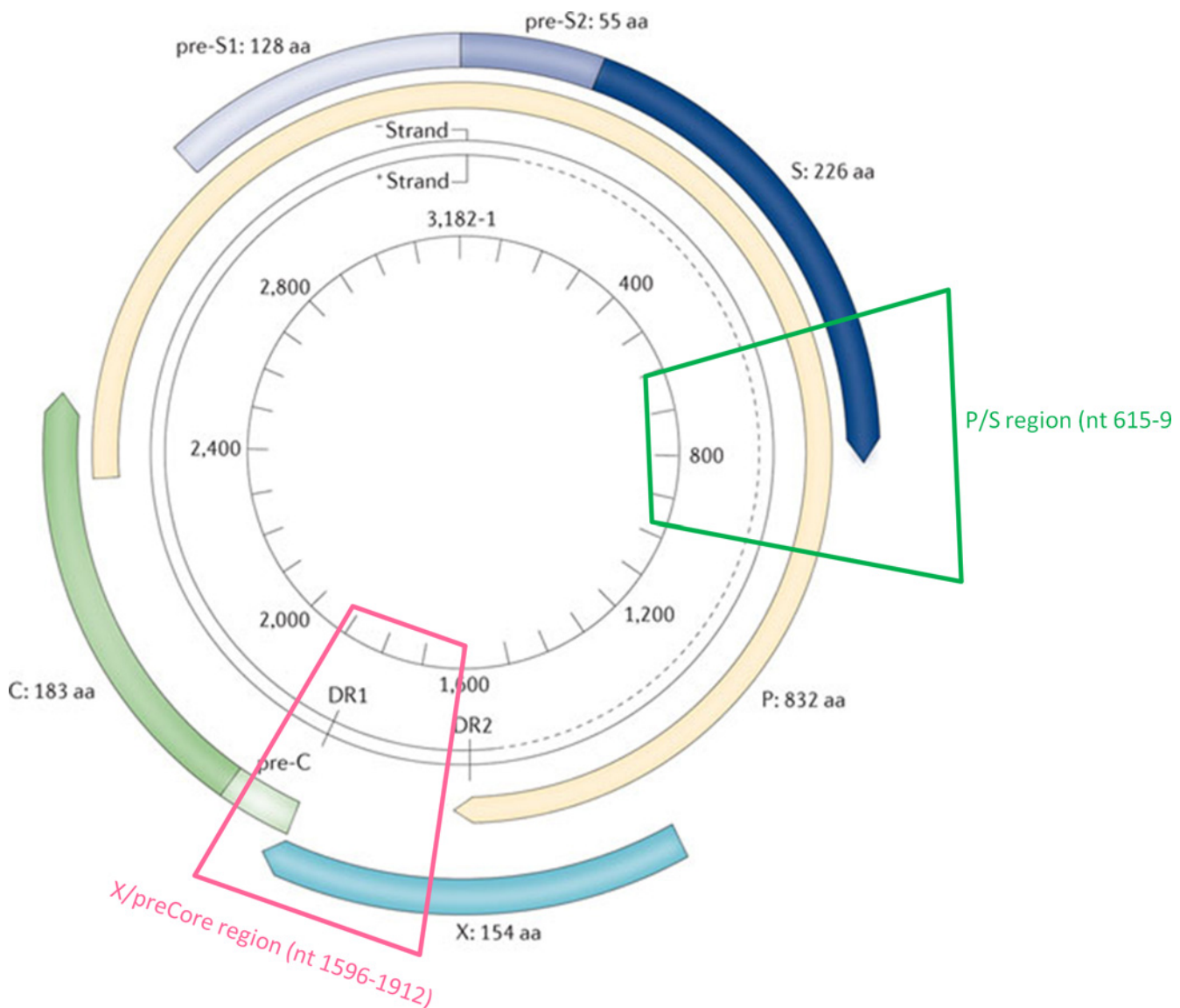

Fig 1. Regions of the HBV genome amplified for the present study: P/S is indicated in green and X/preCore in pink.

doi:10.1371/journal.pone.0144816.g001

treatment), but under intrinsic factors, such as the host immune response. Genotype mixtures were defined as the presence of haplotypes classified as belonging to different genotypes in a single quasispecies sample. As only two fragments of the HBV genome were studied, the information obtained did not suffice to determine whether this was the result of a mixed infection, the presence of recombinants, or both these circumstances. Whatever the case, all haplotypes detected were consistently classified, and with high confidence, using a collection of reference sequences for each HBV genotype, as described in Genotyping in S1 Protocol.

Statistical analysis

Statistical analyses were carried out with IBM SPSS 20 (SPSS Inc., Chicago, USA). Categorical data were tested by the Fisher exact test, and independent variables were compared using the Mann-Whitney U test. Significance was set at p≤0.05.

Results

Validation of P/S and X/preCore regions for genotyping

To check the capacity to differentiate between genotypes in the two regions analyzed, a set of whole HBV genome sequences from GenBank was evaluated by the leave-one-out method using the DB rule discriminant analysis. Each whole genome sequence in the set was classified using the remaining sequences. The genotyping capacity of these reference sequences was explored using UPGMA trees and plotting on the first three components of a multidimensional scaling analysis of the matrix of genetic distances between pairs of reference sequences. All sequences could be perfectly classified except for the genotype J sequence in GenBank (S1 File).

The set of reference sequences was multiple-aligned and trimmed to positions 615 to 969, covering the P/S amplicon (S2 File), and positions 1596 to 1912, covering the X/preCore region (S3 File). Sequences that were equal after trimming were removed. The consensus P/S and X/preCore haplotypes were clustered at identities of 96% and 97%, respectively. The P/S region showed high genotyping capacity, despite the limited amplicon length. Only two Gen/C reference sequences were misclassified as Gen/A using the P/S region (S2 File). In contrast, the X/preCore regions showed clear intermixing of Gen/D and E and Gen/C and I (S3 File).

Genotyping

Routine HBV genotyping by LiPA defined seven cases of Gen/A and one of Gen/D. Two genotype mixtures were detected by LiPA in the first sample (Pt 2, Gen/D and F; Pt 8, Gen/A and F) (Table 1).

UDPS Genotyping in the P/S region. A total of 379,438 sequences were obtained from the P/S region (median 10,960 reads, IQR 16,647). P/S showed high discriminating capacity for the eight main HBV genotypes (Gen/A-H) (S2 File). Among the 379,438 sequences analyzed, three different genotypes were observed (A, D, and F).

Twenty-one of the 30 samples (70%) showed single-genotype infection: 17 samples (81%) were Gen/A and four (19%) Gen/D (Table 2). Genotype mixtures were observed in nine samples (30%): the A/D mixture in five (56%) and the A/D/F mixture in four (44%).

Five of the ten (50%) patients (Pts 1, 2, 3, 8, and 9) showed HBV genotype mixtures in the first sample. Genotype mixtures were also detected in two patients (Pts 2 and 8) before starting treatment (second sample) and in two patients (Pt 2 and 4) under treatment (third sample). Four patients had genotype mixtures in percentages theoretically detectable by LiPA, but the method only detected mixtures in two patients (Pts 2 and 8). Sequence coverage was low in the treatment samples from Pts 9 and 10 (53 and 15, respectively); hence, a possible presence of

Table 2. Percentage of HBV genotypes obtained from the polymerase/surface (P/S) and X/preCore regions of the ten patients. Three samples per patient were analyzed: the first available at diagnosis (1st), the second before starting treatment (2nd), and the third under treatment (3rd).

Patient	Samples analyzed	P/S region			X/preCore region		
		A (%)	D (%)	F (%)	A (%)	DE* (%)	F (%)
	1st	93.0	3.6	3.4	92.2	5.3	2.5
1	2nd	100.0	0	0	98.9	1.1	0
	3rd	100.0	0	0	84.5	14.4	1.1
	1st	6.0	16.2	77.8	86.0	8.5	5.5
2	2nd	99.0	1	0	86.9	8.4	4.6
	3rd	96.4	3.6	0	93.5	5.4	1.1
	1st	98.5	1.5	0	66.9	30.5	2.6
3	2nd	100.0	0	0	93.5	4.3	2.2
	3rd	100.0	0	0	83.6	15.0	1.3
	1st	100.0	0	0	87.8	10.7	1.5
4	2nd	0	100.0	0	86.9	9.9	3.2
	3rd	9.4	89.9	0.7	97.6	2.1	0.3
	1st	100.0	0	0	100.0	0	0
5	2nd	100.0	0	0	87.5	12.2	0.3
	3rd	100.0	0	0	90.0	7.4	2.6
	1st	100.0	0	0	98.9	1.1	0
6	2nd	100.0	0	0	97.1	2.1	0.9
	3rd	100.0	0	0	82.9	9.9	7.2
	1st	100.0	0	0	93.2	2.6	4.3
7	2nd	100.0	0	0	95.9	2.2	1.8
	3rd	100.0	0	0	71.7	9.9	18.4
	1st	77.0	3.0	19.9	89.4	9.7	0.9
8	2nd	93.0	7.0	0	86.5	10.3	3.2
	3rd	100.0	0	0	90.4	8.3	1.3
	1st	87.2	12.8	0	86.8	7.9	5.4
9	2nd	0	100	0	88.7	10.2	1.1
	3rd	100.0	0	0	76.5	19.3	4.2
	1st	0	100	0	83.3	15.2	1.6
10	2nd	0	100	0	94.5	1.1	4.4
	3rd	100.0	0	0	89.2	4.0	6.8

* Results are presented as DE because the X/preCore region showed an intermixing between Gen/D and E (S3 File).

doi:10.1371/journal.pone.0144816.t002

mixtures was not excluded. Of note, these samples also yielded low coverage results in the P/S region. UPGMA trees and MDS plots of the first three most common haplotypes in the P/S region per sample are represented in S4 File with the reference sequences.

UDPS Genotyping in the X/preCore region. In total, 944,184 sequences were obtained in the X/preCore region (median 11,595 reads, IQR 14,682). X/preCore showed intermixing between Gen/D and E and Gen/C and I (S3 File). Gen/E is restricted to Africa [36], and none of the ten patients included in the study were from Africa; however, to be precise, the Gen/D results are presented as Gen/DE. None of the sequences were Gen/C, so GenC/I intermixing in the X/preCore region did not affect our study. Among the 944,184 sequences analyzed, three different genotypes were observed (A, DE, and F).

Gen/A infection alone was detected in only one baseline sample (Pt 5). The remaining 29 samples (96.7%) (baseline, second, and third samples from all patients) showed HBV genotype

Table 3. Comparison of the number of samples with mixed genotypes, the main genotypes and the genotype prevalence between the polymerase/surface (P/S) and X/preCore region.

		P/S region		X/preCore region		p-value
		Number samples (%)		Number samples (%)		
Presence of mixed genotypes		9	(30.0%)	29	(96.7%)	<0.001
Main genotype	A	24	(80.0%)	30	(100.0%)	
	D	5	(16.7%)	0		0.036
	F	1	(3.3%)	0		
Genotype prevalence	A	17	(56.7%)	1	(3.3%)	
	D	4	(13.3%)	0		<0.001
	AD	5	(16.7%)	2	(6.7%)	
	ADF	4	(13.3%)	27	(90.0%)	

doi:10.1371/journal.pone.0144816.t003

mixtures (Table 2). Triple mixtures (A/DE/F) were observed in 27 samples (90%) and double mixture (A/DE) in two samples (10%). UPGMA trees and MDS plots of the first three most common haplotypes in the X/preCore region per sample with the reference sequences are shown in S5 File.

P/S region vs. X/preCore region for genotyping. Comparison of the HBV genotyping results showed several discrepancies between the P/S and X/preCore regions (Table 3). First, a significantly larger number of samples with genotype mixtures were detected in X/preCore (96.7%) than in P/S (30%) (p<0.001). Second, distribution of the genotypes differed: X/pre-Core genotyped all samples as A, whereas P/S genotyped 24 samples as A (80%), five as D (16.7%) and one as F (3.3%) (p = 0.036). Third, genotypes D and F were more commonly detected in the X/preCore (27 samples, 90%) than the P/S region (four samples, 13.3%, p<0.001). Finally, in six samples from four patients (Pts 2, 4, 9, and 10) the main genotype in P/S differed from the one obtained in X/preCore (Table 2). In the baseline sample from Pt 2, genotyping in the P/S region yielded Gen/F, whereas in the X/preCore region it was Gen/A. In the second and third samples from Pt 4, the P/S region mainly yielded Gen/D whereas in the X/preCore region, Gen/A predominated (Fig 2). The second sample from Pt 9 showed Gen/D in P/S and Gen/A in X/preCore. In Pt 10, X/preCore yielded a predominant Gen/A, whereas P/S showed Gen/D in the first and second samples, and Gen/A alone had been selected at the third (under treatment) sample.

HBV genotype changes over the study period

Genotype dynamics. Genotype dynamics were defined as changes in the main genotype (selection of genotypes at percentages of 50%-100%) in each patient (Table 4). During the natural quasispecies dynamics of the infection (percentage difference between the second and first sample), the P/S region showed HBV genotype changes in three patients (Pts 2, 4, and 9) (Table 4). Pt 2 changed from a D/F mixture to Gen/A. Pts 4 and 9 changed from Gen/A to Gen/D. In contrast, no changes in the main genotype were observed with genotyping of the X/preCore region.

During the treatment period (percentage difference between the third and second sample), genotype changed from D to A in two patients (Pts 9 and 10). This change was detected by LiPA and UDPS genotyping of the P/S region, but not the X/preCore region.

Changes in genotype proportions. Changes in genotype percentage that did not imply selection of a new genotype (changes from 0%-49%) were detected in both P/S and X/preCore

([Table 4](#)). Fluctuations in genotype percentage in the P/S region were observed in three (30%) patients during natural quasispecies dynamics (Pts 1, 3, and 8) and in three (30%) during treatment (Pts 2, 4, and 8). In contrast, in the X/preCore region, changes in genotype percentages were observed in all patients during natural quasispecies dynamics and under treatment ([Table 4](#)).

In the analysis of the P/S region during natural quasispecies dynamics, the Gen/A population was found to be enriched in two patients (Pt 1, 7%, and Pt 3, 1.5%). In Pt 2, Gen/A sequences also increased, with 93% ([Table 4](#)), resulting from a decrease in genotypes F and D from the first sample ([Table 2](#)). In Pt 8, genotypes A and D were enriched, with 16% and 4%, respectively. The natural quasispecies dynamics of Pt 9 showed an 87.2% increase in Gen/D. During treatment, there was a 2.6% increase in Gen/D in Pt 2, a 9.4% increase in Gen/A and 0.7% increase in Gen/F in Pt 4, and a 7% increase in Gen/A in Pt 8. Two patients (Pt 9 and 10) changed from Gen/D in the second sample to Gen/A in the third ([Table 4](#)).
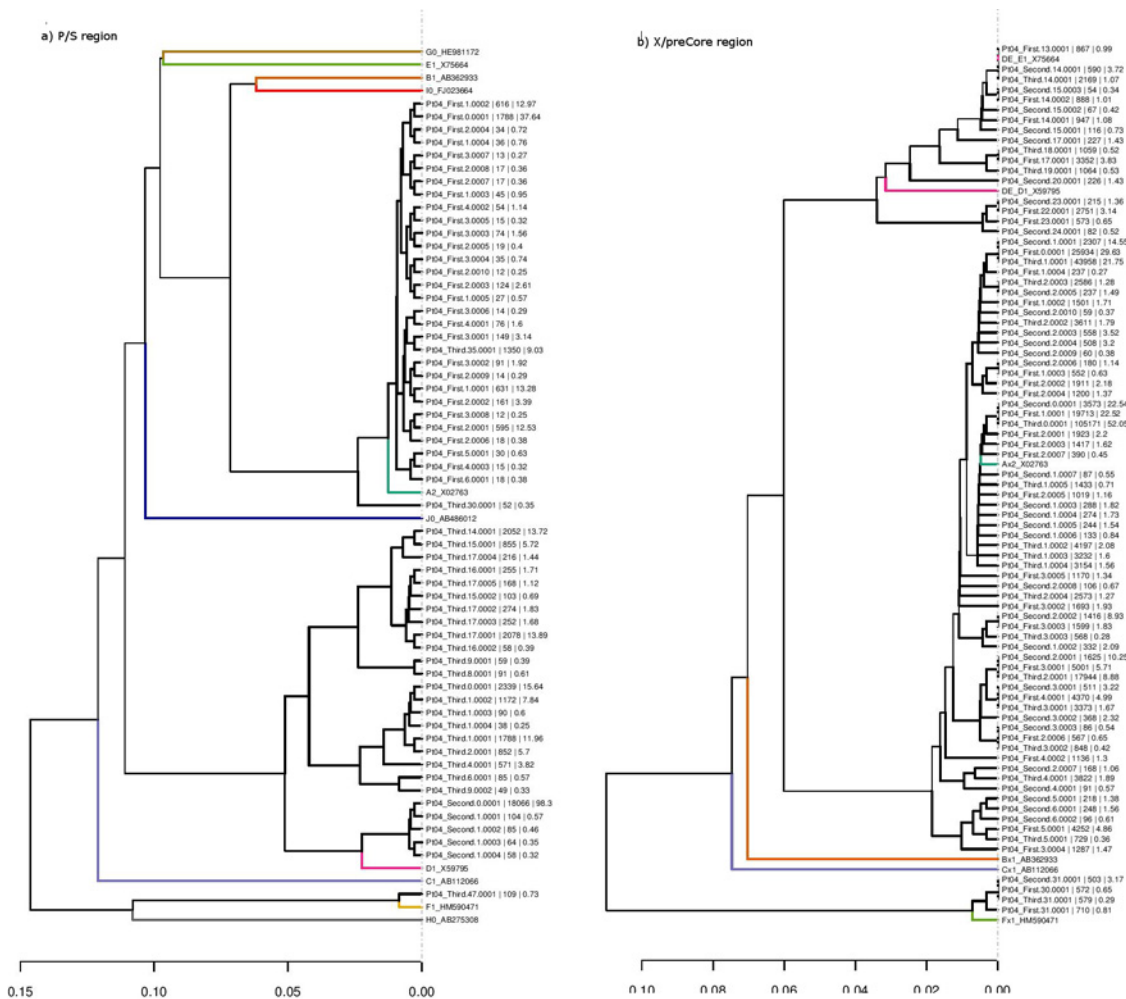

Fig 2. UPGMA tree obtained from the polymerase/surface (a) and the X/preCore (b) regions of the three samples from Pt 4. The main haplotype of the first sample (First.0.0.0001) was genotyped as A, whereas the main haplotype of the second and third sample (Second.0.0.0001 and Third.0.0.0001) was genotype as D. At the X/preCore region, the main genotype was A, but Gen/DE and F haplotypes were also observed. Genotype reference sequences for the two regions: A2_X02763; B1_AB362933; C1_AB112066; D1_X59795 (for P/S region); DE_ D1_X59795 (for X/preCore region); E1_X75664 (for P/S region); DE_E1_X75664 (for X/preCore region); F1_HM590471; G0_HE981172; H0_AB275308; I0_FJ023664; J0_AB486012.

doi:10.1371/journal.pone.0144816.g002

The natural quasispecies dynamics of the X/preCore region showed genotype percentage changes in all samples, but no changes in the main genotype. The percentage of Gen/A increased in four patients (Pt 1 6.7%, Pt 2 0.9%, Pt 3 26.69%, and Pt 7 2.7%). Pt 4 showed an increase in Gen/F (1.7%) and Pts 5 and 6 showed enrichment of both Gen/D and F (Pt 5, 12% D and 0.3% F and Pt 6, 1% D and 0.9% F). Gen/DE increased in Pt 8 (0.6%), Gen/A (1.9%) and D (2.3%) in Pt 9, and Gen/A (11.2%) and F (2.8%) in Pt 10. During treatment, Gen/D and F increased in six patients (Pts 1, 5, 6, 7, 9, and 10), and Gen/D (10.7%) in one (Pt 3) (Table 3). In three patients (Pts 2, 4, and 8) there was Gen/A enrichment (6.6%, 10.7%, and 3.9%, respectively) (Table 3).

Quasispecies complexity

Viral quasispecies complexity in the P/S and X/preCore regions was evaluated in all 30 samples using the mutation frequency (Mf) and nucleotide diversity (Pi) measures (S2 Table). The two indexes of quasispecies complexity were significantly higher in the X/preCore region than the P/S region (Mf, $p = 0.004$ and Pi, $p = 0.003$) (Fig 3), in accordance with the higher overlapping of P/S than X/preCore.

Discussion

Hepatitis B virus genotype has been implicated in HBeAg seroconversion, in the clinical outcome of the infection, and in the response to treatment [8]. The accuracy of genotype

Table 4. Percentage of enriched genotypes (%, type of genotype) in the polymerase/surface (P/S) and X/preCore regions during natural quasispecies dynamics (QD) and under treatment.

Patient	Period*	P/S region	X/preCore region[#]
1	Natural QD	7%, A	6.7%, A
	Treatment		13.3%, D; 1.1%, F
2	Natural QD	93%, A	0.9%, A
	Treatment	2.6%, D	6.6%, A
3	Natural QD	1.5%, A	26.6%, A
	Treatment		10.7%, D
4	Natural QD	100%, D	1.7%, F
	Treatment	9.4%, A; 0.7%, F	10.7%, A
5	Natural QD	-	12%, D; 0.3%, F
	Treatment	-	2.5%, D; 2.3%, F
6	Natural QD	-	1%, D; 0.9%, F
	Treatment	-	7.8%, D; 6.3%, F
7	Natural QD	-	2.7%, A
	Treatment	-	7.7%, D; 16.6%, F
8	Natural QD	16%, A; 4%, D	0.6%, D
	Treatment	7%, A	3.9%, A
9	Natural QD	87.2%, D	1.9%, A; 2.3%, D
	Treatment	100%, A	9.1%, D; 3.1%, F
10	Natural QD		11.2%, A; 2.8%, F
	Treatment	100%, A	2.9%, D; 2.4%, F

*Natural quasispecies dynamics is defined by the percentage difference between the second sample (before starting treatment) and first sample (at diagnosis) analyzed. The treatment period is established by the percentage difference between the third sample (under treatment) and second sample (before starting treatment) analyzed.

[#]X/preCore region showed an intermixing between Gen/D and E (S3 File)
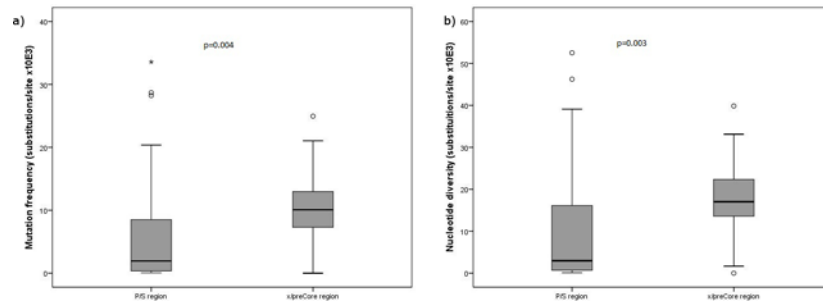
doi:10.1371/journal.pone.0144816.t004

Fig 3. Distribution of quasispecies complexity in all the samples analyzed according to mutation frequency (a) and nucleotide diversity (b) in the polymerase/surface (P/S) and X/preCore regions.

doi:10.1371/journal.pone.0144816.g003

classification is highly dependent on selection of appropriate reference sequences. Our references (S1, S2 and S3 Files) were validated as described in Results, and their information content was evaluated by sliding windows (SW) of 400 nucleotides in steps of 40 (S6 and S7 Files). Although whole genome analysis by SW shows that the PreS region has slightly higher discriminating power for genotyping HBV, we selected two amplicons in the P/S and X/preCore regions for this study. The X/preCore region was chosen because it includes the ENHII region and the main motifs involved in HBV replication. The P/S amplicon was selected to include the fragment generally used for HBV genotyping by LiPA, which was used in a previous step to genotype all samples in the study. The P/S fragment showed adequate genotype discrimination power, correctly classifying the eight main HBV genotypes (A-H).

Genotypes A, D, and F are the most prevalent in our population, and mixed infections with these three genotypes were found in both regions of the HBV genome studied. A significant prevalence of mixed HBV genotype infections has been reported in Asia, and it is suggested that these mixtures may correlate with clinical outcome and viral load [37]. Mixed genotypes (A/D, A/D/F, and D/F) have been described in Central and Eastern Europe [15], and mixed GenB/C has been reported in Taiwan [12]. In a previous study, we found a considerable rate (22%) [11] of mixed HBV genotypes (A/D and A/D/F) by LiPa and clonal methodologies in patients treated with nucleoside(t)e analogue treatment [7,16]. In the present study, genotype mixtures were observed in two (Pt 2 and 8) and nine samples (Pt 1, 2, 3, 4, 8, and 9) by LiPA and UDPS, respectively. Two major reasons may be the cause of these discrepancies. First, the fragment amplified for NGS is different from the fragment analyzed by LiPA; and second, the two techniques have methodological differences: NGS is a sequencing method and the definition of the genotype is obtained by phylogenesis, whereas LiPA is a hybridization method and therefore, an indirect measurement. Inconsistencies between LiPA and UDPS have been described in HCV genotyping [38] and HBV variant resistance detection [39].

The results of the present study were obtained by massive sequencing, a method widely used to study HBV and other viral infections [18–24,26,27]. Based on the experience acquired in previous efforts that confirmed the reproducibility of UDPS and its good correlation with molecular cloning for evaluating HBV and HCV quasispecies [26,27], we used UDPS to quantitatively analyze the presence and distribution of viral genotype mixtures. UDPS data treatment and filtering enabled detection of mixed infections, which were more frequent in X/preCore than in P/S ($p < 0.001$). The high presence of mixed infection in X/preCore is a surprising result that warrants further study. Although this finding remains unexplained, it may be a result of high variability in this region, which encompasses positions (DR1, gap of the negative strand in the conformation of dsHBV-DNA) that may be points for recombination, which could be interpreted as mixed genotype infection. Furthermore, the presence of minor variants

and genotype mixtures may have been underestimated previously because of the limited sensitivity of the methods used.

Selection of one genotype over another may occur due to immunological conditions or the effect of antiviral therapy. Selection or maintenance of Gen/A in the P/S region under treatment in eight of our patients (Pts 1, 3, 5, 6, 7, 8, 9, and 10) suggests that this region is under evolutive pressure during treatment and agrees with our previous observations [16]. HBV genotype fluctuations may be a common phenomenon in P/S, whereas genotype changes were not seen in the X/preCore region. In addition to the paradigm of mixed genotype infections, comparison of the genotypes found in the two regions studied showed discrepancies, suggesting intergenotypic recombination. This phenomena has been suggested previously, based on inconsistent genotyping between different regions of the HBV genome [38].

It cannot be excluded that the discrepancies between regions may have been due to UDPS variability or low coverage. However, in the remaining samples, there were notable differences in the percentages of HBV genotypes observed in the two regions, suggesting a high, complex recombination pattern, likely involving double and triple recombinants [38]. Intergenotypes tend to include the most frequent genotypes coexisting in a human population [13], and the putative intergenotypes observed in our study comprise the most common genotypes in the Spanish population [4]. This fact, added to co-infection with various genotypes and the long life of infected hepatocytes (2.4 to >70 days under LMV), may facilitate multiple infections in a single liver cell by different genotypes [40]. Nevertheless, recombination should be confirmed by UDPS using long sequencing fragments that cover both regions in the same sequence.

The study of quasispecies complexity is limited by the individual patterns that seem to evolve in each patient due to the host immune system. This observation is consistent with the dynamics of consecutive expansion and contraction of quasispecies complexity in response to changes in the host environment. An emerging mutant with greatly improved fitness for the situation would lead to homogenization of the quasispecies, in detriment to the less fit. On the other hand, long periods without environmental changes would produce synonymous mutants of the dominant haplotype, thus contributing to complexity expansion and providing possible escapes for future changes. In this respect, the observed diversity values would correspond to random samplings on sawtooth signals. Despite this limitations, the higher complexity of X/preCore than P/S is in accordance with the higher overlapping of P/S than X/preCore.

In conclusion, this UDPS study provides evidence of HBV genotype mixtures that change over time and illustrates the complex dynamics of the HBV quasispecies as an additional mechanism when adapting to new situations, such as host immune response and/or antiviral treatment. Discrepancies between genotypes in the P/S and X/preCore regions suggest phenomena of intergenotype recombination and indicate the need for an international expert consensus effort to set an HBV genotyping approach that will lead to a more comprehensive understanding of the clinical significance of HBV genotype classification.

Supporting Information

S1 File. Genotyping capacity of the reference sequences using UPGMA trees and plotting (complete HBV genome).
(PDF)

S2 File. Genotyping capacity of the reference sequences trimmed to positions 615 to 969 (covering the P/S amplicon).
(PDF)

S3 File. Genotyping capacity of the reference sequences trimmed to positions 1596 to1 912 (covering the P/S amplicon).
(PDF)

S4 File. UPGMA trees and MDS plots of the first three most common haplotypes in the P/S region per sample with the reference sequences.
(PDF)

S5 File. UPGMA trees and MDS plots of the first three most common haplotypes in the X/ preCore region per sample with the reference sequences.
(PDF)

S6 File. UPGMA tree of the sliding windows analysis of 400 nucleotides in steps of 40 for the eight HBV genotypes.
(PDF)

S7 File. Sliding windows analysis of 400 nucleotides in steps of 40 for the eight different genotypes.
(PDF)

S1 Protocol. Supplementary Materials and methods, Genotyping and Examination of HBV genotypes reference sequence. Detailed description of amplification of the P/S and X/preCore regions, UDPS data treatment, and genotyping, description of genotyping and examination of HBV genomtypes reference sequence.
(DOCX)

S1 Table. Biosample accession numbers for each sample and region analyzed.
(DOCX)

S2 Table. Complexity parameters obtained for the P/S and X/preCore regions. These parameters were obtained for the three samples per patient analyzed: the first available at diagnosis (1st), the second before starting treatment (2nd), and the third while under treatment (3rd).
(DOCX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: AC MH FRF. Performed the experiments: AC CG M. Blasi RC. Analyzed the data: AC JG MH M. Buti FRF. Contributed reagents/materials/analysis tools: JQ DT LN. Wrote the paper: AC JG MH FRF. Review of the manuscript: MRB RE M. Buti FRF.

References

1. WHO. WHO | Hepatitis B [Internet]. World Health Organization; 2015 [Accessed 4 Jun 2015]. Available: http://www.who.int/mediacentre/factsheets/fs204/en/
2. Okamoto H, Tsuda F, Sakugawa H, Sastrosoewignjo RI, Imai M, Miyakawa Y, et al. Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. J Gen Virol. 1988; 69 (Pt 10: 2575–83. PMID: 3171552

3. Kramvis A. Genotypes and genetic variability of hepatitis B virus. Intervirology. 2014; 57: 141–50. doi: 10.1159/000360947 PMID: 25034481

4. Rodriguez-Frias F, Jardi R, Buti M, Schaper M, Hermosilla E, Valdes A, et al. Hepatitis B virus genotypes and G1896A precore mutation in 486 Spanish patients with acute and chronic HBV infection. J Viral Hepat. 2006; 13: 343–350. doi: 10.1111/j.1365-2893.2005.00691.x PMID: 16637866

5. Hsieh T-H, Tseng T-C, Liu C-J, Lai M-Y, Chen P-J, Hsieh H-L, et al. Hepatitis B virus genotype B has an earlier emergence of lamivudine resistance than genotype C. Antivir Ther. 2009; 14: 1157–63. doi: 10.3851/IMP1454 PMID: 20032545

6. Lin C-L, Kao J-H. The clinical implications of hepatitis B virus genotype: Recent advances. J Gastroenterol Hepatol. 2011; 26 Suppl 1: 123–30. doi: 10.1111/j.1440-1746.2010.06541.x PMID: 21199523

7. Buti M, Elefsiniotis I, Jardi R, Vargas V, Rodriguez-Frias F, Schapper M, et al. Viral genotype and baseline load predict the response to adefovir treatment in lamivudine-resistant chronic hepatitis B patients. J Hepatol. 2007; 47: 366–72. doi: 10.1016/j.jhep.2007.04.011 PMID: 17561304

8. Kramvis A, Kew MC. Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy. J Viral Hepat. 2005; 12: 456–64. doi: 10.1111/j.1365-2893.2005.00624.x PMID: 16108759

9. Osiowy C, Giles E. Evaluation of the INNO-LiPA HBV genotyping assay for determination of hepatitis B virus genotype. J Clin Microbiol. 2003; 41: 5473–7. PMID: 14662927

10. Araujo NM, Araujo OC, Silva EM, Villela-Nogueira CA, Nabuco LC, Parana R, et al. Identification of novel recombinants of hepatitis B virus genotypes F and G in human immunodeficiency virus-positive patients from Argentina and Brazil. J Gen Virol. 2012; 94: 150–158. doi: 10.1099/vir.0.047324–0 PMID: 23079380

11. Georgi-Geisberger P, Berns H, Loncarevic IF, Yu ZY, Tang ZY, Zentgraf H, et al. Mutations on free and integrated hepatitis B virus DNA in a hepatocellular carcinoma: footprints of homologous recombination. Oncology. 1992; 49: 386–95. PMID: 1326735

12. Shi W, Carr MJ, Dunford L, Zhu C, Hall WW, Higgins DG. Identification of novel inter-genotypic recombinants of human hepatitis B viruses by large-scale phylogenetic analysis. Virology. 2012; 427: 51–59. doi: 10.1016/j.virol.2012.01.030 PMID: 22374235

13. Yang J, Xing K, Deng R, Wang J, Wang X. Identification of Hepatitis B virus putative intergenotype recombinants by using fragment typing. J Gen Virol. 2006; 87: 2203–15. doi: 10.1099/vir.0.81752–0 PMID: 16847116

14. Shi W, Zhu C, Zheng W, Carr MJ, Higgins DG, Zhang Z. Subgenotype reclassification of genotype B hepatitis B virus. BMC Gastroenterol. 2012; 12: 116. doi: 10.1186/1471-230X-12-116 PMID: 22925657

15. Deterding K, Constantinescu I, Nedelcu FD, Gervain J, Nemecek V, Srtunecky O, et al. Prevalence of HBV genotypes in Central and Eastern Europe. J Med Virol. 2008; 80: 1707–11. doi: 10.1002/jmv.21294 PMID: 18712830

16. Jardi R, Rodriguez-Frias F, Schaper M, Giggi E, Tabernero D, Homs M, et al. Analysis of hepatitis B genotype changes in chronic hepatitis B infection: Influence of antiviral therapy. J Hepatol. 2008; 49: 695–701. doi: 10.1016/j.jhep.2008.03.030 PMID: 18490074

17. Kao JH, Chen PJ, Lai MY, Chen DS. Acute exacerbations of chronic hepatitis B are rarely associated with superinfection of hepatitis B virus. Hepatology. 2001; 34: 817–23. doi: 10.1053/jhep.2001.28188 PMID: 11584381

18. Ghedin E, Fitch A, Boyne A, Griesemer S, DePasse J, Bera J, et al. Mixed infection and the genesis of influenza virus diversity. J Virol. 2009; 83: 8832–41. doi: 10.1128/JVI.00773-09 PMID: 19553313

19. Bull RA, Luciani F, McElroy K, Gaudieri S, Pham ST, Chopra A, et al. Sequential Bottlenecks Drive Viral Evolution in Early Acute Hepatitis C Virus Infection. Ou JJ, editor. PLoS Pathog. 2011; 7: e1002243. doi: 10.1371/journal.ppat.1002243 PMID: 21912520

20. Bunnik EM, Swenson LC, Edo-Matas D, Huang W, Dong W, Frantzell A, et al. Detection of inferred CCR5- and CXCR4-using HIV-1 variants and evolutionary intermediates using ultra-deep pyrosequencing. PLoS Pathog. 2011; 7: e1002106. doi: 10.1371/journal.ppat.1002106 PMID: 21731496

21. Hoffmann C, Minkah N, Leipzig J, Wang G, Arens MQ, Tebas P, et al. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. Nucleic Acids Res. 2007; 35: e91. doi: 10.1093/nar/gkm435 PMID: 17576693

22. Homs M, Buti M, Quer J, Jardí R, Schaper M, Tabernero D, et al. Ultra-deep pyrosequencing analysis of the hepatitis B virus preCore region and main catalytic motif of the viral polymerase in the same viral genome. Nucleic Acids Res. 2011; 39: 8457–71. doi: 10.1093/nar/gkr451 PMID: 21742757

23. Homs M, Caballero A, Gregori J, Tabernero D, Quer J, Nieto L, et al. Clinical application of estimating hepatitis B virus quasispecies complexity by massive sequencing: correlation between natural

evolution and on-treatment evolution. PLoS One. 2014; 9: e112306. doi: 10.1371/journal.pone.0112306 PMID: 25393280

24. Solmone M, Vincenti D, Prosperi MCF, Bruselles A, Ippolito G, Capobianchi MR. Use of massively parallel ultradeep pyrosequencing to characterize the genetic diversity of hepatitis B virus in drug-resistant and drug-naive patients and to detect minor variants in reverse transcriptase and hepatitis B S antigen. J Virol. 2009; 83: 1718–26. doi: 10.1128/JVI.02011-08 PMID: 19073746

25. European Association For The Study Of The Liver. EASL Clinical Practice Guidelines: management of chronic hepatitis B. J Hepatol. European Association for the Study of the Liver; 2009; 50: 227–42. doi: 10.1016/j.jhep.2008.10.001

26. Ramírez C, Gregori J, Buti M, Tabernero D, Camós S, Casillas R, et al. A comparative study of ultra-deep pyrosequencing and cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. Antiviral Res. 2013; 98: 273–83. doi: 10.1016/j.antiviral.2013.03.007 PMID: 23523552

27. Gregori J, Esteban JI, Cubero M, Garcia-Cehic D, Perales C, Casillas R, et al. Ultra-Deep Pyrosequencing (UDPS) Data Treatment to Study Amplicon HCV Minor Variants. PLoS One. 2013; 8: e83361. doi: 10.1371/journal.pone.0083361 PMID: 24391758

28. R: The R Project for Statistical Computing [Internet]. [Accessed 14 Apr 2015]. Available: http://www.r-project.org/

29. Pages H, Aboyoun P, Gentleman R, DebRoy S. Bioconductor—Biostrings [Internet]. [Accessed 14 Apr 2015]. Available: http://www.bioconductor.org/packages/release/bioc/html/Biostrings.html

30. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004; 20: 289–90. PMID: 14734327

31. Cuadras C. Distance analysis in discrimination and classification using both continuous and categorical variables. In: Dodge Y, editor. Statistical AData analysis and Interference. Amsterdam: Elsevier; 1989. pp. 459–473.

32. Cuadras C. A distance approach to discriminant analysis and its properties. Mathematics preprint series. Barcelona; 1991.

33. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980; 16: 111–20. PMID: 7463489

34. Kumar S, Gadagkar SR. Efficiency of the neighbor-joining method in reconstructing deep and shallow evolutionary relationships in large phylogenies. J Mol Evol. 2000; 51: 544–53. PMID: 11116328

35. Borg I, PJF G. Modern multidimensional scaling: theory and applications. 2nd ed. New York: Springer; 2005.

36. Bekondi C, Olinger CM, Boua N, Talarmin A, Muller CP, Le Faou A, et al. Central African Republic is part of the West-African hepatitis B virus genotype E crescent. J Clin Virol. 2007; 40: 31–7. doi: 10.1016/j.jcv.2007.05.009 PMID: 17689139

37. Zhang Q, Cao G. Genotypes, mutations, and viral load of hepatitis B virus and the risk of hepatocellular carcinoma: HBV properties and hepatocarcinogenesis. Hepat Mon. 2011; 11: 86–91. PMID: 22087123

38. Simmonds P, Midgley S. Recombination in the genesis and evolution of hepatitis B virus genotypes. J Virol. 2005; 79: 15467–76. doi: 10.1128/JVI.79.24.15467–15476.2005 PMID: 16306618

39. Mese S, Arikan M, Cakiris A, Abaci N, Gumus E, Kursun O, et al. Role of the line probe assay INNO-LiPA HBV DR and ultradeep pyrosequencing in detecting resistance mutations to nucleoside/nucleotide analogues in viral samples isolated from chronic hepatitis B patients. J Gen Virol. 2013; 94: 2729–38. doi: 10.1099/vir.0.053041–0 PMID: 24045109

40. Neumann AU. Hepatitis B viral kinetics: A dynamic puzzle still to be resolved. Hepatology. 2005; 42: 249–254. doi: 10.1002/hep.20831 PMID: 16025494