



## Basic Study

# Cross-sectional evaluation of circulating hepatitis B virus RNA and DNA: Different quasispecies?

Selene Garcia-Garcia, Maria Francesca Cortese, David Tabernero, Josep Gregori, Marta Vila, Beatriz Pacín, Josep Quer, Rosario Casillas, Laura Castillo-Ribelles, Roser Ferrer-Costa, Ariadna Rando-Segura, Jesús Trejo-Zahinos, Tomas Pumarola, Ernesto Casis, Rafael Esteban, Mar Riveiro-Barciela, Maria Buti, Francisco Rodríguez-Frías

### ORCID number:

Selene Garcia-Garcia 0000-0001-7697-2058; Maria Francesca Cortese 0000-0002-4318-532X; David Tabernero 0000-0002-1146-4084; Josep Gregori 0000-0002-4253-8015; Marta Vila 0000-0001-9303-5189; Beatriz Pacín 0000-0001-8718-6589; Josep Quer 0000-0003-0014-084X; Rosario Casillas 0000-0002-6758-6734; Laura Castillo-Ribelles 0000-0001-9380-7839; Roser Ferrer-Costa 0000-0002-8925-3172; Ariadna Rando-Segura 0000-0003-4555-7286; Jesús Trejo-Zahinos 0000-0001-6049-555X; Tomas Pumarola 0000-0002-5171-7461; Ernesto Casis 0000-0002-6855-1488; Rafael Esteban 0000-0001-5280-392X; Mar Riveiro-Barciela 0000-0001-9309-2052; Maria Buti 0000-0002-0732-3078; Francisco Rodríguez-Frías 0000-0002-9128-7013.

**Author contributions:** Rodríguez-Frías F, Pumarola T and Esteban R designed the research; Cortese MF and Tabernero D equally coordinated the research; Garcia-Garcia S and Quer J designed the experiments; Garcia-Garcia S, Vila M, Pacín B and Casillas R and Castillo-Ribelles L performed the experiments; Cortese MF, Tabernero D, Garcia-Garcia S and Gregori J analyzed data acquired during the experiments and

**Selene Garcia-Garcia, Maria Francesca Cortese, Marta Vila, Beatriz Pacín, Rosario Casillas, Laura Castillo-Ribelles, Roser Ferrer-Costa, Ernesto Casis, Francisco Rodríguez-Frías**, Clinical Biochemistry Research Group, Department of Biochemistry, Vall d'Hebron Institut Recerca-Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona 08035, Spain

**Selene Garcia-Garcia, Maria Francesca Cortese, David Tabernero, Josep Gregori, Beatriz Pacín, Josep Quer, Rafael Esteban, Mar Riveiro-Barciela, Maria Buti, Francisco Rodríguez-Frías**, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Instituto de Salud Carlos III, Madrid 28029, Spain

**José Gregori, José Quer**, Liver Unit, Liver Disease Laboratory-Viral Hepatitis, Vall d'Hebron Institut Recerca-Hospital Universitari Vall d'Hebron, Vall d'Hebron Barcelona Hospital Campus, Barcelona 08035, Spain

**Ariadna Rando-Segura, Francisco Rodríguez-Frías**, Liver Pathology Unit, Departments of Biochemistry and Microbiology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona 08035, Spain

**Ariadna Rando-Segura, Jesús Trejo-Zahinos, Tomas Pumarola**, Department of Microbiology, Hospital Universitari Vall d'Hebron, Vall d'Hebron Barcelona Hospital Campus, Barcelona 08035, Spain

**Rafael Esteban, Mar Riveiro-Barciela, Maria Buti**, Liver Unit, Department of Internal Medicine, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona 08035, Spain

**Corresponding author:** David Tabernero, PhD, Postdoc, Research Scientist, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Instituto de Salud Carlos III, Calle de Melchor Fernández Almagro, 3, Madrid 28029, Spain.

[david.tabernero@ciberehd.org](mailto:david.tabernero@ciberehd.org)

## Abstract

### BACKGROUND

interpreted the results; Garcia-Garcia S and Tabernero D drafted the manuscript; Ferrer-Costa R, Rando-Segura A, Trejo-Zahinos J, Casis E, Riveiro-Barciela M, Buti M and Rodríguez-Frías F critically reviewed the manuscript.

**Supported by** Instituto de Salud Carlos III, No. PI18/01436; and European Regional Development Fund (ERDF).

#### Institutional review board

**statement:** The study was approved by the Ethics Committee of the Vall d'Hebron Research Institute (PR(AG)146/2020) and all patients provided written informed consent for participation.

#### Conflict-of-interest statement:

Authors have no conflict of interest for this manuscript.

**Data sharing statement:** Next-generation sequencing data were submitted to the GenBank SRA database (BioProject accession number PRJNA625435).

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

**Specialty type:** Gastroenterology and hepatology

**Country/Territory of origin:** Spain

#### Peer-review report's scientific quality classification

Grade A (Excellent): 0  
Grade B (Very good): B  
Grade C (Good): 0  
Grade D (Fair): 0  
Grade E (Poor): 0

**Received:** July 8, 2021

Different forms of pregenomic and other hepatitis B virus (HBV) RNA have been detected in patients' sera. These circulating HBV-RNAs may be useful for monitoring covalently closed circular DNA activity, and predicting hepatitis B e-antigen seroconversion or viral rebound after nucleos(t)ide analog cessation. Data on serum HBV-RNA quasispecies, however, is scarce. It is therefore important to develop methodologies to thoroughly analyze this quasispecies, ensuring the elimination of any residual HBV-DNA. Studying circulating HBV-RNA quasispecies may facilitate achieving functional cure of HBV infection.

#### AIM

To establish a next-generation sequencing (NGS) methodology for analyzing serum HBV-RNA and comparing it with DNA quasispecies.

#### METHODS

Thirteen untreated chronic hepatitis B patients, showing different HBV-genotypes and degrees of severity of liver disease were enrolled in the study and a serum sample with HBV-DNA > 5 Log<sub>10</sub> IU/mL and HBV-RNA > 4 Log<sub>10</sub> copies/mL was taken from each patient. HBV-RNA was treated with DNase I to remove any residual DNA, and the region between nucleotides (nt) 1255-1611 was amplified using a 3-nested polymerase chain reaction protocol, and analyzed with NGS. Variability/conservation and complexity was compared between HBV-DNA and RNA quasispecies.

#### RESULTS

No HBV-DNA contamination was detected in cDNA samples from HBV-RNA quasispecies. HBV quasispecies complexity showed heterogeneous behavior among patients. The Rare Haplotype Load at 1% was greater in DNA than in RNA quasispecies, with no statistically significant differences ( $P = 0.1641$ ). Regarding conservation, information content was equal in RNA and DNA quasispecies in most nt positions [218/357 (61.06%)]. In 102 of the remaining 139 (73.38%), HBV-RNA showed slightly higher variability. Sliding window analysis identified 4 hyper-conserved sequence fragments in each quasispecies, 3 of them coincided between the 2 quasispecies: nts 1258-1286, 1545-1573 and 1575-1604. The 2 hyper-variable sequence fragments also coincided: nts 1311-1344 and 1461-1485. Sequences between nts 1519-1543 and 1559-1587 were only hyper-conserved in HBV-DNA and RNA, respectively.

#### CONCLUSION

Our methodology allowed analyzing HBV-RNA quasispecies complexity and conservation without interference from HBV-DNA. Thanks to this, we have been able to compare both quasispecies in the present study.

**Key Words:** Hepatitis B virus RNA; Hepatitis B X gene; Quasispecies; Next-generation sequencing; Quasispecies conservation; Quasispecies complexity

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.

**Core Tip:** Hepatitis B virus (HBV) quasispecies composition and its evolution are related to liver disease progression. HBV-RNA in serum is potentially useful for analyzing viral quasispecies, even in patients with low levels of or suppressed serum HBV-DNA. Few studies have analyzed circulating HBV-RNA quasispecies, and similarities and differences with DNA quasispecies should be assessed. We used next-generation sequencing to analyze RNA quasispecies variability/conservation and complexity, without interference of HBV-DNA, in untreated chronic hepatitis B patients. Comparison of both quasispecies showed similar results between them. DNA quasispecies tended toward greater complexity, while RNA quasispecies tended toward higher variability.

**Citation:** Garcia-Garcia S, Cortese MF, Tabernero D, Gregori J, Vila M, Pacín B, Quer J, Casillas R, Castillo-Ribelles L, Ferrer-Costa R, Rando-Segura A, Trejo-Zahinos J, Pumarola T,

**Peer-review started:** July 8, 2021**First decision:** August 19, 2021**Revised:** August 20, 2021**Accepted:** October 18, 2021**Article in press:** October 18, 2021**Published online:** November 7, 2021**P-Reviewer:** Wu W**S-Editor:** Gao CC**L-Editor:** A**P-Editor:** Gao CC

Casis E, Esteban R, Riveiro-Barciela M, Buti M, Rodríguez-Frías F. Cross-sectional evaluation of circulating hepatitis B virus RNA and DNA: Different quasispecies? *World J Gastroenterol* 2021; 27(41): 7144-7158

**URL:** <https://www.wjgnet.com/1007-9327/full/v27/i41/7144.htm>

**DOI:** <https://dx.doi.org/10.3748/wjg.v27.i41.7144>

## INTRODUCTION

The hepatitis B virus (HBV) is the etiological agent of hepatitis B and can cause both acute and chronic liver disease. In 2015, roughly 257 million people worldwide were estimated to suffer from chronic hepatitis B (CHB) infection, the major complications of which are cirrhosis and hepatocellular carcinoma[1].

The HBV genome is a 3.2 kb-long partially double-stranded relaxed circular DNA (rcDNA), with a complete minus-strand and an incomplete plus-strand. After cytoplasmic release of the viral nucleocapsid containing the rcDNA, it is actively transported into the nucleoplasm, where it will be repaired and converted to covalently closed circular DNA (cccDNA). This form of viral genome remains as an episomal minichromosome in the hepatocyte nuclei for their entire life and serves as the template for the transcription of viral messenger RNAs by the cellular RNA polymerase II. One of them, the pregenomic RNA (pgRNA) is encapsidated and then reverse-transcribed to generate new rcDNA (or double stranded linear DNA), which is released as new infective DNA-containing virions[2]. In addition, full and spliced pgRNA, as well as X open reading frame (HBX) transcripts have been detected in patients' serum[3]. Serum HBV pgRNA has been reported to be useful for monitoring cccDNA activity in hepatitis B e-antigen (HBeAg)-positive (+ve) entecavir-treated patients[4] and to be predictive of HBeAg seroconversion during polymerase inhibitor treatment[5-7]. In addition, its detection in the absence of detectable serum HBV DNA in patients receiving nucleos(t)ide analogs (NA) therapy may allow inference in cccDNA transcription and has been shown to be a predictive biomarker for viral rebound after treatment cessation[8].

In a single HBV infection, viral genomes are present as a complex mutant spectrum constituted by closely related, but not identical, viral populations (genetic variants), termed viral quasispecies[9]. Those variants can occur through numerous host-virus interactions during the viral replication cycle, encompassing not only cellular factors but also viral factors, essentially the error-prone viral polymerase[10]. Serum circulating HBV-RNA is potentially useful for analyzing viral quasispecies, even in patients with low levels of or suppressed serum HBV-DNA. However, unlike DNA quasispecies, HBV-RNA quasispecies have not been subjected to reverse transcription, which is the main source of variability in the HBV genome[11]. Therefore, both quasispecies may be different in terms of nucleic acid variability/conservation and genetic variant complexity. Few studies to date have analyzed circulating HBV-RNA quasispecies[4,6] and similarities and differences with DNA quasispecies should therefore be assessed in more detail. In addition, serum HBV-RNA has been reported to be genetically homogenous to intrahepatic HBV-RNA[4], the main target for targeted gene therapy, which may favor achieving functional cure of HBV infection, the ideal end point of therapy based on sustained hepatitis B virus surface antigen (HBsAg) loss[12]. In order to interfere with the synthesis of all the viral proteins, these therapies should be directed to hyper-conserved sequence targets shared by all the HBV mRNAs. HBX thus provides an ideal target, as its sequence is located at the 3' end of all HBV mRNAs[13,14]. Taking this into account, analysis of serum HBV-RNA quasispecies may be an especially useful tool when looking for hyper-conserved targets for this kind of antiviral therapy. To confirm this, it is necessary to establish a reliable methodology to thoroughly analyze serum HBV-RNA quasispecies.

Taking this in mind, the aim of this study was to establish a methodology, based on next-generation sequencing (NGS), for thoroughly analyzing serum HBV-RNA quasispecies without interference from circulating HBV-DNA. This methodology has been tested in a group of untreated CHB patients to gain a preliminary insight into comparison of genetic variability/conservation and complexity between circulating HBV-RNA and DNA quasispecies.

## MATERIALS AND METHODS

### **Patients and samples**

The study was approved by the Ethics Committee of the Vall d'Hebron Research Institute (PR(AG)146/2020) and all patients provided written informed consent for participation. No animals were used.

In this cross-sectional study, we included a serum sample from 13 well-characterized untreated CHB patients attending the outpatient clinic of Vall d'Hebron University Hospital (Barcelona, Spain). All samples were selected taking into account HBV-DNA levels  $> 5 \text{ Log}_{10} \text{ IU/mL}$  and HBV-RNA levels  $> 4 \text{ Log}_{10} \text{ copies/mL}$  to ensure sufficient levels of both HBV-DNA and RNA to study their quasispecies. In addition, heterogeneity in terms of HBV genotypes and degrees of severity of liver disease were also taken into account in patient selection, in order to obtain an overall picture of differences between HBV-DNA and RNA quasispecies. Exclusion criteria were positive testing for hepatitis D virus, hepatitis C virus, and human immunodeficiency virus.

### **Serological and virological determinations**

HBV serological markers such as HBsAg, HBeAg, anti-HBe antibodies were tested using commercial chemiluminescent immunoassays on a COBAS 8000 analyzer (Roche Diagnostics, Rotkreuz, Switzerland). HBV-DNA was quantified on a COBAS 6800 analyzer (Roche Diagnostics, Mannheim, Germany) with a lower detection limit of 10 IU/mL. HBV genotypes were determined as previously explained[13]. HBV-RNA was quantified using an in-house method. The standard RNA to create the standard curve for this quantification was obtained as follows: 1.1× HBV genome from pTRIEx1.1-HBV[15] was subcloned downstream of a T7 promoter in a pEF6/V5-His TOPO TA vector (Thermo Fisher Scientific-Life Technologies, Austin, TX, United States) following the manufacturer's instructions, to ensure its *in vitro* transcription (pEF6-HBV). Once cloned, the correct orientation and insertion of HBV genome was analyzed by Sanger Sequencing, and pEF6-HBV plasmid was isolated from bacteria using the NucleoBond Xtra Midi Kit (Macherey-Nagel, Dueren, Germany). The pEF6-HBV plasmid was then linearized by NotI restriction enzyme digestion (New England Biolabs, Beverly, MA, United States), and used as template for *in vitro* transcription reaction using the MEGAscript T7 Transcription Kit (Thermo Fisher Scientific-Life Technologies, Austin, TX, United States), by adding the plasmid pEF6-HBV concentration of 0.5 µg/µL diluted in water. At the same time, the *in vitro* transcription of pTRI-Xef positive control DNA provided with the kit was also performed. The RNA resulting from transcription was then purified using the MEGAclear Transcription Clean-up Kit (Thermo Fisher Scientific-Life Technologies, Austin, TX, United States) following the manufacturer's instructions. A DNase I treatment (Life Technologies, Austin, TX, United States) was then carried out for 15 min at room temperature, followed by heat-inactivation at 65 °C for 10 min and adding 2 µL of 25 mmol/L EDTA solution to the reaction mixture, in order to remove the template DNA. The purity of DNase I-treated RNA was checked by measuring the absorbance of 1 µL of the sample at 260 nm using the NanoDrop spectrophotometer (Thermo Fisher Scientific-Life Technologies, Austin, TX, United States). The concentration of this RNA was quantified from 3 µL of the sample by Qubit fluorimeter using the Qubit RNA HS Assay Kit (Thermo Fisher Scientific-Life Technologies, Austin, TX, United States). The absence of DNA in DNase I-treated RNA template was verified using the DNA Master PLUS HybProbe kit in a LightCycler 480 Instrument II system (Roche, Mannheim, Germany) using primers and probe specified in the [Supplementary Table 1](#).

For absolute quantification analyses, the standard curve was defined using the quantified retrotranscribed RNA, taken to a concentration of 7.51  $\text{Log}_{10} \text{ copies/mL}$ , and serial 1:10 dilutions of this standard covering concentrations until 1.51  $\text{Log}_{10} \text{ copies/mL}$ . The points of the standard curve were defined as the mean Ct of a triplicate measurement of the original 7.51  $\text{Log}_{10} \text{ copies/mL}$  standard and its serial dilutions. The standard curve obtained was saved and imported as an external standard curve in each HBV-RNA quantification experiment. In those experiments, at least one dilution of standard RNA used to define one of the standard curve points must be included, in order to adjust the standard curve. Creation of the RNA standard curve and experiments to quantify HBV-RNA levels quantification were performed using one-step quantitative reverse transcription polymerase chain reaction (RT-qPCR) reaction, using the LightCycler 480 RNA Master Hydrolysis Probes kit (Roche, Mannheim, Germany). RT-qPCR program in the Light Cycle 480 instrument was



programmed as described in the [Supplementary Table 1](#).

### **Amplification of HBV region of interest using NGS**

In this study, we analyzed the region between nucleotides (nt) 1255 to 1611, located at the *HBX* 5' end. In this region, we previously identified hyper-conserved sequence stretches in the circulating HBV-DNA quasispecies[13,14], which we aimed at analyzing at circulating HBV-RNA level.

For each serum sample, HBV-DNA was extracted from 200  $\mu$ L of serum using QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. The region of interest was amplified through a 3-round polymerase chain reaction (PCR) protocol. The first-round PCR covered a 1338-bp region and was performed as previously described by our group[16] ([Supplementary Table 1](#)). The second-round and third-round PCRs, were also performed as previously detailed[13] ([Supplementary Table 1](#)). The second-round PCR amplified the region of interest between nt 1255 and 1611 of the HBV genome, yielding final products flanked by universal M13 sequences at both ends. In the third-round PCR, a specific pair of primers was used for each sample, consisting of the same M13 universal sequences (forward and reverse) at their 5' ends and a multiplex identifier (MID) or barcode sequence at their 3' ends. Each individual patient sample required a different MID. The PCR products obtained in this amplification, also known as amplicons, were visualized as single bands on a 1.5% agarose electrophoresis gel, stained with SybrSafe DNA gel Stain (Life Technologies, Carlsbad, CA, United States) with 1x TAE running buffer (Roche Diagnostics, Mannheim, Germany). PCR products were then purified from agarose gel using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). Amplicon quality was verified using the Agilent 2200 TapeStation System and D1000 ScreenTape kit (Agilent Technologies, Waldbronn, Germany). Purified DNA from each sample was quantified by means of fluorescence, using the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Carlsbad, CA, United States) adjusted to the same concentration, and pooled.

In the same serum samples, HBV-RNA was isolated from 140  $\mu$ L serum using the QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. To remove any residual DNA present, isolated HBV-RNA was subsequently treated with DNase I (Life Technologies, Austin, TX, United States) for 15 min at room temperature, followed by heat-inactivation at 65 °C for 10 min and adding 25 mmol/L EDTA solution to the reaction mixture, in keeping with manufacturer's instructions. After DNase I treatment, RNA samples were retro-transcribed into cDNA in 2 steps. The first step involves denaturation of HBV-RNA and Oligo(dT)<sub>17</sub> primer, which binds to polyA sequence, by incubating at 65 °C for 5 min and 20 °C for 5 min. The second is reverse transcription itself involving Accuscript enzyme (Stratagene, Agilent Technologies, Santa Clara, United States) and RNase OUT (Thermo Fisher Scientific-Life Technologies, Austin, United States). A 60-min incubation at 42 °C for the first-strand synthesis reaction is required, followed by a short incubation time of 15 min at 70 °C.

At the same time, elimination of any residual HBV-DNA after DNase I treatment was verified by means of qPCR, as described for the DNase I-treated RNA obtained from *in vitro* transcription of pEF6-HBV vector, described above. This qPCR experiment included the DNase I-treated RNA samples and their respective cDNA obtained after reverse transcription. After obtaining the cDNA, the region of interest between nt 1255 and 1611 was amplified through the same 3-round PCR protocol as HBV-DNA samples, with some modifications according to Agilent's PfuUltra II Fusion HS DNA polymerase insert for cDNA targets.

Finally, amplicon pools obtained from both HBV-DNA and RNA were sequenced using the MiSeq platform (Illumina, San Diego, CA, United States). Those purified pools were processed using the DNA library preparation kit Kapa Hyper Prep kit (Kapa Biosystems-Roche, Cape Town, South Africa), with which amplicon ends were repaired and A-tailed. Each pool was then indexed using the SeqCap Adapter Kit A (Roche Sequencing, Pleasanton, CA, United States). After index ligation pools had been purified with KAPA Pure Beads (Kapa Biosystems-Roche, Cape Town, South Africa) and re-amplified to increase indexed amplicon concentration using the KAPA HiFi HotStart ReadyMix PCR Kit (Roche Sequencing, Pleasanton, CA, United States). PCR products were repurified with another round of KAPA Pure Beads and quantified using LightCycler 480 Instrument II with the Kapa Library Quantification kit (Kapa Biosystems-Roche, Cape Town, South Africa). Each DNA pool was then adjusted to a 4 nmol/L concentration and appropriate volumes of each pool were added to a final pool of all DNA pools, the DNA concentration of which was verified using LightCycler 480 Instrument II (Kapa Library Quantification kit). Finally, this

final pool was sequenced using MiSeq Reagent kit v3 (2 × 300 bp mode with the 600 cycle kit) (Illumina, San Diego, CA, United States).

### **NGS data analyses: Quasispecies complexity and conservation**

Bioinformatics processing and haplotype-centric data analysis pipeline were performed as previously described by our group[16].

The HBV-DNA and RNA quasispecies complexities were evaluated by computation of the Rare Haplotype Load (RHL), a new diversity index that measures enrichment of quasispecies in minority genomes below a given threshold[17]. RHL was calculated using the haplotypes (unique sequences covering the full amplicon observed on the set of sequences) resulting from the intersection of forward and reverse strands without filtering by a minimum abundance. In this study, RHL was computed as the sum of the relative frequencies of all haplotypes obtained from HBV-DNA and RNA quasispecies whose abundance is equal to or less than 1%, as previously described[17].

Sequence conservation at nt level was determined by calculating the information content (IC, bits) of each position in a multiple alignment of all intersected haplotypes obtained in frequencies > 0.25%, as previously explained[13]. Nt IC ranges from 0 bits to 2 bits for a given position, where 2 bits represents the maximum IC value (100% conservation, *i.e.*, no nt changes in a given position in any of all haplotypes analyzed). The mean IC was calculated in windows of 25 nt, moved in steps of 1 nt through nt 1255-1611 (sliding window analysis), the 5% of those windows with the highest mean IC values (the most conserved) and the 5% with the lowest mean IC values (the most variable) were selected. Afterwards, those consecutive windows were connected and their sequences were represented as sequence logos using the R language package motifStack[18]. In each position of the logos, letters representing the nts identified are stacked. The relative sizes of letters indicate the relative frequencies of the nt represented at a given position in the multiple alignments of the haplotypes, and the total height of each stack represents the IC of each position.

### **Statistical analysis**

All statistical analyses were performed in R[19]. Qualitative parameters were expressed as number of cases and percentage. Quantitative parameters were expressed as the median value and interquartile range (IQR) or as mean ± SD, where appropriate. Statistical comparisons between HBV-RNA and DNA quasispecies complexity were performed using the Kruskal-Wallis test, *t* test and two-proportions z-test with Yates continuity correction where appropriate. *P* values < 0.05 were considered significant. The bioinformatics and statistical methods used in this study were reviewed by Dr. Josep Gregori from the Liver Disease Viral Hepatitis Laboratory of Vall d'Hebron Hospital (Barcelona, Spain) and the CIBERehd research group.

## **RESULTS**

### **Patient clinical and virological characteristics**

Clinical, virological, and serological parameters from the 13 patients at the time of obtention of samples included are summarized in Table 1. Most of those patients were HBeAg +ve men, showing clinical characteristics compatible with chronic hepatitis stage of CHB infection[12]. In terms of severity of liver disease, 8 patients were classified as patients with nonsignificant fibrosis [Ishak fibrosis stage (F) < 3 or transient elastography < 7-8.5 kPa], and 5 patients were classified as patients with significant fibrosis (liver biopsy F ≥ 3 and < 5 or transient elastography > 7-8.5 and < 11-14 kPa, *n* = 3) or cirrhosis (liver biopsy F5-6 or transient elastography > 11-14 kPa, *n* = 2), according to World Health Organization Guidelines[20].

### **Comparison of HBV-DNA and RNA quasispecies using NGS**

The qPCR verification of residual HBV-DNA elimination from HBV-RNA isolations confirmed the absence of HBV-DNA from DNase I-treated RNA samples, while showing the presence of cDNA after their reverse transcription. An illustrative example is shown in Figure 1. No HBV-DNA contamination was therefore detected in cDNA samples derived from HBV-RNA quasispecies.

After NGS using the MiSeq platform, Fastq files obtained were filtered by our bioinformatics procedure to obtain the set of haplotypes common to forward and reverse strands. Before subsequent filtering by abundance, those haplotypes represented a total of  $4.35 \times 10^6$  sequence reads with a median of  $2.45 \times 10^5$  reads/sample

**Table 1 Clinical and viral characteristics of chronic hepatitis B patients enrolled in the study**

	Total, n = 13
Age, yr, median (IQR)	41.35 (31.7-47.2)
Male, n (%)	11 (84.6)
Genotype, n (%)	
A	3 (23.1)
C	3 (23.1)
D	2 (15.4)
E	2 (15.4)
F	3 (23.1)
ALT, IU/L, median (IQR)	124 (66-160)
HBeAg +ve, n (%)	11 (84.6)
HBV DNA, Log <sub>10</sub> IU/mL, median (IQR)	8 (8-8.43)
HBV RNA, Log <sub>10</sub> copies/mL, median (IQR)	5.18 (4.94-6.02)
Fibrosis, n (%) <sup>1</sup>	
Nonsignificant	8 (61.64)
Significant	3 (23.08)
Cirrhosis, n (%)	2 (15.39)

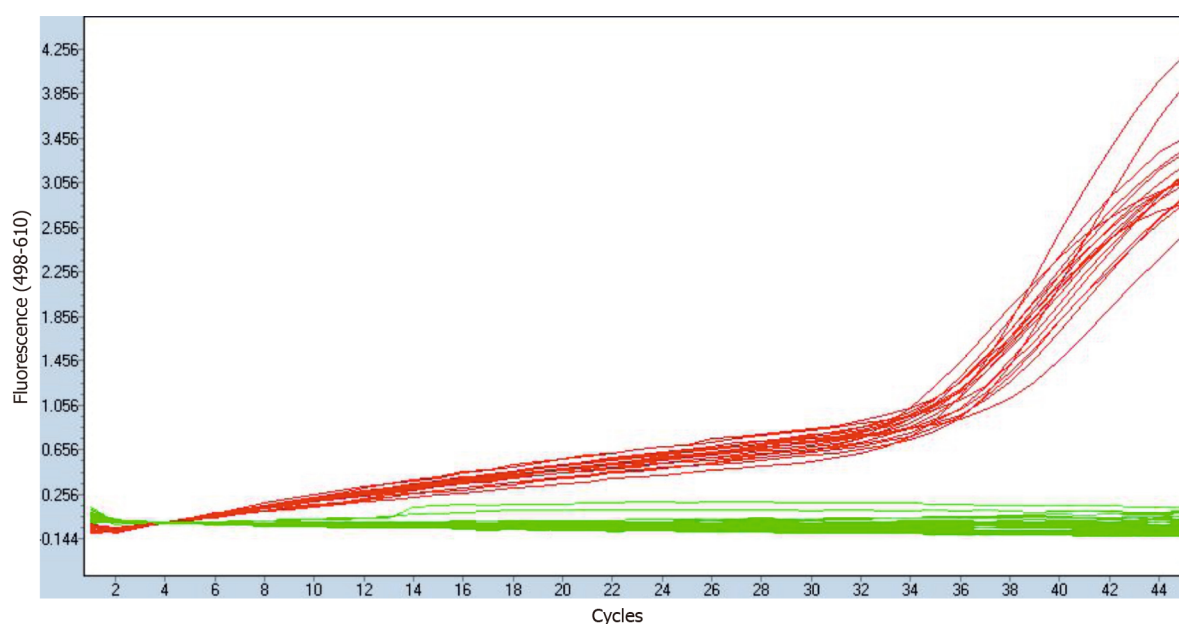
<sup>1</sup>Nonsignificant fibrosis indicates Ishak fibrosis stage < 3, significant fibrosis F ≥ 3 and < 5, and cirrhosis F5-6. Nonsignificant fibrosis by noninvasive markers indicates liver stiffness < 7-8.5 kPa, significant fibrosis > 7-8.5 and < 11-14 kPa and cirrhosis > 11-14 kPa by transient elastography according to World Health Organization Guidelines for the prevention, care and treatment of persons with chronic hepatitis B infection criteria[20].

F: Fibrosis; IQR: Interquartile range; ALT: Alanine aminotransferase; HBeAg +ve: Hepatitis B e-antigen positive; HBsAg: Hepatitis B virus surface antigen; HBV: Hepatitis B virus.

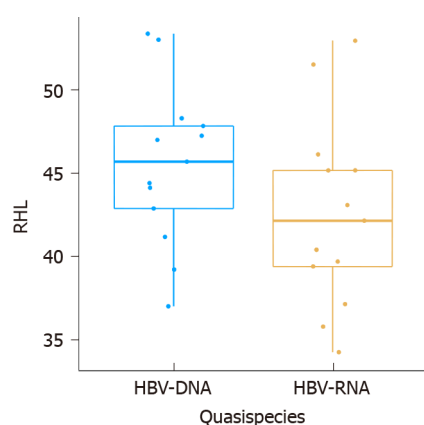
(IQR,  $1.90 \times 10^5$ - $3.19 \times 10^5$ ) for HBV-RNA and  $8.27 \times 10^4$  reads/sample (IQR,  $6.03 \times 10^4$ - $1.01 \times 10^5$ ) for HBV-DNA and were included in the RHL analysis. By setting the threshold of abundance as equal to or less than 1% HBV quasispecies complexity showed a heterogeneous behavior between patients, and although mean RHL was greater in DNA ( $45.45 \pm 4.80$ ) than in RNA ( $42.50 \pm 5.63$ ), the difference was not significant ( $P = 0.1641$ ,  $t$  test) (Figure 2). Similarly, no statistically significant differences were observed when comparing HBV-DNA and RNA quasispecies in the patients with nonsignificant fibrosis ( $46.57 \pm 4.79$  for HBV-DNA and  $43.88 \pm 6.32$  for HBV-RNA,  $P = 0.4642$ , Kruskal-Wallis test), and in the patients with significant fibrosis or cirrhosis ( $43.66 \pm 4.72$  for HBV-DNA and  $40.29 \pm 3.92$  for HBV-RNA,  $P = 0.3055$ , Kruskal-Wallis test).

After RHL calculation, we proceeded to eliminate all haplotypes with abundances below 0.25%, obtaining  $2.91 \times 10^6$  sequence reads with a median of  $1.58 \times 10^5$  reads/sample (IQR,  $1.29 \times 10^5$ - $2.28 \times 10^5$ ) for HBV-RNA and  $5.66 \times 10^4$  reads/sample (IQR,  $3.2 \times 10^4$ - $5.76 \times 10^4$ ) for HBV-DNA. These haplotypes were the basis for conservation analysis through IC calculation position by position in both HBV-RNA and DNA quasispecies (Figure 3). Differences in sequence conservation between the 2 quasispecies were determined by subtracting IC DNA values from IC RNA of 357 nt positions analyzed; between nt 1255-1611. In 218 (61.06%) of these positions, IC values of both HBV-RNA and DNA quasispecies were coincident, while in the remaining 139 (38.93%) nt positions differences between IC of both quasispecies ranged from -0.26 to 0.23, with IC DNA > IC RNA in most of them (102/139, 73.38%) (Figure 4A). These differences were also studied by sliding window analysis of the mean IC RNA-IC DNA values (calculated in windows of 25 nt positions displaced in steps of 1 position between them) confirming that, in general, the HBV-RNA quasispecies was slightly less conserved than the DNA quasispecies (Figure 4B). In keeping with this analysis, HBV-RNA displayed more positions with some variability (IC < 2 bits, 135/357 (38%), than HBV-DNA, 85/357 (24%) ( $P < 0.01$ , two-proportions z-test).

The high degree of similarity between IC values in both HBV-DNA and HBV-RNA quasispecies was also observed in the sliding window analysis of IC in the multiple



**Figure 1** Quantitative polymerase chain reaction verification of residual hepatitis B virus-DNA elimination from hepatitis B virus-RNA isolations after DNase I-treatment. Fluorescence through quantitative polymerase chain reaction amplification samples is shown as green lines for DNase I-treated RNA samples, and as red lines for their respective cDNA retrotranscribed samples.



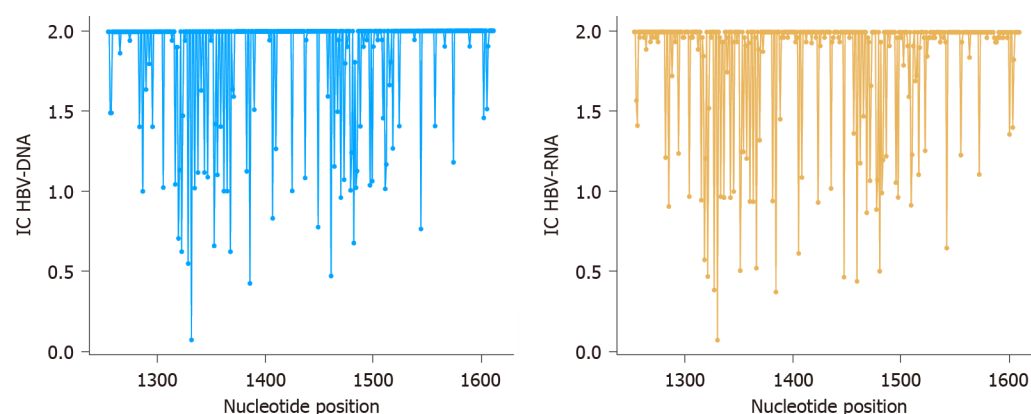
**Figure 2** Comparison of mean hepatitis B virus quasispecies complexity. Hepatitis B virus (HBV) quasispecies complexity analyzed by Rare Haplotype Load of all 13 patients in HBV-DNA quasispecies (blue-framed boxes) and HBV-RNA quasispecies (yellow-framed boxes). Differences were statistically non-significant ( $P = 0.1641$ ,  $t$  test). RHL: Rare Haplotype Load; HBV: Hepatitis B virus.

alignments of haplotypes obtained in both quasispecies. Concatenation of the 5% of windows with higher mean IC values (the most conserved), yielded 4 sequence logos in both quasispecies 3 (75%) of which coincided between both quasispecies: nts 1258-1286 (1283 in RNA quasispecies), 1545-1573 and 1575-1604 (Figure 5). In addition, sliding window analysis also showed the sequence stretch 1519-1543 to be hyper-conserved only in HBV-DNA quasispecies and 1559-1587 only in HBV-RNA quasispecies. Regarding the 5% of windows with lower IC values (the most variable), 2 sequence logos, which coincided in both quasispecies, were obtained: nts 1311-1344 and 1461-1485 (Figure 6). Thus, conservation was highly coincident between HBV-RNA and DNA quasispecies although, interestingly, HBV-RNA quasispecies showed slightly higher variability than HBV-DNA quasispecies.

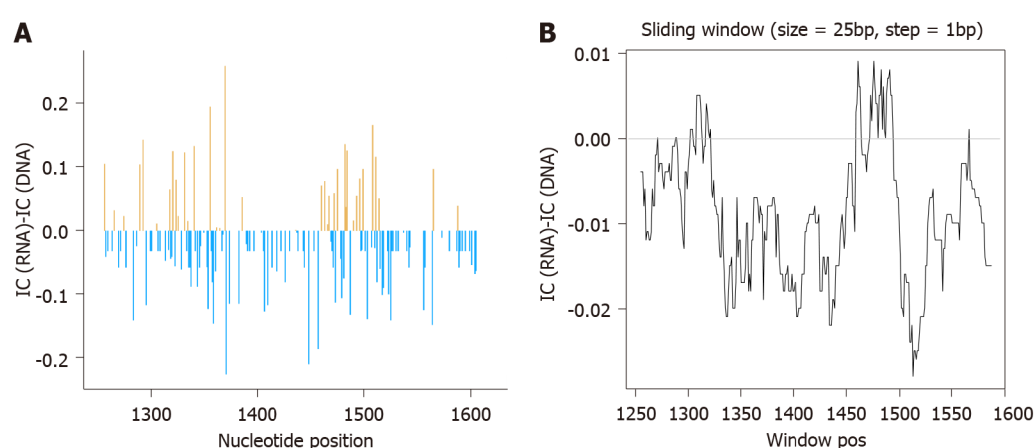
## DISCUSSION

Serum HBV RNA results from a mixture of different viral RNA species: Intact and spliced pgRNA, and truncated and polyA-free RNAs, with varying levels depending on the phase of HBV infection and antiviral treatment[21]. In addition, *HBX* transcripts





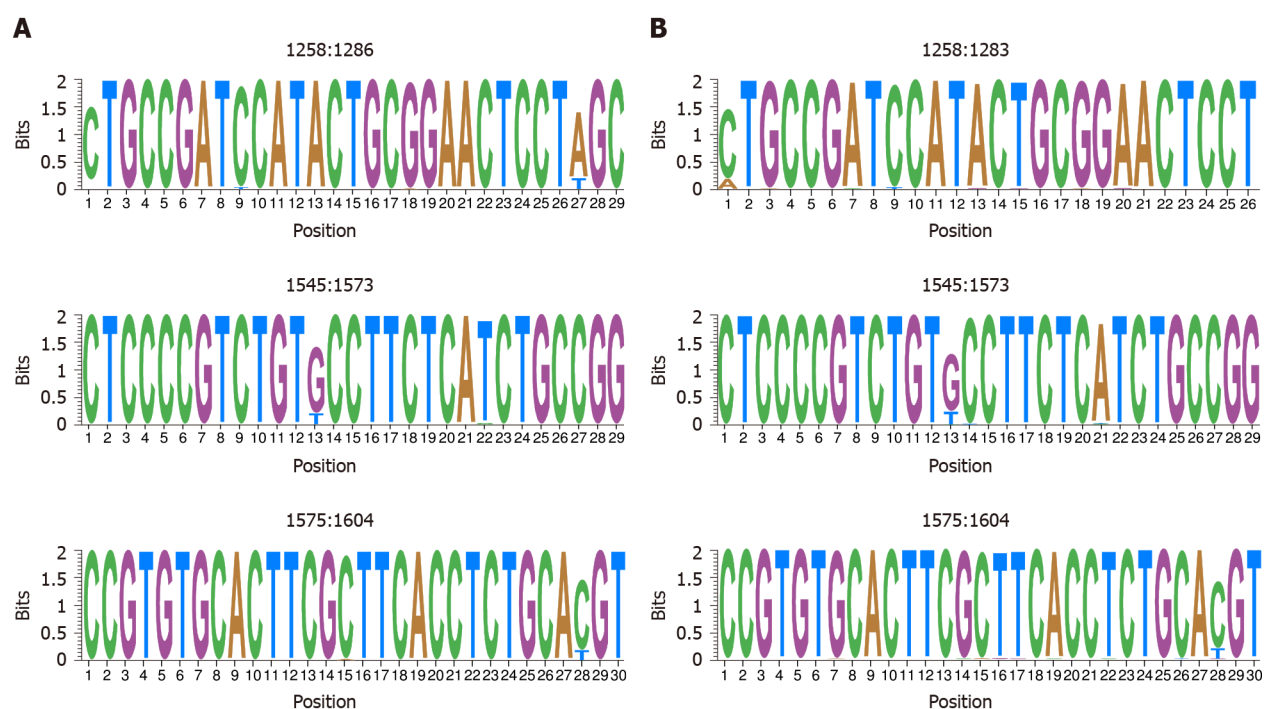
**Figure 3 Conservation and variability of 357 nucleotide positions analyzed.** Information content of nucleotide positions from 1255 to 1611 for hepatitis B virus (HBV)-DNA (blue lines) and HBV-RNA (orange lines) quasispecies. IC: Information content; HBV: Hepatitis B virus.



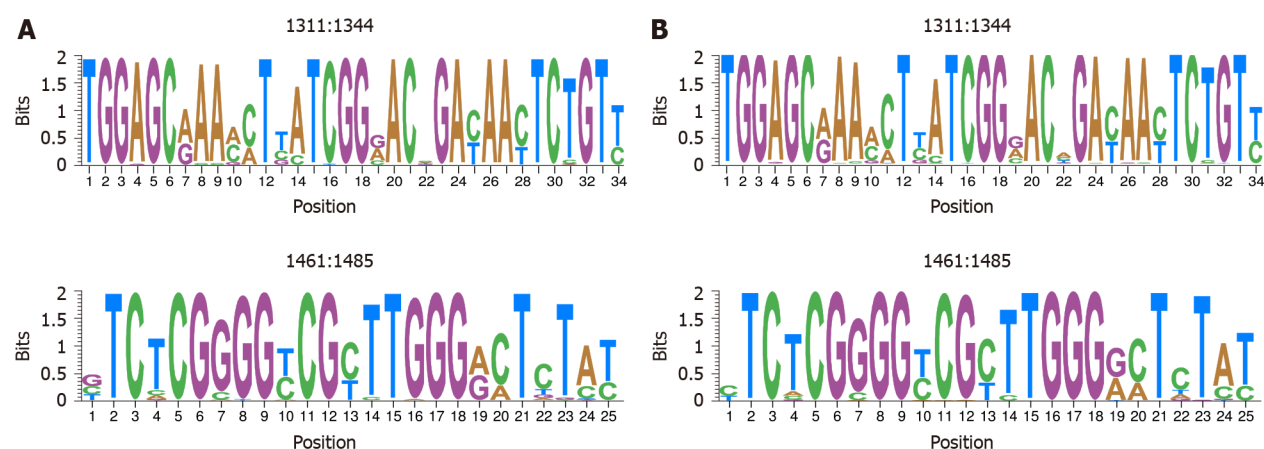
**Figure 4 Differences between information content of hepatitis B virus-DNA and RNA quasispecies.** A: Nucleotide positions in which information content (IC) hepatitis B virus (HBV)-RNA > HBV-DNA are depicted in orange while positions where IC HBV-DNA > HBV-RNA in blue; B: Sliding window analysis of the subtraction of mean IC RNA-IC DNA values, in windows of 25 nucleotide positions, displaced in steps of 1 position between them. IC: Information content.

have also been detected[3]. This serum HBV RNA has been suggested as a potential surrogate marker both to reflect intrahepatic cccDNA levels during the natural course of CHB infection[22,23] and under NA treatment[4], and to predict treatment outcome [4,21] such as HBeAg seroconversion[5-7]. During antiviral treatment, the kinetics of serum HBV-DNA and HBV-RNA seem to be dissociated. While serum HBV-DNA levels rapidly decline after starting treatment, HBV-RNA levels fall more slowly and the HBV-DNA/HBV-RNA ratio increases significantly. This allows inference in cccDNA transcription in the absence of detectable serum HBV DNA and may predict viral rebound after NA cessation[8]. In addition, the slower decline of serum HBV RNA levels may also make it possible to study circulating viral quasispecies, even when serum HBV-DNA levels are too low to do so. However, few studies to date have analyzed circulating HBV-RNA quasispecies[4,6] and little data is therefore available on their similarities and differences. Indeed, it is reasonable to think that both quasispecies may display significant differences between them, as they are subjected to different sources of genetic variability, due to the error-prone viral polymerase, contributing to an estimated error rate of approximately 1 per  $10^4$ - $10^5$  bp for minus-strand DNA synthesis and approximately 1 per  $3.6$ - $15 \times 10^4$  bp for second-strand DNA synthesis[24]. Hence, reverse transcription becomes an outstanding source of variability in new HBV-DNA virion formation. However, HBV-RNA quasispecies are not affected by this source of variability.

The first aim of this study was to establish a reliable methodology to thoroughly analyze serum HBV-RNA quasispecies, without interference from HBV-DNA quasispecies. Previous studies have found RNA levels of between 0.8 and 2.8 Logs lower than DNA levels[5,25,26]. Although we measured HBV-DNA and RNA levels with different units (IU/mL for DNA and copies/mL for RNA) in this study, the results appear to be in line with those previous studies. Even if HBV-DNA is present at low



**Figure 5 Representation by sequence logos of the information content of the most conserved regions in hepatitis B virus-DNA and hepatitis B virus-RNA quasispecies.** The relative sizes of the letters in each stack, each of them representing a nucleotide (nt) position, indicate their relative frequencies at each position within the multiple alignments of nt haplotypes. The total height of each stack of letters depicts the IC of each nt position, measured in bits (Y-axis), therefore 0 bits is the minimum and 2 the maximum conservation. A: Hepatitis B virus-DNA; B: Hepatitis B virus-RNA.



**Figure 6 Representation by sequence logos of the information content of the most variable regions in hepatitis B virus-DNA and hepatitis B virus-RNA quasispecies.** The relative sizes of the letters in each stack, each of them representing a nucleotide (nt) position, indicate their relative frequencies at each position within the multiple alignments of nt haplotypes. The total height of each stack of letters depicts the IC of each nt position, measured in bits (Y-axis), therefore 0 bits is the minimum and 2 the maximum conservation. A: Hepatitis B virus-DNA; B: Hepatitis B virus-RNA.

levels in patient serum, amplicons derived from it may bias NGS-obtained results to some extent, taking into account the high sensitivity of this technology. Thus, to ensure that results obtained from HBV-RNA quasispecies amplification procedure are free from HBV-DNA contamination, we pre-treated RNA isolates with DNase I, and verified the HBV-DNA elimination from these isolates by means of a qPCR assay. This verification step allowed us to confirm that no residual HBV-DNA was present in the HBV-RNA isolates, making it unlikely that reads obtained from HBV-RNA libraries were actually derived from serum HBV-DNA amplification.

Since viral polymerase errors are considered an important source of genetic variability for the HBV genome, we considered RHL to be a diversity index especially suitable for comparing complexities between HBV RNA and DNA quasispecies and assessing the effect of viral polymerase activity on the latter. This is because this

diversity index refers to the fraction of the quasispecies with the lowest fitness, indicative of the intensity with which replication errors accumulate. High viremia and long infection times would tend to result in high RHL values. High values may also be given to mutagen treatments[17]. In our study, HBV-DNA quasispecies showed a higher mean RHL than RNA quasispecies, which suggests an increased presence of low frequency HBV genomes in DNA quasispecies, probably due to the error-prone reverse-transcription origin of HBV-DNA. However, these differences were not statistically significant, either when including all patients or when separating them according to their degree of fibrosis (nonsignificant fibrosis and significant fibrosis or cirrhosis). The RHL shows high correlation with Shannon entropy, an abundance diversity index[17], which was used in a recent study by Yu *et al*[6] to compare HBV-DNA and RNA quasispecies in sequential serum samples in a group of HBeAg+ve patients receiving entecavir or pegylated interferon. Similarly to our results, no significant difference was found between HBV RNA and HBV DNA quasispecies complexity at baseline quasispecies. While the results were similar, we analyzed a different region of the HBV genome (the region of the *HBX* between nt 1255 to 1611) than the study by Yu *et al*[6] (a sequence stretch within the region encoding for the terminal protein domain of the viral polymerase), and our group of patients is more heterogeneous, especially in terms of viral genotype composition. Thus, results from the 2 studies are not directly comparable, and further studies in larger groups of patients are required to confirm the tendency of HBV DNA quasispecies toward higher quasispecies complexity, as shown by mean RHL values.

Comparison of both HBV-DNA and RNA quasispecies conservation yielded similar results, with HBV-DNA slightly more conserved than HBV-RNA despite the effect of reverse-transcription over DNA quasispecies. This observation may indicate that only a fraction of packaged pgRNA is reverse-transcribed and released. In addition, in this study, we identified 3 hyper-conserved regions which coincided in both HBV-RNA and HBV-DNA quasispecies. The first hyper-conserved region identified was between nt 1258-1286, while the second hyper-conserved region consisted of 2 nt fragments (1545-1573 and 1575-1604) spanning a region between nt 1545-1604. Interestingly, these 3 conserved sequence fragments almost overlapped with hyper-conserved sequence fragments described in a previous study carried out by our group at circulating DNA quasispecies level (1255-1286, 1545-1573 and 1575-1603)[13]. Likewise, a recently published study by our group[14] also performed in HBV-DNA quasispecies, reported some of these hyper-conserved regions (1258-1286 and 1575-1605). In addition, these 2 previous studies[13,14] also reported as hyper-conserved the sequence stretch between positions 1519-1543, which we found among the 5% most conserved windows at HBV-DNA quasispecies level but not at HBV-RNA quasispecies level. Conversely, we identified the sequence stretch 1559-1587 as hyper-conserved only in HBV-RNA quasispecies. This fragment was not identified in our previous studies, although it is included in the region between nt 1519 and 1603 defined by 3 conserved nt fragments (1519-1543, 1545-1573, and 1575-1603) described in one of them[13]. Thus, despite the high degree of similarity between sequence conservation in HBV-DNA and RNA quasispecies, this may follow different trends in some sequence stretches. Finally, assessment of the most variable windows (the 5% with the lowest mean IC values) identified the fragments between nt 1311-1344 and 1461-1485 as the most variable in both DNA and RNA quasispecies, similar to the most conserved stretches identified, which confirmed the similarities between both quasispecies.

Given the important and wide role of hepatitis B X protein (HBx), a multifunctional transactivator protein encoded by the *HBX* gene, in the replication of HBV, particularly in intrahepatic cccDNA stability and transcription, it is an interesting target for new therapies to treat CHB-infected patients[8]. Moreover, HBx is located next to the 3' end of all the HBV mRNAs, which means that targeting its sequence may interfere in the synthesis of all the viral proteins[13]. This concept is in line with other strategies aiming for functional cure. Thus, identifying highly conserved regions may suggest valuable targets for pan-genotypic approaches with minimum likelihood of antiviral drug resistance. With this in mind, we analyzed a region of the *HBX* gene where we had previously identified hyper-conserved regions in the circulating HBV-DNA quasispecies[13,14], which we intended to confirm in circulating HBV-RNA. In fact, serum HBV-RNA was found to be genetically homogenous with intrahepatic HBV-RNA[4] the main target for this kind of antiviral therapies. Therefore, given that some sequence fragments may follow different conservation trends in both HBV-DNA and RNA quasispecies, when looking for targets for directed gene therapy, it may be worth verifying their conservation not only at serum HBV-DNA level but also at RNA level. This would make it possible to ensure their effect on their main target. In this study, we confirmed that most of the hyper-conserved regions identified by our group in

previous studies at DNA level[13,14] coincided with those identified in serum HBV-RNA, and may thus also be conserved at intrahepatic HBV-RNA level.

The broad range of HBV genotypes and degrees of severity of liver disease from all patients included in the present study enabled to get an overview of complexity and conservation of HBV-DNA and RNA quasispecies. Nevertheless, the main limitation of the present study was that the comparison with this small and heterogeneous group of patients just enabled to take a preliminary picture of HBV-DNA and RNA quasispecies, which needs to be complemented with further studies exploring both of them in larger and more homogeneous groups of patients in terms of severity of liver disease and CHB clinical stage, in order to verify whether the present results could be extrapolated to any stage of the disease. In addition, the implications of HBV quasispecies parameters analyzed in the context of both treatment and prognosis of CHB are still uncertain. Thus, longitudinal follow-up studies of circulating HBV-RNA quasispecies evolution are necessary to assess its adaptation to different evolutive pressures and to compare with that of HBV-DNA. In this way, much more robust conclusions regarding the comparison between both quasispecies could be drawn. Anyway, the results obtained demonstrate that the methodology established in the present study allows detailed analysis of serum HBV-RNA quasispecies, being a useful tool to perform those studies. Finally, *in vitro* functional studies should be performed to test the potential usefulness of the hyper-conserved regions described here as potential candidates for targeted gene therapy. Gene silencing could be a valuable strategy and *HBX* gene, thanks to its co-terminal localization, an optimal target. These hyper-conserved sequences were highly coincident with the findings of the present study[13,14]. Based on data obtained in those previous studies we have already tested several proposals about gene therapy based on antisense locked nucleic acid Gapmer and small interference RNA (siRNA)[27,28].

## CONCLUSION

In summary, we developed a methodology for analyzing serum HBV-RNA quasispecies without HBV-DNA interference. By including patients with different clinical and virological characteristics we have been able to make a preliminary comparison of complexity and conservation of HBV-DNA and RNA quasispecies. Complexity analysis by the RHL index indicated an increased presence of low frequency HBV genomes in DNA quasispecies, which should be verified in larger groups of patients.

Interestingly, although HBV-DNA and RNA quasispecies showed a similar degree of conservation in the *HBX* 5' region, HBV-RNA quasispecies tended toward higher variability than HBV-DNA. Most hyper-conserved and hyper-variable regions identified were almost overlapped between HBV-DNA and RNA, and were highly coincident with our previous findings. However, we observed different conservation trends between both quasispecies in some sequence fragments, suggesting that serum HBV-RNA quasispecies should also be considered when looking for targets for directed gene therapy.

## ARTICLE HIGHLIGHTS

### Research background

Recent studies show that hepatitis B virus (HBV)-RNA detected in serum is mainly encapsidated pregenomic RNA (pgRNA), which could be a useful biomarker for monitoring covalently closed circular DNA activity. During nucleos(t)ide analogs (NA) therapy it is predictive of hepatitis B e-antigen seroconversion and for following viral rebound after treatment cessation. However, few studies have analyzed the serum HBV-RNA quasispecies, a complex mutant spectrum constituted by closely related, but not identical, viral populations. Analysis of serum circulating HBV-RNA quasispecies in those previous studies showed no significant difference between HBV-RNA and HBV-DNA quasispecies complexity before antiviral treatment and indicated that serum HBV-RNA is genetically homogenous with intrahepatic HBV-RNA, which is the main target for targeted gene therapy. Such therapies should be directed to hyper-conserved sequence targets, ideally located at hepatitis B X gene (*HBX*), located next to the 3' end of all the HBV mRNAs.



### Research motivation

Composition and evolution over time of HBV quasispecies is closely linked to liver disease progression, and serum circulating HBV-RNA is potentially useful for its analysis, even in situations with low or undetectable serum level of HBV-DNA. However, HBV-RNA has not been subjected to reverse transcription, which is the main source of variability in the HBV genome and it may therefore present significant differences with respect to the serum HBV-DNA quasispecies. Nonetheless, little data is available on the comparison of genetic variability/conservation and complexity of both quasispecies. Moreover, analysis of serum HBV-RNA quasispecies may be a very useful tool when looking for hyper-conserved targets for targeted gene therapy, due to its similarities with intrahepatic HBV-RNA, the main target of this kind of antiviral therapy. Thus, in previous studies, we identified hyper-conserved sequence stretches in the circulating HBV-DNA quasispecies, which we aimed to analyze at the circulating HBV-RNA level. Studying circulating HBV-RNA quasispecies may thus favor achieving functional cure of HBV infection. However, it is necessary to establish a reliable methodology to thoroughly analyze this quasispecies, without interference from serum HBV-DNA.

### Research objectives

This study aimed to establish a methodology to achieve an in-depth analysis serum HBV-RNA quasispecies without interference from circulating HBV-DNA, using next-generation sequencing (NGS). With this methodology, we aimed to compare both serum HBV-RNA and DNA quasispecies in a group of untreated chronic hepatitis B (CHB) patients. Considering the potential of HBV-RNA for analyzing HBV quasispecies, even in situations of low or undetectable serum HBV-DNA, this thorough analysis may serve as prognostic factor for clinical follow-up of CHB patients. Furthermore, it may be a useful tool for looking for hyper-conserved targets for targeted gene therapy.

### Research methods

Serum samples were taken from 13 untreated CHB patients attending the outpatient clinic of Vall d'Hebron University Hospital (Barcelona, Spain). HBV-DNA levels  $> 5 \log_{10}$  IU/mL, HBV-RNA levels  $> 4 \log_{10}$  copies/mL and heterogeneity in terms of HBV genotypes and degrees of severity of liver disease were considered as inclusion criteria. HBV-RNA and DNA were extracted differently using specific manual isolation protocols. In addition, HBV-RNA, which was quantified by an in-house method, was treated with DNase I (Life Technologies, Austin, United States) to remove any residual DNA present. The elimination of residual DNA after DNase I digestion was verified by means of quantitative PCR (qPCR). In parallel, these samples were retrotranscribed into cDNA, and along with DNA isolates the 5' end region of *HBX*, between nucleotides 1255-1611 (the amplicon analysed), was amplified by a 3-nested PCR protocol and later sequenced using NGS (MiSeq, Illumina, United States). HBV-RNA and DNA quasispecies complexity was evaluated using Rare Haplotype Load (RHL) index, which measures enrichment of quasispecies in minority genomes. Sequence conservation and variability was determined by calculating the information content (IC) of each position by aligning all unique sequences covering the full amplicon (*i.e.*, haplotypes) in a multiple alignment. After this analysis, the most conserved and variable sequence stretches were represented as sequence logos, which were compared between both quasispecies.

### Research results

After treatment of RNA isolates with DNase I, we confirmed that no residual HBV-DNA was present in the HBV-RNA isolates and that contamination by HBV-DNA in sequences obtained from HBV-RNA was therefore unlikely. HBV quasispecies complexity showed heterogeneous behavior among patients. While RHL was greater in DNA than in RNA quasispecies, differences were not statistically significant. This tendency of HBV DNA quasispecies toward higher quasispecies complexity than HBV-RNA needs to be studied further in larger groups of patients. In general, conservation was highly coincident between HBV-RNA and DNA quasispecies; the majority of nt positions showed the same IC value in both of them. Interestingly, HBV-RNA quasispecies was slightly less conserved than DNA and displayed more positions with some variability ( $IC < 2$  bits). Sliding window analysis in HBV-DNA and RNA quasispecies showed 4 sequence fragments as the most conserved in each of them. Of those fragments 3 coincided between both quasispecies, but 1 was found to be among the most conserved only in HBV-DNA and 1 only in HBV-RNA. The most variable

sequence fragments coincided in both quasispecies.

### Research conclusions

In this study, we describe a methodology for analyzing serum HBV-RNA quasispecies without HBV-DNA interference. This methodology allowed us to compare HBV-DNA and RNA quasispecies. For quasispecies complexity, we analyzed the RHL index, a new reliable diversity index to diagnose mutant spectrum expansions, such as may have occurred with the error-prone reverse transcription of HBV-RNA into DNA. However, although we detected a tendency to greater quasispecies complexity in HBV-DNA than in RNA, the differences were statistically non-significant, which may indicate an increased presence of low-frequency HBV genomes in DNA quasispecies. For this reason, we suggest that further studies in larger groups of patients be performed to confirm this observation. We also found a highly coincident conservation between HBV-RNA and DNA quasispecies in the *HBX* 5' region. Interestingly, HBV-RNA quasispecies showed slightly higher variability than HBV-DNA quasispecies, which may indicate that only a fraction of packaged pgRNA is reverse-transcribed and released. The hyper-conserved sequences identified were highly coincident to what was observed in previous studies by our group in HBV-DNA quasispecies. Nevertheless, we observed that some sequence fragments were differently conserved between HBV-DNA and RNA quasispecies, which suggest that serum HBV-RNA quasispecies should also be considered when looking for targets for directed gene therapy, specially taking into account the similarities between serum and intrahepatic HBV-RNA.

### Research perspectives

In this study, data were obtained from patients with a broad range of HBV genotypes and degrees of severity of liver disease. This allowed us to make a preliminary comparison of complexity and conservation of HBV-DNA and RNA quasispecies. However, further studies with a larger sample size are warranted to confirm our results. The main goal of the methodology for HBV-RNA quasispecies described here is to use it in groups of patients where HBV-DNA levels are usually too low for PCR amplification (*e.g.*, patients under NA treatment, since the generation of HBV-RNA is not inhibited by NA directly), in order to be able to monitor HBV quasispecies in those patients.

## REFERENCES

- 1 **World Health Organization.** Hepatitis B. Fact sheet. 2015; No204. [cited 20 March 2021]. In: World Health Organization [Internet]. Available from: <https://www.who.int/en/news-room/fact-sheets/detail/hepatitis-b>
- 2 **Caballero A,** Tabernero D, Buti M, Rodriguez-Frias F. Hepatitis B virus: The challenge of an ancient virus with multiple faces and a remarkable replication strategy. *Antiviral Res* 2018; **158**: 34-44 [PMID: 30059722 DOI: 10.1016/j.antiviral.2018.07.019]
- 3 **Stadlmayer B,** Diederichs A, Chapus F, Rivoire M, Neveu G, Alam A, Fraise L, Carter K, Testoni B, Zoulim F. Full-length 5'RACE identifies all major HBV transcripts in HBV-infected hepatocytes and patient serum. *J Hepatol* 2020; **73**: 40-51 [PMID: 32087349 DOI: 10.1016/j.jhep.2020.01.028]
- 4 **Wang J,** Yu Y, Li G, Shen C, Meng Z, Zheng J, Jia Y, Chen S, Zhang X, Zhu M, Song Z, Wu J, Shao L, Qian P, Mao X, Wang X, Huang Y, Zhao C, Zhang J, Qiu C, Zhang W. Relationship between serum HBV-RNA levels and intrahepatic viral as well as histologic activity markers in entecavir-treated patients. *J Hepatol* 2017 [PMID: 28870671 DOI: 10.1016/j.jhep.2017.08.021]
- 5 **van Bömmel F,** Bartens A, Mysickova A, Hofmann J, Krüger DH, Berg T, Edelmann A. Serum hepatitis B virus RNA levels as an early predictor of hepatitis B envelope antigen seroconversion during treatment with polymerase inhibitors. *Hepatology* 2015; **61**: 66-76 [PMID: 25132147 DOI: 10.1002/hep.27381]
- 6 **Yu XQ,** Wang MJ, Yu DM, Chen PZ, Zhu MY, Huang W, Han Y, Gong QM, Zhang XX. Comparison of Serum Hepatitis B Virus RNA Levels and Quasispecies Evolution Patterns between Entecavir and Pegylated-Interferon Mono-treatment in Chronic Hepatitis B Patients. *J Clin Microbiol* 2020; **58** [PMID: 32554476 DOI: 10.1128/JCM.00075-20]
- 7 **Ji X,** Xia M, Zhou B, Liu S, Liao G, Cai S, Zhang X, Peng J. Serum Hepatitis B Virus RNA Levels Predict HBeAg Seroconversion and Virological Response in Chronic Hepatitis B Patients with High Viral Load Treated with Nucleos(t)ide Analog. *Infect Drug Resist* 2020; **13**: 1881-1888 [PMID: 32606837 DOI: 10.2147/IDR.S252994]
- 8 **Lok AS,** Zoulim F, Dusheiko G, Ghany MG. Hepatitis B cure: From discovery to regulatory approval. *J Hepatol* 2017; **67**: 847-861 [PMID: 28778687 DOI: 10.1016/j.jhep.2017.05.008]
- 9 **Quer J,** Rodríguez-Frias F, Gregori J, Tabernero D, Soria ME, García-Cehic D, Homs M, Bosch A, Pintó RM, Esteban JI, Domingo E, Perales C. Deep sequencing in the management of hepatitis virus

- infections. *Virus Res* 2017; **239**: 115-125 [PMID: 28040474 DOI: 10.1016/j.virusres.2016.12.020]
- 10 **Revill PA**, Tu T, Netter HJ, Yuen LKW, Locarnini SA, Littlejohn M. The evolution and clinical impact of hepatitis B virus genome diversity. *Nat Rev Gastroenterol Hepatol* 2020; **17**: 618-634 [PMID: 32467580 DOI: 10.1038/s41575-020-0296-6]
- 11 **Cao J**, Luo S, Xiong Y. The Variability of Amino Acid Sequences in Hepatitis B Virus. *Virol Sin* 2019; **34**: 42-49 [PMID: 30610573 DOI: 10.1007/s12250-018-0070-x]
- 12 **European Association for the Study of the Liver**. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J Hepatol* 2017; **67**: 370-398 [PMID: 28427875 DOI: 10.1016/j.jhep.2017.03.021]
- 13 **González C**, Tabernero D, Cortese MF, Gregori J, Casillas R, Riveiro-Barciela M, Godoy C, Sopena S, Rando A, Yll M, Lopez-Martinez R, Quer J, Esteban R, Buti M, Rodríguez-Frías F. Detection of hyper-conserved regions in hepatitis B virus X gene potentially useful for gene therapy. *World J Gastroenterol* 2018; **24**: 2095-2107 [PMID: 29785078 DOI: 10.3748/wjg.v24.i19.2095]
- 14 **Cortese MF**, González C, Gregori J, Casillas R, Carioti L, Guerrero-Murillo M, Riveiro-Barciela M, Godoy C, Sopena S, Yll M, Quer J, Rando A, Lopez-Martinez R, Pacin Ruiz B, García-García S, Esteban-Mur R, Tabernero D, Buti M, Rodríguez-Frías F. Sophisticated viral quasispecies with a genotype-related pattern of mutations in the hepatitis B X gene of HBeAg-ve chronically infected patients. *Sci Rep* 2021; **11**: 4215 [PMID: 33603102 DOI: 10.1038/s41598-021-83762-4]
- 15 **Durantal D**, Carrouée-Durantal S, Werle-Lapostolle B, Brunelle MN, Pichoud C, Trépo C, Zoulim F. A new strategy for studying *in vitro* the drug susceptibility of clinical isolates of human hepatitis B virus. *Hepatology* 2004; **40**: 855-864 [PMID: 15382118 DOI: 10.1002/hep.20388]
- 16 **Godoy C**, Tabernero D, Sopena S, Gregori J, Cortese MF, González C, Casillas R, Yll M, Rando A, López-Martínez R, Quer J, González-Aseguinolaza G, Esteban R, Riveiro-Barciela M, Buti M, Rodríguez-Frías F. Characterization of hepatitis B virus X gene quasispecies complexity in mono-infection and hepatitis delta virus superinfection. *World J Gastroenterol* 2019; **25**: 1566-1579 [PMID: 30983817 DOI: 10.3748/wjg.v25.i13.1566]
- 17 **Gregori J**, Soria ME, Gallego I, Guerrero-Murillo M, Esteban JJ, Quer J, Perales C, Domingo E. Rare haplotype load as marker for lethal mutagenesis. *PLoS One* 2018; **13**: e0204877 [PMID: 30281674 DOI: 10.1371/journal.pone.0204877]
- 18 **Ou J**, Wolfe SA, Brodsky MH, Zhu LJ. motifStack for the analysis of transcription factor binding site evolution. *Nat Methods* 2018; **15**: 8-9 [PMID: 29298290 DOI: 10.1038/nmeth.4555]
- 19 **The R Foundation**. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. [cited 13 July 2020]. In: The R Foundation [Internet]. Available from: <https://www.r-project.org/>
- 20 **World Health Organization**. Guidelines for the prevention, care and treatment of persons with chronic hepatitis B infection. [cited 15 April 2021]. In: World Health Organization [Internet]. Available from: <https://www.who.int/hiv/pub/hepatitis/hepatitis-b-guidelines/en/>
- 21 **Cornberg M**, Lok AS, Terrault NA, Zoulim F; 2019 EASL-AASLD HBV Treatment Endpoints Conference Faculty. Guidance for design and endpoints of clinical trials in chronic hepatitis B - Report from the 2019 EASL-AASLD HBV Treatment Endpoints Conference<sup>‡</sup>. *J Hepatol* 2020; **72**: 539-557 [PMID: 31730789 DOI: 10.1016/j.jhep.2019.11.003]
- 22 **Liu S**, Zhou B, Valdes JD, Sun J, Guo H. Serum Hepatitis B Virus RNA: A New Potential Biomarker for Chronic Hepatitis B Virus Infection. *Hepatology* 2019; **69**: 1816-1827 [PMID: 30362148 DOI: 10.1002/hep.30325]
- 23 **Liu Y**, Jiang M, Xue J, Yan H, Liang X. Serum HBV RNA quantification: useful for monitoring natural history of chronic hepatitis B infection. *BMC Gastroenterol* 2019; **19**: 53 [PMID: 30991954 DOI: 10.1186/s12876-019-0966-4]
- 24 **Park SG**, Kim Y, Park E, Ryu HM, Jung G. Fidelity of hepatitis B virus polymerase. *Eur J Biochem* 2003; **270**: 2929-2936 [PMID: 12846825 DOI: 10.1046/j.1432-1033.2003.03650.x]
- 25 **Jansen L**, Kootstra NA, van Dort KA, Takkenberg RB, Reesink HW, Zaaijer HL. Hepatitis B Virus Pregenomic RNA Is Present in Virions in Plasma and Is Associated With a Response to Pegylated Interferon Alfa-2a and Nucleos(t)ide Analogues. *J Infect Dis* 2016; **213**: 224-232 [PMID: 26216905 DOI: 10.1093/infdis/jiv397]
- 26 **Wang J**, Shen T, Huang X, Kumar GR, Chen X, Zeng Z, Zhang R, Chen R, Li T, Zhang T, Yuan Q, Li PC, Huang Q, Colonna R, Jia J, Hou J, McCrae MA, Gao Z, Ren H, Xia N, Zhuang H, Lu F. Serum hepatitis B virus RNA is encapsidated pregenome RNA that may be associated with persistence of viral infection and rebound. *J Hepatol* 2016; **65**: 700-710 [PMID: 27245431 DOI: 10.1016/j.jhep.2016.05.029]
- 27 **Cortese MF**, Garcia-Garcia S, Casillas R, Lopez-Martinez R, Pacin Ruiz B, Sopena S, Tabernero D, Ferrer-Costa RM, Riveiro-Barciela M, Buti M, Rodríguez-Frías F. Gene Silencing by GAPMERS: A New Therapeutic Tool in HBV Infection. *Hepatology* 2020; **72**: 504A-505A [DOI: 10.1002/hep.31579]
- 28 **Cortese MF**, Garcia-Garcia S, Pacin Ruiz B, Casillas R, Tabernero D, Riveiro-Barciela M, Buti M, Rodríguez-Frías F. A combination of gapmers and/or siRNA as potential gene therapy strategy against HBV infection: *in vitro* results. *J Hepatol* 2021; **75**: S294-S803



Published by **Baishideng Publishing Group Inc**  
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA

**Telephone:** +1-925-3991568

**E-mail:** [bpgoffice@wjgnet.com](mailto:bpgoffice@wjgnet.com)

**Help Desk:** <https://www.f6publishing.com/helpdesk>

<https://www.wjgnet.com>

