

# World Journal of *Gastroenterology*

*World J Gastroenterol* 2020 May 28; 26(20): 2464-2681



**OPINION REVIEW**

- 2464 Percutaneous endoscopic gastrostomy – Too often? Too late? Who are the right patients for gastrostomy?  
*Dietrich CG, Schoppmeyer K*
- 2472 Can adiponectin have an additional effect on the regulation of food intake by inducing gastric motor changes?  
*Idrizaj E, Garella R, Squecco R, Baccari MC*

**REVIEW**

- 2479 Diet in neurogenic bowel management: A viewpoint on spinal cord injury  
*Bernardi M, Fedullo AL, Bernardi E, Munzi D, Peluso I, Myers J, Lista FR, Sciarra T*
- 2498 Role of gut microbiota-immunity axis in patients undergoing surgery for colorectal cancer: Focus on short and long-term outcomes  
*Bartolini I, Risaliti M, Ringressi MN, Melli F, Nannini G, Amedei A, Muiesan P, Taddei A*
- 2514 Metabolomics profile in gastrointestinal cancers: Update and future perspectives  
*Nannini G, Meoni G, Amedei A, Tenori L*
- 2533 Nervous mechanisms of restraint water-immersion stress-induced gastric mucosal lesion  
*Zhao DQ, Xue H, Sun HJ*

**MINIREVIEWS**

- 2550 Treatment of gastrointestinal bleeding in left ventricular assist devices: A comprehensive review  
*Vedachalam S, Balasubramanian G, Haas GJ, Krishna SG*
- 2559 Role of  $\gamma\delta$ T cells in liver diseases and its relationship with intestinal microbiota  
*Zhou QH, Wu FT, Pang LT, Zhang TB, Chen Z*

**ORIGINAL ARTICLE****Basic Study**

- 2570 Serum outperforms plasma in small extracellular vesicle microRNA biomarker studies of adenocarcinoma of the esophagus  
*Chiam K, Mayne GC, Wang T, Watson DI, Irvine TS, Bright T, Smith LT, Ball IA, Bowen JM, Keefe DM, Thompson SK, Hussey DJ*
- 2584 Conservation and variability of hepatitis B core at different chronic hepatitis stages  
*Yll M, Cortese MF, Guerrero-Murillo M, Orriols G, Gregori J, Casillas R, González C, Sopena S, Godoy C, Vila M, Taberner D, Quer J, Rando A, Lopez-Martinez R, Esteban R, Riveiro-Barciela M, Buti M, Rodríguez-Frías F*

- 2599 Sleeve gastrectomy ameliorates endothelial function and prevents lung cancer by normalizing endothelin-1 axis in obese and diabetic rats  
*Ruze R, Xiong YC, Li JW, Zhong MW, Xu Q, Yan ZB, Zhu JK, Cheng YG, Hu SY, Zhang GY*

#### Clinical and Translational Research

- 2618 Clinicopathological features of early gastric cancers arising in *Helicobacter pylori* uninfected patients  
*Sato C, Hirasawa K, Tateishi Y, Ozeki Y, Sawada A, Ikeda R, Fukuchi T, Nishio M, Kobayashi R, Makazu M, Kaneko H, Inayama Y, Maeda S*

#### Case Control Study

- 2632 Anhedonia and functional dyspepsia in obese patients: Relationship with binge eating behaviour  
*Santonicola A, Gagliardi M, Asparago G, Carpinelli L, Angrisani L, Iovino P*

#### Retrospective Study

- 2645 Decreased of BAFF-R expression and B cells maturation in patients with hepatitis B virus-related hepatocellular carcinoma  
*Khlaiphuengsin A, Chuaypen N, Sodsai P, Buranapraditkun S, Boonpiyathad T, Hirankarn N, Tangkijvanich P*

#### Prospective Study

- 2657 Role of dynamic perfusion magnetic resonance imaging in patients with local advanced rectal cancer  
*Ippolito D, Drago SG, Pecorelli A, Maino C, Querques G, Mariani I, Franzesi CT, Sironi S*

#### SYSTEMATIC REVIEWS

- 2669 Association between non-alcoholic fatty liver disease and obstructive sleep apnea  
*Umbro I, Fabiani V, Fabiani M, Angelico F, Del Ben M*

**ABOUT COVER**

Editorial board member of *World Journal of Gastroenterology*, Vito Annese, MD, PhD, Associate Professor, Chief Doctor, Department of Medical and Surgical Sciences, University Hospital Careggi, Florence 50139, Italy

**AIMS AND SCOPE**

The primary aim of *World Journal of Gastroenterology* (*WJG*, *World J Gastroenterol*) is to provide scholars and readers from various fields of gastroenterology and hepatology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

*WJG* mainly publishes articles reporting research results and findings obtained in the field of gastroenterology and hepatology and covering a wide range of topics including gastroenterology, hepatology, gastrointestinal endoscopy, gastrointestinal surgery, gastrointestinal oncology, and pediatric gastroenterology.

**INDEXING/ABSTRACTING**

The *WJG* is now indexed in Current Contents®/Clinical Medicine, Science Citation Index Expanded (also known as SciSearch®), Journal Citation Reports®, Index Medicus, MEDLINE, PubMed, PubMed Central, and Scopus. The 2019 edition of Journal Citation Report® cites the 2018 impact factor for *WJG* as 3.411 (5-year impact factor: 3.579), ranking *WJG* as 35<sup>th</sup> among 84 journals in gastroenterology and hepatology (quartile in category Q2). CiteScore (2018): 3.43.

**RESPONSIBLE EDITORS FOR THIS ISSUE**

Responsible Electronic Editor: *Yan-Liang Zhang*  
 Proofing Production Department Director: *Yun-Xiaojuan Wu*  
 Responsible Editorial Office Director: *Ze-Mao Gong*

<b>NAME OF JOURNAL</b> <i>World Journal of Gastroenterology</i>
<b>ISSN</b> ISSN 1007-9327 (print) ISSN 2219-2840 (online)
<b>LAUNCH DATE</b> October 1, 1995
<b>FREQUENCY</b> Weekly
<b>EDITORS-IN-CHIEF</b> Subrata Ghosh, Andrzej S Tarnawski
<b>EDITORIAL BOARD MEMBERS</b> <a href="http://www.wjgnet.com/1007-9327/editorialboard.htm">http://www.wjgnet.com/1007-9327/editorialboard.htm</a>
<b>PUBLICATION DATE</b> May 28, 2020
<b>COPYRIGHT</b> © 2020 Baishideng Publishing Group Inc

<b>INSTRUCTIONS TO AUTHORS</b> <a href="https://www.wjgnet.com/bpg/gerinfo/204">https://www.wjgnet.com/bpg/gerinfo/204</a>
<b>GUIDELINES FOR ETHICS DOCUMENTS</b> <a href="https://www.wjgnet.com/bpg/GerInfo/287">https://www.wjgnet.com/bpg/GerInfo/287</a>
<b>GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH</b> <a href="https://www.wjgnet.com/bpg/gerinfo/240">https://www.wjgnet.com/bpg/gerinfo/240</a>
<b>PUBLICATION ETHICS</b> <a href="https://www.wjgnet.com/bpg/GerInfo/288">https://www.wjgnet.com/bpg/GerInfo/288</a>
<b>PUBLICATION MISCONDUCT</b> <a href="https://www.wjgnet.com/bpg/gerinfo/208">https://www.wjgnet.com/bpg/gerinfo/208</a>
<b>ARTICLE PROCESSING CHARGE</b> <a href="https://www.wjgnet.com/bpg/gerinfo/242">https://www.wjgnet.com/bpg/gerinfo/242</a>
<b>STEPS FOR SUBMITTING MANUSCRIPTS</b> <a href="https://www.wjgnet.com/bpg/GerInfo/239">https://www.wjgnet.com/bpg/GerInfo/239</a>
<b>ONLINE SUBMISSION</b> <a href="https://www.f6publishing.com">https://www.f6publishing.com</a>

## Basic Study

**Conservation and variability of hepatitis B core at different chronic hepatitis stages**

Marçal Yll, Maria Francesca Cortese, Mercedes Guerrero-Murillo, Gerard Orriols, Josep Gregori, Rosario Casillas, Carolina González, Sara Sopena, Cristina Godoy, Marta Vila, David Taberner, Josep Quer, Ariadna Rando, Rosa Lopez-Martinez, Rafael Esteban, Mar Riveiro-Barciela, Maria Buti, Francisco Rodríguez-Frías

**ORCID number:** Marçal Yll (0000-0002-7030-3360); Maria Francesca Cortese (0000-0002-4318-532X); Mercedes Guerrero-Murillo (0000-0002-5556-2460); Gerard Orriols (0000-0002-1138-5909); Josep Gregori (0000-0002-4253-8015); Rosario Casillas (0000-0002-6758-6734); Carolina González (0000-0002-0169-5874); Sara Sopena (0000-0002-3309-5486); Cristina Godoy (0000-0001-5037-1916); Marta Vila (0000-0001-9303-5189); David Taberner (0000-0002-1146-4084); Josep Quer (0000-0003-0014-084X); Ariadna Rando (0000-0003-4555-7286); Rosa López-Martinez (0000-0002-8450-6986); Rafael Esteban (0000-0001-5280-392X); Mar Riveiro-Barciela (0000-0001-9309-2052); Maria Buti (0000-0002-0732-3078); Francisco Rodríguez-Frías (0000-0001-9058-4641).

**Author contributions:** Rodríguez-Frías F designed the research; Cortese MF coordinated the research; Yll M and Cortese MF equally contributed to design the experiments; Yll M, Orriols G, Godoy C, Sopena S, Casillas R, González C, Vila M and Rando A performed the experiments; Yll M, Cortese MF, Gregori J and Guerrero-Murillo M analyzed data acquired during the experiments and interpreted the results; Yll M and Cortese MF drafted the manuscript; Cortese MF, Taberner D, Lopez-Martinez R, Riveiro-

Marçal Yll, Maria Francesca Cortese, Gerard Orriols, Rosario Casillas, Carolina González, Sara Sopena, Cristina Godoy, Marta Vila, David Taberner, Ariadna Rando, Rosa Lopez-Martinez, Francisco Rodríguez-Frías, Liver Pathology Unit, Departments of Biochemistry and Microbiology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona 08035, Spain

Marçal Yll, Maria Francesca Cortese, Mercedes Guerrero-Murillo, Josep Gregori, Rosario Casillas, Sara Sopena, Marta Vila, Josep Quer, Francisco Rodríguez-Frías, Liver Unit, Liver Disease Laboratory-Viral Hepatitis, Vall d'Hebron Institut Recerca-Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona 08035, Spain

Mercedes Guerrero-Murillo, Department of Microbiology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona 08035, Spain

José Gregori, Cristina Godoy, David Taberner, 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

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:** Maria Francesca Cortese, PhD, Research Scientist, Liver Pathology Unit, Departments of Biochemistry and Microbiology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Passeig Vall d'Hebron 119-129, Barcelona 08035, Spain. [maria.cortese@vhir.org](mailto:maria.cortese@vhir.org)

**Abstract****BACKGROUND**

Since it is currently not possible to eradicate hepatitis B virus (HBV) infection with existing treatments, research continues to uncover new therapeutic strategies. HBV core protein, encoded by the HBV core gene (*HBC*), intervenes in both structural and functional processes, and is a key protein in the HBV life cycle. For this reason, both the protein and the gene could be valuable targets for new therapeutic and diagnostic strategies. Moreover, alterations in the protein sequence could serve as potential markers of disease progression.

Barciela M, Buti M, Quer J, Esteban R and Rodríguez-Frías F critically reviewed the manuscript.

**Supported by** the Instituto de Salud Carlos III, Spain, the European Regional Development Fund, No. PI18/01436.

#### Institutional review board

**statement:** The study was reviewed and approved by the Clinical Research Ethics Committee of Hospital Universitari Vall d'Hebron.

#### Conflict-of-interest statement:

Josep Gregori is an employee of Roche Diagnostics, SL.

#### Data sharing statement:

Next-generation sequencing data were submitted to the GenBank SRA database (BioProject accession number PRJNA625435).

#### ARRIVE guidelines statement:

The authors have read the ARRIVE guidelines, and the manuscript was prepared and revised according to the ARRIVE guidelines.

#### 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/>

**Manuscript source:** Invited manuscript

**Received:** February 28, 2020

**Peer-review started:** February 28, 2020

**First decision:** April 9, 2020

**Revised:** May 8, 2020

**Accepted:** May 19, 2020

**Article in press:** May 19, 2020

**Published online:** May 28, 2020

**P-Reviewer:** Said ZNA

**S-Editor:** Zhang H

**L-Editor:** A

**E-Editor:** Ma YJ



## AIM

To detect, by next-generation sequencing, *HBC* hyper-conserved regions that could potentially be prognostic factors and targets for new therapies.

## METHODS

Thirty-eight of 45 patients with chronic HBV initially selected were included and grouped according to liver disease stage [chronic hepatitis B infection without liver damage (CHB,  $n = 16$ ), liver cirrhosis (LC,  $n = 5$ ), and hepatocellular carcinoma (HCC,  $n = 17$ )]. HBV DNA was extracted from patients' plasma. A region between nucleotide (nt) 1863 and 2483, which includes *HBC*, was amplified and analyzed by next-generation sequencing (Illumina MiSeq platform). Sequences were genotyped by distance-based discriminant analysis. General and intergroup nt and amino acid (aa) conservation was determined by sliding window analysis. The presence of nt insertion and deletions and/or aa substitutions in the different groups was determined by aligning the sequences with genotype-specific consensus sequences.

## RESULTS

Three nt (nt 1900-1929, 2249-2284, 2364-2398) and 2 aa (aa 117-120, 159-167) hyper-conserved regions were shared by all the clinical groups. All groups showed a similar pattern of conservation, except for five nt regions (nt 1946-1992, 2060-2095, 2145-2175, 2230-2250, 2270-2293) and one aa region (aa 140-160), where CHB and LC, respectively, were less conserved ( $P < 0.05$ ). Some group-specific conserved regions were also observed at both nt (2306-2334 in CHB and 1935-1976 and 2402-2435 in LC) and aa (between aa 98-103 in CHB and 28-30 and 51-54 in LC) levels. No differences in insertion and deletions frequencies were observed. An aa substitution (P79Q) was observed in the HCC group with a median (interquartile range) frequency of 15.82 (0-78.88) *vs* 0 (0-0) in the other groups ( $P < 0.05$  *vs* CHB group).

## CONCLUSION

The differentially conserved *HBC* and HBV core protein regions and the P79Q substitution could be involved in disease progression. The hyper-conserved regions detected could be targets for future therapeutic and diagnostic strategies.

**Key words:** Hepatitis B virus; Hepatitis B core gene; Next-generation sequencing; Genetic conservation; Amino acid substitution; Gene therapy; Small interfering RNA

©The Author(s) 2020. Published by Baishideng Publishing Group Inc. All rights reserved.

**Core tip:** New tools for hepatitis B virus infection treatment and follow-up are needed. Hepatitis B virus core protein has a key role in viral replication and persistence. Analysis of viral quasispecies by next-generation sequencing can identify conserved regions in viral genes or proteins that may serve as targets for new therapeutic and diagnostic strategies. Moreover, it may help identify prognostic markers of liver disease progression. Here, we detected hyper-conserved nucleotide and amino acid regions regardless of the clinical stage. Moreover, we observed several group-specific conserved and variable regions and an amino acid substitution that could be indicative of different disease progression.

**Citation:** Yll M, Cortese MF, Guerrero-Murillo M, Orriols G, Gregori J, Casillas R, González C, Sopena S, Godoy C, Vila M, Tabernero D, Quer J, Rando A, Lopez-Martinez R, Esteban R, Riveiro-Barciela M, Buti M, Rodríguez-Frías F. Conservation and variability of hepatitis B core at different chronic hepatitis stages. *World J Gastroenterol* 2020; 26(20): 2584-2598

**URL:** <https://www.wjgnet.com/1007-9327/full/v26/i20/2584.htm>

**DOI:** <https://dx.doi.org/10.3748/wjg.v26.i20.2584>

## INTRODUCTION

Hepatitis B virus (HBV) is a small virus with a specific tropism for the liver. It belongs

to the *Hepadnaviridae* family. Despite the existence of effective preventive vaccines, an estimated 257 million people worldwide live with chronic HBV infection and more than 880000 people die every year of HBV-related complications such as liver cirrhosis (LC) and hepatocellular carcinoma (HCC)<sup>[1]</sup>.

HBV is an enveloped virus equipped with 3.2 kb of partially double-stranded circular DNA produced by the reverse transcription of an RNA intermediate known as pregenomic RNA<sup>[2]</sup>. This ribonucleic intermediate is produced from a viral DNA molecule that interacts with cellular (histone and non-histone) and viral proteins, forming a “mini-chromosome” known as covalently closed circular DNA (cccDNA) that remains in hepatocyte nuclei for the rest of the cell’s life<sup>[3]</sup>. Although current antiviral therapy can control viral replication, it is not capable of interfering with the formation or persistence of cccDNA, rendering HBV infection eradication impossible. This mini-chromosome could even be a source of HBV reactivation after clinical resolution and HBsAg seroclearance<sup>[4]</sup>. Due to persistent infection, up to 1% of Caucasian patients with noncirrhotic chronic HBV infection have been found to develop HCC<sup>[5]</sup>.

Gene therapy has emerged as one of the most promising strategies for blocking disease progression, and results from studies investigating the potential of small interfering RNA (siRNA) systems as adjuvant therapy are encouraging<sup>[6]</sup>. siRNA is a double-stranded noncoding RNA [with an optimal length of 21 nucleotides (nt)] that interacts with target messenger RNA, promoting its degradation and silencing of the gene<sup>[7]</sup>.

HBV reverse transcriptase lacks 3' to 5' proofreading activity, which leads to viral genome variability comparable to that observed in an RNA virus<sup>[8]</sup>. This genetic variability is further increased by inter- and intra-genotype recombination events<sup>[9]</sup>. In short, HBV circulates as a complex mixture of closely related genetic variants (haplotypes) known as quasispecies<sup>[10]</sup>.

The HBV core protein (HBc) [encoded by the HBV core gene (*HBC*) from the PreCore/Core open reading frame (ORF)] is essential for viral replication. It is a structural 21-kDa protein that self-assembles to create dimers that assemble in hexamers forming the icosahedral viral capsid<sup>[11,12]</sup>. It has 183 amino acids (aa) (185 for genotype A) with a N-terminal domain and a C-terminal domain (CTD) connected through a linker region. The N-terminal domain ranges from aa position 1 to 149 (including the linker region aa 140 to 149) and constitutes the  $\alpha$  helix-rich assembly domain<sup>[13]</sup>. The CTD is shorter (aa 150 to 183, or 185 for genotype A) and constitutes the functional domain<sup>[13]</sup>. The CTD allows HBc to intervene in a multitude of processes such as subcellular traffic, viral genome release, capsid assembly and transport, RNA metabolism, and viral pregenomic RNA reverse transcription<sup>[14]</sup>. Considering just how essential this protein is for viral replication, it could be an optimal target for gene therapy. Moreover, mutations in HBc may have different roles in liver disease progression, positioning them as potentially useful prognostic genetic markers.

Next-generation sequencing (NGS) is a highly sensitive technique for studying viral quasispecies; it is capable of detecting highly conserved regions of the HBV genome, regardless of genome or clinical stage<sup>[9]</sup>. Moreover, it supports the identification and quantitative determination<sup>[15]</sup> of specific variants that could be used as markers to predict prognosis and treatment response in patients with HBV infection.

The aim of this study was to apply NGS to analyse HBc conservation and variability at the nt and aa levels in patients with different stages of chronic HBV infection in order to identify hyper-conserved regions of the *HBC* gene that could be a target for gene therapy and to determine possible prognostic factors of disease progression

## MATERIALS AND METHODS

### Patients and samples

The study was reviewed and approved by the Clinical Research Ethics Committee of Hospital Universitari Vall d’Hebron (PR(AG)146/2020). No animals were used.

Forty-five patients with chronic HBV infection were recruited from members of the general population seen at the outpatient clinic at Vall d’Hebron University Hospital in Barcelona, Spain. They tested negative for hepatitis D virus, hepatitis C virus, and human immunodeficiency virus, and had a viral load > 3 log IU/mL, which is the limit of polymerase chain reaction (PCR) amplification sensitivity. HBV serological markers such as the surface antigen (HBsAg), the e antigen (HBeAg), and anti-HBe antibodies were tested using commercial chemiluminescent assays on a COBAS 8000 analyzer (Roche Diagnostics, Rotkreuz, Switzerland). HBV DNA was quantified by

real-time PCR with a detection limit of 10 IU/mL (COBAS 6800, Roche Diagnostics). Patients were divided into 3 clinical groups according to liver disease stage determined by biopsy or diagnostic imaging in line with the EASL guidelines<sup>[16]</sup>: Chronic HBV infection without liver damage (CHB group), chronic HBV infection with liver cirrhosis (LC group), and chronic HBV infection with hepatocellular carcinoma (HCC group).

### **HBC gene amplification and NGS**

HBV DNA was extracted from 200 µL of serum using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. The region of interest was amplified through a 3-step nested PCR protocol (Figure 1). The first step (PCR1) covered a large region between nt 1774-2930 that includes the *HBC* gene (nt 1901-2464 for genotype A and 1901-2458 for other genotypes). As the Illumina MiSeq platform (Illumina, San Diego, CA, United States) allows read lengths of up to 600 bp, the following amplification steps were performed by dividing *HBC* into 2 amplicons (amplicon 1 = nt 1863-2317 and amplicon 2 = nt 2205-2483), which overlapped in a 112 nt-long portion (PCR2). The M13-tail, added in step 2, was used for the last step (PCR3), which introduced a 10 nt-long sample-specific multiplex identifier. All the PCR steps were performed using high-fidelity Pfu Ultra II DNA polymerase (Stratagene, Agilent Technologies, Santa Clara, CA, United States). The primers and protocols are reported in Table 1.

The final PCR products were purified with Agencourt AMPure XP magnetic beads (Beckman Coulter, Beverly, LA, United States) and their quality verified using the Agilent 2200 TapeStation System and D1000 ScreenTape kit (Agilent Technologies, Waldbronn, Germany).

Purified amplicons were quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific-Life Technologies, Austin, TX, United States) and pooled to guarantee that the 2 amplicons for each patient were adequately represented in the analysis (2.5x for amplicon 1 and 1x for amplicon 2, due to their different lengths). The amplicon pool was sequenced by NGS on the Illumina MiSeq platform.

The reads obtained underwent an in-house bioinformatics filtering procedure based on R scripts<sup>[17]</sup>, as previously described by our group<sup>[18]</sup>. For each amplicon, a group of unique sequences (haplotypes) forming the viral quasispecies was obtained. All sequences that did not match in the overlapping 112-nt region between amplicon 1 and 2 were discarded.

The bioinformatics methods used in this study were reviewed by Mercedes Guerrero-Murillo from the Microbiology Department at Vall d'Hebron Hospital (Barcelona, Spain) and by Dr. Josep Gregori from the Liver Disease Viral Hepatitis Laboratory at Vall d'Hebron Hospital (Barcelona, Spain), CIBERehd research group, and Roche Diagnostics SL.

### **Genotyping of the haplotypes**

The amplicons from each patient were aligned with the same region of the respective amplicons extracted from 106 full-length HBV genome sequences representative of genotypes A to J obtained from the NCBI GenBank (Supplementary Table 1). Genotyping was conducted by applying distance-based discriminant analysis (DB rule)<sup>[19,20]</sup>, which considers the inter- and intra-class variability of all genotypes. Genetic distances were computed according to the Kimura-80 model<sup>[21]</sup>.

### **Conservation and mutation analysis**

Sequence conservation at nt and aa levels was determined by calculating the information content (IC) of each position in a multiple alignment of all haplotypes detected with a frequency > 0.25.

This analysis calculates the mean IC for windows of 25 nt (or 10 aa), starting from the first position in the multiple alignment and moving forward in steps of 1<sup>[9]</sup>. The hyper-conserved regions were detected by aligning all haplotypes, regardless of clinical stage. Differences in sequence conservation between the groups were determined by comparing IC values.

To identify specific nt insertions and deletions (indels) and aa substitutions that could discriminate between the groups, haplotype sequences were aligned with their genotype-specific consensus sequence. Consensus was obtained by aligning the sequences of the subgenotypes of interest extracted from the 106 full-length HBV genome sequences. Polymorphisms were identified by aligning haplotype sequences with a population consensus sequence and discarded.

### **Statistical analysis**

Sequence conservation differences between the groups in the sliding windows were analysed using the Wilcoxon-Mann-Whitney test. Frequencies of aa changes detected

**Table 1** Primer design and polymerase chain reaction protocols for each amplified region

	PCR	Primer	Primer sequence (5'->3')	Amplified region	Protocol
1 <sup>st</sup> step	PCR1	Forward	TAGGAGGCTGTAGGC	1774-2930	95 °C 5 min; (95°C 20 s, 49 °C 20 s, 72 °C 15 s) × 35 cycles; 72 °C 3 min
		Reverse	GGAAAGAATCCCAGAGG		
2 <sup>nd</sup> step	PCR2 A.1	Forward A.1	<b>GTTGTAAAACGACG</b> <b>GCCAGTTTCAAGCCT</b> CCAAGCTGT	1863-2317	95 °C 2 min; (95 °C 20 s, 58 °C 20 s, 72 °C 15s) × 35 cycles; 72 °C 3 min
		Reverse A.1	<b>CACAGGAAACAGCT</b> ATGACCGATAGGGG CATTGGTGGTCT		
	PCR2 A.2	Forward1 A.2	<b>GTTGTAAAACGACG</b> <b>GCCAGTGGTTTCATA</b> TTCTTGCC	2205-2483	95 °C 2 min; (95 °C 20 s, 50 °C 20 s, 72 °C 15 s) × 35 cycles; 72 °C 3 min
		Forward2 A.2	<b>GTTGTAAAACGACG</b> <b>GCCAGTGGTTTCACA</b> TTTCCTGIC		
		Forward3 A.2	<b>GTTGTAAAACGACG</b> <b>GCCAGTGGTTTCACA</b> TTTCCTGIC		
		Forward4 A.2	<b>GTTGTAAAACGACG</b> <b>GCCAGTGGTTTCACA</b> TTTCCTGCC		
		Reverse A.2	<b>CACAGGAAACAGCT</b> ATGACCTCCACCTT ATGAGTCCAAG		
3 <sup>rd</sup> step	PCR3	Forward (specific per sample)	<b>GTTGTAAAACGACG</b> <b>GCCAGT</b> +specific 10 nt MID		95 °C 2 min; (95 °C 20 s, 60 °C 20 s, 72 °C 15 s) × 20 cycles; 72 °C 3 min
		Reverse (specific per sample)	<b>CACAGGAAACAGCT</b> <b>ATGACC</b> +specific 10 nt MID		

Bold nucleotides indicate the M13 sequence. Forward primers in PCR2-A2 were multiplexed at the same concentration to cover all HBV genotypes. The protocols of amplification are reported. A.1: Amplicon 1; A.2: Amplicon 2; PCR: Polymerase chain reaction; MID: Multiplex identifier.

were compared with the Kruskal-Wallis test and described as median and interquartile range (IQR). All analyses were performed in R version 3.2.3.  $P < 0.05$  was considered significant.

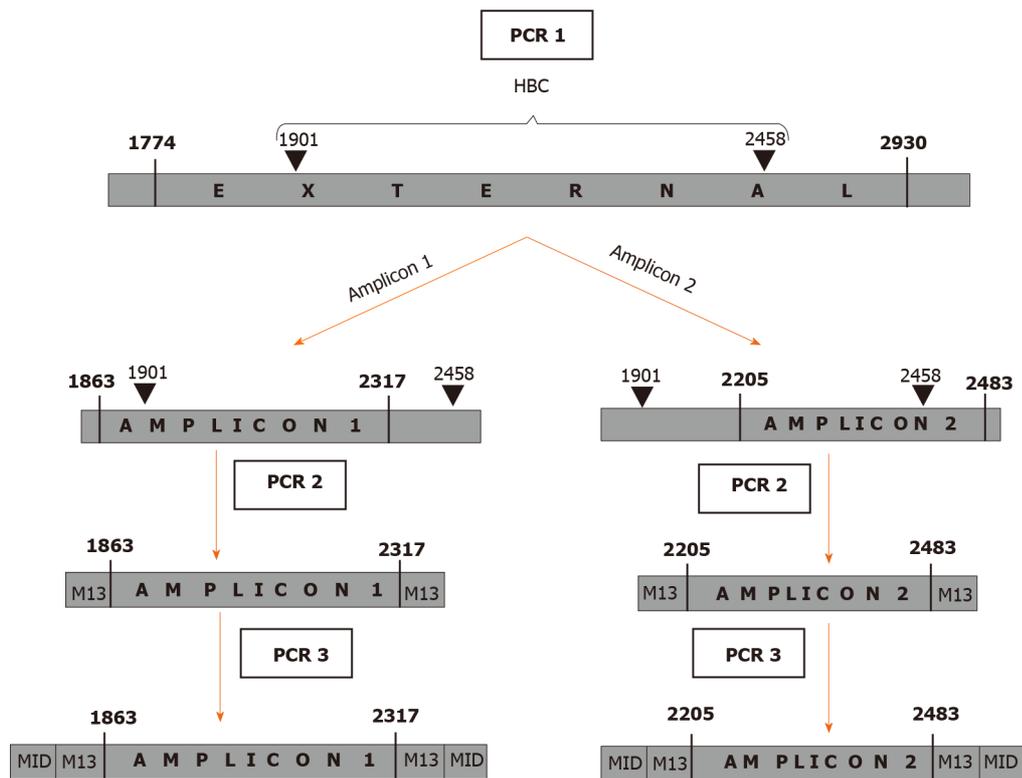
## RESULTS

### *Patients characteristics and NGS results*

Of the 45 patients with chronic hepatitis initially included in the study, 38 passed the sequencing quality filters and had correctly overlapping amplicons 1 and 2. After application of the quality filters, a median (IQR) of 133156.5 (85961.25-605212) and 66571 (25958.5-2301225) sequences per patient were obtained respectively for amplicon 1 and amplicon 2. NGS data were submitted to the GenBank SRA database (BioProject accession number PRJNA625435; BioSample accession numbers are reported in [Supplementary Table 2](#)). In the clinical groups, there were 16 patients with CHB, 5 with LC, and 17 with HCC. The clinical and viral characteristics (including genotypes) are reported in [Table 2](#).

### *Sequence conservation at the nt level*

Sequence conservation was studied by applying a sliding window analysis to the entire *HBC* sequence overlapping the 2 amplicons at the common 112 nt-long portion. No differences in IC were observed on analyzing the sequences by haplotype considering or not their relative frequency ([Figure 2A](#)). Considering the IC of all the nt-sequence haplotypes obtained (regardless of clinical group), we identified 3 hyper-conserved regions (nt 1900-1929, 2249-2284, and 2364-2398, [Figure 2B](#)). Most of the nt positions within these regions yielded the maximum IC value of 2 bits (100% conservation).



**Figure 1 Schematic summary of the 3 amplification steps.** In the first amplification step (PCR1), a large region was amplified. In the following step (PCR2), the region was divided into 2 amplicons that overlapped in a 112 nucleotide-long portion. In the third step (PCR 3) a sample identifier (MID) was added. PCR: Polymerase chain reaction; MID: Multiplex identifier.

On comparing the IC of each clinical group by haplotype, the HCC and LC groups showed similar conservation patterns; CHB was notably associated with the lowest level of conservation, mainly evident in 5 regions: nt 1946-1992, 2060-2095, 2145-2175, 2230-2250, and 2270-2293 ( $P < 0.05$ , **Figure 3A**). Three group-specific conserved regions were detected: 1 in the CHB group (nt 2306-2334) and 2 in the LC group (nt 1935-1976 and 2402-2435; **Figure 3B**). Most of the nt positions within these regions yielded the maximum IC value of 2 bits (100% conservation).

### Sequence conservation at the aa level

The aa sequences of the haplotypes were translated from their respective nt sequences using the *HBC* reading frame.

Sliding window analysis of the aa haplotypes of the 38 patients by haplotype and haplotype frequency (**Figure 4A**) showed that the HBc protein was highly conserved throughout its sequence except for the central region (between aa 50 and 100), where conservation was slightly decreased. Two common hyper-conserved regions were detected: 1 between aa 117-120 and 1 between aa 159-167 (**Figure 4B**). All the aa in these regions had a conservation of around 100% (4.32 bits).

On analyzing aa conservation by haplotype in relation to clinical stage, the 3 groups showed a similar pattern, except for a region between aa 140 and 160, which was less conserved in the LC group compared with the CHB and HCC groups ( $P < 0.05$ , **Figure 5A**). Again, 3 group-specific conserved aa regions were detected: 1 in the CHB group (aa 98-103) and 2 in the LC group (aa 28-30 and 51-54, **Figure 5B**). All the aa in these regions had a conservation of around 100% (4.32 bits).

### nt indels and aa changes

nt indels and aa changes were identified by aligning the patients' haplotypes with their genotype-specific consensus sequence.

In the CHB group, 8/16 patients had indels in *HBC*, vs 2/17 in the HCC group and 1/5 in the LC group. The indels consisted of the insertion or deletion of one nt at positions 1951 or 2085 (a thymine in 1951 and a guanine in 2085; **Table 3**). In all cases, a truncated HBc protein was produced. However, due to the limited number of patients, no statistical differences were observed on comparing the frequencies between the groups.

On analysing the presence of aa changes, we identified the aa substitution P79Q

**Table 2** Main clinical and viral characteristics of hepatitis B virus-infected patients enrolled in the study

Median [IQR]	CHB (n = 16)	HCC (n = 17)	LC (n = 5)	P value
Age	38.5 [33.5-46.5]	67 [58-69]	56 [48-66]	0.002
Viral load (log IU/mL)	6.8 [5.7-8.0]	5.5 [4.7-6.7]	5.7 [4.8-6.2]	NS
ALT	56.5 [41.25-180.5]	70 [47-212]	46 [43-79]	NS
AST	56 [34.75-124]	120.5 [59-163.5]	66.45 [48.675-84.225]	NS
Platelets (10 <sup>9</sup> /L)	183 [161.5-226]	136 [98.5-255]	81.5 [61.25-101.75]	NS
Proportion				
Gender (male)	11/16	15/17	3/5	
HBeAg (positive)	8/16	3/17	0/5	
Genotype, % (n)				
A	18.8 (3)	5.9 (1)	20.0 (1)	
C	37.5 (6)	5.9 (1)	20.0 (1)	
D	25.0 (4)	64.7 (11)	40.0 (2)	
D/A	(0)	5.9 (1)	0.0 (0)	
D/E	6.3 (1)	11.8 (2)	0.0 (0)	
E	6.3 (1)	0.0 (0)	20.0 (1)	
F	6.3 (1)	5.9 (1)	0.0 (0)	

D/E and D/A indicate mixtures of the 2 genotypes. The frequency of each genotype within the clinical groups is reported as percentage (%) and number of patients (n). CHB: Chronic hepatitis B infection without liver damage; HCC: Hepatocellular carcinoma; LC: Liver cirrhosis; ALT: Alanine aminotransferase (normal value < 40 IU/mL); AST: Aspartate aminotransferase (normal value < 40 IU/mL); IQR: Interquartile range; NS: No-statistical P value.

(proline to glutamine) in the HCC group with a median (IQR) frequency of 15.82 (0-78.9) *vs* (0-0) in the CHB group ( $P < 0.05$ ) and 0 (0-0) in the LC group (Figure 6).

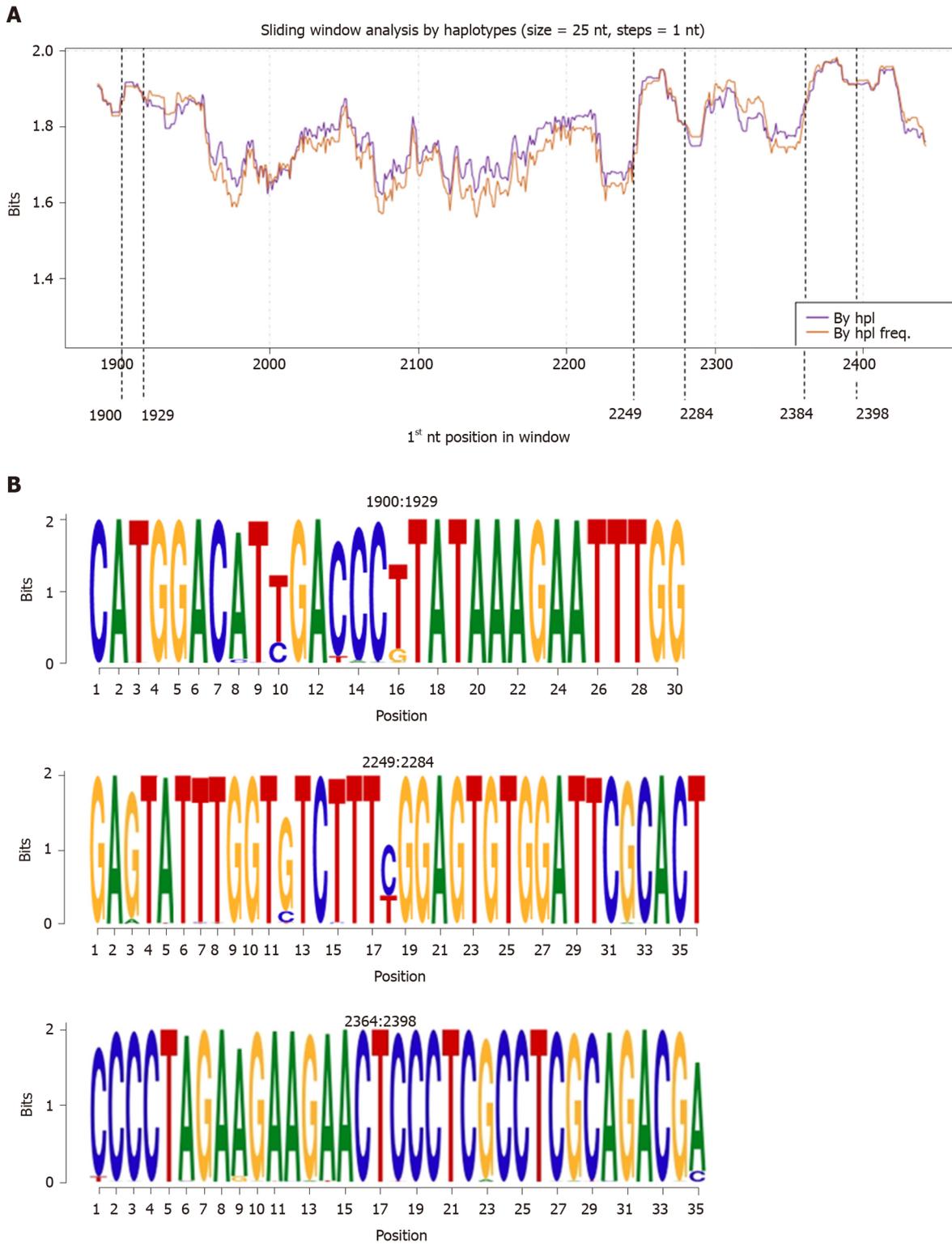
## DISCUSSION

The HBc protein, encoded by the *HBC* gene, is a key element in viral replication and disease progression and is involved in both structural and functional processes. Studying gene and protein sequences in patients with different clinical stages of HBV infection could provide important information on the pathogenic role of this protein. Moreover, the identification of hyper-conserved regions at both nt and aa levels could help develop new therapeutic approaches, including gene therapy. In this study, we used NGS to analyse *HBC* quasispecies in a group of patients with chronic HBV infection stratified by liver disease stage.

First, we studied quasispecies conservation to search for hyper-conserved nt and aa regions regardless of clinical stage or viral genotype. Current treatment based on nucleos(t)ide inhibitors does not affect cccDNA levels or transcriptional activity and therefore cannot eliminate HBV infection. This viral mini-chromosome supports the continuous expression of viral antigens that possibly contribute to disease progression, even in the presence of drug-induced viral suppression<sup>[22]</sup>.

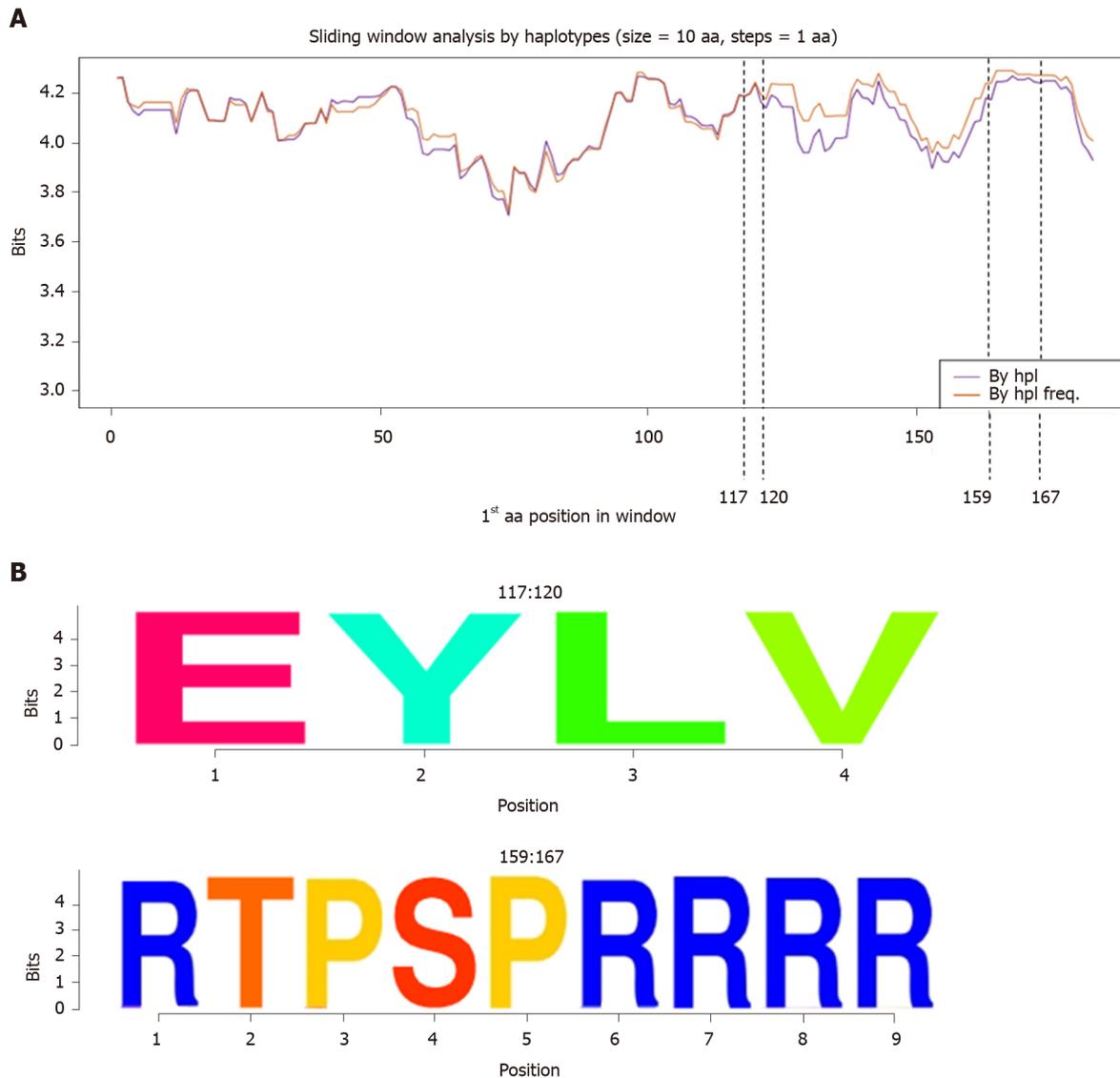
New therapeutic approaches are thus required to control HBV expression, and the targeted delivery of siRNA is one of the most promising approaches under investigation<sup>[23]</sup>. Several siRNAs are currently being tested against X and S ORFs. A study conducted in chimpanzees showed that multiple injections of ARB-1467 (a mixture of 3 interfering RNAs targeting both X and S ORFs<sup>[24]</sup>) led to a 90% reduction in HBsAg levels and a 50% reduction in cccDNA within 28 d of treatment<sup>[25]</sup>. None of the molecules currently available, however, target *HBC*, which considering its role in viral replication could be a valuable target for siRNA-based therapies.

In this study, we analysed quasispecies conservation of the entire *HBC* gene in patients infected by different HBV genotypes and with different clinical stages of disease in order to identify hyper-conserved regions that might be useful for pangenotypic and panclinical RNA silencing strategies. On analyzing nt conservation for the group of 38 patients, we detected 3 shared hyper-conserved regions, namely the start codon of *HBC* expression (nt 1900-1929), a portion with 2 CD8 epitopes (HLA-A24 and A3303) (nt 2249-2284)<sup>[26]</sup>, and an arginine-rich portion of the CTD (nt 2364-2398). All 3 sequences could be valuable targets for a new gene silencing



**Figure 2 Information content analysis at nucleotide level.** A: Sliding window analysis of Hepatitis B core gene performed by aligning the quasispecies haplotypes for all 38 patients with and without considering their relative frequency. Each point on the graph represents the mean information content (in bits) of the 25-nucleotides windows, with forward displacement of 1 nucleotide step between windows. The purple line shows the analysis by haplotype (By hpl), which is the mean information content obtained from the multiple alignments of all quasispecies haplotypes. The orange line represents the analysis by haplotype frequency (By hpl freq), which is the mean information content from the multiple alignments of all the patients' quasispecies haplotypes considering their relative frequency. The dashed lines indicate the 3 common hyper-conserved regions observed, with reporting of their positions. B: Representation of detected hyper-conserved regions as sequence logos (with reporting of nucleotide positions). The relative sizes of the letters in each stack indicate their relative frequencies at each position within the multiple alignments of nucleotide haplotypes. The total height of each stack of letters depicts the information content of each nucleotide position, measured in bits (Y-axis): from minimum (0) to maximum conservation (2). By hpl: Analysis by haplotype; By hpl freq: Analysis by haplotype frequency; nt: Nucleotide.



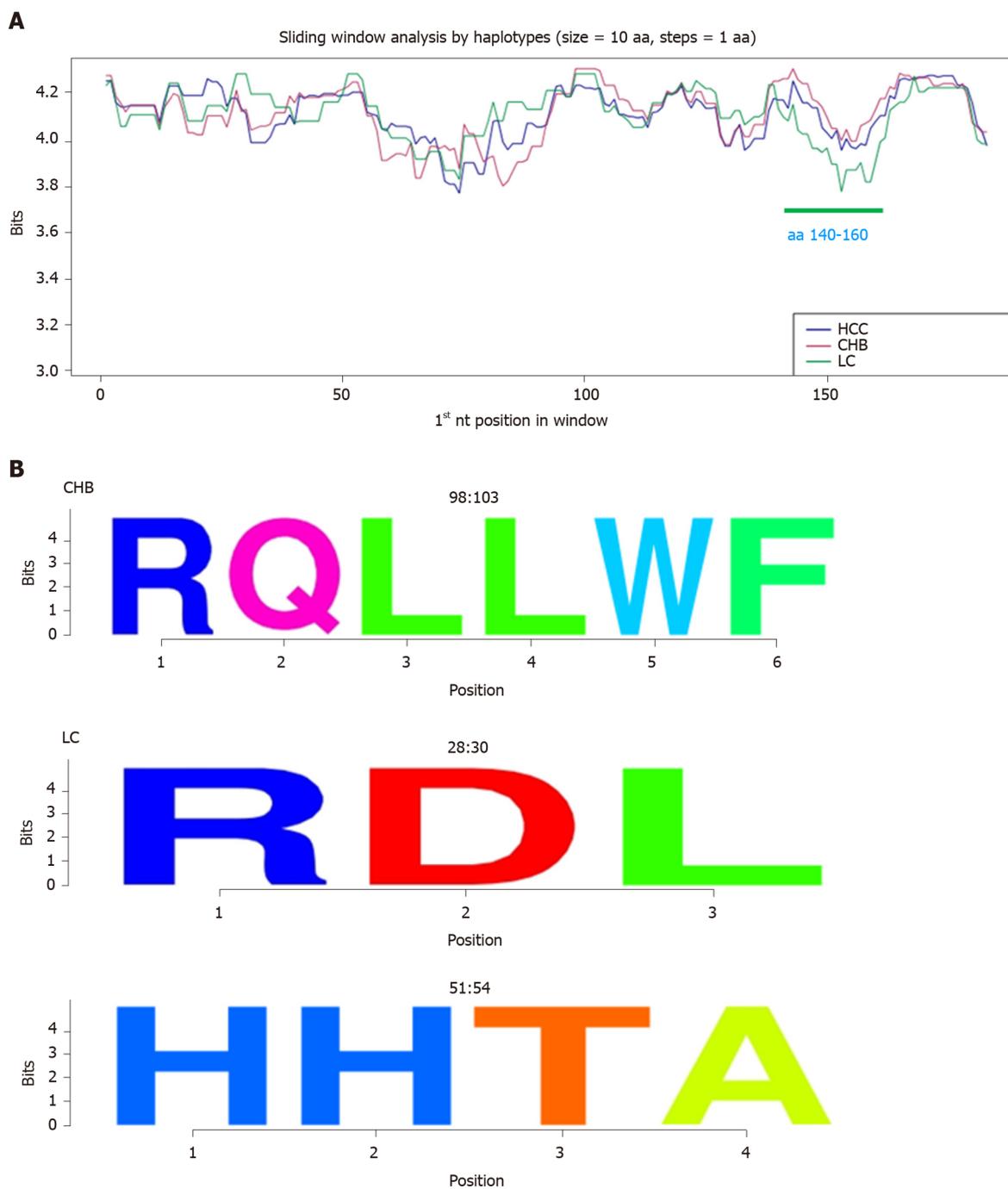


**Figure 4 Information content analysis at amino acid level.** A: Sliding window analysis of the Hepatitis B core protein sequence for all 38 patients with and without consideration of relative frequency. Each point on the graph is the result of the mean information content (in bits) of the 10-amino acid in size windows, with forward displacement between them of 1 amino acid step. The purple line represents the information content of all the quasispecies haplotypes (By hpl) whereas the orange line indicates the information content considering haplotype frequency (By hpl freq.). The dashed lines show the 2 common amino acid hyper-conserved regions observed, with reporting of their positions. B: Representation of amino acid hyper-conserved regions detected as sequence logos (with reporting of amino acid positions). The relative sizes of the letters in each stack indicate their relative frequencies at each position within the multiple alignments of amino acid haplotypes. The total height of each stack depicts the information content of each amino acid position, measured in bits (Y-axis); range: 0 bits (0% conservation) to 4.32 bits (100% conservation). By hpl: Analysis by haplotype; By hpl freq.: Analysis by haplotype frequency; aa: Amino acid.

167, and RRRR aa 172-175) that guarantee adequate protein subcellular localization acting as nuclear or cytoplasmic localization signals<sup>[27]</sup>. The second hyper-conserved aa region (aa 159-167) included one of these arginine-rich domains.

The high degree of sequence conservation observed in HBc may be indicative of its importance in protein function, positioning it as a possible target for diagnostic and therapeutic strategies. Recent studies have defined HBV core-related antigen (HBcrAg, which consists of HBc, HBeAg, and HBV p22 protein) as a promising serological viral marker, particularly for patients with low viral loads, such as treated patients<sup>[28]</sup> and patients with chronic HBeAg-negative infection<sup>[29]</sup>. This potential marker, however, has some limitations related to its high limits of detection (2 log IU/mL) and quantification (3-7 log IU/mL). The hyper-conserved regions observed in our study could be used as targets to improve HBc detection technology.

Aptamers are emerging as a promising diagnostic and therapeutic option for different diseases<sup>[30]</sup>. These molecules consist of single-strand DNA or RNA with high affinity and specificity and no toxicity or immunogenicity<sup>[31]</sup>. *In vitro* testing of an aptamer generated using the matrix domain of HBV (located in the large surface protein L and related to the nucleocapsid envelope) resulted in a 50% decrease in



**Figure 5 Information content analysis at amino acid level by clinical group.** A: Sliding window analysis of the Hepatitis B core protein by haplotype between the different clinical groups (HCC in blue, CHB in red, and LC in green). The green horizontal line corresponds to the region where LC group is less conserved compared to the CHB and HCC groups ( $P < 0.05$ ). B: Representation of CHB- and LC-specific conserved amino acid regions as sequence logos. Positions are reported at the top of each logo. CHB: Chronic hepatitis B infection without liver damage; HCC: Hepatocellular carcinoma; LC: Liver cirrhosis; aa: Amino acid;  $P$ :  $P$  value.

HBV titre in treated cell supernatants<sup>[32]</sup>. In another study, an aptamer targeting *HBC* resulted in a reduction in extracellular HBV DNA by interfering with nucleocapsid assembly<sup>[31]</sup>. Again, the hyper-conserved regions detected in our study could be novel targets for aptamer-based strategies that might work independently of clinical stage or HBV genotype. They could be also used to elaborate a new HBV detection system, as has been done with hepatitis C virus<sup>[33]</sup> and syncytial viruses<sup>[34]</sup>.

On analyzing nt and aa conservation in relation to clinical stage of HBV infection, all 3 groups showed similar patterns at the aa level, although the HBV quasispecies in the LC group was slightly less conserved (mainly between aa 140-160). At the nt level, conservation was lower in the CHB group than in the other 2 groups, largely in the 5 regions between nt 1946-1992, 2060-2095, 2145-2175, 2230-2250, and 2270-2293. This finding could be consistent with the high replication rate of HBV during this clinical stage. Moreover, the first variable region (nt 1946-1992) includes three CD8 HLA

**Table 3** Relative frequencies of nucleotide insertions/deletions detected

Clinical stage (n/total)	Patient	Relative frequency (% of mutated haplotypes)		
CHB (8/16)	1	1951 (1 nt: T) 8.36 (8.7)	2085 (1 nt: G)	
	2		17.12 (40)	
	3		3.19 (5)	
	4	0.37 (5.9)		
	9	2.02 (8.82)		
	10		1.34 (50)	
	12		1.04 (10)	
	13	2.79 (22.22)		
	HCC (2/17)	28		0.78 (4)
		33		2.42 (4.8)
LC (1/5)	34		17.42 (19.2)	

The table shows the relative frequency of insertions/deletions, together with the percentage (%) of mutated haplotypes per patient. Only patients carrying these mutations were included in the table. CHB: Chronic hepatitis B infection without liver damage; HCC: Hepatocellular carcinoma; LC: Liver cirrhosis; T: Thymine; G: Guanine; nt: Nucleotide.

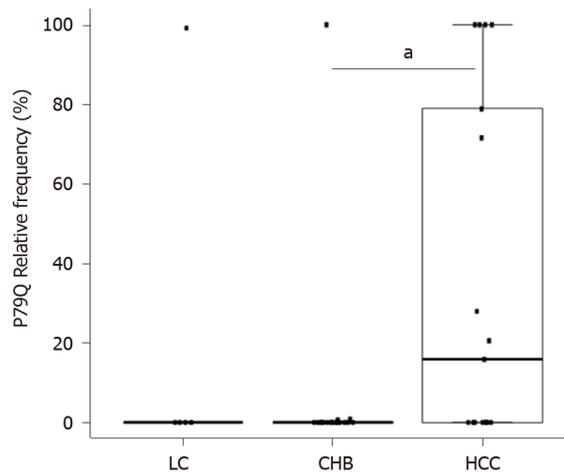
epitopes (epitopes B5101, B3501, and B0702 at nt positions 1958-1982)<sup>[26]</sup>, suggesting an attempt at immune evasion. Although the CHB group had the lowest levels of sequence conservation, we detected 2 group-specific conserved regions: aa 98-103 and nt 2306-2334. The nt region included the first 5 aa of the linker region, suggesting thus an important role for this region, which is involved in capsid assembly<sup>[35,36]</sup> and viral DNA synthesis<sup>[36]</sup>. In the LC group we detected 2 exclusively conserved nt regions (nt 1935-1976 and 2402-2435, which would translate respectively to aa 11-25 and 167-178) and 2 exclusively conserved aa regions (aa 28-30 and 51-54). The first related regions (nt 1935-1976 and aa 28-30) included portions of HBc (aa 14-18 and aa 23-39 respectively) that are involved in capsid assembly and envelopment and virion production<sup>[37]</sup>, highlighting the importance of these functions in LC. The second LC-specific nt region (nt 2402-2435) contained an arginine-rich domain of the CTD when translated.

The identification of group-specific conserved regions suggests different evolutionary histories that may have different effects on disease progression. Further studies, however, are needed to prove the association between these regions and different clinical stages and to investigate their role in liver disease progression.

Considering the risk and severity of disease progression, identification of prognostic factors would be of great help. A number of studies have focused on detecting aa changes possibly related to different clinical stages. The mutations T1753C and A1762T/G1764A (K130M/V131I in HBx) of basal core promoter, for example, were identified as possible prognostic markers for HCC<sup>[38,39]</sup>, while HBc aa mutations F24Y, E64D, E77Q, A80I/T/V, L116I, and E180A were linked to the development of cirrhosis and HCC<sup>[40]</sup>. In our study, one of the aa changes detected, P79Q, was exclusively observed in the HCC group. Mutations at this position have been found to be slightly associated with tumour relapse after resection<sup>[41]</sup>. More *in vitro* studies are required to investigate the role of the P79Q mutation in liver disease progression.

One limitation of our study is that we were not able to include large numbers of patients with different stages of liver disease due to the limits of PCR detection. This was particularly evident in the LC group, which was very small. Larger samples are needed to confirm our results. Moreover, although the Illumina MiSeq platform offers long read lengths, they are not sufficient to cover the entire *HBc* gene, making it necessary to divide it into 2 partially overlapping amplicons. Nonetheless, these 2 fragments were treated as independent samples during sequencing and subsequently analysed as such.

In summary, we have identified a number of nt and aa hyper-conserved regions that could be valuable targets for new therapeutic and diagnostic strategies. The role of group-specific conserved regions in liver disease progression requires further analysis. The P79Q substitution could be a possible prognostic factor for HCC. *In vitro* studies, however, are required to determine whether this change might affect viral



**Figure 6 Relative frequency of P79Q substitution in the 3 clinical groups.** Each dot represents a patient. The Bonferroni-corrected *P* value was calculated by Kruskal-Wallis test with posthoc Dunn multiple comparison test. (<sup>a</sup>*P* < 0.05). CHB: Chronic hepatitis B infection without liver damage; HCC: Hepatocellular carcinoma; LC: Liver cirrhosis; *P*: *P* value; P79Q: Proline to glutamine in position 79.

replication and to investigate associations between cellular damage and onset of HCC.

## ARTICLE HIGHLIGHTS

### Research background

Despite the existence of effective preventive vaccines, an estimated 257 million people worldwide live with chronic hepatitis B virus (HBV) infection and more than 880000 people die due to the development of liver cirrhosis and/or hepatocellular carcinoma. Although infection can be controlled with existing treatment, eradication is currently impossible due to the persistence of covalently closed circular DNA in hepatocyte nuclei that acts as a template for viral expression. New therapeutic approaches are needed, and gene therapy has been proposed as one of the most promising options. HBV core protein [encoded by the HBV core gene (*HBC*)] is a structural protein with functional activity that has a key role in viral replication and disease progression. Accordingly, it could be a potential target for new therapeutic and diagnostic strategies, and its variability could be a valuable prognostic factor for disease progression.

### Research motivation

As eradication of HBV infection is currently unachievable, new therapeutic strategies are necessary. Moreover, current treatments cannot interfere with the expression of viral proteins that can favor disease progression. Gene therapy based on silencing RNA is one of the most promising therapeutic approaches currently under investigation. The identification of hyper-conserved regions in key viral genes and proteins (such as *HBC*) is essential to orchestrate an effective strategy regardless of clinical stage or viral genotype.

### Research objectives

This study aimed to identify, by next-generation sequencing, hyper-conserved regions in *HBC* quasispecies of patients with different clinical stages of chronic HBV infection that could be a valuable target for gene therapy. Considering the essential role of the *HBC* gene and its encoded protein HBV core protein in HBV infection, changes in gene and protein conservation in specific clinical groups could be determining factors in disease progression and hence serve as prognostic factors for clinical follow-up.

### Research methods

The *HBC* gene was amplified by a 3-nested PCR protocol and later sequenced by next-generation sequencing (MiSeq, Illumina, United States) in 38 HBV-monoinfected chronic patients [16 with chronic hepatitis B infection without liver damages (CHB group), 5 with liver cirrhosis (LC group) and 17 with hepatocellular carcinoma (HCC group)]. Quasispecies sequences were genotyped by distance-based discriminant analysis, and general and intergroup nucleotide (nt) and amino acid (aa) conservation was determined by sliding window analysis. The presence of nt insertion and deletions and/or aa substitutions in the different groups was determined by aligning the sequences with a genotype-specific consensus sequence.

### Research results

Three nt (nt 1900-1929, 2249-2284, 2364-2398) and two aa (aa 117-120, 159-167) hyper-conserved regions shared by all the clinical groups were identified. By comparing gene and protein conservation between the different clinical groups, a similar pattern of conservation was observed, although CHB showed five nt less conserved regions (nt 1946-1992, 2060-2095, 2145-

2175, 2230-2250, 2270-2293) and LC one aa less conserved region (between aa 140 and 160). Moreover, some group-specific conserved regions were detected at both nt (nt 2306-2334 in CHB and 1935-1976 and 2402-2435 in LC) and aa (aa 98-103 in CHB and 28-30 and 51-54 in LC) levels. No differences in indel frequency were observed between the clinical groups. Contrarily, we identified an aa substitution (P79Q) that was more frequent in HCC [median (interquartile range) frequency of 15.82 (0-78.9) *vs* 0 (0-0) for the other groups;  $P < 0.05$  *vs* the CHB group].

### Research conclusions

We have identified a number of nt and aa regions that were highly conserved in the presence of different viral genotypes and clinical stages. These could be valuable targets for future pangenotypic and panclinical therapeutic and diagnostic strategies. The different clinically related conserved regions and the P79Q aa substitution could potentially be used as prognostic factors for disease progression.

### Research perspectives

Our findings could guide the creation of a new gene therapy strategy based on RNA silencing. In-depth analysis of group-specific conserved or variable regions and their role in disease progression is needed. Further *in vitro* studies are required to determine whether the P79Q aa substitution might affect viral replication and to investigate associations between cell damage and onset of HCC.

## ACKNOWLEDGEMENTS

The statistical and bioinformatics methods used in this study were reviewed by Mercedes Guerrero-Murillo from the Microbiology Department at Vall d'Hebron Hospital (Barcelona, Spain) and by Dr. Josep Gregori from the Liver Disease Viral Hepatitis Laboratory of Vall d'Hebron Hospital (Barcelona, Spain), CIBERehd research group, and Roche Diagnostics SL.

## REFERENCES

- 1 **World Health Organization.** Global Hepatitis Report, 2017. World Heal Organ [Internet] 2017; 62. Available from: <http://www.who.int/hepatitis>
- 2 **Yuen MF,** Chen DS, Dusheiko GM, Janssen HLA, Lau DTY, Locarnini SA, Peters MG, Lai CL. Hepatitis B virus infection. *Nat Rev Dis Primers* 2018; 4: 18035 [PMID: 29877316 DOI: 10.1038/nrdp.2018.35]
- 3 **Allweiss L,** Dandri M. The Role of cccDNA in HBV Maintenance. *Viruses* 2017; 9: 156 [PMID: 28635668 DOI: 10.3390/v9060156]
- 4 **Bowden S,** Locarnini S, Chang TT, Chao YC, Han KH, Gish RG, de Man RA, Yu M, Llamoso C, Tang H. Covalently closed-circular hepatitis B virus DNA reduction with entecavir or lamivudine. *World J Gastroenterol* 2015; 21: 4644-4651 [PMID: 25914474 DOI: 10.3748/wjg.v21.i15.4644]
- 5 **Papathodoridis GV,** Chan HL, Hansen BE, Janssen HL, Lampertico P. Risk of hepatocellular carcinoma in chronic hepatitis B: assessment and modification with current antiviral therapy. *J Hepatol* 2015; 62: 956-967 [PMID: 25595883 DOI: 10.1016/j.jhep.2015.01.002]
- 6 **Ren GL,** Huang GY, Zheng H, Fang Y, Ma HH, Xu MC, Zhang HB, Zhang WY, Zhao YG, Sun DY, Hu WK, Liu J. Changes in innate and permissive immune responses after hbv transgenic mouse vaccination and long-term-siRNA treatment [corrected]. *PLoS One* 2013; 8: e57525 [PMID: 23472088 DOI: 10.1371/journal.pone.0057525]
- 7 **Gish RG,** Yuen MF, Chan HL, Given BD, Lai CL, Locarnini SA, Lau JY, Wooddell CI, Schlupe T, Lewis DL. Synthetic RNAi triggers and their use in chronic hepatitis B therapies with curative intent. *Antiviral Res* 2015; 121: 97-108 [PMID: 26129970 DOI: 10.1016/j.antiviral.2015.06.019]
- 8 **Lin YY,** Liu C, Chien WH, Wu LL, Tao Y, Wu D, Lu X, Hsieh CH, Chen PJ, Wang HY, Kao JH, Chen DS. New insights into the evolutionary rate of hepatitis B virus at different biological scales. *J Virol* 2015; 89: 3512-3522 [PMID: 25589664 DOI: 10.1128/JVI.03131-14]
- 9 **González C,** Tabernero D, Cortese MF, Gregori J, Casillas R, Riveiro-Barciela M, Godoy C, Sopena S, Rando A, Yil 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]
- 10 **Locarnini S,** Zoulim F. Molecular genetics of HBV infection. *Antivir Ther* 2010; 15 Suppl 3: 3-14 [PMID: 21041899 DOI: 10.3851/IMP1619]
- 11 **Nassal M,** Rieger A, Steinau O. Topological analysis of the hepatitis B virus core particle by cysteine-cysteine cross-linking. *J Mol Biol* 1992; 225: 1013-1025 [PMID: 1613786 DOI: 10.1016/0022-2836(92)90101-o]
- 12 **Diab A,** Foca A, Zoulim F, Durantel D, Andrisani O. The diverse functions of the hepatitis B core/capsid protein (HBc) in the viral life cycle: Implications for the development of HBc-targeting antivirals. *Antiviral Res* 2018; 149: 211-220 [PMID: 29183719 DOI: 10.1016/j.antiviral.2017.11.015]
- 13 **Zlotnick A,** Venkatakrishnan B, Tan Z, Lewellyn E, Turner W, Francis S. Core protein: A pleiotropic keystone in the HBV lifecycle. *Antiviral Res* 2015; 121: 82-93 [PMID: 26129969 DOI: 10.1016/j.antiviral.2015.06.020]
- 14 **Ning X,** Luckenbaugh L, Liu K, Bruss V, Sureau C, Hu J. Common and Distinct Capsid and Surface Protein Requirements for Secretion of Complete and Genome-Free Hepatitis B Virions. *J Virol* 2018; 92: e00272-18 [PMID: 29743374 DOI: 10.1128/JVI.00272-18]
- 15 **Teng CF,** Huang HY, Li TC, Shyu WC, Wu HC, Lin CY, Su JJ, Jeng LB. A Next-Generation Sequencing-Based Platform for Quantitative Detection of Hepatitis B Virus Pre-S Mutants in Plasma of Hepatocellular Carcinoma Patients. *Sci Rep* 2018; 8: 14816 [PMID: 30287845 DOI: 10.1038/s41598-018-33051-4]

- 16 **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]
- 17 **Team RC.** A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria. 2016; Available from: <https://www.r-project.org/>
- 18 **Ramírez C,** Gregori J, Buti M, Tabernero D, Camós S, Casillas R, Quer J, Esteban R, Homs M, Rodríguez-Frías F. 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**: 273-283 [PMID: 23523552 DOI: 10.1016/j.antiviral.2013.03.007]
- 19 **Cuadras C.** A distance approach to discriminant analysis and its properties. In: Mathematics preprint series. Barcelona: 1991
- 20 **Cuadras C.** Distance analysis in discrimination and classification using both continuous and categorical variables. In: Statistical AData analysis and Interference. Amsterdam, 1989: 459-473
- 21 **Kimura M.** A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980; **16**: 111-120 [PMID: 7463489 DOI: 10.1007/bf01731581]
- 22 **Nassal M.** HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. *Gut* 2015; **64**: 1972-1984 [PMID: 26048673 DOI: 10.1136/gutjnl-2015-309809]
- 23 **Soriano V,** Barreiro P, Benitez L, Peña JM, de Mendoza C. New antivirals for the treatment of chronic hepatitis B. *Expert Opin Investig Drugs* 2017; **26**: 843-851 [PMID: 28521532 DOI: 10.1080/13543784.2017.1333105]
- 24 **Flisiak R,** Jaroszewicz J, Lucejko M. siRNA drug development against hepatitis B virus infection. *Expert Opin Biol Ther* 2018; **18**: 609-617 [PMID: 29718723 DOI: 10.1080/14712598.2018.1472231]
- 25 **Durantel D.** New treatments to reach functional cure: Virological approaches. *Best Pract Res Clin Gastroenterol* 2017; **31**: 329-336 [PMID: 28774415 DOI: 10.1016/j.bpg.2017.05.002]
- 26 **Zhang Y,** Wu Y, Deng M, Xu D, Li X, Xu Z, Hu J, Zhang H, Liu K, Zhao Y, Gao F, Bi S, Gao GF, Zhao J, Liu WJ, Meng S. CD8<sup>+</sup> T-Cell Response-Associated Evolution of Hepatitis B Virus Core Protein and Disease Progress. *J Virol* 2018; **92**: e02120-17 [PMID: 29950410 DOI: 10.1128/JVI.02120-17]
- 27 **Li HC,** Huang EY, Su PY, Wu SY, Yang CC, Lin YS, Chang WC, Shih C. Nuclear export and import of human hepatitis B virus capsid protein and particles. *PLoS Pathog* 2010; **6**: e1001162 [PMID: 21060813 DOI: 10.1371/journal.ppat.1001162]
- 28 **Hosaka T,** Suzuki F, Kobayashi M, Fujiyama S, Kawamura Y, Sezaki H, Akuta N, Suzuki Y, Saitoh S, Arase Y, Ikeda K, Kobayashi M, Kumada H. Impact of hepatitis B core-related antigen on the incidence of hepatocellular carcinoma in patients treated with nucleos(t)ide analogues. *Aliment Pharmacol Ther* 2019; **49**: 457-471 [PMID: 30663078 DOI: 10.1111/apt.15108]
- 29 **Riveiro-Barciela M,** Bes M, Rodríguez-Frías F, Tabernero D, Ruiz A, Casillas R, Vidal-González J, Homs M, Nieto L, Sauleda S, Esteban R, Buti M. Serum hepatitis B core-related antigen is more accurate than hepatitis B surface antigen to identify inactive carriers, regardless of hepatitis B virus genotype. *Clin Microbiol Infect* 2017; **23**: 860-867 [PMID: 28288829 DOI: 10.1016/j.cmi.2017.03.003]
- 30 **Kaur H,** Bruno JG, Kumar A, Sharma TK. Aptamers in the Therapeutics and Diagnostics Pipelines. *Theranostics* 2018; **8**: 4016-4032 [PMID: 30128033 DOI: 10.7150/thno.25958]
- 31 **Zhang Z,** Zhang J, Pei X, Zhang Q, Lu B, Zhang X, Liu J. An aptamer targets HBV core protein and suppresses HBV replication in HepG2.2.15 cells. *Int J Mol Med* 2014; **34**: 1423-1429 [PMID: 25174447 DOI: 10.3892/ijmm.2014.1908]
- 32 **Orabi A,** Bieringer M, Geerlof A, Bruss V. An Aptamer against the Matrix Binding Domain on the Hepatitis B Virus Capsid Impairs Virion Formation. *J Virol* 2015; **89**: 9281-9287 [PMID: 26136564 DOI: 10.1128/JVI.00466-15]
- 33 **Pleshakova TO,** Kaysheva AL, Shumov ID, Ziborov VS, Bayzhanova JM, Konev VA, Uchaikin VF, Archakov AI, Ivanov YD. Detection of Hepatitis C Virus Core Protein in Serum Using Aptamer-Functionalized AFM Chips. *Micromachines (Basel)* 2019; **10**: 129 [PMID: 30781415 DOI: 10.3390/mi10020129]
- 34 **Percze K,** Szakács Z, Scholz É, András J, Szeitner Z, Kieboom CH, Ferwerda G, Jonge MI, Gyuresányi RE, Mészáros T. Aptamers for respiratory syncytial virus detection. *Sci Rep* 2017; **7**: 42794 [PMID: 28220811 DOI: 10.1038/srep42794]
- 35 **Ludgate L,** Liu K, Luckenbaugh L, Streck N, Eng S, Voitenleitner C, Delaney WE 4th, Hu J. Cell-Free Hepatitis B Virus Capsid Assembly Dependent on the Core Protein C-Terminal Domain and Regulated by Phosphorylation. *J Virol* 2016; **90**: 5830-5844 [PMID: 27076641 DOI: 10.1128/JVI.00394-16]
- 36 **Liu K,** Luckenbaugh L, Ning X, Xi J, Hu J. Multiple roles of core protein linker in hepatitis B virus replication. *PLoS Pathog* 2018; **14**: e1007085 [PMID: 29782550 DOI: 10.1371/journal.ppat.1007085]
- 37 **Zheng CL,** Fu YM, Xu ZX, Zou Y, Deng K. Hepatitis B virus core protein dimer-dimer interface is critical for viral replication. *Mol Med Rep* 2019; **19**: 262-270 [PMID: 30387827 DOI: 10.3892/mmr.2018.9620]
- 38 **Ge Z,** Tian T, Meng L, Song C, Yu C, Xu X, Liu J, Dai J, Hu Z. HBV mutations in EnhI/BCP/PC region contribute to the prognosis of hepatocellular carcinoma. *Cancer Med* 2019; **8**: 3086-3093 [PMID: 31033235 DOI: 10.1002/cam4.2169]
- 39 **Chiu AP,** Tschida BR, Sham TT, Lo LH, Moriarity BS, Li XX, Lo RC, Hinton DE, Rowlands DK, Chan CO, Mok DKW, Largaespada DA, Warner N, Keng VW. HBx-K130M/V131I Promotes Liver Cancer in Transgenic Mice via AKT/FOXO1 Signaling Pathway and Arachidonic Acid Metabolism. *Mol Cancer Res* 2019; **17**: 1582-1593 [PMID: 30975706 DOI: 10.1158/1541-7786.MCR-18-1127]
- 40 **Al-Qahtani AA,** Al-Anazi MR, Nazir N, Abdo AA, Sanai FM, Al-Hamoudi WK, Alswat KA, Al-Ashgar HI, Khan MQ, Albenmoussa A, El-Shamy A, Alanazi SK, Dela Cruz D, Bohol MFF, Al-Ahdal MN. The Correlation Between Hepatitis B Virus Precore/Core Mutations and the Progression of Severe Liver Disease. *Front Cell Infect Microbiol* 2018; **8**: 355 [PMID: 30406036 DOI: 10.3389/fcimb.2018.00355]
- 41 **Jia J,** Li H, Wang H, Chen S, Wang M, Feng H, Gao Y, Wang Y, Fang M, Gao C. Hepatitis B virus core antigen mutations predict post-operative prognosis of patients with primary hepatocellular carcinoma. *J Gen Virol* 2017; **98**: 1399-1409 [PMID: 28640739 DOI: 10.1099/jgv.0.000790]



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: <http://www.f6publishing.com/helpdesk>  
<http://www.wjgnet.com>

