




**ADVERTIMENT.** L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  <https://creativecommons.org/licenses/?lang=ca>

**ADVERTENCIA.** El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <https://creativecommons.org/licenses/?lang=es>

**WARNING.** The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



**Universitat Autònoma  
de Barcelona**

Department of Biochemistry, Molecular Biology and Biomedicine

Doctoral thesis

**Study of similarities and differences between  
circulating hepatitis B virus RNA and DNA  
quasispecies in different regions of viral genome  
and development of an antiviral gene therapy  
strategy**

**Thesis autor**

Selene García García

**Thesis directors**

Dr. Francisco Rodríguez Frías

Dr. David Tabernero Caellas

Dra. Maria Francesca Cortese

**Thesis tutor**

Dr. Francisco Rodríguez Frías

**Barcelona, 2023**



## AGRADECIMIENTOS

En primer lugar, me gustaría agradecer a mis tres directores de tesis Paco, David y María Francesca. A Paco, por su ayuda infinita, su charlas e historietas y la cercanía que tanto le caracteriza. Gracias por contagiarnos a todos tu pasión, dedicación y esfuerzo por todo aquello que se pretende conseguir en la vida. David, moltes gràcies per tota l'ajuda al llarg d'aquests tres anys i mig de tesi. Has sabut escoltar-me i donar-me els millors consells per tirar endavant. Francesca, gracias por toda la ayuda y por haberme apoyado como lo has hecho durante mi recorrido, por confiar siempre tanto en mí y mostrarme la importancia de discernir sobre cuáles son las batallas por las que vale la pena luchar. Voldria agrair-li també al Dr. Josep Gregori per tota l'ajuda proporcionada. Sense els teus coneixements i anàlisis informàtics aquesta tesi no hagués estat possible. Agradecer también a la Dra. Ariadna Rando y a todo el equipo de la Dra. Maria Buti por su colaboración. Agrair al Dr. Josep Quer i al seu equip per estar sempre disposats a ajudar-me.

Moltes gràcies a la Marta pel seu suport. Gracias a Adrián, Gerard, Unai, Guillem, Chantal y Susana. Gracias Adrián por compartir conmigo tantos momentos de risas y memes, son mi alegría en los perezosos viajes de Renfe y metro. Quiero agradecer también a Irene por haber sido la alegría del laboratorio durante el tiempo que coincidimos. Empezamos contigo esta aventura y aunque te bajaste del barco en mitad de nuestro camino nos has seguido apoyando y animando siempre. Vull agrair també al Gerard. Gràcies per tots els moments i riures viscuts. Gracias también a Àngels, Elena, Isa, Pablo y Noelia.

Quiero agradecerle con especial interés a Bea, mi *partner in crime* en este viaje. No tengo ninguna duda de que sin ti esto no hubiera sido lo mismo. Gracias por todo lo que has hecho y haces por mí, por saber escucharme y aconsejarme poniéndote siempre en mi piel, dándome los consejos genuinos y des del corazón. Gracias, amiga.

A los chicos de prácticas. En especial a Sofía. También a Alicia, Dani y Aleix. Quiero destacar la complicidad con Aleix, una persona maravillosa, entusiasta y muy trabajadora. Esta recta final ha sido mucho más amena con todas las historietas, la mayoría inverosímiles, que llevamos acumuladas en tan pocos meses.

Los secres, Merche, Ferrán, Julia, Tina, David... y en especial a Jose y Manu con quienes ha sido imposible aguantar un día sin reír. De corazón, sois maravillosos. Nunca podré olvidar la noche del Enigma.

Quiero dedicar unas palabras también a Adriana, Elena, Meritxell, Mar y Aurora. Ha sido precioso encontrarme en este camino con gente tan espectacular como vosotras. Adriana

y Elena, me llevo esos dos años de Barceló guardaditos en mi corazón. También a los chicos de HIV, a Pamela y a los celadores Rai y Pablo por su amabilidad y trato exquisito, ha sido un placer coincidir con vosotros.

Darle las gracias también a todo el equipo de Sant Joan de Déu con el que realicé mis prácticas de Máster. Carles, Cris, Assumpta, Marta... Amaresh, Natàlia, Jessi, Desi fuisteis mis alas para seguir creciendo, gracias por todos los consejos y por tantos recuerdos llenos de carcajadas. Fue una época maravillosa. Y Ana Doncel, la complicidad que nos unió en pocas personas la he encontrado.

A mis amigos de toda la vida, en especial a Carlos Maestre por ser una de las personas con la que más me río, por sus ocurrencias, sus ideas buenísimas proponiendo cualquier plan, por todas las veces que me has preguntado cómo iba la tesis, por preocuparte, por ser el mejor organizador que pudiera tener un grupo de amigos, y por junto con Pau, enseñarnos el mejor invento que pocos conocen, el SNOFT. Gracias a Blanca por ser mi ejemplo de fortaleza y lucha diaria. Gràcies al Marc Obón per tots els consells que m'has donat sempre, per ser-hi i per comptar amb mi quan has necessitat algú amb qui parlar. Gracias también a Efrén, sin tus viajes con la nave hacia Sant Joan De Déu llegar hasta donde estoy hoy no hubiera sido posible. Gracias por sacarme una sonrisa cuando más lo he necesitado con tus videos y audios que, sí, efectivamente, los tengo todos guardados y acudo a ellos siempre que lo necesito. Gracias por aquel abrazo en segundo de carrera en mitad del pasillo cuando eché a llorar al verte por lo agobiada que iba con los estudios. A Jose, por haber retomado el contacto después de tanto tiempo sin saber el uno del otro. A Ana Belén, gracias por ser una hermana, te quiero.

Quiero agradecer también a lo mejor que me ha dado la carrera, mis chicas. Ari, Gemma, Laura y Natalia. Adoro reírnos recordando siempre las mismas anécdotas con la gracia del primer día. Al record de totes nosaltres queden tantíssimes hores d'estudi a la biblioteca de socials, finde si, finde també. De tot cor us dic que sou les millors. Os estimo.

A mi kor... cuánto te quiero. Qué fuerte eres y qué ejemplo de valentía. Prometimos no dejarte nunca sola y así será. Nos tendremos eternamente. Te quiero mucho. Marc y Carol... la luz más brillante que podría iluminar mi camino. Hemos llegado hasta aquí juntos porque, junto con mi hermana, habéis sido mi fuerza y mis ganas de seguir adelante. Gracias por no dejarme nunca sola, por querer siempre lo mejor para mí, por ser mi apoyo incondicional, y por ser ejemplo de constancia y perseverancia. Os admiro. Sé con certeza absoluta lo orgullosos que estáis de mí. Al igual que yo de vosotros. Os quiero.

Gracias a mi hermana por la que siento debilidad y a la que quiero más que a nadie. Eres el mejor regalo que la vida podría haberme hecho. Aunque no te lo diga muy a menudo sabes que te admiro y quiero muchísimo. Jordi, gracias por ser una persona tan auténtica y entregada con la familia. A Lorena, Luís y mi sobrino Martín. Os quiero. A Alana, a la que adoro. A mis padres, y a Salva y Juani. Gracias por ser mi motor, por confiar y apostar por mí. Gracias por entenderme y darme mi espacio en los momentos más difíciles. Gracias por comprender mi nivel de auto exigencia y no juzgarlo nunca. Gracias por lo fácil que me habéis puesto el camino y no ponerme nunca límites. Mama y papa, vuestra dedicación, entrega y pasión por nosotras es inhumana. Si llego a ser madre espero que mis hijos se sientan igual de orgullosos que yo de vosotros. A ti papa decirte que, tantos viernes que no podíamos vernos durante la carrera porque tenía que estudiar, aquí tienes el fruto de tanto esfuerzo. Deciros que esto no se acaba aquí, que esto es otro peldaño más que subo gracias a toda vuestra ayuda, y que seguimos subiendo con la misma actitud y ganas del principio.

A mis abuelos, a los que amo. En especial a mi yayo, también mi debilidad, por el que siento una admiración absoluta y amor eterno. La persona más luchadora que jamás he conocido. Te quiero muchísimo.

Y a todo aquel que no haya mencionado y que haya contribuido en el desarrollo directa o indirectamente de esta tesis, muchas gracias.

## ABBREVIATIONS

### A

---

A	adenine
aa	amino acid
AASLD	American Association for the Study of Liver Diseases
ADV	adefovir dipivoxil
ALT	alanine aminotransaminase
anti-HBc	hepatitis B core antibody
anti-HBe	hepatitis B e antibody
anti-HBs	hepatitis B surface antibody
ApoE	apolipoprotein E
APRI	aspartate aminotransferase to platelet ratio index
ASGPR	asialoglycoprotein receptor
ASO	antisense oligonucleotide
AST	aspartate aminotransferase

### B

---

bp	base pair
BCP	basal core promoter
BFV	besifovir

### C

---

C	cytosine
cDNA	complementary DNA
CHB	chronic hepatitis B
C <sub>m</sub>	commons index
CP	core promoter
Ct	cycle threshold
CTD	C-terminal domain
CURS	core upstream regulatory sequence

## D

---

d	days
D <sub>A</sub>	net genetic distance
DB-rule	distance-based discriminant rule
Del	deletions
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotide
DMEM	dulbecco's modified eagle medium
DMSO	dimethyl sulfoxide
DR1/DR2	direct repeat 1/direct repeat 2
dsDNA	double-stranded DNA
dsIDNA	double-stranded lineal DNA
DTT	dithiothreitol

## E

---

EASL	European Association for the Study of the Liver
EB	elution buffer
ENH 1/ENH 2	enhancer 1/enhancer 2
ESCRT	endosomal sorting complexes required for transport
ETV	entecavir

## F

---

FBS	fetal bovine serum
FIB-4	fibrosis index based on four factors
FLASH	fast length adjustment of short reads



## G

---

G	guanine
GalNac	N-acetylgalactosamine
GP	Gapmer
GRE	glucocorticoid-responsive element
GSH	glutathione
G418	geneticin

## H

---

HBcAg	hepatitis B core antigen
HBcrAg	hepatitis B core-related antigen
HBeAg	hepatitis B e antigen
HBsAg	hepatitis B surface antigen
HBV	hepatitis b virus
HBV-cDNA	hepatitis B virus complementary DNA
HBV-DNA	hepatitis B virus DNA
HBV-RNA	hepatitis B virus RNA
<i>HBX</i>	hepatitis B X gene
HBx	hepatitis B X protein
HCC	hepatocellular carcinoma
HNP	Hill numbers profile
hNTCP	human NTCP
H <sub>s</sub>	Shannon entropy
H <sub>SI</sub>	Simpson index
HSPGs	heparin sulfate proteoglycans

## I

---

IC	information content
IFN- $\alpha$	interferon alpha
Ins	insertions
Indel	insertions/deletion
IQR	interquartile range
iPSCs	induced pluripotent stem cells
IT	immune tolerance
IU	international units

## K

---

kb	kilobase
kDa	kiloDalton

## L

---

LAM	lamivudine
L-HBsAg	large hepatitis B surface antigen
LLOQ	lower limit of quantification
LNA	locked nucleic acid
LNPs	lipid nanoparticles
LOD	limit of detection

## **M**

---

MGE	multiplicities of genome equivalents
MGP	magnetic glass particle
M-HBsAg	middle hepatitis B surface antigen
MHR	major hydrophilic region
min	minutes
mL	milliliters
mm	millimeters
MOE	2'-O-methoxyethyl
MVBs	multivesicular bodies

## **N**

---

NAs	nucleos(t)ide analogues
NES	nuclear export signal
NGS	next-generation sequencing
NLS	nuclear localization signal
NPC	nuclear pore complex
NPs	nanoparticles
Ns	indeterminations
nt	nucleotide
NTCP	sodium taurocholate co-transporting polypeptide
$\Pi$	nucleotide diversity

## **O**

---

ORF	open reading frame
$O_v$	overlap index

## P

---

P	patient
PCR	polymerase chain reaction
pegINF- $\alpha$	pegylated INF- $\alpha$
PEG	polyethylene glycol
Pen/Strep	penicillin and streptomycin
pgRNA	pregenomic RNA
pi	post infection
poly A	poly adenylation signal
preC RNA	precore RNA
PS	phosphorothioate
PU	polyurethane
p22cr	22 kiloDalton precore protein
$\varphi$	phi signal

## Q

---

QFF	quasispecies fitness fraction
qPCR	quantitative PCR
QS	quasispecies

## R

---

R1/R2	read 1/read 2
RH	RNase H
RHL	rare haplotype load
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAPII	RNA polymerase II
RT	reverse transcriptase
RT-qPCR	quantitative reverse transcription PCR

## S

---

SBS	sequencing by synthesis
ssDNA	single-stranded DNA
SD	standard deviation
S-HBsAg	small hepatitis B surface antigen
SVPs	subviral particles
siRNA	small interfering RNA
STD	standard

## T

---

T	thymine
TAE	tris acetate EDTA
TA	TATA-like boxes
TBV	telbivudine
TDF	tenofovir disoproxil fumarate
TAF	tenofovir alafenamide
TLM	translocation motif
TCID <sub>50</sub>	median tissue culture infectious dose
TP	terminal protein

## U

---

IU	international units
URR	upper regulatory region
UPGMA	unweighted pair group method with arithmetic

## W

---

WHO	World Health Organization
-----	---------------------------

# INDEX

---

ABSTRACT.....	1
RESUMEN .....	3
RESUM.....	5
1. INTRODUCTION.....	8
1.1. Hepatitis B virus.....	8
1.2. History of hepatitis B virus.....	8
1.2.1. Hepatitis B virus taxonomy and genotypes/sub-genotypes.....	9
1.2.2. Hepatitis B virus epidemiology and prevalence .....	11
1.3. Hepatitis B virus genome composition .....	12
1.3.1. preS/S ORF: hepatitis B virus envelope.....	15
1.3.2. preC/C ORF: hepatitis B virus core antigen and hepatitis B virus e antigen .....	16
1.3.3. P ORF: hepatitis B virus polymerase .....	16
1.3.4. X ORF: hepatitis B virus X protein .....	18
1.4. Hepatitis B virus life cycle .....	20
1.4.1. Molecular mechanisms of viral polymerase involved in reverse transcription.....	24
1.5. Hepatitis B virus genetic variability and quasispecies: definition and clinical significance	24
1.6. Natural history of chronic hepatitis B infection .....	25
1.7. Serum hepatitis B virus RNA: biomarker of hepatitis B infection.....	27
1.8. Current hepatitis B therapy.....	29
1.9. Cell culture models for studying hepatitis B virus infection .....	32
1.10. Gene therapy against hepatitis B virus .....	33
1.11. Next-generation sequencing.....	34
2. HYPOTHESIS .....	41
3. AIMS .....	44
4. MATERIALS AND METHODS .....	46
4.1. Patients and samples.....	46
4.2. Hepatitis B virus DNA and RNA extraction.....	46
4.2.1. Automated extraction .....	46
4.2.2. Manual extraction.....	46
4.3. Hepatitis B virus DNA and RNA quantification .....	48

4.4.	Serological determinations .....	53
4.5.	Removal of hepatitis B virus DNA from RNA isolations and reverse transcription .....	54
4.6.	Hepatitis B virus DNA and RNA amplification to obtain amplicon libraries for next-generation sequencing analysis.....	58
4.7.	Verification of hepatitis B virus DNA amplification .....	62
4.8.	PCR product purification .....	63
4.9.	Quantification and normalization of the purified PCR product .....	63
4.10.	Amplicon sequencing by next-generation sequencing using Illumina MiSeq system ..	64
4.10.1.	End repair and A-tailing.....	64
4.10.2.	Adapter ligation .....	66
4.10.3.	Post-ligation clean-up (1 <sup>st</sup> clean-up) .....	67
4.10.4.	Library amplification .....	68
4.10.5.	Post-amplification clean-up (2 <sup>nd</sup> clean-up) .....	69
4.10.6.	Quality verification of the purified amplicon pools .....	69
4.10.7.	qPCR-based library quantification (1st qPCR) .....	69
4.10.8.	Pool of amplicon libraries pools and quantification (2nd qPCR) .....	71
4.10.9.	Denaturation and dilution of pool of pools and PhiX .....	71
4.11.	Next-generation sequencing data analyses obtained using Illumina MiSeq system....	72
4.12.	Haplotype genotyping.....	74
4.13.	Quasispecies conservation and variability .....	75
4.14.	Quasispecies complexity .....	76
4.14.1.	Incidence-based indexes .....	76
4.14.2.	Abundance-based indexes .....	76
4.14.2.1.	Hill numbers profile.....	76
4.14.3.	Rare haplotype load.....	78
4.14.4.	Quasispecies fitness fractions.....	79
4.14.5.	Functional indexes .....	80
4.14.5.1.	Nucleotide diversity .....	80
4.14.5.2.	Sample size and rarefaction.....	81
4.15.	Similarity between quasispecies distributions .....	81
4.16.	Genetic distance between quasispecies .....	82
4.17.	Statistical analysis .....	83
4.18.	Antisense locked nucleic acid Gapmers and/or siRNA-based therapy against hepatitis B virus infection: <i>in vitro</i> testing .....	84

4.18.1.	Hepatitis B virus production: preparation of hepatitis B viral stock from HepG2.2.15	84
4.18.2.	Hepatitis B virus infection assay in HepG2-hNTCP DMSO-treated cells .....	86
4.18.3.	Antisense locked nucleic acid Gapmers and/or siRNA design.....	86
4.18.4.	Antisense locked nucleic acid Gapmers and/or siRNA transfection in HepG2-hNTCP DMSO-treated and HBV-infected cells .....	89
4.18.5.	Assessment of antisense locked nucleic acid Gapmers entry and dose-response curve	90
5.	RESULTS .....	92
5.1.	First study: Cross-sectional evaluation of circulating hepatitis B virus RNA and DNA: Different quasispecies? .....	92
5.1.1.	Aims .....	92
5.1.2.	Study design .....	92
5.1.3.	Verification of residual hepatitis B virus DNA from RNA isolations .....	94
5.1.4.	Next-generation sequencing results.....	94
5.1.4.1.	Hepatitis B virus DNA and RNA quasispecies complexity analysis.....	94
5.1.4.2.	Hepatitis B virus DNA and RNA quasispecies conservation analysis .....	96
5.2.	Second study: Comparison of circulating hepatitis B virus DNA and RNA quasispecies in the surface/polymerase overlapping region .....	100
5.2.1.	Aims .....	100
5.2.2.	Study design .....	100
5.2.3.	Verification of residual hepatitis B virus DNA from RNA isolations .....	102
5.2.4.	Next-generation sequencing results.....	104
5.2.4.1.	Comparison of master haplotypes.....	105
5.2.4.2.	Quasispecies complexity assessment with quasispecies fitness fractions and Hill number profiles.....	106
5.2.4.2.1.	Baseline samples .....	107
5.2.4.2.2.	Comparison of HBV-DNA and HBV-RNA quasispecies evolution .....	112
5.2.4.3.	Haplotype distribution similarities: Indexes of Commons and Overlap .....	116
5.2.4.3.1.	Baseline samples .....	116
5.2.4.3.2.	Changes in hepatitis B virus DNA and RNA quasispecies during longitudinal follow-up .....	118
5.2.4.4.	Genetic distance between quasispecies .....	119
5.3.	Third study: The combination of antisense locked nucleic acid Gapmers and/or siRNA as a potential hepatitis B virus gene therapy: preliminary in vitro results .....	123



5.3.1. Antisense locked nucleic acid Gapmers entry into HepG2-hNTCP 2.5% DMSO-treated cells and dose-response curve.....	123
5.3.2. Antisense locked nucleic acid Gapmers inhibition of hepatitis B virus expression at early and late treatment.....	125
5.3.3. Increasing silencing efficiency by using a antisense locked nucleic acid Gapmers and siRNA combination strategy .....	126
6. DISCUSSION.....	130
6.1. Discussion first study: Cross-sectional evaluation of circulating hepatitis B virus RNA and DNA: Different quasispecies? .....	132
Complexity comparison of HBV-DNA and HBV-RNA quasispecies using next-generation sequencing.....	133
Conservation comparison of HBV-DNA and HBV-RNA quasispecies using next generation sequencing.....	134
6.2. Discussion second study: Comparison of circulating hepatitis B virus DNA and RNA quasispecies in the surface/polymerase overlapping region .....	136
Comparison of master haplotypes .....	137
Quasispecies complexity by quasispecies fitness fraction and Hill numbers profile .....	139
Similarity between common haplotypes distributions .....	142
Nucleotide diversity in hepatitis B virus DNA and RNA quasispecies and genetic distances between them.....	144
6.3. Discussion third study: The combination of antisense locked nucleic acid Gapmers and/or siRNA as a potential hepatitis B virus gene therapy: preliminary in vitro results .....	147
7. CONCLUSIONS .....	153
8. LIMITATIONS .....	156
9. FUTURE PERSPECTIVES .....	158
10. REFERENCES .....	161
11. APPENDIXES .....	183
Appendix 1.....	183
Appendix 2.....	199
Appendix 2. Table 1.....	199
Appendix 3.....	201
Appendix 3. Table 1.....	201
Appendix 3. Table 2.....	204
Appendix 3. Table 3.....	206
Appendix 3. Table 4.....	209
Appendix 4.....	212

# ABSTRACT

The World Health Organization (WHO) estimated that, in 2019, 296 million people worldwide were living with chronic hepatitis B virus (HBV) infection, with 1.5 million new infections each year despite the availability of a safe and effective vaccine. Although highly effective antiviral treatments with nucleos(t)ides analogues are available, they do not cure the infection or guarantee that the liver disease will not progress because they do not interfere with the expression of covalently closed DNA (cccDNA) or viral DNA integrated into the genome of the hepatocytes.

HBx protein and viral reverse transcriptase (RT) are two key elements in the HBV life cycle. The former, encoded by the hepatitis B X gene (*HBX*), plays an important role in intrahepatic cccDNA stability and its transcription. The latter, is responsible for HBV being the most variable among DNA viruses, as RT is an error-prone HBV polymerase domain that lacks error-proofreading activity and provides genetic variability in catalysing reverse transcription of encapsidated pregenomic RNA (pgRNA) into relaxed circular DNA.

Serum HBV-DNA levels decline rapidly after the start of treatment, whereas those of HBV-RNA decline more slowly. It allows the study of viral quasispecies (QS) through circulating HBV-RNA in patients under treatment with even undetectable HBV-DNA. Regarding QS, serum HBV-RNA was dynamic and more genetically homogenous to simultaneously sampled intrahepatic HBV-RNA than to the cccDNA. For this reason, many of the mechanisms of viral inhibition based on gene therapy approaches are focus on HBV-derived transcripts as the main target. However, few studies to date have analysed circulating HBV-RNA QS and therefore there is little data on its similarities and differences relative to that of HBV-DNA QS. Since HBV-RNA QS has not been subjected to reverse transcription process, it is reasonable to assume that there may be significant differences between HBV-DNA and HBV-RNA QS.

According to this, in the first study we analysed a region of the 5' end of *HBX* where we had previously identified hyper-conserved regions in the circulating HBV-DNA [2,3], which we also intended to confirm in circulating HBV-RNA. We aimed to compare both serum HBV-DNA and HBV-RNA QS in a group of 13 well-characterised untreated patients with different HBV genotypes and degrees of severity of liver disease, with increased complexity and sequence conservation observed in HBV-DNA. In addition, it was confirmed that most of

the previous hyper-conserved regions identified coincided with those revealed in serum HBV-RNA, and may thus also be conserved at intrahepatic HBV-RNA level.

In the second study, we aimed to compare similarities and differences between HBV-RNA and HBV-DNA QS in the surface/polymearse (S-RT) overlapping region of HBV genome, using a set of indexes of genetic diversity, similarity, and genetic distances to provide detailed information. A serum sample from 12 untreated patients was included. A follow-up sample of three out of 12 patients was also included. The possibility of analysing changes between both QS allowed us to get an idea of the independent evolutive trends of each viral QS, which may be different according to the patient.

In the third study, *in vitro* functional studies were performed in a cellular infection model to test the potential usefulness of gene therapy targeting the hyper-conserved sequences reported in the first study and even confirmed at circulating HBV-RNA level. Taking all this into account, we tested several candidate oligonucleotides for gene therapy based on LNA Gapmer and small interfering RNA.

# RESUMEN

En 2019 la Organización Mundial de la Salud (OMS) estimó que 296 millones de personas en todo el mundo vivían con infección crónica por el virus de la hepatitis B (VHB), con 1,5 millones de nuevas infecciones por año, a pesar de la disponibilidad de una vacuna segura y efectiva. A pesar de la existencia de los análogos de nucleós(t)idos, estos antivirales no logran curar la infección ni garantizan que la hepatopatía no progrese, ya que no eliminan ni interfieren en la expresión del ADN circular covalentemente cerrado (ADNccc) o del ADN viral integrado en el genoma de los hepatocitos.

La proteína HBx y la transcriptasa reversa viral (RT) son dos elementos clave en el ciclo vital del VHB. La primera, codificada por el gen X del VHB (*HBX*), juega un papel importante en la estabilidad del ADNccc intrahepático y su transcripción. La segunda, es responsable de que, entre los virus de ADN, el VHB sea uno de los más variables, ya que es un dominio de la polimerasa viral que carece de actividad correctora de errores, proporcionando variabilidad genética durante la transcripción reversa del ARN pregenómico encapsulado a ADN circular relajado.

Los niveles séricos ADN-VHB disminuyen rápidamente tras el inicio del tratamiento, mientras que los de ARN-VHB lo hacen más lentamente. Esto permite estudiar las quasispecies (QS) virales mediante el ARN-VHB circulante en pacientes en tratamiento, con niveles incluso indetectables de ADN-VHB. La QS del ARN-VHB circulante es dinámica y genéticamente más homogénea a la del ARN intrahepático que a la del ADNccc. Por ello, muchos de los mecanismos de inhibición viral basados en terapia génica se centran en los transcritos derivados del VHB como diana principal. Pocos estudios han analizado la QS circulante de ARN-VHB, por lo que existen pocos datos sobre sus similitudes y diferencias con respecto a la QS de ADN-VHB. Dado que la QS de ARN-VHB no se ha sometido a un proceso de transcripción reversa, es razonable suponer que puede haber diferencias significativas entre ambas QS.

En el primer estudio se analizó una región del extremo 5' del *HBX* en la que previamente se identificaron regiones híperconservadas en el ADN-VHB circulante, que pretendíamos confirmar en el ARN-VHB circulante. Se compararon ambas QS en un grupo de 13 pacientes no tratados con diferentes genotipos del VHB y grados de gravedad de la enfermedad hepática observándose una mayor complejidad y conservación de la secuencia en el ADN-VHB. Además, se confirmó que la mayoría de las regiones híperconservadas

anteriormente identificadas coincidían con las reveladas en el ARN-VHB sérico, por lo que es posible que también se conserven a nivel de ARN-VHB intrahepático.

En el segundo estudio, se compararon ambas QS en la región de solapamiento de superficie/polimerasa (S-RT) del genoma del VHB, utilizando un conjunto de índices de diversidad, similitud y distancias genéticas. Se incluyó una muestra de suero de 12 pacientes no tratados y también una muestra de seguimiento de tres de los 12 pacientes. El análisis de los cambios entre ambas QS nos permitió observar las tendencias evolutivas independientes de cada QS, que pueden ser diferentes según el paciente.

En el tercer estudio, se utilizó un modelo *in vitro* de infección celular para comprobar la posible utilidad de la terapia génica dirigida a las secuencias híperconservadas mencionadas en el primer estudio e incluso confirmadas a nivel del ARN-VHB circulante. Se probaron varios oligonucleótidos candidatos para terapia génica basados en LNA Gapmer y *small interfering RNA*.

# RESUM

Al 2019, la Organització Mundial de la Salut (OMS) va estimar que 296 milions de persones a tot el món vivia amb infecció crònica pel virus de la hepatitis B (VHB), amb 1,5 milions de noves infeccions per any malgrat la disponibilitat d'una vacuna segura i eficaç. Tot i l'existència dels anàlegs de nucleòs(t)ids, aquests antivirals no aconsegueixen curar la infecció ni garantir que l'hepatopatia no progressi, ja que no eliminen ni interfereixen en l'expressió del ADN circular covalentment tancat (ADNccc) o del ADN viral integrat al genoma dels hepatòcits.

La proteïna HBx i la transcriptasa reversa viral (RT) són dos elements clau al cicle vital del VHB. La primera, codificada pel gen X del VHB (*HBX*), juga un paper important en l'estabilitat del ADNccc intrahepàtic i la seva transcripció. La segona, és responsable de què, entre els virus d'ADN, el VHB sigui un dels més variables, ja que es un domini de la polimerasa viral que manca d'activitat correctora d'errors, proporcionant variabilitat genètica durant la transcripció reversa del ARN pregenòmic encapsulat a ADN circular relaxat.

Els nivells sèrics de ADN-VHB disminueixen ràpidament amb l'inici del tractament, mentre que els d'ARN-VHB ho fan més lentament. Això permet estudiar les quasispècies (QS) virals a través de l'ARN-VHB circulat en pacients amb tractament, amb nivells inclús indetectables d'ADN-VHB. La QS d'ARN-VHB circulat és dinàmica i genèticament més homogènia a la de l'ARN intrahepàtic que a la del ADNccc. Per aquest motiu, molts dels mecanismes d'inhibició viral basats en teràpia gènica es centren en els transcrits derivats del VHB com a diana principal. Hi ha pocs estudis que hagin analitzat la QS circulat d'ARN-VHB, motiu pel qual existeixen poques dades sobre les seves similituts i diferències amb la QS d'ADN-VHB. Donat que la QS d'ARN-VHB no ha estat sotmesa a un procés de transcripció reversa, és raonable suposar que poden haver diferències significatives entre ambdues QS.

Al primer estudi es va analitzar una regió de l'extrem 5' del *HBX* on previament es van identificar regions hiperconservades a l'ADN-VHB circulat, que preteniem confirmar a l'ARN-VHB circulat. Es van comparar ambdues QS en un grup de 13 pacients no tractats amb diferents genotips del VHB i graus de gravetat de la malaltia hepàtica observant-se una major complexitat i conservació de la seqüència a l'ADN-VHB. A més, es va confirmar que la majoria de les regions hiperconservades anteriorment identificades coincidien amb

les revelades a l'ARN-VHB sèric, pel que és possible que també es conservin a nivell d'ARN-VHB intrahepàtic.

Al segon estudi, es van comparar ambdues QS a la regió de solapament entre l'antigen de superfície i el domini RT de la polimerasa viral (S-RT) del genoma del VHB, utilitzant un conjunt d'índexs de diversitat, similitut i distàncies genètiques. Es va incloure una mostra de sèrum de 12 pacients no tractats. També es va incloure una mostra de seguiment de tres dels 12 pacients. L'anàlisi dels canvis entre ambdues QS ens va permetre observar les tendències evolutives independents de cada QS, que poden ser diferents segons el pacient.

Al tercer estudi, es va utilitzar un model *in vitro* d'infecció cel·lular per comprovar la possible utilitat de la teràpia gènica dirigida a les seqüències hiperconservades esmentades al primer estudi i també confirmades a nivell d'ARN-VHB circulant. Es van provar diversos oligonucleòtids candidats per teràpia gènica basats en LNA Gapmer i *small interfering RNA*.

## **INTRODUCTION**

---



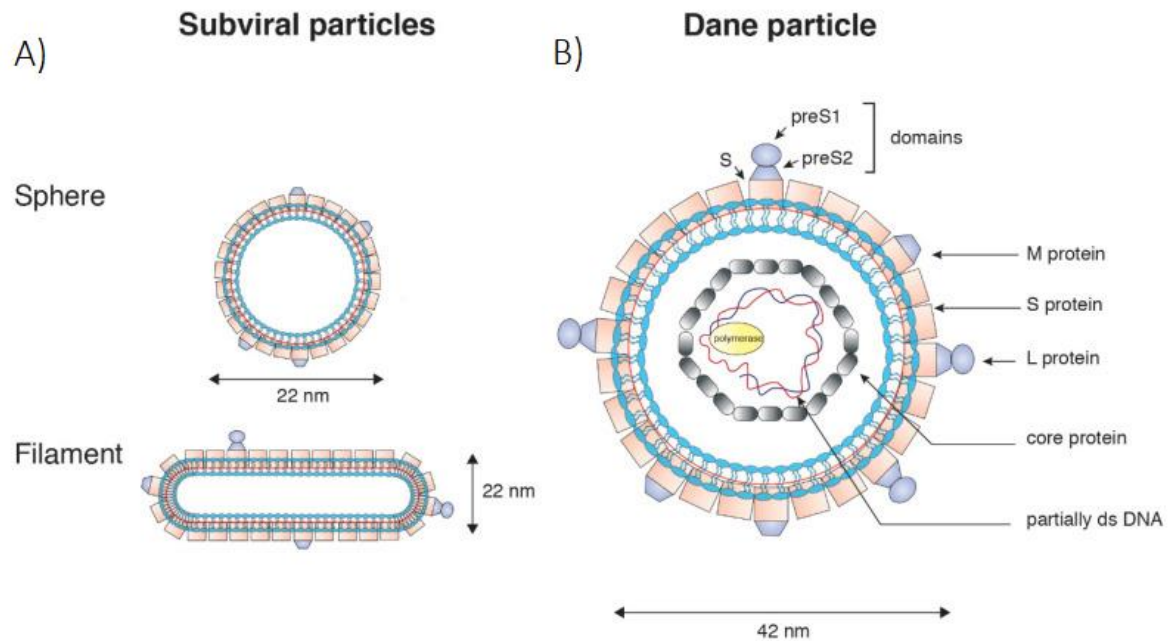
# 1. INTRODUCTION

## 1.1. Hepatitis B virus

The hepatitis B virus (HBV) is the etiological agent of hepatitis B and can cause both acute and chronic liver disease. In 2019, World Health Organization estimates that 296 million people worldwide were estimated to suffer from chronic hepatitis B (CHB) infection, with 1.5 million new infections each year [1]. It is estimated that around 15-40% of CHB patients will develop liver cirrhosis, liver failure or hepatocellular carcinoma (HCC), the major complications of CHB, with mortality rate of around 15-25% [1]. There exists a safe and effective vaccine against HBV that offers 98% to 100% protection. Nevertheless, few of all 296 million people living with the virus are diagnosed and not enough children are vaccinated against it [4]. Africa is the least vaccination region against HBV, and the Western Pacific Region the most [4].

## 1.2. History of hepatitis B virus

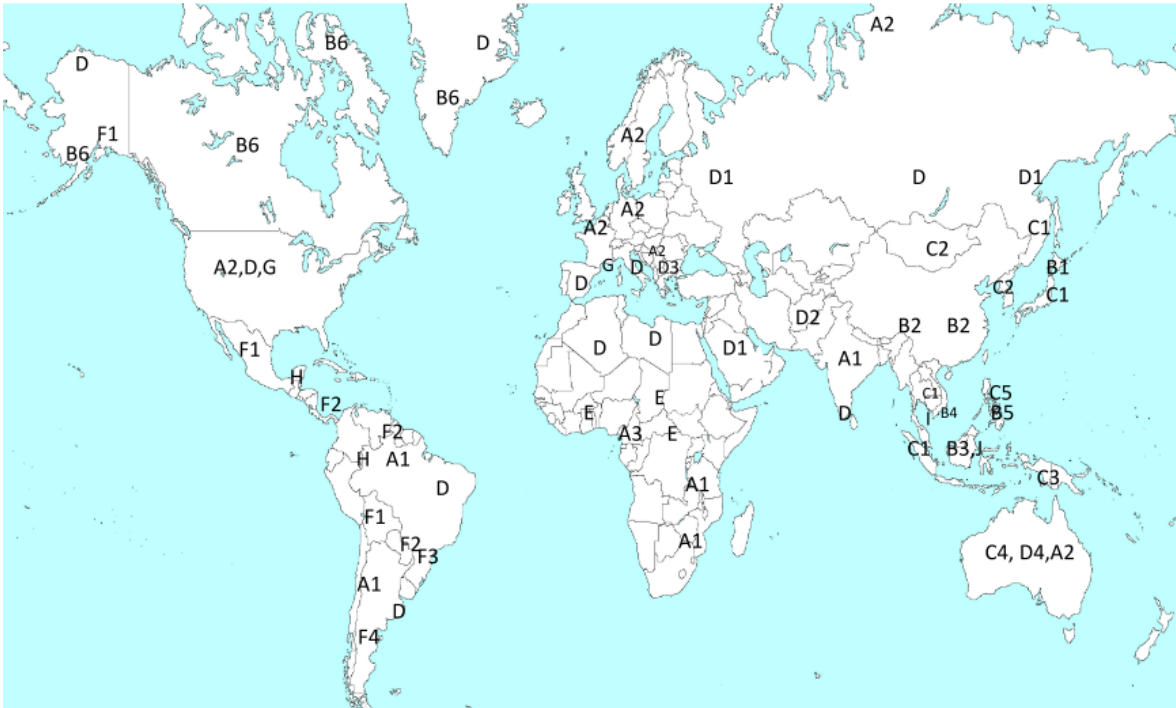
HBV was discovered in 1965 by Nobel Prize winner Dr. Baruch Blumberg (1925-2011) originally called as "Australia Antigen" because it was isolated from an Australian aborigine's blood sample [5]. Later, Australia Antigen began to be related to an infectious agent capable of causing viral hepatitis in man [6]. Bayer et al. [7] studied Australia Antigen with electron microscopy by means of the isolation method established by Alter and Blumberg [8]. They observed small and spherically shaped isolated particles about 22 millimetres (nm) in diameter (**Figure 1; A**), and other with the same diameter but filamentous in shape (**Figure 1; A**). Spherical and filamentous particles are known today as subviral particles (SVPs), defined as non-infectious particles lacking a genome-containing capsid [9]. Later, Dane, Cameron, and Briggs recognized a larger particle about 42 nm in diameter with an electron-opaque core of about 27 nm [10]. It represents the whole virus particle (known as "Dane particles," **Figure 1; B**). Finally, Australia Antigen was seen to be on the surfaces of both virus-like particles associated with HBV (22 nm and 42 nm), which is what is known today as the hepatitis B surface antigen (HBsAg) [11]. It was not until 10 years that the entire HBV genome was sequenced [12] and viral structure and cycle replication was understood [13].



**Figure 1. Representation of hepatitis B virus particles.** Figure available from Herrscher C., Roingeard P., and Blanchard E [14].

### 1.2.1. Hepatitis B virus taxonomy and genotypes/sub-genotypes

HBV belongs to Hepadnaviridae family, which includes five genera: *Parahepadnavirus*, *Metahepadnavirus*, *Herpetohepadnavirus*, *Avihepadnavirus* and *Orthohepadnavirus* [15]. Based on genome sequences, HBV has been phylogenetically classified into 10 genotypes (A-J), which differ from each other by 7.5-15% nucleotide (nt) differences [16]. Genotype J is considered as a putative genotype as it has only been isolated from one individual [17]. Genotypes A-D, F, H, and I are further classified into around 40 sub-genotypes with 4-8% genome divergence [17]. HBV genotypes have a distinct geographical distribution, as depicted in **Figure 2**. Genotype A is commonly encountered in Europe, North America, Southeast Africa, and India. Genotypes B and C are found in Asia and Oceania. Genotype D, the most widespread, is prevalent in North America, North Africa, Europe, the Middle East and Oceania. Genotype E is found in West Africa. Genotype F is dominant in South and Central America. Genotype G is reported in France and the United States. Genotype H is present in Central and South America [18]. Genotype I is present in Vietnam and Laos. Finally, genotype J has only been reported in a single person from the Ryukyu Islands of Japan [19].



**Figure 2. Hepatitis B virus genotypes and sub-genotypes distribution throughout the world.** Hepatitis B virus sub-genotypes are named with the genotype letter followed by a number *Figure available from N. Rajoriya, C. Combet, F. Zoulim et al. [18].*

The clinical importance of HBV genotype lies in the fact that each one is associated with different clinical outcomes, disease progression, and response to antiviral treatment [20][21]. The **Table 1** compares clinical and virological characteristics among HBV genotypes. For example, genotype A is common in CHB patients [20]. Genotype B is more common in chronically infected individuals, previously defined as asymptomatic carriers due to the absence of liver lesions, but also in patients with liver failure. Genotype C is associated with late hepatitis B e antigen (HBeAg) seroconversion and advanced liver disease [22]. Genotype C has a higher incidence of liver cirrhosis and HCC than genotype B [23]. Genotype D is predominant in patients with acute hepatitis [20]. Genotype H is characterised by low levels of HBV-DNA and HBeAg, and by occult HBV and less severe liver disease [24].

<b>GENOTYPE</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E-J</b>
<b>Transmission</b>	Horizontal	Perinatal	Perinatal	Horizontal	Horizontal
<b>Tendency of chronicity</b>	Higher	Lower	Higher	Lower	ND
<b>Positivity of HBeAg</b>	Higher	Lower	Higher	Lower	ND
<b>HBeAg seroconversion</b>	Earlier	Earlier	Later	Later	ND
<b>HBsAg seroclearance</b>	More	More	Less	Less	ND
<b>Histological injury</b>	Lower	Lower	Higher	Higher	ND
<b>Clinical outcomes</b>	Better	Better	Worse	Worse	Worse in genotype F
<b>Response to INF-<math>\alpha</math></b>	Higher	Higher	Lower	Lower	Lower in genotype G
<b>Response to nucleos(t)ide analogs</b>	No significant differences among genotypes A to D				ND
<b>HBV-DNA load</b>	ND	Lower	Higher	ND	ND

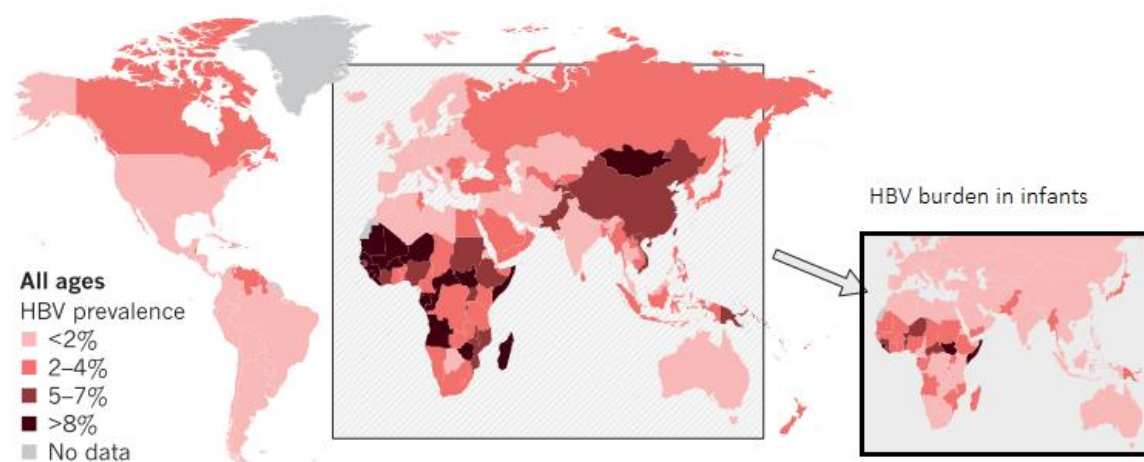
**Table 1. Comparison of clinical and virological characteristics among hepatitis B virus genotypes.** Table adapted from Sunbum M [25].

### 1.2.2. Hepatitis B virus epidemiology and prevalence

HBV is mainly transmitted from an infected mother to her new born during at birth (perinatal transmission) or by horizontal transmission. The latter includes both contact with infected blood through exposure to needles and other contaminated or unsterilised instruments, and contact with infected body fluids (sexual transmission) associated with unprotected sex,

such as saliva and menstrual, vaginal and seminal fluids [1]. HBV can live on the surface of an inert material for at least 7 days. During this time, HBV can cause infection, especially in unvaccinated people [1].

The global prevalence and burden of HBV infection is represented in **Figure 3**. The highest burden of HBV infection is in Western Pacific and African regions, with 116 million and 81 million chronically infected people, respectively. These parts of the world are followed by the Eastern Mediterranean region with 60 million infected people, South-East Asia region with 18 million, European region with 14 million and Americas region with 5 million [1].



**Figure 3. Global prevalence of hepatitis B virus infection.** Despite HBV is most prevalent in Africa and Western Pacific Regio, HBV burden in infants is clearly dominant in Africa (inset). *Abbreviations: HBV, hepatitis B virus. Figure adapted from I. Graber-Stiehl [4].*

### 1.3. Hepatitis B virus genome composition

The infectious form of HBV, the Dane particle, consists of a nucleocapsid formed by the hepatitis B virus core antigen (HBcAg), which encloses the viral genome bound by polymerase (**Figure 4**). The viral nucleocapsid is enveloped by a lipoprotein coat composed of cellular lipids and HBsAg [26]. The enclosed viral genome is partially double-stranded circular DNA (dsDNA), also known as relaxed circular DNA (rcDNA), 3.2 kilobases (kb) in length, with a complete strand (minus) and an incomplete strand (plus) of variable length. The minus strand is characterised by viral polymerase covalently bound to its 5' end and by a terminal redundancy sequence consisting of a ten-nt DNA flap [27]. The plus strand is distinguished by a 5' capped RNA primer and a single-stranded DNA (ssDNA) gap [27]. The viral genome contains four highly overlapping open reading frames (ORF) that encode the different viral proteins (**Figure 4**):

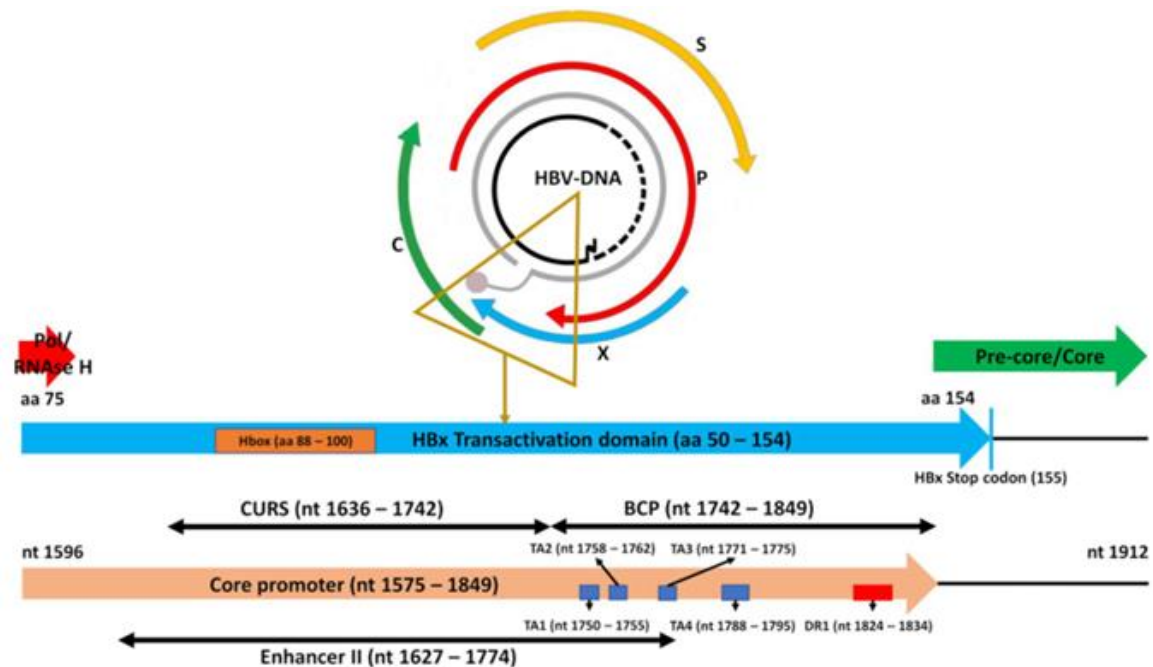


**Figure 4. The overlapping genome structure of hepatitis B virus.** The numbers indicate the localization of first and last nucleotides for each open reading frame. *Abbreviations: dsDNA, double-stranded DNA; ORF, open reading frame; TP, terminal protein; RT, reverse transcriptase; RH, RNase H; GRE, glucocorticoid-responsive element; ENH1, enhancer 1; ENH2, enhancer 2; DR1, direct repeat 1; DR2, direct repeat 2; poly A, polyadenylation.* Figure adapted from Garcia-Garcia, S, et al [26].

- **P ORF:** It encodes the viral polymerase, which is covalently linked to the 5' end of the minus strand [28].
- **preS/S ORF:** It completely overlaps ORF P and encodes the three envelope proteins forming the viral envelope.
- **preC/C ORF:** It encodes both the HBcAg and HBeAg, a non-structural protein secreted in the serum of patients with a high virus load, indicating high infectivity [29]. In addition, HBeAg is essential for defining the HBV infection status (see section 1.6. *Natural history of chronic hepatitis B infection*).

- **X ORF:** It encodes the hepatitis B X protein (HBx), a viral oncoprotein that acts as a promiscuous transactivator and is strongly associated with the development of HCC [30].

The HBV genome also contains two direct repeats: DR1, which located within the preC ORF, is the initiation site for minus strand DNA synthesis, and DR2, which is located where the P and X ORFs overlap, is the initiation site for plus strand DNA synthesis. These repeated sequences are highly conserved in all mammalian hepadnavirus genomes [31]. It also contains two enhancer elements: ENH1, which is located within the P ORF and acts as the promoter of the X ORF and ENH2, which is located within the X ORF, overlapping the core promoter (CP) and acts as the promoter for both the preC/C ORF and the P ORF [32]. The CP, which overlaps with the 3' end of the X ORF and the 5' end of the preC region, controls the transcription of pregenomic RNA (pgRNA), an intermediate in viral replication that is translated to HBcAg and the viral polymerase, and precore RNA (preC RNA), which is translated into HBeAg. The CP includes the core upstream regulatory sequence (CURS) and the basic core promoter (BCP). It also contains sequence and structural motifs essential for viral replication such as the four TATA-like boxes (TA) 1 to 4, the ENH2, and the DR1 (Figure 5) [33].



**Figure 5. Core promoter region of hepatitis B virus genome.** Abbreviations: HBx, hepatitis B X protein; CURS, core upstream regulatory sequence; BCP, basal core promoter; DR1, direct repeat 1; TA, TATA-like boxes. Figure available from Garcia-Garcia, S, et al [33].

A glucocorticoid-responsive element (GRE) is located where the P and preS/S ORFs overlap (**Figure 4**). It acts by increasing the transcriptional level of a given gene upon hormone receptor binding. The U5-like region, which is located between the preC/C ORF, is the most highly conserved sequence in the HBV genome. Finally, the polyadenylation (poly A) signal sequence, which is located within the C ORF, is involved in the addition of the poly A tail to all mRNAs [31].

### 1.3.1. preS/S ORF: hepatitis B virus envelope

It encodes the three envelope proteins: large (L-HBsAg), middle (M-HBsAg) and small (S-HBsAg), which are translated from 2.4/2.1 kb mRNAs and transcribed from both cccDNA and the integrated viral genome [34]. The HBV virus can randomly integrate into the host genome as a replication-deficient form of the virus, allowing only the expression of HBsAg [35][36]. Because integrated HBV-DNA also produces HBsAg, HBsAg may not be an appropriate marker of cccDNA transcriptional activity. The L-HBsAg is translated from preS1 + preS2 + S domains (400 amino acids (aa) in length); the M-HBsAg from preS2 + S domains (281 aa) and the S-HBsAg from the S domain (226 aa) only [34]. The S-HBsAg drives the release of viral particles, M-HBsAg enriches the secretion of the virion, and L-HBsAg is involved in the interaction with HBcAg in the packaging of the virion at the endoplasmic reticulum (ER) where envelope proteins are synthesised [37]. The C-terminal S domain, which is common to all three envelope proteins, contains four transmembrane domains and the “a” determinant, the main target of HBsAg antibodies (anti-HBs), in the major hydrophilic region (MHR) [30][34]. The HBsAg variants with clinical significance are mainly located at the dominant epitope of HBsAg within the “a” determinant in the MHR [38]. Furthermore, in the preS1 domain, which is only present in L-HBsAg, the glycine 2 residue is myristoylated, a modification considered essential for hepatocyte infection [34]. Thus, the HBV life cycle begins with viral attachment to hepatocytes, mediated by L-HBsAg (see section “1.4. Hepatitis B virus life cycle”).

As mentioned above, a peculiarity of HBV is its ability to produce not only infectious virions, but also empty non-infectious SVPs formed by different forms of surface proteins. In the absence of L-HBsAg, only S-HBsAg is secreted as non-infectious SVPs. L-HBsAg can abolish SVP secretion depending on the L-/S-HBsAg ratio [39]. The L-HBsAg, M-HBsAg and S-HBsAg are reported to be present in the envelope of an infectious Dane particle in a ratio of 1:1:4 [40]. Spherical SVPs are composed of 90% S-HBsAg and 10% M-HBsAg, whereas filamentous SVPs contain 80% S-HBsAg, 10% M-HBsAg, and 10% L-HBsAg [14] (**Figure 1**). SVPs are secreted in large excess relative to infectious Dane particles [41]. During natural infection with HBV, these empty non-infectious SVPs are secreted in excess,



approximately 1,000 to 100,000 fold higher than the infectious particles [42]. There are several theories as to the biological function of this excess. SVPs are immunogenic and therefore interact with host antibodies, facilitating the path of infectious virions to hepatocytes [42]. They may also contribute to the “immune-tolerant” phase, which is the early phase of persistent infection (see section 1.6. *Natural history of chronic hepatitis B infection*).

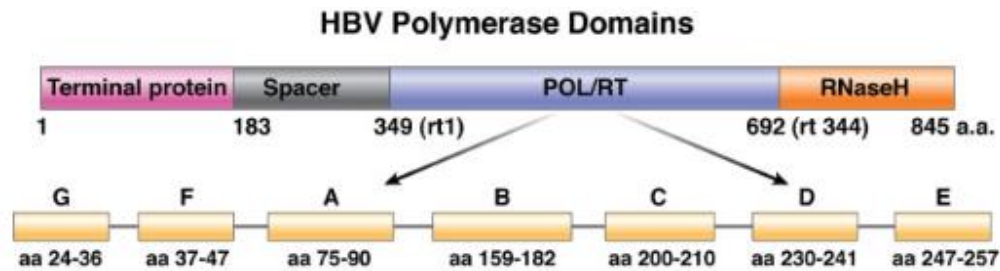
Interestingly, the RT domain of the P ORF is overlapped by the *preS/S* gene. Consequently, some substitutions in the *preS/S* gene, which may occur both naturally and therapeutically, could lead to a serious event in terms of RT drug susceptibility [43]. In addition, mutations in the *preS/S* gene could also result in deletion of HBsAg or synthesis of variants of truncated proteins, which are commonly found in CHB patients. Only the C-terminal truncation of L-HBsAg and M-HBsAg have transactivating properties [44]. For this reason, the relationship between *preS/S* gene mutations and HCC has been clearly investigated in the preS region (including both preS1 and preS2 domains), the most variable region of the HBV genome [45]. In particular, these mutations may be immune-selected, triggering a dysregulated synthesis of L-HBsAg, which accumulate in the ER of the hepatocytes [46][37] because they can neither be properly folded [47] nor efficiently secreted. This event may activate ER stress-signalling pathways with consequent induction of oxidative DNA damage and genomic instability, significantly increasing the risk of HCC development in CHB patients [46][47].

### 1.3.2. *preC/C* ORF: hepatitis B virus core antigen and hepatitis B virus e antigen

The *preC/C* ORF encodes the HBcAg and HBeAg proteins. It is partially overlapped by the X and P ORFs and its transcription is regulated by the Cp. The Cp allows the formation of two 3.5 kb mRNAs: the pgRNA and the preC RNA. The *preC/C* ORF consists of two translation start codons in the same reading frame [48]. When transcription starts from the first start codon of the *preC/C* ORF, the preC protein or the HBeAg precursor p25 is synthesised from the preC RNA translation. Through proteolysis events, p25 leads to the HBeAg precursor p22, which eventually forms HBeAg, also known as p17 [48]. On the other hand, if transcription starts from the second start codon, the HBcAg (p21) is synthesised from the C ORF on the pgRNA translation [48].

### 1.3.3. P ORF: hepatitis B virus polymerase

The P ORF encodes the RNA-dependent DNA viral polymerase, which consists of four domains [49] (**Figure 6**):



**Figure 6. Schematic of the hepatitis B polymerase open reading frame illustrating the four functional domains and the seven catalytic subdomains (A-G).** Domains A, C and perhaps D participate directly in dNTP binding and catalysis. Domains B and E are involved in the precise positioning of the primer-template relative to the active site. *Figure from Ghany, M and Liang, T.J [50].*

- 1) The terminal protein (TP) domain.
- 2) The polymerase/reverse transcriptase (RT) domain.
- 3) The C-terminal RNase H (RH) domain,
- 4) The spacer domain between the RT domain and the TP domain.

Human immunodeficiency virus (HIV) and HBV RT share approximately 20% similarity in the RT domain [51], and 33% similarity in the RNase H domain [52]. However, the TP and spacer domains of the HBV polymerase are unique and have no structural homologues.

The TP domain, located at the N-terminal of the viral polymerase, is an essential component due to the presence of a tyrosine residue to which the first nt of the newly synthesised minus strand of DNA from pgRNA is attached during the replication process [49]. The RT and RH domains provide the DNA polymerase and the RNA hydrolysis activities, respectively to the viral polymerase. The spacer domain is thought to make the TP domain more flexible, facilitating its interaction with the viral RNA element [49] ( $\epsilon$  signal, see section 1.4.1. *Molecular mechanisms of viral polymerase involved in reverse transcription*).

The catalytic region can be subdivided into seven domains: A-G (**Figure 6**) [50]. Domain A and C are part of the dNTP binding pocket. The catalytic centre of RT included in domain C contains a sequence of four aa: tyrosine (Y), methionine (M), aspartate (D) and aspartate (D) (YMDD motif), which represents the active site of the enzyme [30]. The YMDD motif, which is involved in nt binding in the catalytic site of the polymerase [53], is highly conserved among viral RT [30]. Domain B and E are involved with positioning of the primer-template strand to the catalytic region [50]. Mutations in the domain D may indirectly affect the geometry of the dNTP binding site, since residues of this domain lie outside dNTP binding site. Domains F and G may be involved in interactions with the incoming dNTP and with the template nt.

The viral polymerase lacks proof-reading ability and introduces approximately  $3.2 \times 10^{-5}$  –  $7.9 \times 10^{-5}$  nt substitutions per replication cycle, contributing to an estimated error rate of  $\sim 1$  per  $10^4$ – $10^5$  bp for minus-strand DNA synthesis and  $\sim 1$  per  $3.6$ – $15 \times 10^4$  bp for second-strand DNA synthesis [54]. In addition, HBV is a highly replicative virus, with an average of  $10^{11}$  viral particles produced per day in an infected liver [55] and 200-1000 virions per day in each cell during acute infection [56]. As a result, HBV mutates at a rate similar to that of RNA viruses [30]. Those mutants with sufficient fitness, or those mutants resistant to immune or antiviral pressure, are selected and characterise the viral population.

HBV resistance to lamivudine, a RT inhibitor for the treatment of CHB (see section 1.6. *Current hepatitis B virus therapy*), has been reported to have aa substitutions in the M residue of the conserved YMDD motif, showing increased replication competence compared to the wild type and resistance to lamivudine [53]. Since antiviral treatment affects this region of the genome, we focused our second study on investigating the evolution of QS in the short term under the evolutionary pressure of treatment and without treatment.

#### 1.3.4. X ORF: hepatitis B virus X protein

HBx, encoded by the smallest ORF of HBV, consists of 154 aa [57]. This protein is characterised by surprisingly pleiotropic function due to its direct interactions with multiple cellular proteins [33]. Like HBsAg, the subgenomic RNA encoding HBx can also be transcribed from integrated HBV-DNA. As a result, either the C-terminal domain of *HBX* can be deleted, resulting in C-terminal deletion mutants of HBx with aberrant function, or the viral enhancer upstream of *HBX* can be placed in the proximity of cellular genes that contribute to carcinogenesis [58]. The X RNA forms the 3' end of all viral transcripts [59], so perturbation at this level could abolish the synthesis of all the viral antigens.

HBx has an outstanding role in regulating the HBV life cycle, host-virus interactions, and HBV-related HCC. For example, HBx may bind to CUL4-DDB1-SMC forming the HBx-CUL4-DDB1-SMC complex by inhibiting the expression of SMC5/6 needed for the structural maintenance of chromosomes (SMC). SMC5/6 is a host restriction factor that suppresses HBV transcription from the cccDNA template that cannot act due to the action of HBx. Furthermore, HBx is able to facilitate the interaction of stimulating cellular transcriptional factors to the cccDNA, in order to regulate the transcription of host genes, disrupt protein degradation, modulate signalling pathways, manipulate cell death, and deregulate the cell cycle [60]. HBx is considered as the major carcinogenesis-related protein in HBV infection. HBx is required for promoting viral replication by epigenetic stimulation of cccDNA transcription [61]. The presence of *HBX* transcripts encoding HBx have been observed in

the supernatant of infected primary hepatocytes and in the plasma of CHB patients [62], suggesting that the early expression of the HBx at the time of infection is dependent on the extracellular presence of such a transcript [62]. This early expression is necessary for the degradation of the Smc5/6 complex, which restricts viral replication by inhibiting HBV gene expression [63]. The  $\alpha$ -helical motif termed the H-box in the HBx targets SMC5/6 complex for ubiquitination and degradation by the DDB1-CUL4-ROC1 E3 ubiquitin ligase [64]. In addition, the presence of a population of shorter *HBX* transcripts, which may originate from an alternative promoter located in the protein coding region, has also been reported [65].

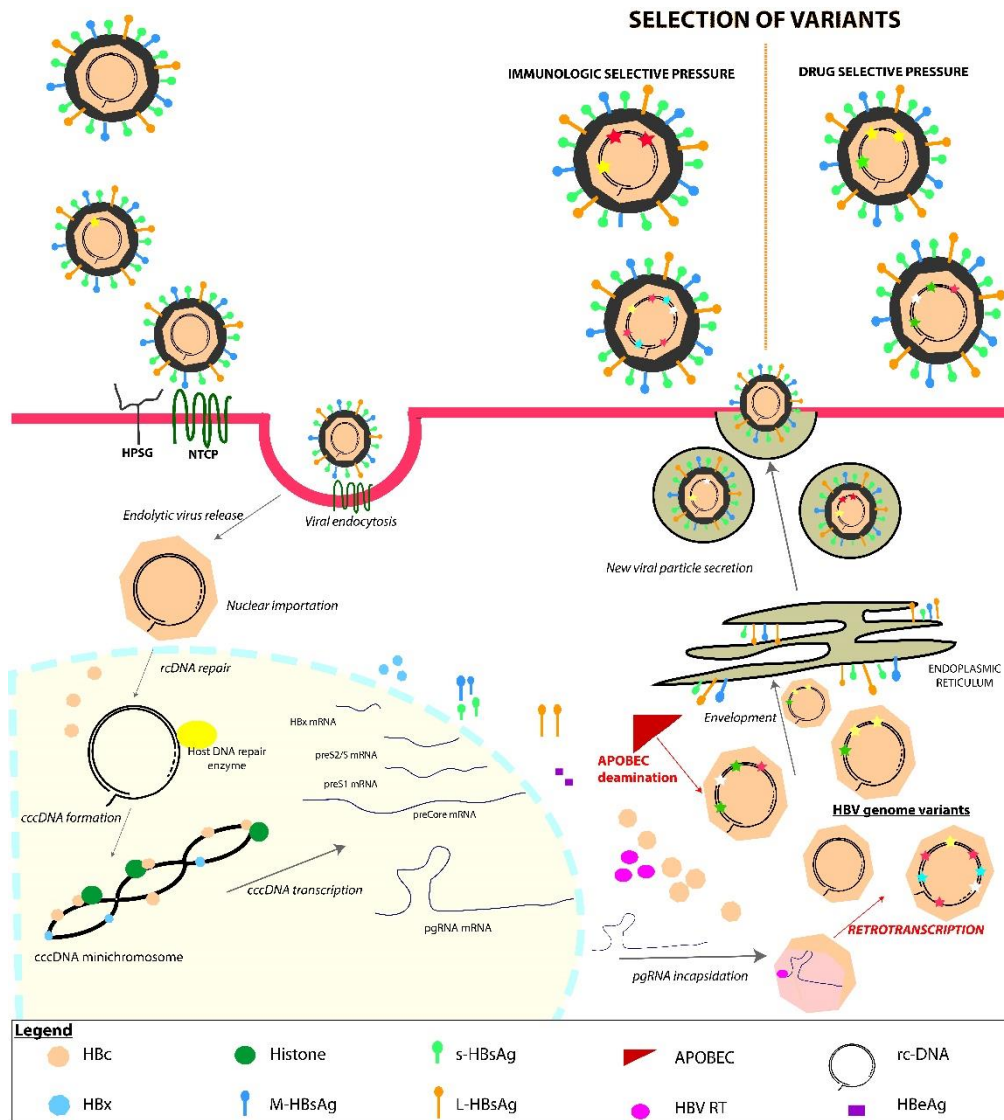
HBx is located in both the cytoplasm and the nucleus of hepatocytes [66][67][68] depending on its expression level: HBx mainly accumulates in the cytoplasm when highly expressed, whereas nuclear localisation dominates when HBx is low expressed [67][69]. HBx is targeted to the nucleus by a nuclear localisation signal (NLS) and exported by a nuclear export signal (NES) [70]. Nonetheless, it is not well understood whether the intracellular localisation of HBx is responsible for promoting viral replication. It is thought that cytoplasmic HBx induces stimulation of the viral genome. Several controversial *in vitro* studies on the function of HBx have been published due to the variability in its expression levels. Klein et al., added an NLS to the HBx (NLS-HBx) and showed no support for viral genome replication [71]. In contrast, Tang et al. suggested that nuclear HBx is responsible for the activation of viral replication [72]. Leupin et al. also reported the pivotal role of nuclear HBx by adding both NLS-HBx and NES-HBx. While NLS-HBx activated viral replication, NES-HBx was unsuccessful [73]. In the nucleus, HBx would act as a promiscuous transactivator, regulating transactivation through a direct interaction with several transcription factors. Furthermore, cytoplasmic HBx has also been associated with mitochondrial dysfunction in hepatocytes [74]. Differences remain in the relative contribution of cytoplasmic and nuclear HBx to the stimulation of viral replication.

With respect to HBV replication, HBx promotes viral transcription from the cccDNA template and enhances viral polymerase activity [75]. The importance of HBx in transcription and replication has been demonstrated using HBx-deficient woodchuck livers, which are unable to establish a productive HBV infection [76][77]. Conversely, transfection with exogenous HBx restores transcriptional and replication properties [78][79][80]. HBx is also involved in pgRNA encapsidation by stimulating HBcAg phosphorylation [81]. Structurally, HBx has an N-terminal negative regulatory domain (from aa 1 to 50) and the C-terminal transactivation domain (from aa 51 to 154) [82]. Variability in the HBx transactivating C-terminal domain has been described with multiple insertions and deletions. This HBx C-terminal domain plays a key role in controlling multiple functions due to its transactivating activity [83]. This

region of HBx is encoded by the *HBX* 3' end region, where deletions (Del) have been detected more frequently in the tumour than in the adjacent non-tumour tissues of HCC patients [84]. These Del, along with insertions (Ins) in this region of *HBX*, can result in significantly altered HBx due to *HBX* frameshifts. In addition, these Indels may not only affect HBx, but also the properties of the many important regulatory and structural motifs (CP, ENH2, DR1, etc.) that overlap with it. Nonetheless, given the important role of the HBx protein in viral replication, it would be reasonable to assume that the gene encoding this protein would have a conserved region. In this context, the 3' end region was excluded and we focused our first study on the 5' end region of HBX and its upstream non-coding region (nt 1255-1611).

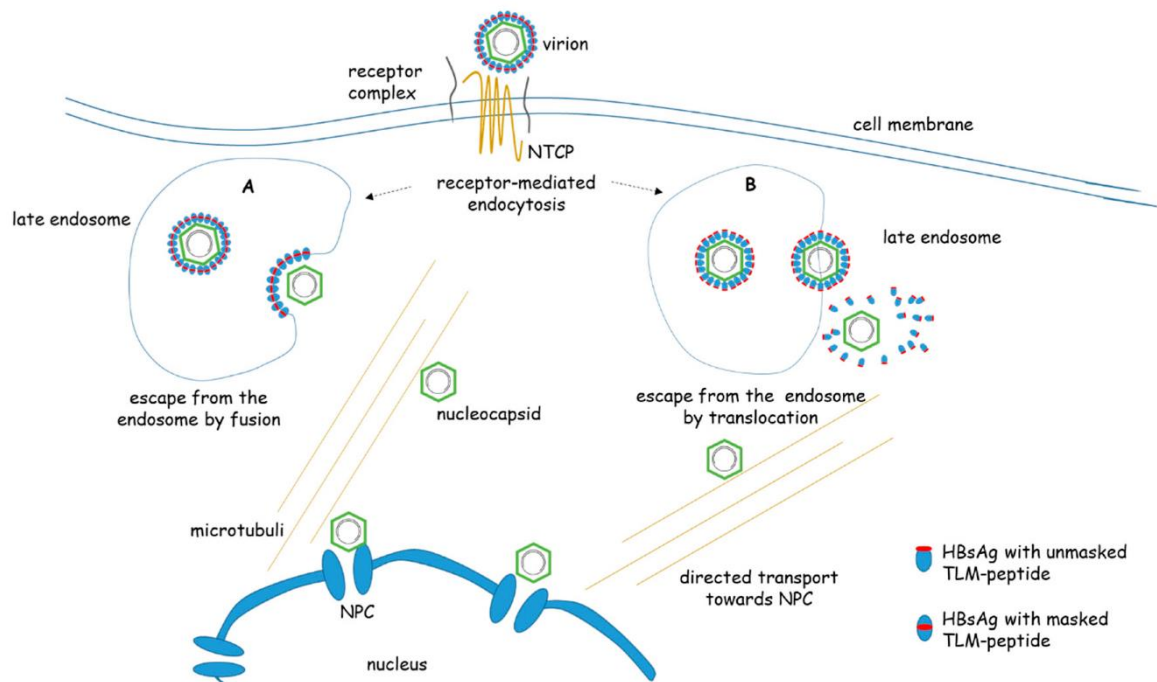
#### 1.4. Hepatitis B virus life cycle

The L-HBsAg protein is crucial for viral entry into the hepatocyte. Entry is initially initiated by low-affinity binding of HBV to heparin sulfate proteoglycans (HSPGs), which are found on the cell surface and in the extracellular matrix of almost all cells [14] (**Figure 7**). High-affinity binding then occurs between the N-terminal aa in the preS1 domain and the hepatocyte surface bile acid transporter sodium taurocholate co-transporting polypeptide (NTCP) to mediate viral entry via endocytosis [9][27].



**Figure 7. Schematic representation of hepatitis B virus life cycle.** Figure available from Garcia-Garcia, S, et al [26].

Two models have been proposed for the escape of HBV from the late endosome [85] (**Figure 8**): i) membrane fusion leading to the release of the nucleocapsid into the cytoplasm or ii) translocation across the endosomal membrane into the cytoplasm mediated by a translocation motif (TLM) present in HBsAg that is unmasked during proteolytic processing in the endosomal compartment. The viral nucleocapsid containing the HBV genome, rcDNA, is then released into the hepatocyte cytoplasm. The nucleocapsid is actively transported by microtubules to the nuclear pore complex (NPC), where the rcDNA is released into the nucleoplasm.



**Figure 8. Speculation of two models on post-entry processing of hepatitis B virus. A)** By membrane fusion and, consequently, the nucleocapsid is released into the cytoplasm. **B)** By translocation into the cytoplasm, a process involving surface proteins. Finally, nucleocapsid is actively transported to nuclear pore complex. *Abbreviations: NTCP, sodium taurocholate co-transporting polypeptide, NPC, nuclear pore complex; HBsAg, hepatitis B surface antigen; TLM, translocation motif. Figure from Bingfu Jiang and Eberhard Hildt [85].*

Once in the nucleus, the ssDNA gap in plus strand is repaired by nuclear host DNA repair enzyme [86] converting the rcDNA into a complete double stranded covalently closed circular DNA molecule (cccDNA) [14]. The cccDNA is compacted with viral proteins (HBx and HBcAg) and cellular histones, resulting in a kind of viral ‘minichromosome’ that remains in the nucleus throughout the entire life of the infected hepatocyte and acts as a template for all HBV viral mRNAs [35][87]. Like HBx, HBcAg is a nucleocytoplasmic shuttling protein, containing both NLS and NES [88][89]. Two types of viral mRNAs are produced by host RNA polymerase II: the subgenomic mRNAs and the pg/preC RNA. The former includes 3 viral transcripts, 2.4kb, 2.1kb and 0.7kb, which are translated into L, M, S-HBsAg and HBx, respectively, while the latter includes longer and shorter 3.5kb transcripts (pgRNA and preC RNA, respectively) encoding either HBcAg and viral polymerase or HBeAg, respectively [90]. All 5 viral mRNAs share the same 3’ end and are polyadenylated using the polyA signal. The pgRNA is encapsidated into newly formed capsids and RT by the viral polymerase into a new rcDNA molecule (or double stranded linear DNA – dsIDNA) [91]. Viral nucleocapsids dsIDNA-contained can be transported to the nucleus contributing to the further replenishment of cccDNA pool via homologous recombination [35]. Furthermore,

dsDNA is the major substrate for HBV-DNA integration [92]. Integration of HBV-DNA sequences into the host genome as well as prolonged production of HBx and HBsAg and/or of modified forms of HBx and preS/S envelope proteins may contribute to the persistence of viral infection and liver damage via multiple mechanisms, ultimately leading to the development of liver cancer [92]. In terms of retrotranscription, during this process the HBV genome can acquire some mutations (represented by different coloured-stars in HBV rcDNA in **Figure 7**) due to the lack of proof-reading activity of the viral polymerase increasing the pool of genomic mutations. A maturation signal in the nucleocapsid must be detected by the envelope machinery to prevent the release of immature nucleocapsids [85]. The difference between mature and immature nucleocapsids lies in the phosphorylation/dephosphorylation status of the C-terminal domain (CTD) of HBcAg, where several serine phosphorylation sites are found [85]. Phosphorylation of the CTD alters the structure of the nucleocapsid, generating a particular structural signal that triggers RNA packaging and RT. Otherwise, dephosphorylation of the nucleocapsid again alters its structure, this time promoting envelopment and release of viral particles [85]. A high level of CTD phosphorylation would impair pgRNA packaging, resulting in empty virions lacking the HBV-DNA genome [93]. Mature nucleocapsids cross the ER to acquire their envelope. Complete virions are then formed and released by budding into multivesicular bodies (MVBs) via the endosomal sorting complex required for transport (ESCRT) pathway. Alternatively, nucleocapsids containing rcDNA can also be transported into the nucleus to replenish cccDNA [94]. Based on cell-culture experiments, HBV-DNA integration into the human genome is an early event that can occur within a week after HBV infection [95], although it can also occur throughout CHB infection. It occurs randomly and can cause genetic damage and chromosomal instability promoting carcinogenesis [35]. In the case of HBeAg-negative CHB patients, HBV-DNA integration may be the cause of HBsAg production, which occurs despite complete cccDNA silencing, thus impeding the so-called “functional cure,” the surrogate marker of which is HBsAg loss [96]. The *in vitro* HBV-DNA integration rate is reported to be 1 integration per  $10^4$  infected cells in the early viral life cycle and remains stable between 3 and 9 days post infection [95]. After HCC development, HBV-DNA integration is more common in tumour liver tissue (86.4%) than in nearby liver tissue (30.7%) [97].

Other enveloped nucleocapsids containing dsDNA, HBV-RNA, or even lacking HBV-DNA genome (empty virions) can also be formed in the HBV life cycle and be released as enveloped virions [98]. In addition, the presence of other HBV-RNAs, such as spliced pgRNA variants and *HBX* transcripts, has been reported in patients [59]. Non-enveloped



nucleocapsids (naked nucleocapsids) have also been described [99]. In terms of subviral particles, filamentous particles are secreted via the same pathway as virions, whereas sphere subviral particles are released via the secretory pathway [85]. Non-enveloped particles are thought to be released in an ESCRT-independent pathway [100].

All these features – the different circulating particles, the persistence of cccDNA and integrated DNA – make HBV infection a highly dynamic process that can lead to various hepatic complications.

#### 1.4.1. Molecular mechanisms of viral polymerase involved in reverse transcription

The encapsidation signal is a cis-acting element, known as epsilon signal ( $\epsilon$ ), a stem-loop structure located at the 5' end of the pgRNA, in the preC region. The interaction between the viral polymerase and the  $\epsilon$  stem-loop triggers both encapsidation and RT with help of cellular chaperones [49]. The  $\epsilon$  structure is therefore essential for viral replication. The viral polymerase uses the  $\epsilon$  bulge as template to synthesise the first 3-4 nt of the minus strand DNA using the tyrosine residue of the TP domain as a primer. The viral polymerase with these 3-4 nt covalently is then translocated from  $\epsilon$  at the 5' end of the pgRNA to a complementary sequence of the DR1 element located at the 3' end of the pgRNA [101]. Thanks to the interaction of  $\epsilon$  with phi signal ( $\phi$ ) located at the 3' end of the pgRNA, translocation of the viral polymerase occurs [102]. After translocation, the RT domain of the viral polymerase completes the synthesis of the minus strand DNA. As the minus strand DNA is synthesised, the pgRNA is degraded by the RH domain, leaving approximately 17 nt of the 5' end of the pgRNA including the DR1 region [101]. The 17 nt are translocated to the DR2 element on the minus strand DNA, which serves as a primer for plus strand DNA synthesis. Synthesis of plus strand DNA from DR2 to the 5' end of the minus strand DNA occurs. The 3' end of the plus strand DNA is translocated to the 3' end of the minus strand DNA using the terminal redundancy sequence consisting of a ten nt DNA flap present in the minus strand DNA to complete the circularisation. The plus-strand is then extended to produce the 3.2kb partially sdDNA, i.e., the rcDNA, found in the mature virion [101].

#### 1.5. Hepatitis B virus genetic variability and quasispecies: definition and clinical significance

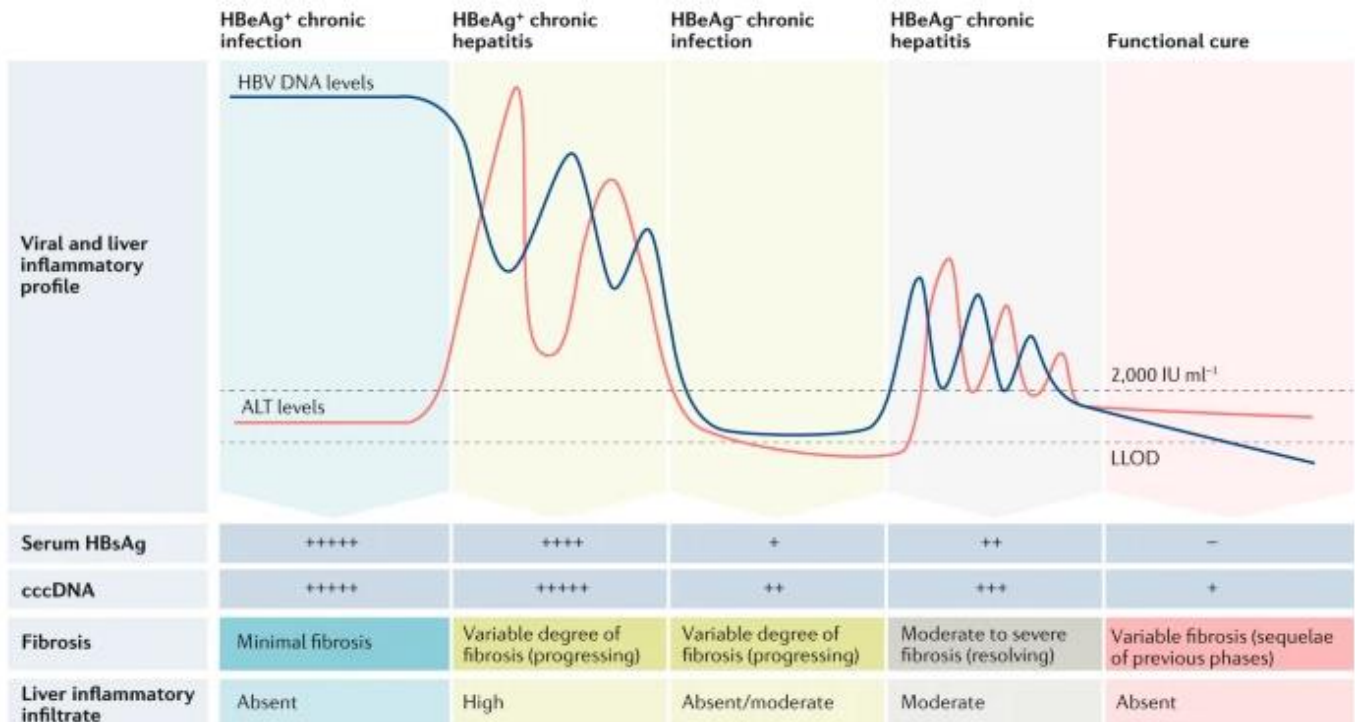
HBV infection is characterised by the high complexity of the viral population. The main source of genetic variability is the viral polymerase, which lacks of proof-reading activity during reverse transcription. Moreover, genetic variability in the viral genome is further increased by other molecular mechanisms. Host DNA and RNA deaminases, which attack viruses for their mutagenic activity, contribute by increasing viral complexity (represented

by different coloured-stars in HBV rcDNA in **Figure 7**) [103][104]. At this point, mature HBV nucleocapsids containing both wild-type and mutant HBV-DNA are secreted as a mixture of closely related genomic variants defined as quasispecies (QS) [105]. A QS is defined as the different genomes (haplotypes) present and their frequencies (the number or fraction of molecules with the same sequence), known as the haplotype distribution [106]. HBV QS are subject to evolution, allowing HBV to adapt to multiple selective pressures, such as the immune system, antiviral treatment and vaccination, and only the fittest are selected [107]. These mutations can be observed in HBV QS during long-term infection, allowing evolution to be studied through changes in haplotype distribution over time [106]. A high degree of variability in the viral population has important implications for epidemiology, diagnosis, pathogenesis and therapeutic management [107].

### 1.6. Natural history of chronic hepatitis B infection

HBV infection is considered acute if the patient is infected for the first time and the infection lasts less than six months. If the infection lasts longer than six months, i.e. if the patient tests positive for HBV six months after the first blood test, the acute infection has not cleared naturally and is considered to be chronic [108]. The risk of HBV becoming chronic depends on the age at which the first infection occurs. If infection occurs at birth, 90% of babies will develop CHB infection. If it occurs in childhood, about 50% of children will become chronically infected. If the first infection occurs in adulthood, up to 10% of people will progress from acute to chronic infection [108][1].

In 2016, the American Association for the Study of Liver Diseases (AASLD) practice guidelines defined CHB as a chronic necroinflammatory disease of the liver caused by persistent HBV infection with HBsAg positivity for more than 6 months. Nonetheless, in 2017, the European Association for the Study of the Liver (EASL) published updated guidelines that take into account a new nomenclature for the phases of HBV infection. According to the EASL 2017 clinical practice guidelines [96], the natural history of CHB infection is divided into five phases according to the interaction between HBV replication and the host immune response. Detection of HBeAg is critical for defining infection status and disease management (**Figure 9**):



**Figure 9. Natural history of chronic hepatitis B infection.** The natural history of chronic hepatitis B is described by five clinical or virological phases of infection and/or hepatitis (according to European Association for the Study of the Liver 2017 clinical practice guidelines [96]). Alanine transferase levels are normal in the chronic infection phases (HBeAg positive or HBeAg negative). During the HBeAg-negative phases, HBsAg is thought to be preferentially produced by integrated viral genome, although cccDNA is detectable. Depending on the balance between viral replication and immune control, patients may move from one phase to another, and these phases are not necessarily sequential. Levels of cccDNA decline over time, but cccDNA is never completely eliminated. However, serum HBsAg levels decrease over time. Viral genome integration events are thought to increase with the duration of infection, as does the proportion of HBsAg. The extent of liver fibrosis observed at each phase is shown. *Abbreviations: -, not detected; +, detected; ++, detected with high-frequency and increasing for each additional + added; HBsAg, hepatitis B surface antigen; cccDNA, covalently closed circular DNA; LLOD, lower limit of detection. Figure adapted from Fanning GC et al [109].*

- 1) **HBeAg-positive chronic infection** (“immune-tolerant” phase). In this phase HBV is actively replicating, so patients in this phase are highly infectious. It is characterised by very high HBV-DNA ( $>10^7$  IU/ml) and HBsAg levels, normal ALT levels ( $<19$  U/L in women females and  $<30$  U/L in men), and no or minimal histological liver damage. This phase is prolonged in those patients infected at birth.
- 2) **HBeAg-positive chronic hepatitis** (“immune-active” phase). It is characterised by high/intermediate HBV-DNA ( $10^4$ - $10^7$  IU/ml) and HBsAg levels, the presence of HBeAg levels, elevated ALT levels, and moderate to severe liver necroinflammation progressing to fibrosis. This stage may occur after several years after the first stage, and patients infected in adulthood enter this stage

earlier. Patients who control HBV and achieve HBeAg seroconversion and HBV-DNA suppression enter the third phase.

- 3) **HBeAg-negative chronic infection** (“inactive carrier” phase). It is characterised by undetectable or low levels of HBV-DNA (<2,000 IU/ml), low levels of HBsAg, the presence of HBeAg antibodies (anti-HBe), normal ALT levels and no or minimal histological liver damage. Spontaneous HBsAg loss and/or seroconversion may occur in 1–3% of cases per year.
- 4) **HBeAg-negative chronic hepatitis** (“immune-reactive” phase). It is characterised by persistent or fluctuating intermediate/high HBV-DNA levels, fluctuating or persistently elevated ALT levels, the absence of HBeAg with detectable anti-HBe, and histological evidence of liver necroinflammation and fibrosis. Most of these patients accumulate mutations in the preC and/or the basal core promoter regions that impair or eliminate HBeAg expression.
- 5) **HBsAg-negative phase** (“occult HBV infection” phase). It is characterised by usually, but not always, undetectable HBV-DNA levels, normal ALT levels, positive antibodies to HBcAg (anti-HBc), serum negative HBsAg with or without detectable anti-HBs. If HBsAg loss occurs before the development of cirrhosis, there is a minimal risk of cirrhosis, decompensation and HCC. However, if HBsAg loss occurs after the development of cirrhosis, a high risk of HCC is reported. In addition to host-related risk factors related to the host (older age, male sex, alcohol consumption, underlying diseases such as diabetes, co-infection with other viruses, smoking, several human leukocyte antigen genetic variants that play a key role in the host immune response to viral infection through antigen presentation [110], etc.), there are also risk factors related to HBV itself, such as certain genotypes (as mentioned in section 1.2.1. *Hepatitis B virus taxonomy and genotypes/sub-genotypes*) and specific mutations that may contribute to liver disease progression and HCC development.  
If this phase ends with a sustained loss of circulating HBsAg and HBV-DNA levels, with or without anti-HBsAg, a functional cure is achieved.

### 1.7. Serum hepatitis B virus RNA: biomarker of hepatitis B infection

Quantification of intrahepatic cccDNA levels and their level of transcriptional activity would be valuable tools for estimating the risk of developing HCC [111][112][113], but requires liver biopsy, which is an invasive technique. Furthermore, liver biopsy requires only a small section of the liver, and HBV is unevenly distributed throughout the tissue. Quantitative PCR (qPCR) protocols exist for cccDNA, but the presence of rcDNA and integrated HBV-DNA in

the infected hepatocyte interferes with accurate cccDNA quantification. The search for plasma biomarkers that allow the study of the evolution of the viral population is important, especially in patients with undetectable viremia, in order to monitor the progression of liver disease.

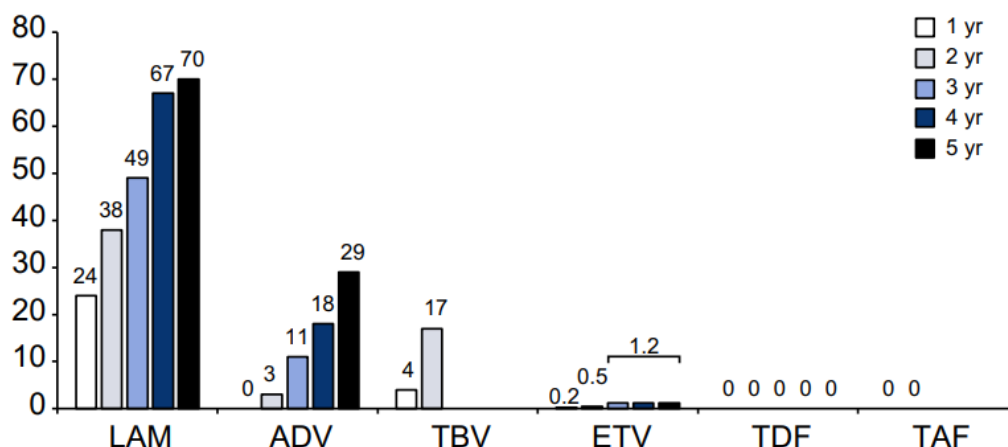
In circulating virus particles, the viral capsid has been reported to contain mainly non-retrotranscribed pgRNA, although it can also encapsulate other forms of RNA, both viral (such as *HBX* transcripts) and cellular [98]. Circulating HBV-RNA is a valid alternative for monitoring patients even in the undetectable viremic phase. Viral RNA does not circulate freely but is found in encapsidated virus-like particles and naked capsids in serum [114]. Its origin is thought to depend on an alteration in the RT process [115]. Unlike serum HBsAg, the 3.5 kb serum pgRNA is produced exclusively from cccDNA, so that it would be a direct evidence of cccDNA transcriptional activity in infected hepatocytes [115][116]. The amount of HBV-RNA-containing particles has been shown to increase under NAs *in vitro* and in transgenic mice [115]. Hence, the amount of serum HBV-RNA should reflect the amount of pgRNA present in the liver and therefore the level of transcriptional activity of cccDNA. Serum HBV-RNA has been described in both HBsAg-positive and seroconverted patients [117]. Recently, it has been suggested that circulating viral RNA may be an early predictor of HBeAg seroconversion in HBeAg-positive patients on nucleos(t)ides analogues (NAs) treatment [118][119][120][121] and treated with pegylated (PEG)-interferon (IFN)-2a [122]. In addition, the kinetics of serum HBV-DNA and HBV-RNA appear to be dissociated during antiviral treatment, with HBV-RNA levels declining more slowly than HBV-DNA. This allows conclusions to be drawn about cccDNA transcription in the absence of detectable serum HBV-DNA, and is therefore a possible predictor of viral rebound after treatment interruption [115]. Circulating HBV-RNA has been shown to be related to intrahepatic RNA (even more so than circulating DNA) both quantitatively and in terms of QS complexity [119]. This evidence points to circulating HBV-RNA as a valuable plasma marker of intrahepatic HBV expression and the study of HBV QS complexity. In this sense, the study of circulating HBV-RNA QS would allow the identification of the most relevant variants associated with disease progression, allowing the study of viral evolution even in patients with undetectable levels of circulating DNA (without having to resort to the study of biopsies), and the identification of possible prognostic variants in the development of cirrhosis or HCC.

To date, serum HBV-RNA levels have been quantified using various in-house real-time RT-PCR assays. However, they are characterised by both low sensitivity, approximately 1.6 log<sub>10</sub> U/mL or 2 log<sub>10</sub> copies/mL [123][124][125], and low reproducibility. A new fully automated and highly reproducible real-time RT-PCR assay was used to quantify circulating

HBV-RNA with increased sensitivity [lower limit of quantification (LLOQ) of 10 copies/mL] [126].

### 1.8. Current hepatitis B therapy

There are currently two types of therapy available for CHB infection, IFNs and NAs, including a total of eight approved antiviral drugs: i) IFN- $\alpha$ -2b (paediatric), ii) PEG-IFN-2a (adult), iii) lamivudine (LAM), iv) telbivudine (TBV), v) entecavir (ETV), vi) adefovir dipivoxil (ADV), vii) tenofovir disoproxil fumarate (TDF), and viii) tenofovir alafenamide (TAF) [96]. According to the European Association for the Study of the Liver (EASL), NAs can be classified into those with a lower frequency of HBV resistance, including ETV, TDF and TAF, and those with a higher frequency of HBV resistance, including LAM, ADV and TBV (**Figure 10**).



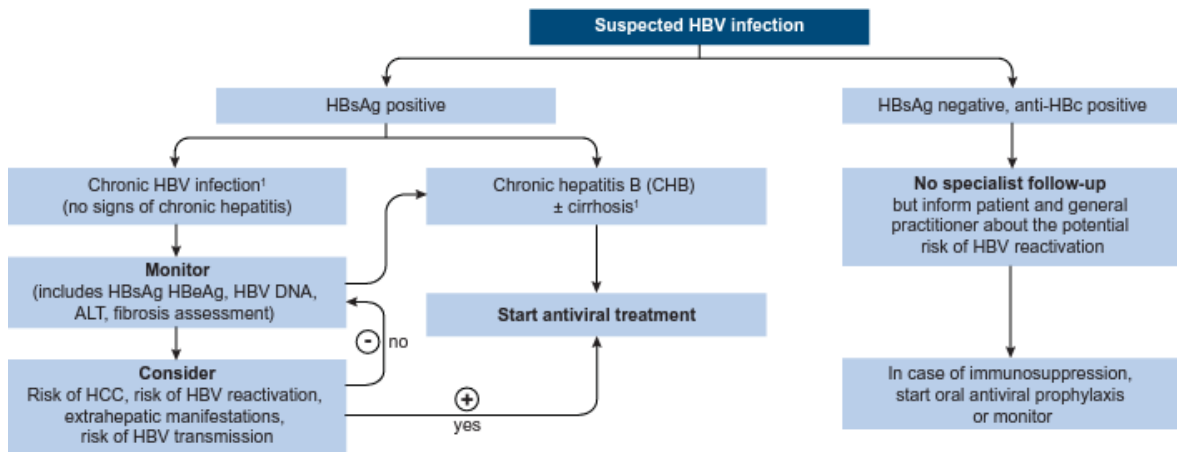
**Figure 10. Cumulative incidence of hepatitis B virus resistance to lamivudine, adefovir, entecavir, telbivudine, tenofovir and tenofovir alafenamide in pivotal trials in nucleos(t)ide-naïve patients with chronic hepatitis B.** Abbreviations: LAM, lamivudine; ADV, adefovir; ETV, entecavir; TBV, telbivudine; TAF, tenofovir alafenamide. Figure from EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection [96].

Side effects are more common with IFN $\alpha$  treatment than with NAs. For this reason, NAs are the first-line oral treatment for CHB [127] because they achieve viral suppression in a greater number of patients [128]. However, to date, no treatment can guarantee non-progression of liver disease [129][130]. In HBV infection, the cumulative risk of developing HCC (even in the absence of detectable viral load) at 5 and 10 years has been estimated to be 1.6-5.9%, with higher risks in cirrhotic patients, men and those over 50 years of age [131]. In Caucasian patients, although lower than in Asians, the risk of developing HCC

during treatment is still relevant and remains <1% in non-cirrhotic patients and between 1.8-5.2% in cirrhotic patients [132].

TDF is the most widely used therapeutic option for CHB infection because of the low likelihood of TDF resistance [96][133]. In fact, no signature mutation for TDF resistance has yet been identified in prospective patient cohorts, even after seven years of therapy [134], although cases of inadequate response to TDF have been reported [135][136]. Nevertheless, some extremely low frequent variants have been described, including the multidrug resistant rtS78T variant selected during a TDF-ETV combination therapy. This variant simultaneously generates a premature stop codon at sC69, deleting almost the entire S-HBsAg. Since May 2017, besifovir (BFV), a new NA, has been approved for the treatment of CHB patients in South Korea [137]. BFV is clinically similar to ETV and TDF, but has fewer side effects [138][139]. Nevertheless, the two major LVM resistance mutations, rtL180M and rtM204V, have also been associated with BFV resistance [140]. Therefore, further studies on BFV are needed to consider it as a good option against drug-resistance [141]. NAs have been shown to be safe and effective in significantly reducing viral replication by improving inflammation and fibrosis, preventing liver failure, reducing the rate of HCC, and improving the overall quality of life of patients with CHB. However, HBV eradication is currently unattainable due to the persistence of cccDNA. NAs compete with HBV polymerase and inhibit RT from pgRNA to rcDNA by terminating rcDNA synthesis. They are >99% effective in significantly reducing HBV-DNA levels to undetectable, but do not affect the expression of cccDNA or integrated DNA [127]. However, in the absence of treatment-induced viral replication, viral transcription is maintained, providing a source of residual viral replication and viral antigen production, that contributes to liver disease progression [128][129][142]. Hence, NAs would rarely achieve the main therapeutic goal of HBV “functional cure”, defined by sustained HBsAg loss and undetectable HBV-DNA serum levels with or without seroconversion to anti-HBs [96]. This suggests that low levels of cccDNA, even undetectable levels, may still allow viral replication, contributing to reactivation of infection and lifelong risk of liver disease progression [143]. This leads to lifelong treatment with NAs to prevent viral rebound [128]. The indications for treatment, which are the same for both HBeAg-positive and HBeAg-negative CHB patients, are based on the combination of both high serum HBV-DNA (>2,000 IU/ml) and elevated ALT levels and the severity of liver injury, generally assessed by liver biopsy, the gold standard method for assessing of liver histology, with at least moderate necroinflammation and/or fibrosis [143][96]. Nonetheless, non-invasive methods can also be used to assess the severity of liver histology [96], including measurements of the liver stiffness (transient elastography

using Fibroscan) and serum biomarkers related to the stage of liver fibrosis, such as transaminases, and platelet levels (**Figure 11**).



**Figure 11. Assessment of disease management of HBV infection.** Abbreviations: HBV, hepatitis B virus; HBsAg, hepatitis B surface antigen; HBeAg, hepatitis B e antigen; ALT, alanine transferase; HCC, hepatocellular carcinoma; anti-HBc, hepatitis B core antigen antibodies. Figure from EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection [96].

The combination of these biomarkers results in non-invasive scores for fibrosis, the so-called aspartate aminotransferase (AST) to platelet ratio index (APRI) and the four-factor fibrosis index (FIB-4) [144][145]. The presence of anti-HBe, i.e., HBeAg seroconversion, during treatment is consistent with loss of HBV infectivity [96], but a direct marker of viral replication is essential to avoid viral rebound after treatment interruption. HBV-DNA levels and serum HBsAg titres are two of the well-established serological viral markers for predicting the risk of developing cirrhosis and HCC. Nevertheless, these events can occur even in the presence of undetectable HBV-DNA and HBsAg seroclearance [146][131]. Besides, serum HBV-DNA levels become undetectable during treatment, so other markers are needed to monitor or predict HBV disease progression and response to treatment. HBV core-related antigen (HBcrAg) is considered a promising viral marker for disease monitoring and early prediction of HBeAg seroconversion [147]. HBcrAg is thought to be a good candidate for reflecting intrahepatic cccDNA transcription activity in CHB patients [148][149] with low HBV-DNA levels (< 2000 IU/mL) [150], formerly referred to as “inactive carriers”. HBcrAg levels can be used to differentiate between HBeAg positive chronic infection and hepatitis. HBcrAg levels are lowest in HBeAg-negative infection, but increase in HBeAg-negative hepatitis. HBcrAg is therefore a good candidate for differentiating HBeAg-negative chronic infection from hepatitis, detecting HBeAg seroconversion and monitoring antiviral therapy [151]. High serum levels of HBcrAg have been associated with HCC [152]. HBcrAg



is a combination of three related proteins, all products of the preC/C ORF: HBcAg, HBeAg and the truncated 22 kilodalton (kDa) precore-derived protein (p22cr). Like HBeAg, p22cr is derived from the preC ORF, but with further processing at both the N- and C-terminal domains [153]. However, HBcrAg has no predictive value for HBsAg loss and relapse in patients who stop treatment [151].

### 1.9. Cell culture models for studying hepatitis B virus infection

The main system used to study HBV infection *in vitro* is the use of primary human hepatocytes. Although this system allows the complete and natural life cycle of the virus, it has some limitations. For example, they are not easy to obtain and can rapidly lose their phenotype in culture [154]. The HepG2 cells are a hepatoma-derived cell line commonly used as surrogate hepatocyte model [155]. However, in the context of HBV infection, although HepG2 can support productive HBV infection, they lack the NTCP receptor and are therefore not susceptible to viral infection.

For this reason, this model has been modified by inducing the exogenous expression of human NTCP (hNTCP) polypeptide by the transduction with lentivirus encoding hNTCP and the puromycin resistance gene [156]. However, although the expression of hNTCP allows HBV entry, it is not sufficient to maintain the production of the infection for long, since these cells are line cells with a high proliferation rate and they rapidly dilute the cccDNA and the viral expression. The dimethyl sulfoxide (DMSO), a differentiation inducer, is also required as it induces a reversible G1 arrest in HepG2-hNTCP cells and polarises them to hepatocyte-like phenotype [157]. As the G1 arrest by DMSO is gradual, approximately two weeks are required for hepatocytes to block replication under DMSO treatment [158]. The expression of hNTCP in HepG2 cells can also be induced and maintained by DMSO treatment [159]. Therefore, the addition of DMSO helps to improve the efficiency of HBV infection at various levels. However, high concentrations of DMSO are cytotoxic to hepatocytes and thus interfere with viral infection [160].

A further development of the HepG2 cells are the HepG2.2.15. This human hepatoblastoma cell line is a suitable model to study the pathogenesis of HBV infection [161], as these cells have been stably transformed with the HBV genome, allowing them to produce high viral titres. Plasmids carrying an over-length HBV genome, such as pCHT-9/3091 or P26, are used [161]. They were transduced with a plasmid carrying the gene that confers resistance to geneticin (G418) and four 5'-3' tandem copies of the HBV genome [161].

### 1.10. Gene therapy against hepatitis B virus

As previously mentioned, the NA-based strategy allows HBV infection to be controlled and 1-5% of treated patients achieve functional cure. However, these therapeutic tools do not act on intrahepatic cccDNA, which supports the continuous expression of the viral proteins and is thus the source of viral relapses when treatment is interrupted. In fact, only 30% of these patients achieve durable HBV suppression [162][163]. Moreover, it has been reported that the cumulative risk of developing HCC after 5 or 10 years of HBV DNA undetectability could be around 1% in non-cirrhotic patients, reaching a risk of more than 5% in the presence of cirrhosis [132][131]. New treatment tools are therefore needed to control cccDNA expression even in successfully treated patients.

The first strategy is to reduce cccDNA copies within the hepatocyte. The intrahepatic cccDNA pool can be replenished either by intracellular recycling of newly produced rcDNA, which is reimported into the nucleus, or by *de novo* reinfection. The half-life and maintenance/replenishment of cccDNA is still under debate. However, there is an inverse relationship between hepatocyte turnover and cccDNA copies [164], as the pool is controlled and naturally reduced by cell killing, dilution by cell proliferation and cell “healing” mediated by inflammatory cytokines [165]. Identification of the molecular mechanisms involved in cccDNA stability and its transcriptional activity may contribute to the development of new therapies aimed at reducing, silencing or eradicating the cccDNA reservoir [165]. Several strategies have been considered, such as degrading cccDNA and/or killing of infected cells to promote hepatocyte turnover [166]. To date, these approaches are still under investigation and there is no evidence that a monotherapy based on them will result in functional cure.

Another highly valuable strategy is to suppress cccDNA expression by silencing viral mRNAs. Several oligonucleotide-based therapies for CHB infection have been reported, depending on the mechanism of viral suppression [167]. Viral suppression can occur either at the cytoplasmic level of viral mRNAs (HBV-derived transcripts) and proteins, or at the nuclear level targeting cccDNA [168].

Among the various classes of oligonucleotides [167], the small interfering RNAs (siRNAs) are double-stranded oligoribonucleotides of 21-23 nt in length that are recognised by the cytoplasmic RNA-induced silencing complex (RISC) [169]. The RISC takes up one strand of the siRNA depending on the thermodynamic stability of the base at the 5' end. The strand with a less thermodynamic stability at the 5' end is loaded into the RISC complex, while the remaining strand is degraded as a substrate for the RISC complex. This allows the bound

strand of the siRNA to be released and hybridise to the sequence-specific mRNA target. This interaction re-activates RISC and promotes mRNA cleavage [170]. Various siRNAs have already been tested against HBV infection [171][172].

Another class are antisense oligonucleotides (ASOs), which are single-stranded oligonucleotides of 15-22 nt in length designed to bind complementary mRNA. The ASO/mRNA (DNA/RNA) heteroduplex is recognised by cytoplasmic and nuclear RNase H [173], which cleaves the RNA strand of the DNA-RNA heteroduplex [169]. To date, Bepirovirsen (GSK3228836) is the most promising investigational ASO with very promising results in curing CHB [174]. It is currently in phase 3 clinical trials evaluating Bepirovirsen in NA-treated participants with CHB. However, two main problems limit the clinical use of ASOs: the difficulty of oligonucleotide delivery to the target cells and rapid degradation by nucleases [175]. However, the latter can be reduced by designing Gapmers (GPs). Modifications to internucleotide linkages and the ribose sugar are introduced into the backbone of the GPs to improve serum stability, target binding, potency and reduce immunogenicity [176]. Most ASOs with therapeutic benefit currently in clinical development are Gapmer ASOs. To date, three GP ASOs, mipomersen (Kynamro®) [177], inotersen (Tegsedi®) [178], and volanesorsen (Waylivra®) [179], have been approved for clinical use.

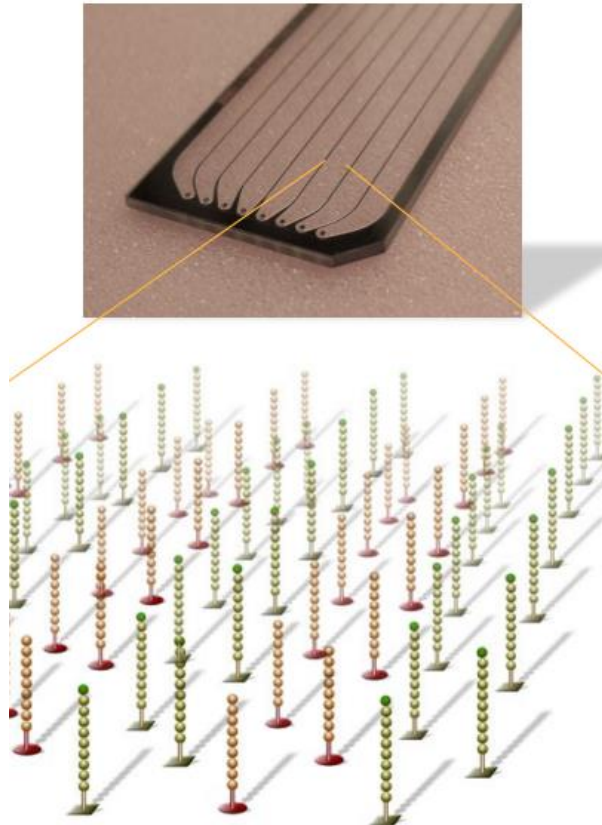
Although the gene silencing is a promising strategy that could ensure the inhibition of cccDNA expression and thus limit the expression of liver disease, the extremely high rate of HBV variability could be a strong limitation. To overcome this, NGS could be extremely helpful in identifying hyper-conserved regions in the HBV QS that could be targeted by silencing molecules. In fact, a preliminary approach based on this strategy has recently been reported [2].

### 1.11. Next-generation sequencing

Next-generation sequencing (NGS) is based on the parallel sequencing of many small pieces of DNA. NGS allows in-depth studies of viral QS because it can read thousands of clonal sequences, providing a large number of individual sequences from the same fragment [26]. This high-throughput technology has been a breakthrough in the speed of sequencing. This, in turn, has been accompanied by a reduction in cost [180].

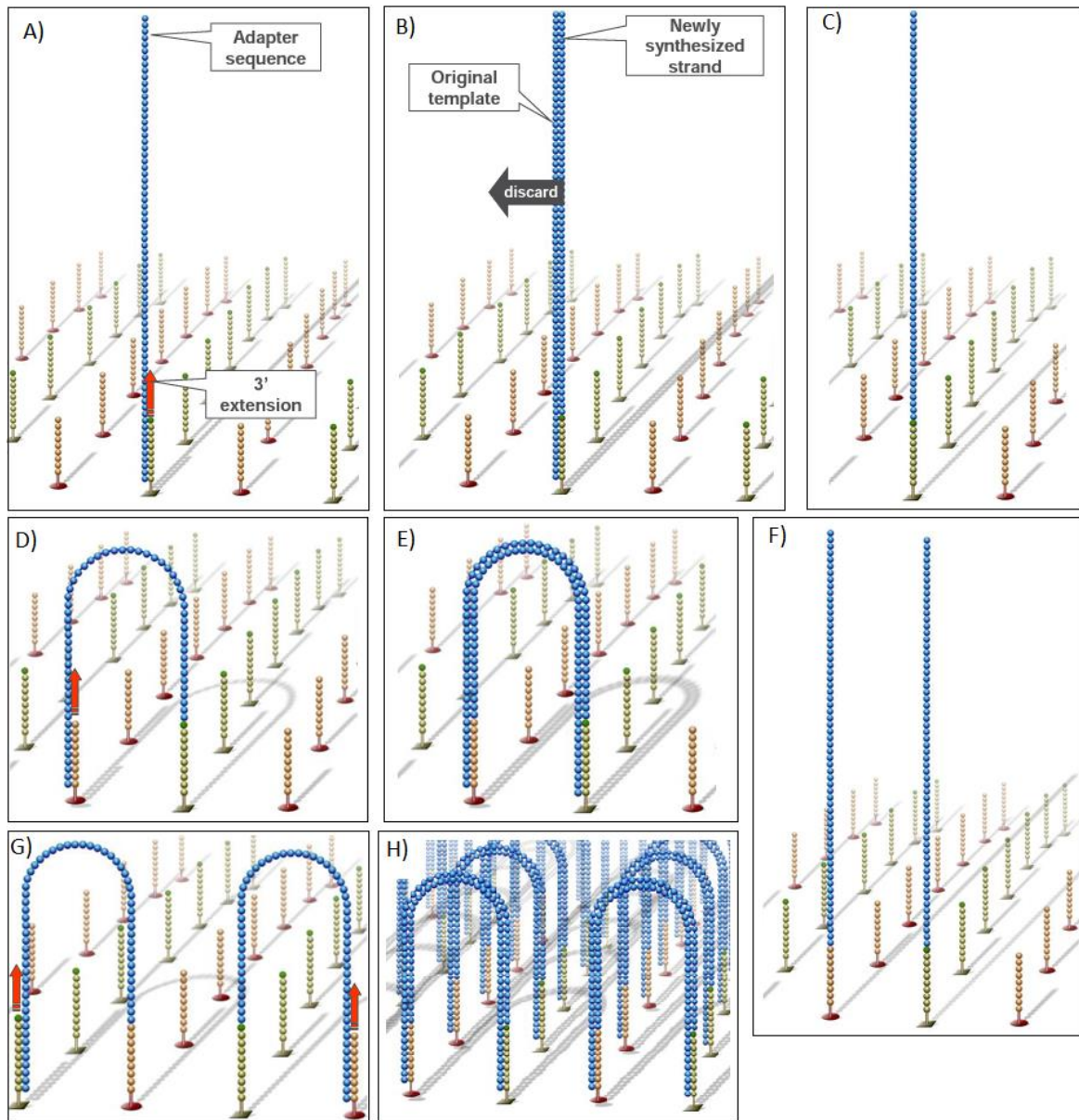
Viral QS analysis was performed using MiSeq Illumina sequencing technology (Illumina, San Diego, CA, United States). The overall Illumina sequencing workflow includes 3 steps: 1) library preparation, 2) amplification to generate clonal clusters, and 3) parallel sequencing of massive DNA fragments. The system used by the Illumina platform for amplification and

detection of incorporated nt is based on the sequencing by synthesis (SBS) approach [26]. It consists of repeated synthesis cycles that result in small clusters of identical DNA fragments [181]. Thus, a cluster contains thousands of identical sequences from clonal amplification. Cluster formation takes place in a flow cell, a thick glass slide with lanes. Each lane is coated with a lawn of oligonucleotides complementary to the library adapters [182] (**Figure 12**).



**Figure 12. Illumina flow cell representation.** Figure available on [https://www.ogc.ox.ac.uk/wp-content/uploads/2017/09/Illumina\\_Sequencing\\_Overview\\_15045845\\_D.pdf](https://www.ogc.ox.ac.uk/wp-content/uploads/2017/09/Illumina_Sequencing_Overview_15045845_D.pdf)

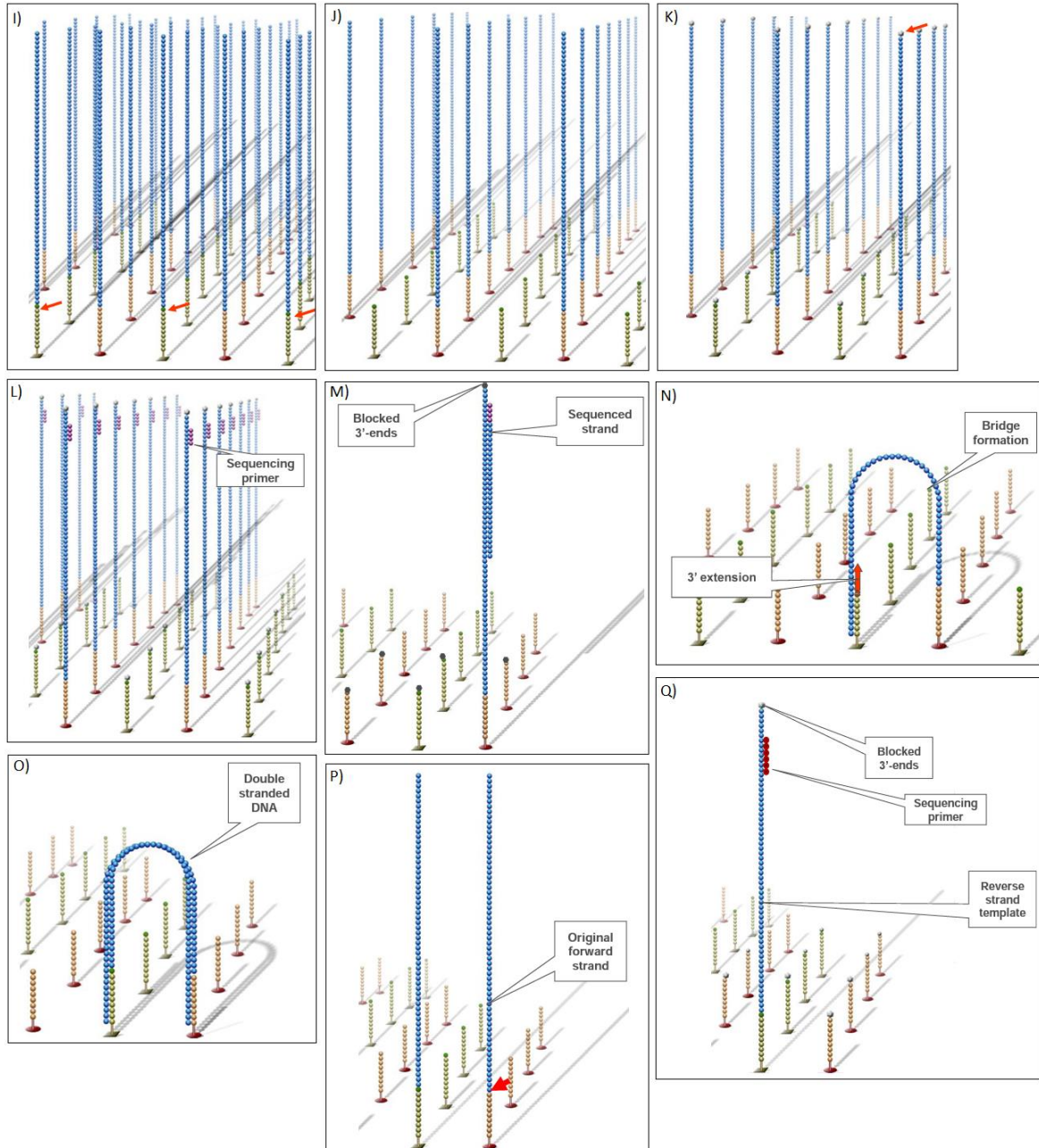
The formation of clusters by SBS is described in detail below according to information provided by the manufacturer [182]. As mentioned in section 4.12.3. *Adapter ligation*, once adapter ligation has occurred, each amplicon is flanked on each side by adapters containing P5 and P7 sequences at their ends. Due to the complementarity of these sequences to oligonucleotides present on the surface of the flow cell, single-stranded DNA (ssDNA) libraries hybridise to them. The 3' extension is then performed by 3'-5' polymerases (**Figure 13; A**), and the complementary strand to the original DNA template is formed (**Figure 13; B**).



**Figure 13. Part I of the schematic representation of the different sequencing steps using the Illumina platform.** Figure adapted from [https://www.ogc.ox.ac.uk/wp-content/uploads/2017/09/Illumina\\_Sequencing\\_Overview\\_15045845\\_D.pdf](https://www.ogc.ox.ac.uk/wp-content/uploads/2017/09/Illumina_Sequencing_Overview_15045845_D.pdf)

The double-stranded DNA (dsDNA) molecule is denatured and the original DNA template, which in fact it is no longer attached to the surface of the flow cell, is removed by washing (**Figure 13; C**). The new ssDNA molecule, which is indeed covalently bound to the surface of the flow cell, flips over and its 5' end hybridises to the adjacent lawn primer, forming a bridge (**Figure 13; D**). 5'-3' extension of the hybridised primer occurs, forming a double-stranded bridge (bridge amplification) (**Figure 13; E**). The double-stranded bridge is denatured (linearisation) resulting in two copies (forward 5'-3' and reverse 3'-5') of covalently bound ssDNA molecules (**Figure 13; F**). These two steps, bridge amplification

and linearisation, are repeated (**Figure 13; G**) until multiple bridges are formed (**Figure 13; H**). Reverse strands (3'-5') are discarded (**Figure 14; I**), leaving a cluster formed with only forward strands (5'-3') (**Figure 14; J**).

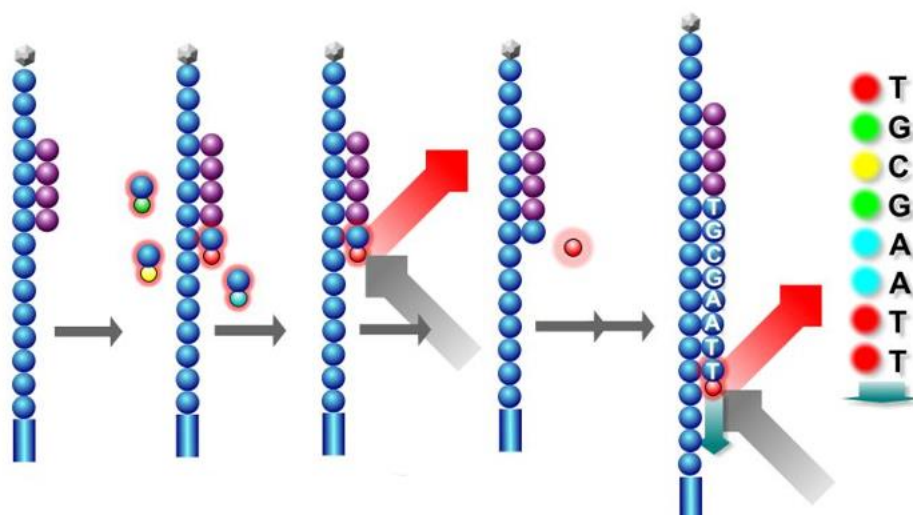


**Figure 14. Part II of the schematic representation of the different sequencing steps using the Illumina platform.** Figure adapted from [https://www.ogc.ox.ac.uk/wp-content/uploads/2017/09/Illumina\\_Sequencing\\_Overview\\_15045845\\_D.pdf](https://www.ogc.ox.ac.uk/wp-content/uploads/2017/09/Illumina_Sequencing_Overview_15045845_D.pdf)

At this point, the free 3' ends of both the ssDNA forward molecules (5'-3') and the lawn primers are blocked to prevent unwanted DNA extension (**Figure 14; K**). The sequencing primer hybridises to the adapter sequence located at the blocked 3' end of the ssDNA

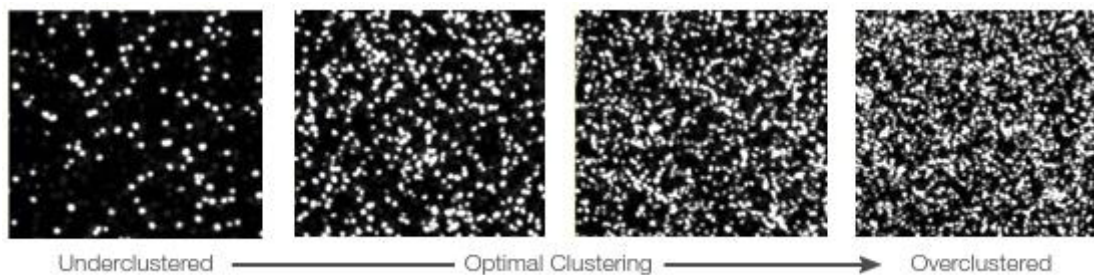
forward molecules (**Figure 14; L**) and the first sequencing cycle begins. Hence, forward strands (5'-3') are sequenced by SBS (**Figure 14; M**). Newly sequenced strands are denatured by SBS and consequently removed by washing. At this point, the forward ssDNA template molecules remain covalently attached to the lawn primers. The 3' ends of both the template ssDNA forward molecules and the lawn primers are then unblocked. The template ssDNA forward molecules will loop over and their 3' end will hybridise to the adjacent lawn forming a bridge (**Figure 14; N**). 3'-5' extension of the hybridised primer occurs, forming a double-stranded bridge (bridge amplification) (**Figure 14; O**). The double-stranded bridge is denatured (linearisation) resulting in two copies (forward 5'-3' and reverse 3'-5') of covalently bound ssDNA molecules (**Figure 14; P**). The forward strands (5'-3'), which were just sequenced in the previous step, are discarded by washing away, leaving a cluster formed with only strands (3'-5') (**Figure 14; Q**). As shown in **Figure 14; Q**, the free 3' ends of both ssDNA reverse molecules (3'-5') and lawn primers are blocked. The sequencing primers hybridise to the adapter sequence, resulting in the formation of forward strands that are complementary to those obtained in the first sequencing cycle. This process is called paired-end sequencing.

The four deoxynucleotides (dNTPs) added during strand extension (Adenine, A; Thymine, T; Guanine, G; Cytosine, C) after primer sequence hybridisation are labelled with a fluorochrome to facilitate their identification. When the nt is incorporated, the dye is imaged and cleaved, allowing the addition of the following fluorescence-labelled dNTP to be added [183] (**Figure 15**).



**Figure 15. Schematic illustration of the sequence by synthesis approach.** Abbreviations: T, thymine; G, guanine; C, cytosine; A, adenine. Figure available from [https://www.ogc.ox.ac.uk/wp-content/uploads/2017/09/Illumina\\_Sequencing\\_Overview\\_15045845\\_D.pdf](https://www.ogc.ox.ac.uk/wp-content/uploads/2017/09/Illumina_Sequencing_Overview_15045845_D.pdf)

The density of clonal clusters has a major impact on sequencing in terms of data quality, reads passing filter, Q30 scores and total data output [184]. Both underclustering and overclustering have disadvantages. While underclustering can lead to high data quality, it results in low output data. Conversely, overclustering leads directly to impoverished runs due to overloaded signal intensities [184]. Therefore, it is important to find a balance between the two (**Figure 16**).



**Figure 16. Images illustrating the different cluster densities, from underclustering to overclustering.** Figure available from <https://www.illumina.com/content/dam/illumina-marketing/documents/products/other/miseq-overclustering-primer-770-2014-038.pdf>

As mentioned above, a suboptimal cluster density results in both lower Q30 scores and a lower percentage of clusters passing filter. On the one hand, low Q30 scores occur when there is inaccuracy in base calling. This is due to a decrease in the ratio of base intensity to background for each base [184]. On the other hand, the percentage of clusters that pass the filter is taken as evidence of the cleanliness of the signal from each cluster [184]. This is associated with lower data output, *i.e.*, reduced yield.



## **HYPOTHESIS**

---

## 2. HYPOTHESIS

Chronic hepatitis B (CHB) infection cannot be completely eradicated due to the persistence of covalently closed circular DNA (cccDNA) and hepatitis B virus (HBV) integrated DNA in the host genome, which remain in the nucleus of infected liver cells for life [35][87]. The cccDNA produces the pregenomic RNA (pgRNA), which is reverse transcribed into negative-strand DNA within the newly formed capsids, followed by asymmetric synthesis of plus-strand DNA to produce viral genomic relaxed circular DNA (rcDNA) [91]. These capsids are normally enveloped and released into the bloodstream. However, circulating viral particles containing non-retrotranscribed pgRNA have been described [98]. In addition, the presence of other HBV-RNAs such as spliced pgRNA variants and HBV X gene (*HBX*) transcripts have been reported in HBV patients (4). Recently, circulating HBV-RNA has been proposed as a marker of cccDNA activity and it has also been associated with intrahepatic HBV-RNA, both in terms of quantity and quasispecies (QS) complexity [119]. Nonetheless, little is known about the variability/conservation and complexity of circulating HBV-RNA QS compared to serum HBV-DNA QS. Moreover, the analysis of serum HBV-RNA QS may be highly important to validate the applicability of hyper-conserved regions in viral transcripts as targets for gene therapy. The latter represents a promising new antiviral therapeutic strategy that could allow the control of the intrahepatic expression of viral proteins that could promote liver disease progression and viral reactivation after treatment interruption. In this sense, the similarity of circulating HBV-RNA to the viral transcripts targeted by silencing molecules would make the study of serum HBV-RNA QS a valuable tool to achieve the functional cure of HBV infection, the ideal endpoint of therapy based on sustained suppression of viral replication and loss of hepatitis B virus surface antigen (HBsAg) loss [96]. However, it is necessary to establish a reliable methodology for the thorough analysis of this QS.

In keeping with this, the following hypotheses are proposed:

- Circulating serum HBV-RNA has not been subjected to reverse transcription, which is the main source of variability in the HBV genome, and it may therefore show some differences in terms of QS variability/conservation and complexity with respect to the serum HBV-DNA in:

- A region including the overlapping retro transcriptase (RT) and surface (S) domains of the HBV genome, covering most of the functional domain B and all of functional domains C-E.
- A region containing the 5' end negative regulatory domain of the HBx protein.
- In a previous study, the group identified two highly conserved regions, encompassing nucleotides 1255-1286 and 1519-1603 within the *HBX*, in circulating HBV-DNA from 27 CHB patients [2]. In relation to these previous findings:
  - Confirmation of the conservation observed at the serum HBV-DNA level also in circulating viral transcripts would provide information on whether these targets could also be conserved at the intrahepatic HBV-RNA level, given its closer relationship to circulating HBV-RNA than to HBV-DNA.
  - The HBx protein plays a key role in viral replication and the progression of liver damage. In addition, thanks to the co-terminal localisation of *HBX* in the viral genome, its sequence is comprised into the 3' end of all the viral transcripts. Therefore, a gene therapy strategy based on antisense oligonucleotides, such as Gapmers, and siRNA could be a valuable and effective strategy to inhibit HBV expression. In particular, as the targeted regions are hyper-conserved at both viral HBV-DNA and HBV-RNA QS, this therapeutic strategy may be suitable regardless of HBV genotypes or the clinical stage of the patient.



### 3. AIMS

Based on the hypotheses considered above, two main objectives are proposed for this thesis project:

- 1) To establish a reliable methodology for in-depth analysis of HBV-RNA QS.
- 2) To study the genetic variability/conservation and complexity of circulating HBV-RNA QS in two distinct regions of the HBV genome: the *HBX*, which is common to all viral RNA transcripts, and the S-RT region, one of the major targets of antiviral therapies. The study will use next-generation sequencing to analyse HBV-RNA QS in depth and compare the results with those of circulating HBV-DNA QS.
- 3) To design and test in an *in vitro* cellular infection model a new therapeutic strategy based on gene silencing with Gapmers and siRNA targeting hyper-conserved *HBX* regions previously identified by our group and confirmed on HBV-RNA QS.

## **MATERIALS AND METHODS**

---

## 4. MATERIALS AND METHODS

### 4.1. Patients and samples

Considering the first and second studies, a total of 28 serum samples were included from 25 chronic hepatitis B (CHB) patients attended in outpatient clinics of both Vall d'Hebron University Hospital and Hospital Clinic (Barcelona, Spain). Exclusion criteria were positive testing for hepatitis D virus, hepatitis C virus, and HIV. The patients were classified according to World Health Organization Guidelines for the prevention, care and treatment of persons with chronic hepatitis B infection criteria [185].

### 4.2. Hepatitis B virus DNA and RNA extraction

Given the variety of manual and automated extraction methods currently on the market, using different protocols and technologies, the choice of product could have important implications for the diagnosis of viral infections and monitoring of response to antiviral therapy, as well as for experimental work and research [186].

#### 4.2.1. Automated extraction

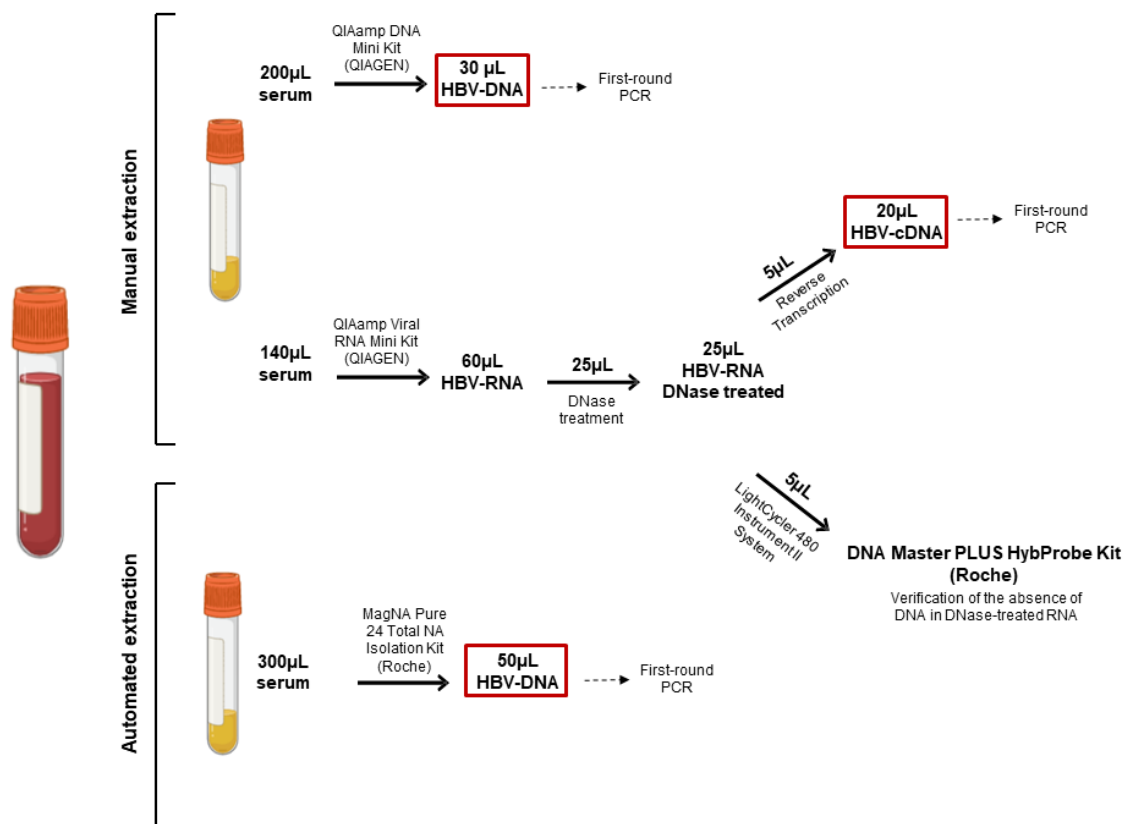
Automated nucleic acid extraction platforms, which require little hands-on time, have recently gained popularity in both clinical and research laboratories [186]. The main advantage of automation is the minimisation of user intervention and variability from extraction to extraction. Fully-automated nucleic acid extraction system was performed with MagNA Pure System using the MagNA Pure 24 Roche Diagnostics, Mannheim, Germany). The nucleic acid isolation procedure is based on the MagNA Pure Magnetic Glass Particle (MGP) technology [187]. Total nucleic acid was extracted from 300µL of serum sample using the MagNA Pure LC Total Nucleic Acid Isolation Kit (Roche Diagnostics, Mannheim, Germany). The serum sample is lysed, nucleic acids are released, and nucleases are denatured. Nucleic acids then bind to the silica surface of the added MGP due to the chaotropic salt conditions and the high ionic strength of the lysis binding buffer. MGP with bound nucleic acids are magnetically separated from the residual lysed sample and unbound substances are removed by several washing steps [187]. Finally, purified nucleic acids are eluted from MGP in 50µl of nuclease-free water and stored at -80° C until use.

#### 4.2.2. Manual extraction

Specific protocols are available for manual DNA and RNA extraction. The workflow for processing serum samples is shown in **Figure 17**. Two different kits from QIAGEN were

used for rapid, efficient, and simple silica-membrane-based nucleic acid purification from serum. For each serum sample, HBV-DNA was extracted from 200 $\mu$ L of serum or plasma using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) and HBV-RNA was extracted from 140 $\mu$ L of serum or plasma using the QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany), both according to the manufacturer's instructions. For viral nucleic acid purification, all buffers and components of both kits are RNase-free. The purification procedure consists of 4 steps. It is performed using QIAamp spin columns in a standard microcentrifuge [188]. Samples are then lysed and loaded onto the QIAamp spin column. DNA or RNA is absorbed onto the QIAamp silica during a brief centrifugation. Proteins and other contaminants are not retained on the QIAamp membrane due to salt and pH conditions. DNA or RNA bound to the QIAamp membrane is washed with two different buffers in two centrifugation steps. Purified DNA or RNA is eluted from the QIAamp spin column [188]. While HBV-DNA is eluted in a double elution step using 2 x 15 $\mu$ L of nuclease-free water and stored at -20°C until use, HBV-RNA is eluted also in a double elution step using 2 x 30  $\mu$ L of nuclease-free water and stored at -80°C until use.





**Figure 17. Serum samples processing workflow.** The diagram shows the main steps in the sample processing workflow steps. It shows which assays were used to finally start the amplification process of the DNA or cDNA extract with the first round of PCR. *Abbreviations: HBV, hepatitis B virus; cDNA, cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription.*

### 4.3. Hepatitis B virus DNA and RNA quantification

HBV-DNA detection and viral load measurement are essential for treatment decisions and patient monitoring [96]. In both studies, HBV-DNA was quantified by qPCR on a COBAS 6800 System (Roche Diagnostics, Mannheim, Germany) throughout the assay *cobas HBV Test* (Roche Diagnostics, Mannheim, Germany). The system uses a universal sample preparation protocol and reagents to isolate, purify and extract the total nucleic acids from 675  $\mu$ L of plasma sample. The qPCR allows for the amplification and simultaneous absolute quantification of the amplified product. The lower limit of quantification (LLOQ) of this assay is 10 international units (IU)/mL, with a linear range from 10 to  $1 \times 10^9$  IU/mL, and the limit of detection (LOD) is 6.6 IU/mL [125].

HBV-RNA is a new biomarker of HBV infection that has not yet been incorporated into routine clinical practice. This biomarker is also quantified by qPCR, but requires an additional step of reverse transcription prior to DNA amplification and reverse transcriptase (RT) activity of the polymerase. In the first study, HBV-RNA was quantified using an in-house method. First, a standard curve was generated for this qPCR quantification as follows:

1. The 1.1 x HBV genome from pTRiEx1.1-HBV [189] was subcloned downstream of a T7 promoter into a pEF6/V5-His TOPO TA vector (Thermo Fisher Scientific-Life Technologies, Austin, United States) according to the manufacturer's instructions to ensure its *in vitro* transcription (pEF6-HBV).
2. After cloning, the correct orientation and insertion of the HBV genome was analysed by Sanger sequencing and the pEF6-HBV plasmid was isolated from bacteria using the NucleoBond Xtra Midi Kit (Macherey-Nagel, Dueren, Germany).
3. The pEF6-HBV plasmid was then linearised by NotI restriction enzyme digestion (New England Biolabs, Beverly, United States) and used as template for the *in vitro* transcription reaction using the MEGAscript T7 Transcription Kit (Thermo Fisher Scientific-Life Technologies, Austin, United States) by adding the plasmid pEF6-HBV at a concentration of 0.5 µg/µL diluted in water. At the same time, the *in vitro* transcription of the pTRI-Xef positive control DNA supplied with the kit was performed.
4. The RNA resulting from transcription was then purified using the MEGAclean Transcription Clean-up Kit (Thermo Fisher Scientific-Life Technologies, Austin, United States) following the manufacturer's instructions. A DNase I treatment (Life Technologies, Austin, United States) was then carried out for 15 min at room temperature, followed by heat-inactivation at 65°C for 10 min and adding 2µL of 25 mmol/L EDTA solution to the reaction mixture, in order to remove the template DNA.
5. The purity of the DNase I-treated RNA was checked by measuring the absorbance of 1µL of the sample at 260 nm using the NanoDrop spectrophotometer (Thermo Fisher Scientific-Life Technologies, Austin, United States). The concentration of this RNA was quantified from 2µL of the sample by Qubit fluorimeter using the Qubit RNA HS Assay Kit (Thermo Fisher Scientific-Life Technologies, Austin, United States).
6. The absence of DNA in the DNase I-treated RNA template was verified by means of a qPCR-based absolute quantification assay using the DNA Master PLUS HybProbe Kit (Roche Diagnostics, Rotkreuz, Switzerland) in a LightCycler 480

Instrument II System (Roche Diagnostics, Rotkreuz, Switzerland). The primers and the probe used are listed in **Table 2**.

Primer sequence (5'->3')	Amplified region
Forward:	
<b>GATCCATACTGCGGAAC</b>	
<b>TCCT</b>	
Reverse:	
<b>GHAGGATCCAGTTGGCA</b>	from nt 1263-1408
<b>GYAC</b>	
TaqMan probe:	
<b>LC610-</b>	
<b>CTTGTTTTGCTCGCAGC</b>	
<b>MGGTCTGG-BBQ</b>	

**Verification of DNase  
treatment**

Step	Temperature	Time	Acquisition Mode	Analysis Mode
<b>Denaturation</b>	95°C	10 min	None	None
	95°C	5 s	None	
<b>Amplification</b>	65°C	30 s	None	Quantification
(45 cycles)	72°C	15 s	Single	
<b>Hold</b>	37°C	20 s	None	None

HBV-RNA				
Quantification				
Step	Temperature	Time	Acquisition Mode	Analysis Mode
RT	63°C	3 min	None	None
Denaturation	95°C	30 s	None	None
Amplification (45 cycles)	95°C	10 s	None	Quantification
	65°C	30 s	None	
	72°C	1 s	Single	
Hold	40°C	10 s	None	None

**Table 2. Primer design and polymerase chain reaction protocols for verification of DNase treatment and hepatitis B virus RNA quantification.** LC610 is a fluorescent dye for labelling oligonucleotide hybridization probes used in LightCycler qPCR assays. BBQ is a non-fluorescent chromophore that absorbs fluorescence energy from a neighbouring fluorophore, thereby preventing emission of fluorescent light. *Abbreviations: LightCycler 610, LC610; BBQ, BlackBerry Quencher, RT: reverse transcription.*

- For absolute quantification analyses, the standard curve was defined using the quantified reverse transcribed RNA (i.e., complementary DNA [cDNA]), taken to a concentration of  $7.51 \log_{10}$  copies/mL, and serial 1:10 dilutions of this standard covering concentrations until  $1.51 \log_{10}$  copies/mL. The points of the standard curve were defined as the mean cycle threshold (Ct) of a triplicate measurement of the concentration of the original  $7.51 \log_{10}$  copies/mL standard and its serial dilutions, using the LightCycler 480 RNA Master Hydrolysis Probes Kit (Roche Diagnostics, Rotkreuz, Switzerland), in a LightCycler 480 Instrument II System (Roche Diagnostics, Rotkreuz, Switzerland), as described in **Table 2**. The Ct is the number of the PCR cycles at which the fluorescence emitted by the sample crosses a threshold line, from which amplification is considered to be exponential. Ct values are inversely proportional to the amount of cDNA present in the sample.
- The obtained standard curve was saved and imported as an external standard curve in each HBV-RNA quantification experiment. In these experiments, at least one

dilution of the standard RNA used to define one of the standard curve points must be included to adjust the standard curve.

As with the RNA standard curve, HBV-RNA levels were quantified by a qPCR-based absolute quantification assay using one-step quantitative reverse transcription polymerase chain reaction (RT-qPCR) with the LightCycler 480 RNA Master Hydrolysis Probes Kit in the LightCycler 480 Instrument II System (Roche Diagnostics, Mannheim, Germany) as described in **Table 2**. Prior to quantification, HBV-RNA isolated from serum samples was treated with DNase I as previously described (point 4) to remove any residual DNA present. The concentration of each sample included was then calculated by plotting its Ct against the standard curve imported at the end of the experiment and adjusted by the Ct of the dilution of standard RNA included in the experiment.

In the second study, HBV-RNA was also quantified by RT-qPCR, but the assay used was automated on the COBAS 6800 System (Roche Diagnostics, Mannheim, Germany) [190][191]. This assay had a LLOQ of 10 copies/mL and a linear range of 10 to  $10^9$  copies/mL with synthetic RNA [191]. HBV-RNA was considered undetectable (LOD) if the result was less than 3.2 copies/mL.

#### 4.4. Serological determinations

HBV serological markers such as HBsAg and HBeAg were tested from 150µL of plasma sample using commercially available chemiluminescent immunoassays (Elecsys HBsAg II Kit and Elecsys HBeAg Kit, respectively) on a COBAS 8000 analyser (Roche Diagnostics, Rotkreuz, Switzerland). According to the manufacturer's protocol, the Elecsys HBsAg II assay uses monoclonal and polyclonal anti-HBs from mouse and sheep for HBsAg detection. The test is based on the sandwich principle. A first incubation is performed with 50µL of sample and two biotinylated monoclonal anti-HBs and a mixture of monoclonal anti-HBs antibody and polyclonal anti-HBs labelled with ruthenium. A sandwich complex is formed. In the second incubation, streptavidin-coated microparticles are added and the complex is bound to the solid phase by the interaction of biotin and streptavidin. The reaction mixture is then aspirated into the measuring cell, where the microparticles are magnetically captured on the surface of the electrode. According to the manufacturer's protocol, the Elecsys HBeAg assay is based on the same test principle but uses monoclonal anti-HBe (mouse) for the detection of HBeAg. A first incubation is performed with 21µL of sample and a biotinylated monoclonal HBeAg-specific antibody, and a monoclonal HBeAg-specific antibody labelled with ruthenium.

#### 4.5. Removal of hepatitis B virus DNA from RNA isolations and reverse transcription

To remove residual DNA from RNA preparations, isolated HBV-RNA was treated with DNase. Two different DNases were used in each study. In the first study, HBV-RNA was treated with DNase I (Life Technologies, Austin, United States) as described above (see section 4.2. *Hepatitis B virus DNA and RNA quantification, point 4*). In the second study, a more rigorous DNase treatment was performed using the Turbo DNA-free Kit (Thermo Fisher Scientific-Life Technologies, Austin, United States). The TURBO DNase enzyme is an engineered version of wild-type DNase I with 350% greater catalytic efficiency and a markedly higher affinity for DNA than conventional DNase I, making it more effective in removing trace quantities of DNA contamination [192]. In addition, the DNase treatment was enhanced by performing a two-step incubation at 37°C for 30 min each one. Half the amount of TURBO DNase Enzyme was added to the sample in the first step. Then, the remaining half of the TURBO DNase Enzyme from the previous step was added in the second step. The TURBO DNase Enzyme was inactivated by adding DNase Inactivation Reagent for 5 min at room temperature. Incubated samples were centrifuged at 10,000 x g for 1.5 min. Finally, the supernatant containing the RNA was transferred to a fresh tube without disturbing the pellet of DNase Inactivation Reagent.

After DNase treatment, RNA samples were reverse transcribed into cDNA in 2 steps. The first step involves denaturation of HBV-RNA and Oligo(dT)<sub>17</sub> primer, which binds to the polyA sequence at the 3' end of mRNAs in the serum or plasma samples, among which there will be the circulating HBV-RNAs. The second is the RT itself, involving RT enzyme and dithiothreitol (DTT), a reducing agent that helps to reduce disulphide bonds which will loosen the secondary structure of the RNA. As an additional step for protection of RNA during the synthesis reaction, the RNase OUT Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific-Life Technologies, Austin, United States) was added. In the first study, AccuScript High Fidelity Reverse Transcriptase (Stratagene, Agilent Technologies, Santa Clara, United States) was chosen for RT (**Table 3**). In the second study SuperScript III Reverse Transcriptase (Thermo Fisher Scientific-Life Technologies, Austin, United States) was used (**Table 4**).

First Step (HBV-RNA and Oligo(dT)<sub>17</sub> denaturation)

COMPONENT	VOLUME	CONCENTRATION
Add Water, PCR-Grade	To 16.5 $\mu$ L	-
AccuScript Reverse Transcriptase Buffer	2 $\mu$ L	10x
Oligo(dT) <sub>17</sub> primer (50 $\mu$ M)	1 $\mu$ L	100 ng/ $\mu$ L
dNTP mix	0.8 $\mu$ L	40 mM (10 mM of each dNTP)
DNase-treated RNA	10 $\mu$ L	

**THERMAL-CYCLING PROGRAM**

Temperature	Time
65°C	5 min
20°C	5 min

Second Step (RT)

COMPONENT	VOLUME	CONCENTRATION
DTT	2 $\mu$ L	100 mM
RNAse OUT	0.5 $\mu$ L	40U/ $\mu$ L
AccuScript High Fidelity Reverse Transcriptase	1 $\mu$ L	-
First step product	16.5 $\mu$ L	



---

**THERMAL-CYCLING  
PROGRAM**

---

<b>Temperature</b>	<b>Time</b>
42°C	60 min
70°C	15 min
4°C	∞

---

**Table 3. Reverse transcription protocol used in the first study with AccuScript High Fidelity Reverse Transcriptase following manufacturer's protocol. Abbreviations: dNTP, Deoxynucleotide; RT, reverse transcription; DTT, dithiothreitol.**

First Step (HBV-RNA and Oligo(dT)<sub>17</sub> denaturation)

COMPONENT	VOLUME	CONCENTRATION
Add Water, PCR-Grade	To 13 $\mu$ L	-
Oligo(dT) <sub>17</sub> primer (50 $\mu$ M)	1 $\mu$ L	-
dNTP mix	1 $\mu$ L	40 mM (10 mM of each dNTP)
DNase-treated RNA	5 $\mu$ L	

**THERMAL-CYCLING  
PROGRAM**

Temperature	Time
65°C	5 min
Ice	1 min

Second Step (RT)

COMPONENT	VOLUME	CONCENTRATION
First-Strand Buffer	4 $\mu$ L	5X
DTT	1 $\mu$ L	0.1 M
RNAse OUT	1 $\mu$ L	40U/ $\mu$ L
SuperScript III Reverse Transcriptase	1 $\mu$ L	200U/ $\mu$ L
First step product	13 $\mu$ L	

THERMAL-CYCLING PROGRAM	
Temperature	Time
50°C	60 min
70°C	15 min
4°C	∞

**Table 4. Reverse transcription protocol used in the second study with SuperScript III Reverse Transcriptase following manufacturer's protocol.** *Abbreviations: dNTP, Deoxynucleotide; RT, reverse transcription; DTT, dithiothreitol.*

At the same time, elimination of any residual HBV-DNA after DNase treatment was verified by means of a qPCR-based absolute quantification assay, as described for the DNase I-treated RNA obtained from in vitro transcription of pEF6-HBV vector (**Table 2**).

#### 4.6. Hepatitis B virus DNA and RNA amplification to obtain amplicon libraries for next-generation sequencing analysis

Library preparation is crucial to the success of NGS workflow. This step relies on converting DNA or cDNA samples into a library of fragments compatible to be sequenced on Illumina sequencing platforms.

In the first study, the first-round PCR covered a 1338 bp region, from positions nt 599 to 1936 of HBV genome and was performed as previously described by our group [193] (**Table 5**).

PCR	Primer sequence (5'->3')	Amplified region	Protocol
First-round PCR	Forward: <b>TGTATTCCCATCCCATCATC</b>	from nt	95 °C for 5 min followed by 35 cycles of 95 °C for 20 s, 53°C for 20 s, and 72°C for 15 s, and finally 72 °C for 3 min
	Reverse: <b>AGWAGCTCCAAATTCTTTATAAGG</b>	599 to 1936	
Second-round PCR	Forward: <b><u>GTTGTAAAACGACGGCCAGT</u></b>	from nt	95 °C for 2 min followed by 30 cycles of 95 °C for 15 s, 60°C for 20 s, and 72°C for 15 s, and finally, 72 °C for 3 min
	Reverse: <b><u>CACAGGAAACAGCTATGACC</u></b>	1234 to 1631	
Third-round PCR	Forward: MID- <b><u>GTTGTAAAACGACGGCCAGT</u></b>	Complementary to universal M13 sequences	95°C for 2 min followed by 20 cycles of 95°C for 15 s, 60°C for 20 s, and 72°C for 15 s, and finally, 72°C for 3 min
	Reverse: MID- <b><u>CACAGGAAACAGCTATGACC</u></b>		

**Table 5. Primer design and polymerase chain reaction protocols for the amplified 5'X region of the first study.** MID sequences are underlined. *Abbreviations: nt, nucleotide; MID, multiplex identifier.*

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 (20 bp oligonucleotide) at both ends (**Table 5**). In the third-round PCR, a specific pair of primers were used for each sample. It consists of the same M13 universal sequences (forward and reverse) at their 3' ends, which hybridize with the M13 universal sequences located at the ends of the previous amplicons, and a multiplex identifier (MID) or barcode sequence (10

bp oligonucleotide) at their 5' ends [2] (**Table 5**). MID allows demultiplexing sequences obtained from amplicons from each individual sample, once pooled with other amplicons obtained in the same region from different samples.

In the second study, the first-round PCR covered a 746bp region from positions nt 246 to 992 of HBV genome (**Table 6**).

PCR	Primer sequence (5'→3')	Amplified region	Protocol
First-round PCR	Forward: <b>GTCTAGACTCGTGGTGGACTTCTCTC</b>	from nt 246 to 992	95 °C for 2 min followed by 30 cycles of 95 °C for 20 s, 48°C for 20 s, and 72°C for 15s, and finally 72 °C for 3 min
	Reverse: <b>TGACADACTTTCCAATCAAT</b>		
Second-round PCR	Forward: <b><u>GTTGTAAAACGACGGCCAGT</u></b>	from nt	<b>For DNA target:</b> 95 °C for 2 min followed by 30 cycles of 95 °C for 20 s, 48°C for 20 s, and 72 °C for 15s, and finally, 72 °C for 3 min
	- <b>ACYTGTATTCCCATCCCAT</b>		<b>For cDNA target:</b>
	Reverse: <b><u>CACAGGAAACAGCTATGACC</u></b>	596 to 992	95 °C for 1 min followed by 35 cycles of 95 °C for 20 s, 48°C for 20 s, and 72 °C for 30s, and finally, 72 °C for 3 min
	- <b>TGACADACTTTCCAATCAAT</b>		
Third-round PCR	Forward: <b>MID-<u>GTTGTAAAACGACGGCCAGT</u></b>	Complementary to universal M13 sequences	95°C for 2 min followed by 35 cycles of 95°C for 20 s, 60°C for 20 s, and 72°C for 30s, and finally, 72°C for 3 min
Reverse: <b>MID-<u>CACAGGAAACAGCTATGACC</u></b>			

**Table 6. Primer design and polymerase chain reaction protocols for the amplified S-RT region of the second study.** MID sequences are underlined. *Abbreviations: nt, nucleotide; MID, multiplex identifier.*

The second-round PCR amplified the region of interest between nt 615 and 972 of the HBV genome, yielding final products flanked by universal M13 sequences (**Table 6**) at both ends. The third-round PCR was performed using the same primers as in the first study.

All PCR steps were performed using high-fidelity Pfu Ultra II DNA polymerase (Agilent Technologies, Santa Clara, United States). In the case of HBV-DNA, pTRiEx1.1-HBV plasmid was included as positive control. For HBV-RNA, pEF6-HBV plasmid (originating from 1.1 x HBV genome from pTRiEx1.1-HBV subcloned in a pEF6/V5-His TOPO TA vector as detailed in section 4.2. *HBV-DNA and RNA quantification*) was used as positive control. Once *in vitro* transcribed to RNA, it was treated like any other sample, i.e., DNase treated and reverse transcribed to cDNA.

#### 4.7. Verification of hepatitis B virus DNA amplification

For both studies, the amplified PCR products, also known as amplicons, were visualized as single bands on a 1.5% agarose electrophoresis gel. Agarose is a type of heteropolysaccharide that forms a viscous solution when dissolved in hot 1x Tris-acetate-EDTA (TAE) running buffer (Roche Diagnostics, Mannheim, Germany), but it solidifies as this solution cools down acquiring a gel consistency [194]. Electrophoresis is based on the migration of electrically charged molecules through the agarose gel under the effect of an electrical field [194]. Nucleic acids, which are negatively charged because of phosphate groups in nucleotides, always migrate to the positive pole (anode). The gel was stained with SybrSafe DNA Gel Stain (Life Technologies, Carlsbad, CA, United States), a highly sensitive dye for staining double stranded (ds) DNA, which is detected under ultraviolet illumination.

Visualization of PCR products on an agarose gel, allows checking the amplicon length. To this end, a molecular weight marker or ladder [DNA Molecular Weight Marker XIV (100bp ladder), 100-2642 bp (Roche Diagnostics, Mannheim, Germany), or Lambda DNA/EcoRI + HindIII Marker, 3, 125-21226 bp (Thermo Fisher Scientific-Life Technologies, Austin, United States)] was always loaded, along with the amplicons. In addition, loading PCR positive and negative controls allowed assessing PCR performance and checking for potential PCR contaminations, respectively. Only 2  $\mu$ L of PCR product are usually enough to simply verify in a qualitative way the presence or absence of DNA amplification in each PCR. However, depending on how the purification of the amplified product is to be carried out (see section 4.8. *PCR product purification*), a different amount of sample could be loaded.

#### 4.8. PCR product purification

A critical point when preparing NGS libraries was size selection and clean-up. It consisted of selecting the intended DNA fragments size range (that of amplicon molecules), while discarding all the smaller fragments, such as primer dimers, which will negatively interfere on sequencing efficiency. This procedure was performed differently in the two studies presented.

In the first study, the total volume of the third-round PCR was loaded into an agarose gel and each band, corresponding to an amplicon, was carefully cut out of the agarose gel, and manually purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany), following the manufacturer's protocol. Briefly, this kit provides all needed fungibles and reagents for silica-membrane-based purification DNA fragments from gels. The protocol is based on melting the cut agarose gel slices in a buffer containing a pH indicator to determine the optimal pH for DNA binding, followed by a bind-wash-elute procedure with a final double elution of 2 x 25  $\mu$ L of EB Buffer (10mM Tris-Cl, pH 8.5), which ensures a DNA recovery up to 80%.

In the second study, the third-round PCR amplification products were automatically purified using paramagnetic beads [KAPA Pure Beads (Kapa Biosystems-Roche, Cape Town, South Africa)] through the Freedom EVO 100 instrument (Tecan, Mannedorf, Switzerland). Beads must be equilibrated at room temperature for 30 min before use and well vortexed. Then, beads were mixed with DNA using a beads-to-sample volumetric ratio of 0.64x, by adding 32  $\mu$ L of beads, 25  $\mu$ L PCR-grade water and 25  $\mu$ L of amplification products to a well from a 96-well PCR plate. The DNA binds to the beads through electrostatic interaction, and this beads-DNA mixture was incubated, washed with 80% ethanol and amplicons were finally eluted. Up to 40 amplicons can be purified in a single run following manufacturer's instructions for size selection in NGS workflows.

#### 4.9. Quantification and normalization of the purified PCR product

Finally, purified PCR amplicons were quantified by means of fluorescence, using the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific - Life Technologies, Carlsbad, United States), following the manufacturer's protocol, automated in the Freedom EVO 100 instrument (Tecan, Mannedorf, Switzerland) coupled to the fluorescence microplate reader Infinite 200 Pro (Tecan, Mannedorf, Switzerland). Briefly, this kit is based on a highly sensitive fluorescent nucleic acid stain reagent used to quantitate dsDNA. A standard curve was defined using the lambda DNA standard, provided at 100  $\mu$ g/mL, which is diluted 50-fold in TE Buffer to make the 2  $\mu$ g/mL working solution. This working solution is the first



point of the standard curve, the other points were defined using serial dilutions of this solution covering concentrations until 50 pg/mL. All these dilutions, along with purified PCR products, were excited at 480 nm and the fluorescence emission intensity was measured at 520 nm, following manufacturer's instructions.

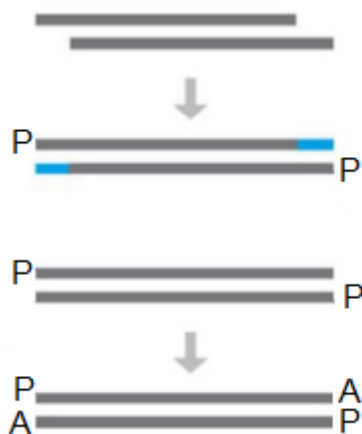
Once quantified, amplicons from each PCR product were adjusted to the same concentration, 1.00E+9 molecules/ $\mu$ L, with EB Buffer, a procedure known as normalization. This procedure allows each of the amplicons to be represented in equimolecular amounts (5.00E+10 molecules) when pooled together.

#### 4.10. Amplicon sequencing by next-generation sequencing using Illumina MiSeq system

In both studies the amplicon libraries were sequenced using MiSeq platform (Illumina, San Diego, CA, United States). Preparation of libraries for DNA sequencing for Illumina systems involves multiple steps. The workflow comprises conversion of amplicon ends (end-repaired and A-tailed) and ligation of adapters, clean-ups, library amplification (if necessary), quantification, and finally library normalization before loading onto a flow cell.

##### 4.10.1. End repair and A-tailing

After normalization and pooling of the amplicon libraries, each pool was processed with the KAPA HyperPrep kit (KAPA Biosystems Roche, Cape Town, South Africa), according to the manufacturer's instructions, which includes all the enzymes and reaction buffers required to prepare amplicon libraries for Illumina sequencing. The first step of this procedure consisted of end repairing and A-tailing pooled amplicons.



**Figure 18. End-repair and A-tailing procedure.** *Figure adapted from manufacturer (Illumina, San Diego, CA, United States).*

The end-repair and A-tailing process consist of the following steps (**Figure 18**):

- 1) The ends of each of the pooled amplicons must be transformed from cohesive (sticky) ends to blunt ends.
- 2) The 5' ends of the blunted DNA pooling amplicons are then phosphorylated.
- 3) The 3' ends of the blunted DNA pooled amplicons are adenylated (A-tailing), thus an A is added.
- 4) This phosphorylation and extra "A" addition is needed to T-A ligation with Illumina adapters (see the following section *4.10.2. Adapter ligation*).

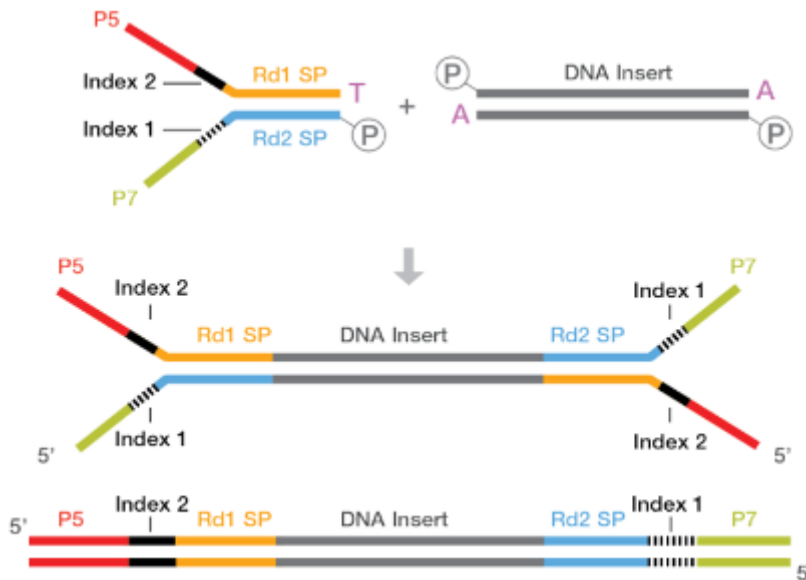
The PCR protocol and thermal-cycling program are detailed below (**Table 7**).

COMPONENT	VOLUME	
Amplicon pool	50 $\mu$ L	
End Repair & A-Tailing Buffer	7 $\mu$ L	
End Repair & A-Tailing Enzyme	3 $\mu$ L	
<b>Total Volume</b>	<b>60 <math>\mu</math>L</b>	
THERMAL-CYCLING PROGRAM		
Step	Temperature	Time
End Repair	20°C	30 min
A-tailing	65°C	30 min
Hold	4°C	1 min

**Table 7. Thermal-Cycling program of end conversion step following manufacturer's protocol.**

#### 4.10.2. Adapter ligation

The second step of amplicon library preparation with the KAPA HyperPrep kit consisted of ligation of Illumina adapters at both ends to 5'-phosphorylated, 3'-dA-tailed amplicons obtained in the previous step. These adapters are a pair of 50-60 annealed nt long that have several regions [195] (**Figure 19**):



**Figure 19. Graphic illustration of Illumina adapters.** *Figure available on <https://support.illumina.com/bulletins/2020/12/how-short-inserts-affect-sequencing-performance.html>*

- 1) **The P5 and P7 region**, which is complementary to the lawn of oligonucleotides present on the flow cell surface.
- 2) **The index 2 and index 1 sequences**, which are about specific 6-8 nt that function as an identifier for each pool of amplicons, enabling multiplexing of many pools in a single sequencing run.
- 3) **The read 1 (R1) and read 2 (R2) specific primers**, which will allow sequencing to start.
- 4) **The phosphate on the 5' end of the P7 strand**, which will ligate with the 3' end of pooled amplicons.
- 5) **An extra T presents in the 3' end of the P5 strand** by which T-A ligation may occur along with the “A” added in the previous step at the 3' end of the pooling amplicons.

Thanks to these adapters the amplicons pools will be identified on the flow cell for sequencing after the sequencing procedure. To perform ligation of the adapters, the first

step was to prepare the mix detailed in **Table 8** and then added it together with the end repair and A-tailing reaction product (60  $\mu\text{L}$ ) and 2.5  $\mu\text{L}$  of the different adapter per each library (SeqCap Adapter Kit A; Roche Sequencing, Pleasanton, United States). The ligation reaction was carried out by incubation at 20°C for 15 min.

COMPONENT	VOLUME
Ligation Buffer	30 $\mu\text{L}$
DNA Ligase	10 $\mu\text{L}$
PCR-grade water	7.5 $\mu\text{L}$
<b>Total Volume</b>	<b>47.5 <math>\mu\text{L}</math></b>

**Table 8. PCR protocol for adapter ligation step following manufacturer's protocol.**

#### 4.10.3. Post-ligation clean-up (1<sup>st</sup> clean-up)

The clean-up was intended to eliminate excess unligated adapters and other remaining small DNA fragments, and was performed with KAPA Pure Beads (Kapa Biosystems-Roche, Cape Town, South Africa), as described in section 4.7. *PCR product purification*. However, due to the lowest number of samples to be purified, this procedure was performed manually at a beads-to-sample volumetric ratio of 0.8x (**Table 9**).

COMPONENT	VOLUME
Adapter ligation reaction product	110 $\mu\text{L}$
KAPA Pure Beads	88 $\mu\text{L}$
<b>Total Volume</b>	<b>198 <math>\mu\text{L}</math></b>

**Table 9. First clean-up step after ligation following manufacturer's protocol.**

So that, those fragments longer than ~ 200 bp will bind to the beads and will be captured. Fragments below 200 bp will remain in the supernatant and will be discarded [196].

#### 4.10.4. Library amplification

The final step of preparation of amplicon libraries with the KAPA HyperPrep kit consisted of performing an additional short PCR amplification of these libraries as described in **Table 10**. Although this is an optional step, it was applied to the preparation of the amplicon libraries of the two studies presented in this thesis, since it is highly recommended to those pools with amplicons at low concentration.

COMPONENT	VOLUME
KAPA HiFi HotStart Ready Mix (2x)	20 $\mu$ L
KAPA Library Amplification Primer Mix (10x)	25 $\mu$ L
Adapter-ligated library	5 $\mu$ L
Total Volume	50 $\mu$ L

THERMAL-CYCLING PROGRAM			
Step	Temperature	Time	Cycles
Initial denaturation	98°C	45 s	1
Denaturation	98°C	15 s	
Annealing	60°C	30 s	5
Extension	72°C	30 s	
Final extension	72°C	1 min	1
Hold	4°C	1 min	

**Table 10.** PCR protocol for library amplification step following manufacturer's protocol.

#### 4.10.5. Post-amplification clean-up (2<sup>nd</sup> clean-up)

This second clean-up was carried out in the same way as the first clean-up [see 4.9.3. *Post-ligation clean-up (1<sup>st</sup> clean-up)*]. It was intended to remove excess of salts, primers or primer dimers created during the library amplification.

#### 4.10.6. Quality verification of the purified amplicon pools

Once pooled amplicon libraries (with Illumina adapters ligated) have been amplified and purified, their quality was electrophoretically verified using the Agilent 2200 TapeStation instrument and D1000 ScreenTape Kit (Agilent Technologies, Waldbronn, Germany). This step was intended to ensure that no short DNA fragments remained. Briefly, up to 16 lanes can be loaded into the ScreenTape Kit following manufacturer's instructions. Reagents needed to be equilibrated at room temperature for 30 min. Then 1  $\mu$ L of purified amplicons or ladder was mixed with 3  $\mu$ L D100 Sample Buffer and loaded into the 2200 TapeStation instrument. The D1000 ScreenTape system is designed for analysing DNA molecules from 35-1000 bp. The built-in software allows visualization of the electrophoresis results, both as electrophoretic peaks and as virtual electrophoresis gel lanes. The use of a standard DNA ladder provides with highly detailed information on the presence and length of DNA fragments present in each sample, with higher sensitivity as an agarose gel.

If previous purifications performed well, a single electrophoretic peak would be expected, shown as a single band in a virtual electrophoresis gel, corresponding to the size of the pooled amplicons, with no signal-to-noise of other fragments.

#### 4.10.7. qPCR-based library quantification (1st qPCR)

After quality assessment of purified pooled amplicon libraries, they were quantified. A qPCR-based absolute quantification assay was performed using the KAPA Library Quantification Kit (KAPA Biosystems Roche, Cape Town, South Africa) in a LightCycler 480 II Instrument (Roche Diagnostics, Mannheim, Germany). This kit includes the Primer Premix (10x), which specifically bind to the P5 and P7 regions of the adapters, therefore only correctly ligated amplicons will be quantified by this procedure. Hence, it is an accurate quantification of the amplicon libraries capable of interacting with the lawn of oligonucleotides on the flow cell surface and be sequenced in the MiSeq platform. Purified pooled libraries must be diluted 1:10000 to fall within the linear range of the assay (0.0002 pM – 20 pM). Fresh dilutions were prepared with EB Buffer and loaded in triplicate. Likewise, quantification of six DNA standards (STD1, 20 pM; STD2, 2 pM; STD3, 0.2 pM; STD4 0.02 pM; STD5, 0.002 pM; and STD6, 0.0002 pM) supplied in the kit and negative control were performed in duplicate. The qPCR was carried out using the protocol specified in **Table 11**.

COMPONENT	VOLUME
KAPA SYBRE FAST qPCR Master Mix (2x) with Primer Premix (10x) added	6 $\mu$ L
PCR-grade water	2 $\mu$ L
Purified pooled amplicons or DNA standards	2 $\mu$ L
<b>Total Volume</b>	10 $\mu$ L

**THERMAL-CYCLING  
PROGRAM**

Step	Temperature	Time	Acquisition Mode	Analysis Mode
<b>Denaturation</b>	95°C	5 min	None	None
<b>Amplification</b>	95°C	30 s	None	
(35 cycles)	60°C	45 s	Single	Quantification
<b>Hold</b>	37°C	2 min	None	None

**Table 11. PCR protocol for library quantification step.**

To quantify the concentration of amplicon libraries in each pool, the average Ct value for each DNA standard was plotted against its concentration in  $\log_{10}$  pM, to generate a standard curve. Then the concentration of each sample was calculated by plotting its Ct against the standard curve.

This step could be simplified by quantifying the concentration of amplicon libraries in each pool using the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific-Life Technologies, Austin, United States).

#### 4.10.8. Pool of amplicon libraries pools and quantification (2nd qPCR)

The purpose behind this step was to pool together multiple pools of amplicon libraries to include them in a single sequencing experiment. Each pool had a particular adapter ligated to its amplicons, which allowed its subsequent demultiplexing. Illumina sequencing chemistry requires denaturation of amplicon libraries in a subsequent step (see next section 4.10.9. *Denaturation and dilution of pool of pools and PhiX*) with a preferred starting concentration ranging from 2 – 4 nM. Hence, once the concentration of each pool was known, each of them were normalized to 4nM with EB Buffer. Normalized amplicon libraries were pooled into one (pool of pools). To ensure that final concentration was 4nM, a second qPCR was performed following the same procedure as the previous qPCR quantification of amplicon libraries pools [see section 4.10.7. *qPCR-based library quantification (1st qPCR)*]. Normalized individual pools and the pool of pools were diluted to 1:400 and 1:4000 with EB Buffer. These dilutions were included in the qPCR along with the 1:10000 dilutions of individual pools, prepared for the previous qPCR, which served as positive control. In addition, the commercial quality control ready-to-use PhiX library (Illumina, San Diego, CA, United States) was also diluted to 1:400 and 1:4000 and included as internal control.

#### 4.10.9. Denaturation and dilution of pool of pools and PhiX

To ensure that amplicons hybridise to the lawn of oligonucleotides on the flow cell, it was essential its conversion from double stranded to single stranded DNA fragments. Denaturation was carried out incubating 5 µL of the pool of amplicon libraries pools with 5 µL of NaOH 0.2 N during 5 min at RT followed by neutralization with Tris-HCl 200 mM pH 7-8. The commercial quality control ready-to-use PhiX library (Illumina, San Diego, CA, United States) was also denatured, and neutralized by the same procedure. PhiX library, which was created from the bacteriophage PhiX DNA sequences, served as control for Illumina quality runs [197] providing artificial diversity when it comes to libraries with low sequence diversity. Then, 990 µL of HT1 hybridization buffer were added to both neutralized library and PhiX, achieving a final concentration of 20 pM. In both studies library and PhiX were diluted from 20 pM to 12-15 pM with HT1 buffer, with the aim to obtain an amplicon cluster density between 600 and 800 clusters/mm<sup>2</sup> in the flow cell. At this point, library and PhiX were mixed in a 4:1 ratio (library:PhiX) to a final volume of 600 µL, whose mix is known as the loading pool. The loading pool was then incubated at 95°C for 2 min to ensure denaturalization of amplicon libraries. MiSeq Reagent Kit v3 (600 cycles) was selected for sequencing.



#### 4.11. Next-generation sequencing data analyses obtained using Illumina MiSeq system

For optimal results sequencing with MiSeq platform (Illumina, San Diego, CA, United States), the whole amplicon (without Illumina adapters but with MID and M13, both explained below) should ideally be between 350-450 bp in length, not exceeding 500 bp. Hence the importance of choosing a fragment of appropriate length. In addition, before starting a sequencing experiment it is crucial to know the depth of sequencing to be achieved, expressed by the coverage metric. It is defined as the number of times a genome has been sequenced. Illumina does not have an official recommendation for sequencing coverage level. Previous studies from our group have defined a sequence depth of at least 10000 reads per DNA strand as the minimum number of reads to assure with high confidence that no false polymorphic site is retained, and that the probability of missing a real variant at 1% abundance is extremely low [198].

The v3 MiSeq chemistry and imaging technology limits the readable portion of the DNA fragments to 2 x 301 bp. Thus, the first 300 bases were read on the MiSeq, generating the R1, and in R2 an equivalent number of bases were read from the opposite direction, which is known as paired-end reading. For a paired-end sequencing run, one R1 and one R2 fastq file was created for each sample. Hence, a fastq file incorporated the sequence data from the clusters that passed filters [199]. Sequences created after imaging process are known as raw reads. To all these reads, a haplotype-centric data analysis pipeline was applied, aimed at preserving full-length read sequence integrity, to obtain haplotypes (i.e., unique sequences) that completely covered the full amplicon. All the filtering steps described below were performed with in-house developed scripts [193][200] using R language [201]. The libraries used were Biostrings [202], ShortRead [203] and ape [204].

The steps to be followed were:

- 1) Quality control of fastq files: quality profiles by position of each sample were analysed.
- 2) Fast Length Adjustment of Short reads (FLASH) software tool was used to overlap paired-end reads (R1 and R2) reconstructing full amplicon reads [205]. FLASH parameters were established as a minimum overlap between R1 and R2 of 20 bp with a maximum of 10% mismatches. Reads not matching this condition were discarded.
- 3) Exclusion of all reads with more than 5% bp with Phred score below Q30 [206].

- 4) Demultiplexing reads by MID and template specific sequences within primers. Firstly, the sequence of the first 10 nt was read to identify a specific MID sequence, which allowed to distinguish among different samples. Only one mismatch with the expected sequences was allowed. Secondly, template specific primer sequences helped identifying both strands; forward and reverse and could also be used to identify different regions of viral genome. Up to three mismatches with the expected sequences were allowed. Finally, primers were trimmed and a fasta file was obtained by each specific MID-template specific primer-forward/reverse, and strand combination. Identical reads were collapsed into haplotypes with their respective frequencies in corresponding fasta headers.

Once demultiplexed, the bioinformatics pipeline proceeded differently in both studies:

- 5) In the first study:
  - a) Haplotypes of each fasta file were aligned with the corresponding master sequence, and a quality filter was applied. Haplotypes containing more than two indeterminations (Ns), three gaps or 99 differences with respect to the master sequence were discarded. Finally, accepted Ns and gaps were repaired in keeping with dominant haplotype.
  - b) Haplotypes with an abundance below 0.1% were discarded, and those common to both strands were selected and called consensus haplotypes, their frequencies were computed as the sum of reads in each strand.
  - c) Consensus haplotypes were the basis for subsequent diversity calculations.
  - d) Substitutions and conservation were analysed on the haplotypes resulting after an ulterior abundance filter at 0.25%.
- 6) In the second study:
  - a) Haplotypes of each fasta file were aligned with the corresponding master sequence, and a quality filter was applied. Haplotypes with more than two Ns, three gaps or 99 differences with respect to the master sequence were discarded. Finally, accepted Ns and gaps were repaired in keeping with dominant haplotype.
  - b) Haplotypes common to both strands with no previous abundance filter were selected, and their frequencies were computed as the sum of reads in both strands. Haplotypes unique to one strand were discarded.

Bioinformatics pathways and statistical methods used in both studies were supervised by Dr. Josep Gregori from Liver Disease Viral Hepatitis Laboratory of Vall d'Hebron Institut de Recerca (Barcelona, Spain).

#### 4.12. Haplotype genotyping

In the second study, the HBV genotype was determined in the overlapping S-RT region of viral genome amplified (**Table 6**) by the distance-based discriminant rule (DB-rule). This method is based on distances between each haplotype and a set of 134 reference sequences from HBV genotypes A to H (19 genotype A sequences, 21 B, 19 C, 20 D, 20 E, 19 F, 7 G, and 9 H). Accession numbers of these sequences on NCBI GenBank are detailed in **Appendix 2. Table1**. The reference sequences were obtained from 151 full length HBV genome sequences, downloaded from the NCBI GenBank (National Center for Biotechnology Information, USA), representative of the different viral genotypes. These sequences were grouped according to their HBV genotypes, then they were multiple aligned and trimmed to positions 615 to 972 with BioEdit software [207]. Sequences that were equivalent after trimming were removed, leaving only one. At least five non-identical reference sequences from each genotype were selected.

Once the set of reference sequences for each genotype was selected, the DB-rule was applied to genotype each haplotype. DB-rule is based on dissimilarity ( $\delta$ ), taken as genetic distances (computed according to the Kimura-80 model [208]) in a multiple alignment of each haplotype to the set of reference sequences. The distance of each haplotype in each sample to each genotype was determined by defining a proximity function ( $\Phi$ ) between them. Given  $n_1 \dots n_g$  reference sequences drawn from  $g$  genotypes  $\Omega_1 \dots \Omega_g$  and a distance function  $\delta$  between sequences observations, the following proximity functions are defined for each genotype  $k$ :

$$\hat{\Phi}_k(\mathbf{x}) = \frac{1}{n_k} \sum_{i=1}^{n_k} \delta_{i(k)}^2 - \frac{1}{2n_k^2} \sum_{i,j=1}^{n_k} \delta_{ij(k)}^2, \quad k = 1, \dots, g, \quad *$$

\*  $x = \text{haplotype} \in \Omega_1 \cup \dots \cup \Omega_g$ ;  $\delta_i = \text{distance from } x \text{ to the } i\text{-th reference sequence of } \Omega_k$ ;  $\delta_{ij(k)} = \text{distance between two reference sequences } i, j \text{ of } \Omega_k$

This function provides the average squared genetic distance between the haplotype sequence and each of the reference sequences of genotype  $k$ , corrected by the second term, which is a measure of genotype variability.

Assuming that  $\omega$  is an haplotype to be allocated into a genotype, this allocation is done according to the shortest distance:

$$\text{Allocate } \omega \text{ to } \Omega_j \text{ if } \hat{\Phi}_j(\mathbf{x}) = \min \{ \hat{\Phi}_1(\mathbf{x}), \dots, \hat{\Phi}_g(\mathbf{x}) \}$$

#### 4.13. Quasispecies conservation and variability

In the first study, sequence conservation at nt level was determined by calculating the information content (IC) of each position in a multiple alignment of all the different haplotypes obtained in frequencies > 0.25% found in each sample [2]. The IC is defined as the number of binary decisions (i.e., answers to yes/no questions;  $n_q$  required to find a specific element in a set of  $N$  elements, which is calculated as  $n_q = \log_2 N$ ). In a nt sequence not all 4 nt (A, C, T and G) are equally likely to be found in a given sequence position, they have different probabilities ( $p_i$ ). Due to uncertainty while finding a given nt in a specific position ( $j$ ), the IC is calculated according to Shannon entropy ) [209]. Hence, IC is a classical conservation measure defined as:

$$\text{IC}_j = \log_2(N) - \sum_{i=1}^j p_{ij} \log_2(p_{ij}) *$$

\*  $j$  = alignment position;  $p_{ij}$  = frequency of nt  $i$  in position  $j$  and  $N$  = the number of possibilities (4 nt)

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 that position in any of all haplotypes analysed). In addition to position by position, the conservation and variability was also studied as a sliding window analysis. This analysis consists of calculating the mean IC in windows of 25 nt (which corresponds to the length of a possible target for siRNA therapy), starting in the first position of the multiple alignment and moving in steps of 1 nt through nt 1255 to 1611. 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 [210]. In each position of the logos, letters representing the nts identified were 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.

#### 4.14. Quasispecies complexity

QS complexity is an intrinsic QS property that quantifies both the haplotype diversity and frequency regardless the viral population size [211]. QS complexity may be related to pathogenic potential and clinical evolution. This is the reason by which its study is critical. It is currently unattainable to analyse the QS complexity of the complete mutant repertoire of the viral populations. Currently, NGS of regions of interest in viral genomes, allows analysing this property in samples of thousands of sequences of viral genomes of such populations, obtained with low error procedures. This makes NGS the most accurate approach to estimate QS complexity.

QS complexity could be estimated with the same procedures for interpolation (rarefaction) and extrapolation (prediction) species diversity from samples of natural habitats in general ecology or fossil records in palaeontology. Thus, different indexes proven adequate to measure species diversity in ecology, could also be applied to calculate QS complexity, providing different information about that [211]. According to the kind of information which they provide, diversity indexes are classified into incidence, abundance, and functional indexes.

##### 4.14.1. Incidence-based indexes

They are based on the count of entities in a multiple alignment of haplotypes [212]. They are wealth indexes, *i.e.*, the number of species in a community [211], or the number of haplotypes in the context of the present studies.

##### 4.14.2. Abundance-based indexes

They consider both, counts of entities and their frequency in population [212]. They measure either diversity (number and frequency of different haplotypes) or evenness (uniformity of the haplotype distribution) [211]. The most used abundance indexes are the Shannon entropy ( $H_s$ ), the Gini-Simpson index.

##### 4.14.2.1. Hill numbers profile

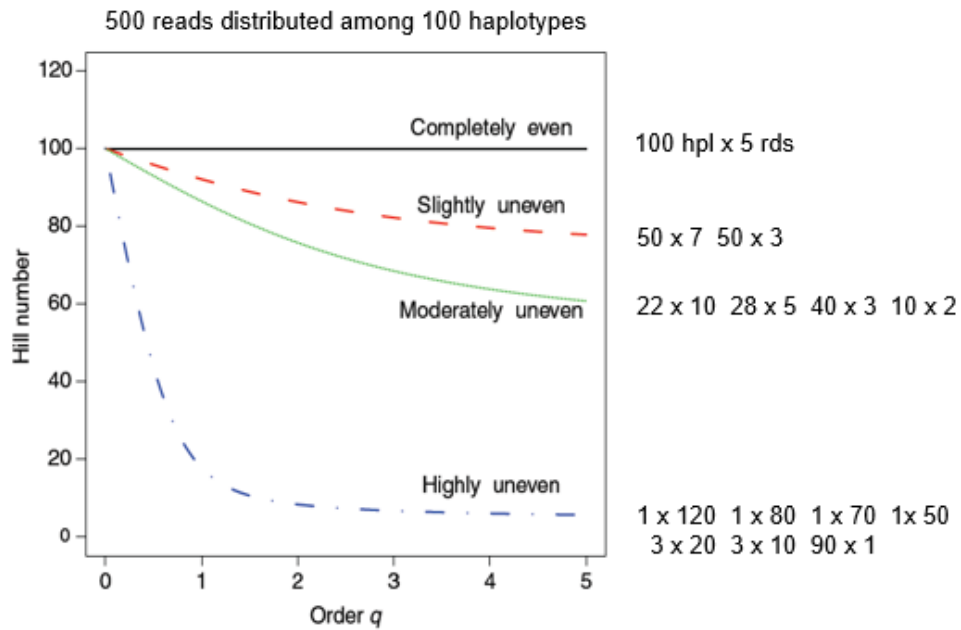
Hill numbers (HN) are a family of diversity measures that incorporates the transformed Shannon and Gini-Simpson measures, along with species richness. HN measure diversity as the effective number of species [213]. Different Hill numbers ( ${}^qD$ ) are defined as:

$${}^qD = \left( \sum_{i=1}^S p_i^q \right)^{1/(1-q)} \quad *$$

\*  $i$  from 1 to  $S$ , where  $S$  = the species number;  $p_i$  = the proportion of the species  $i$ ;  $q$  = the order of Hill number

The order ( $q$  value) represents the contribution of the different haplotypes to the diversity: when  $q$  is 0 all haplotypes have the same weight (their relative abundances do not count at all) and contribute equally to the measurement; with increasing values of  $q$  the measure of diversity becomes progressively less sensitive to rare haplotypes, and at infinity only the abundance of the dominant haplotype matters. Therefore,  $q$  controls the sensitivity of the measure to species (haplotype) relative abundance.

- The HN of order  $q = 0$ ,  ${}^0D$ , is based on the count of entities in a multiple alignment of haplotypes, i.e., the number of haplotypes. In this situation all viral QS would have the same HN value. So that, all QS will consist of the same haplotype number that will not change as  $q$  increases.
- When  $q = 1$ ,  ${}^1D$  is the exponential form of  $H_s$ , which enables QS diversity to be measured taking into account the haplotypes or variants found and their relative frequencies, and can be roughly interpreted as the number of “typical species” [214][215].
- When  $q = 2$ ,  ${}^2D$  is the inverse of the Simpson index ( $H_{Si}$ ),  $1/H_{Si}$ , which measures the probability that two randomly selected genomes from a viral population belong to the same haplotype [211]. Therefore, dominant or most common species receive greater weight than rare ones; the contribution from rare species is seriously discounted.  ${}^2D$  can be roughly interpreted as the number of “abundant species”.
- $q \geq 2$  results in a nearly unbiased estimator for  ${}^qD$ . Estimators for  $q \geq 2$  are almost independent of sample size, because all the higher order HN are mainly dominated by the most abundant species. The estimated diversity profile curve is thus generally slowly varying for  $q \geq 2$  [211] (**Figure X**).  ${}^\infty D$  is the inverse of the relative abundance of the dominant haplotype, while  ${}^{-\infty} D$  is the inverse of the relative abundance of the rarest haplotype.



**Figure 20. Diversity profile for assemblages of differing evenness.** *Figure from Gotelli N.J. and Chao A. (2013) In book: Encyclopedia of Biodiversity adapted by Dr. Josep Gregori.*

As depicted in **Figure 20**, the measure of diversity using HN can be portrayed on a single curve as function of  $q$ . This “diversity profile” of effective haplotypes vs  $q$  portrays all the information about haplotypes abundance distribution. The diversity profile curve is a decreasing function of  $q$ . The more uneven the distribution of relative abundances, the more steeply the curve declines. For a perfectly even distribution among QS haplotypes, the profile curve is a flat line at number of haplotypes in the sample [216]. For a better understanding, if curve is a horizontal line (black line in **Figure 20**), all haplotypes representing those QS would be on the same frequency. If any diversity index based on these frequencies were calculated, the highest diversity values would be obtained. Then, the HN of any  $q$  order would give the maximum reflecting that there is no dominance, i.e., no master haplotype. So, the uncertainty in assigning a chosen read randomly to any haplotype is maximal. It corresponds to a completely even distribution. On the other hand, hardly sloped curves correspond to the minimum diversity (blue dashed line in **Figure 20**). In this situation, some haplotypes dominate over other, so distribution is hardly uneven and when  $q$  increases, HN will decrease.

#### 4.14.3. Rare haplotype load

The rare Haplotype Load (RHL) is a diversity index that measures enrichment of QS in minority genomes below a given abundance threshold [212]. It may be considered intermediate between incidence and abundance indexes.

In both studies, it was calculated using the haplotypes resulting from the intersection of forward and reverse complement reverse strand without filtering by a minimum abundance. RHL was computed as the sum of the relative frequencies of all haplotypes obtained from HBV-DNA and HBV-RNA QS whose abundance is  $\leq 1\%$  [212].

#### 4.14.4. Quasispecies fitness fractions

At a given time, a QS is usually composed of a highly predominant haplotype, a few low- to medium-frequency genomes, different rare haplotypes with very low fitness but still capable of replicating, and some faulty genomes incapable of replicating. This composition can be illustrated using the set of frequencies of all haplotypes as parameters of a multinomial distribution [217].

$$\Pi = \{p_1, p_2, \dots, p_n\} \text{ with } \sum_i^n p_i = 1^*$$

\*  $p_1, p_2 \dots p_n$  = the frequencies of the haplotypes from a given QS in order of decreasing magnitude;  $p_i$  = sorted in decreasing order without a loss of generality

The QS can be divided into fractions restricted by frequency thresholds of interest, where partition into four fractions is depicted.

$$\begin{aligned} \Pi_1 &= \{p_1, p_2, \dots, p_k\}, \forall p_i : p_i \geq p_k \\ \Pi_2 &= \{p_{k+1}, p_{k+2}, \dots, p_l\}, \forall p_i : p_l \leq p_i < p_k \\ \Pi_3 &= \{p_{l+1}, p_{l+2}, \dots, p_m\}, \forall p_i : p_m \leq p_i < p_l \\ \Pi_4 &= \{p_{m+1}, p_{m+2}, \dots, p_n\}, \forall p_i : p_n \leq p_i < p_m \\ p'_1 &= \sum_i^k p_i; p'_2 = \sum_{k+1}^l p_i; p'_3 = \sum_{l+1}^m p_i; p'_4 = \sum_{m+1}^n p_i; \text{ with } \sum_{i=1}^4 p'_i = 1^* \end{aligned}$$

\*  $p'_1, p'_2, p'_3$  and  $p'_4$  = the four fractions

The QS evolution may be represented as the changes observed in the volume (fraction of molecules) of the four QS fitness fractions (QFF) [106]. The four fractions can be defined as follows:

- 1) **Master**: the fraction of molecules belonging to the most frequent haplotype; that is, the one present at the highest percentage ( $p'_1 = p_1$ ).



- 2) **Emerging**: the fraction of molecules presents at a frequency of > 0.1% and less than the master percentage, belonging to haplotypes that can compete with the predominant one and possibly replace it (p'2).
- 3) **Low fitness**: the fraction of molecules presents at frequencies from 1% to 0.1%, belonging to haplotypes that have a low probability of progressing to higher frequencies (p'3).
- 4) **Very low fitness**: the fraction of molecules presents at frequencies < 0.1% belonging to haplotypes with very low fitness and to defective genomes. The likely fate of these molecules individually is degradation, but the fraction is continuously fed with new very low fitness genomes produced by replication errors or by host editing activities (p'4).

The QFF are reads or molecules aggregations. Then, high QFF master values, in general, will correspond to low diversity QS. Conversely, high values in the other three fractions will correspond to high diversities.

The QFF combined with the HNP, which quantifies the effective number of haplotypes, provides a means to visualize and analyse molecular changes in the composition of a QS over time [217]. In the second both QFF and HNP were applied to analyse the composition of both HBV-DNA and HBV-RNA QS to compare their evolutionary trends.

#### 4.14.5. Functional indexes

They are those based on sequence differences among the observed haplotypes [212], and may include or not the frequency of each of them in the population. Among the most used functional indexes there is the nucleotide diversity ( $\pi$ ).

##### 4.14.5.1. Nucleotide diversity

It measures global genetic heterogeneity of a population considering the mean number of mutations between each haplotype pair. It is based on the mean of different nucleotides per site between two random genomes of the QS [218].

$$\pi = \sum_{i=1}^H \sum_{j=1}^H p_i d_{ij} p_j *$$

\* $p_i$  = frequency of haplotype  $i$  in the viral QS;  $p_j$  = frequency of haplotype  $j$  in the viral QS;  $d_{ij}$  = Hamming distance (number of mutations differentiating  $i$  and  $j$  haplotypes).

Therefore, differently from previously commented incidence/abundance indexes, which assess distribution of haplotypes within the QS,  $\pi$  was in the second study to assess sequence heterogeneity in circulating HBV-DNA and HBV-RNA QS.

#### 4.14.5.2. *Sample size and rarefaction*

Give the difficulty to obtain homogeneous coverages with the different samples in an experiment, and the known dependency of diversity measures with respect to sample size, the observed diversity values must be corrected to a common coverage. The process is known as rarefaction and consists in the repeated resampling of each sample to the reference size, to obtain in each cycle the required set of diversity indexes. The rarefied value of each index in each sample is computed as the average of the resampled values. One thousand cycles of resampling with reposition were used in the studies [219]. Despite that sample size corrections are possible, the amount of information conveyed by a study is limited by the size of the smallest sample, and hence a minimum acceptable sample size must be fixed as part of the experimental design.

#### 4.15. *Similarity between quasispecies distributions*

The similarity between the haplotype distribution of two QS, could be computed with similarity indexes. These indexes are useful to study the extent of changes between two highly related QS. In the case of the second study, two different indexes were used to study the similarity between haplotype distributions in circulating HBV-DNA and HBV-RNA QS [106]:

1. Commons ( $C_m$ ): defined as the fraction of reads belonging to haplotypes populated in both QS A and B (i.e., found in common in both QS).

$$C_m = \frac{1}{2} \sum_i \{(p_i + q_i) I(p_i > 0 \wedge q_i > 0)\} *$$

\*  $p_i$  = frequency of haplotype  $i$  in the viral QS A;  $q_i$  = frequency of haplotype  $i$  in the viral QS B

2. Overlap ( $O_v$ ): As the sum of the minimum proportion of common haplotypes.

$$O_v = \sum_i \min(p_i, q_i) *$$

\*  $p_i$  = frequency of haplotype  $i$  in the viral QS A;  $q_i$  = frequency of haplotype  $i$  in the viral QS B

Thus, when the two QS have all their haplotypes identical, the index of  $C_m$  results in a value of 1, regardless of the proportions of those haplotypes, while the  $O_v$  index may result in low values if the proportions of these haplotypes are highly dissimilar [106].

#### 4.16. Genetic distance between quasispecies

Another way of assessing the extent of changes between two QS may be by the genetic distance between them. This measurement provides the net genetic distance between two QS, taken as populations of viruses [106]. The nucleotide distance between two QS [220], X and Y, may be estimated by:

$$D_{XY} = \sum_{i \in X} \sum_{j \in Y} p_i d_{ij} q_j *$$

\*  $p_i$  and  $q_j$  = proportion of the  $i$ -th haplotype in QS X, and that of the  $j$ -th haplotype in QS Y, and  $d_{ij}$  = genetic distance between both haplotypes.

This distance is interpreted as the average number of nucleotide substitutions between the reads from QS X and QS Y.

Considering the nucleotide diversity of each QS [220], defined as the average number of nucleotide substitutions for a couple of arbitrary reads in both  $D_x$  and  $D_y$  QS, which may be estimated by:

$$D_X = \frac{N_X}{N_X - 1} \sum_{i \in X} \sum_{j \in X} p_i d_{ij} p_j$$

$$D_Y = \frac{N_Y}{N_Y - 1} \sum_{i \in Y} \sum_{j \in Y} q_i d_{ij} q_j$$

\*  $N_X$  and  $N_Y$  are the number of reads in each QS

Then, the net nucleotide substitutions between the two QS [220] is estimated by:

$$D_A = D_{XY} - (D_X + D_Y)/2$$

\*  $D_A$  = net genetic distance between two QS

$C_m$ ,  $O_v$ , and  $D_A$  contribute complementary information about QS composition similarity or distance.  $C_m$  and  $O_v$  show the degree to which two haplotype distributions are similar, whereas  $D_A$  takes also into account the load of genetic differences between two QS [106].

To display  $D_A$  among sequences dendrograms generated by Unweighted Pair Group Method with Arithmetic mean (UPGMA) clustering method were used. A dendrogram is a type of tree diagram using different branches that shows hierarchical clustering between the different haplotypes. Clusters are grouped sequences that results from their genetic distances.

#### 4.17. Statistical analysis

All statistical data analysis were performed using R programming language [201]. Qualitative parameters were expressed as number of cases and percentage. Quantitative parameters were expressed as the median value and interquartile range (IQR) or as median  $\pm$  standard deviation (SD), where appropriate.

Statistical comparisons between HBV-RNA and HBV-DNA QS complexity were performed using the Kruskal-Wallis test and t test. Two-proportions z-test was performed to compare HBV-RNA and HBV-DNA QS variability/conservation with Yates continuity correction. A p-value less than 0.05 ( $p < 0.05$ ) was considered statistically significant.

## 4.18. Antisense locked nucleic acid Gapmers and/or siRNA-based therapy against hepatitis B virus infection: *in vitro* testing

### 4.18.1. Hepatitis B virus production: preparation of hepatitis B viral stock from HepG2.2.15

HepG2.2.15 cells were cultured in cell incubators at 37°C in a 5% CO<sub>2</sub> atmosphere in Dulbecco's Modified Eagle Medium (DMEM) medium (Corning, Arizona, United States) supplemented with 10% fetal bovine serum (FBS) (Corning, Arizona, United States), which was previously decomplexed by heat-inactivation at 56°C in water bath for 30 min. In addition, DMEM was supplemented with 1% penicillin and streptomycin (Pen/Strep) 5,000 U/mL (Corning, Arizona, United States), 1% GlutaMAX (Corning, Arizona, United States), 1% sodium pyruvate (Corning, Arizona, United States), 1% MEM Non-Essential Acids Solution (Corning, Arizona, United States), and a final concentration of 0.5 mg/mL G418 (Corning, Arizona, United States).

HepG2.2.15 cells were maintained on collagen-coated flask. The coating solution was prepared by diluting type I collagen from rat tail (Thermo Fisher Scientific-Life Technologies, Austin, United States) with PBS 1X (Corning, Arizona, United States) at a final concentration of 0.01X. The flasks were finally coated by incubation at 37°C for at least 2 h before seeding cells or overnight at room temperature. Before use, the collagen solution was aspirated and rinsed with PBS 1X. HepG2.2.15 cells were expanded by splitting them every 2-3 days to get three T175 flasks with 70% confluence. Once reached that confluence, adherent cells were washed with PBS 1X before being detached from the T175 flask with trypsin/EDTA (Corning, Arizona, United States). Trypsin/EDTA was then inactivated by adding culture medium and HepG2.2.15 were pooled and re-seeded into a HYPERflask (Corning, Arizona, United States). Thanks to their 10 interconnected layers, the HYPERflask vessels have a growth surface of 1720 cm<sup>2</sup>, allowing to obtain a viral production of 560 mL. To induce the viral production, HepG2.2.15 cells were cultured in DMEM supplemented with 2.5% FBS, 1% Pen/Strep 5,000 U/mL, 1% GlutaMAX, 1% sodium pyruvate, 1% MEM Non-Essential Amino Acids Solution and 2% DMSO (Sigma-Aldrich, Oakville, ON, Canada). Supernatants were collected every 3 days for 18 days. The first supernatant collection was discarded.

The collected supernatants (560 mL per each collection time point) were centrifuged at 100 x *g* for 10 min to remove insoluble precipitate such as des-attached cells. Supernatants were then collected and clarified by sterile filtering through a 0.45 µm pore size filter to further eliminate cellular debris. HEPES (1%) was then added in each collected viral

production. For virus precipitation, polyethylene glycol (PEG) MW 8000 (Thermo Fisher Scientific-Life Technologies, Austin, United States) was added into supernatant reaching a final concentration of 5% PEG. The addition of PEG promotes HBV viral particles precipitation during centrifugation and boosts the interaction of the virus with heparansulfate proteoglycans, prerequisite of NTCP-binding, during infection. For the preparation of 40% PEG, 400 gr PEG MW 8000 were dissolved in PBS 1X sterile to a final volume of 1 L and autoclaved. To promote the contact between PEG and the viral particles, the PEG-treated viral productions were maintained at 4°C overnight with an agitation of 200 rpm. The next day, the productions were subjected to centrifugation at 10,000 x g for 1 h with fixed angle rotor at 4°C. The supernatants were discarded, and the viral particles (pellet) were resuspended in DMEM and pooled into one tube resulting in a 100X concentration (viral stock). The concentrated viral stock was incubated again at 4°C overnight with an agitation of 100 rpm, and the next day, the viral stock was aliquot and freeze at -80°C. The HBV load was quantified by qPCR on a COBAS 6800 system (Roche Diagnostics, Mannheim, Germany). For viral load conversion, 1 IU was considered equal to 5.6 copies.

To confirm that all the viral productions had similar infectivity and can be equally used for *in vitro* experiments, HBV infectivity was analysed by titration to evaluate the Tissue culture Infectious Dose (TCID<sub>50</sub>) assay.

HepG2-hNTCP were maintained in DMEM supplemented with 10% FBS, 1% Pen/Strep 5,000 U/mL, 1% GlutaMAX and a final concentration of 5 µg/mL puromycin (Sigma-Aldrich, Oakville, ON, Canada). To use them for *in vitro* infection, cells were treated with 2.5% DMSO for at least 14 days. For HBV titration, HepG2-hNTCP 2.5% DMSO-treated cells were seeded in a 96-well plate at concentration of 4.2x10<sup>4</sup> cells in 100 µL DMEM supplemented with 10% FBS, 1% Pen/Strep 5,000 U/mL, 1% GlutaMAX and 2.5% DMSO). HBV from the viral stocks was added by using a 4-fold serial dilution from 1:4 to 1:2048. Each condition was performed in triplicate. To promote infection, 4% PEG was added to each well and a centrifugation step at 1,000 x g for 1 h at 37°C was performed. The infection was maintained overnight by incubating the plate at 37°C overnight. The following day, the infection media was replaced by 200 µL fresh culture media. Supernatant was collected for HBeAg detection at 5 days (d) post infection (pi). Finally, TCID<sub>50</sub> was calculated according to Spearman-Käber method [221]. The viral stocks with extreme divergence in terms of infectivity were discarded.

#### 4.18.2. Hepatitis B virus infection assay in HepG2-hNTCP DMSO-treated cells

HepG2-NTCP 2.5% DMSO-treated cells were cultured in a 48-well plate at concentration of  $1.25 \times 10^5$  cells in 250  $\mu$ L final volume. The following day, cells were infected with HBV at 500 multiplicities of genome equivalents (MGE) by centrifugation at 1,000  $\times$   $g$  for 1h at 37°C followed by incubation at 37°C overnight. The volume of viral suspension required to infect HepG2-NTCP 2.5% DMSO-treated cells depends on the number of genome equivalents of DNA per cell, i.e., MGE, viral load in cop/mL and on the number of cells. The next day, the medium was replaced with fresh culture media.

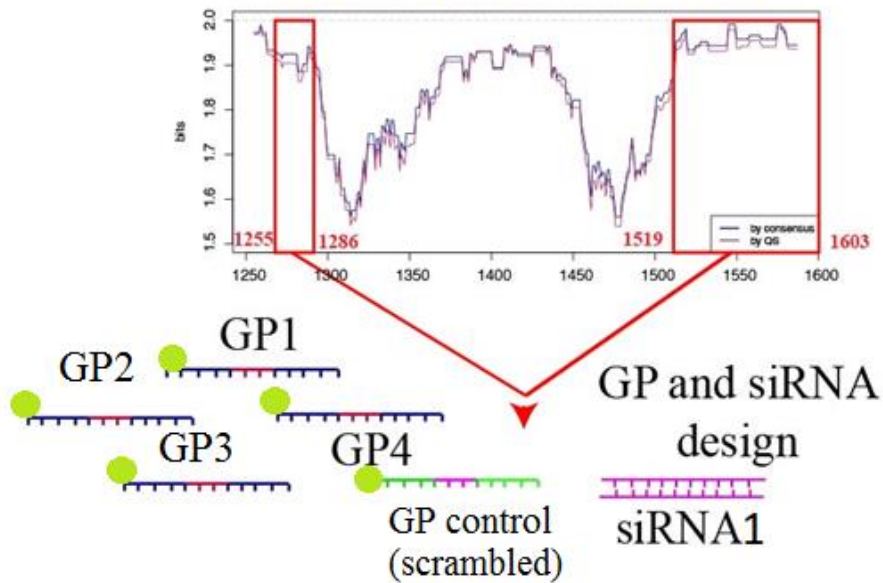
#### 4.18.3. Antisense locked nucleic acid Gappers and/or siRNA design

NGS has been used by our group to develop gene silencing strategies and two hyper-conserved regions in the 5' end region of the *HBX* gene [2]. The conservation of these regions was also confirmed by looking separately at different HBV clinical groups. These two hyper-conserved regions could be potential targets for gene therapy using the siRNA system. Given the importance of circulating viral transcripts as markers of viral expression, confirmation of the proposed conservation not only at the HBV-DNA level but also at the circulating HBV-RNA level would allow validation of this potential treatment even in the absence of detectable viremia. In addition, circulating HBV-RNA levels would allow the efficacy of siRNA treatment to be assessed.

Four different types of GPs and one siRNA were designed to target the aforementioned hyper-conserved regions of *HBX* previously identified by our group [2]. Notably, the hyper-conserved regions observed in the HBV-DNA were also confirmed when looking circulating HBV-RNA conservation in the same region. GPs were designed using an algorithm property of Qiagen. GPs contained an internal stretch of unmodified DNA bases with modified bases at each end of the oligonucleotide [169], also known as locked nucleic acids (LNA, also known as RNA-like segments). The LNA flanking regions confer on the oligonucleotide both resistance to nucleases degradation and increased affinity to its complementary mRNA thanks to both 2'-O-methoxyethyl (MOE) modification at the 2'-position of the ribose sugar in the LNA flanking regions and to modified phosphorothioate (PS) bonds. The modification of the phosphate group to PS was introduced throughout the entire oligonucleotide backbone, not only in the LNA flanking regions. The PS modifications were assigned in unknown positions. To ensure that the observed inhibition effect was sequence-specific, a scramble GP (carrying an unspecific sequence) was also used as control in each experiment. All GPs were designed with fluorescein at their 5' end (**Figure 21**).

## Hyper-conserved regions

Information content nt 1255-1611



**Figure 21. Gapmers and siRNA design.** Four different types of Gapmers and one siRNA were designed targeting hyper conserved regions of hepatitis B X gene previously found by our group [2]. A scrambled Gapmer was used as control in each experiment. *Abbreviations: GP, Gapmer; siRNA, small interfering RNA.*

Differently, the siRNA was designed by using the online tool iScore Designer, and the one with the highest score rank was selected. Obtained from GPs and siRNA sequences as well as specific positions in HBV genome are depicted in **Table 12**.



	Sequence	Length (bp)	Position (nt)
Antisense LNA Gappers	GP1: /5'-FAM/*C*G*C*A*G*T*A*T*G*G*A*T*C*G*G*C	16	1259-1275
	GP2: /5'-FAM/*C*G*G*G*A*C*G*T*A*A*A*C*A*A*A*G	16	1419-1434
	GP3: /5'-FAM/*A*C*G*G*G*A*C*G*T*A*A*A*C*A*A*A	16	1420-1435
	GP4: /5'-FAM/*G*C*A*G*A*G*G*T*G*A*A*G*C*G*A*A	16	1583-1599
	GP control (scrambled): /5'-FAM/*A*A*C*A*C*G*T*C*T*A*T*A*C*G*C	15	-
siRNA	GCCGAUCCAUCUGCGGAA	19	1260-1278

**Table 12. Gappers and siRNA sequences and their respective positions in hepatitis B virus genome.** Noted that fluorescent Gappers were designed with a fluorescein at their 5'end. Phosphorothioate backbone modifications are indicated by “\*”. Abbreviations: LNA, locked nucleic acids; GP, Gapper; siRNA, small interfering RNA.

The GP1 and siRNA sequences are in the promoter region of *HBX*. Despite four GPs are targeted against a hyper-conserved region of *HBX* (nt 1255 to 1611), the hyper-conservation degree is not the same throughout the entire region. GP1 and GP4 were targeted against much more hyper-conserved sequence compared to GP2 and GP3.

Every oligo was obtained after analysis by capillary electrophoresis or high-performance liquid chromatography, and the identity of the compound was verified by mass spectrometry. Oligo quantity was validated by UV absorbance at 260 nm. Oligos are shipped dry and ready for use upon resuspension. Before the first use, 300 µL TE Buffer were added to 15 nmol of GP for a 50 µM stock solution. One hundred µL sterile nuclease-free water were added to 10 nmol of siRNA for a 100 µM stock solution. For both oligos, an

additional dilution was done to achieve a 10  $\mu$ M working solution. After resuspension all the aliquoted solutions were stored at -20°C.

#### 4.18.4. Antisense locked nucleic acid Gapmers and/or siRNA transfection in HepG2-hNTCP DMSO-treated and HBV-infected cells

HepG2-hNTCP DMSO-treated and HBV-infected cells were transfected with GPs and siRNA at 48 h (hereinafter known as “early condition”) or 5 d pi (hereinafter known as “late condition”) at a final concentration of 50 nM using TransIT-X2 (Mirus Bio LLC, Wisconsin, United States), a non-liposomal polymeric system. Each experiment was performed at least three times in duplicate. A combination strategy based on GP+GP or GP+siRNA was also tested. For this purpose, each molecule was added at 50nM to have an additive or synergic effect.

Cells and supernatants were collected after 72 h of treatment. To obtain the cellular pellet, the infected and treated cells were des-attached from the well by adding 100  $\mu$ L trypsin/EDTA and 200  $\mu$ L culture media to inactivate trypsin. The resulting 300  $\mu$ L were then centrifuged at 1,300 rpm for 5 min, and pellets were maintained with 20uL of RNA later at 80°C before being used for the intracellular pgRNA quantification.

The intracellular pgRNA was extracted from pellet through the RNAeasy Plus Mini Kit (Qiagen, Hilden, Germany), and DNase I treated (Life Technologies, Austin, United States), as described in section 4.2. *HBV-DNA and RNA quantification, point 4*. The DNase-treated pgRNA was quantified from 2  $\mu$ L of the sample by Qubit fluorimeter using the Qubit RNA HS Assay Kit (Thermo Fisher Scientific-Life Technologies, Austin, United States). All pgRNA samples were normalized to 20ng by adding sterile nuclease-free water. After normalization, pgRNA was quantified by means of a qPCR-based absolute quantification assay using one-step RT-qPCR, with the LightCycler 480 RNA Master Hydrolysis Probes Kit in the LightCycler 480 Instrument II System (Roche Diagnostics, Mannheim, Germany). The primers and the probe used are specified in **Table 2**.

HBsAg and HBeAg were quantified in supernatants using commercial chemiluminiscent immunoassays (Roche Diagnostics, Rotkreuz, Switzerland), as described above (see section 4.3. *Serological determinations*). Notably, in HBeAg quantification, a standard curve was used. The curve was obtained through a 2-fold serial dilution of the HBeAg standard from 10 IU/mL to 0.15 IU/mL [222].

Notably, to avoid that results were biased by the inhibition of the expression of HBV caused by transfection, for both proteins and pgRNA, the percentage of inhibition was calculated

related to the scrambled oligonucleotide, although an infected-untreated condition was included in each experiment.

#### 4.18.5. Assessment of antisense locked nucleic acid Gapmers entry and dose-response curve

The entry of the 5'-FAM scrambled GP into the differentiated cells was tested by analysing the percentage of green positive cells through FACS Calibur flow cytometer.

A dose-response curve was generated to assess the relationship between the dose of the four GPs versus the effects that they exert on the cells, depicting the magnitude of the response of the cells, either therapeutic or toxic. Inhibition percentage was calculated considering the inhibition of HBeAg since it is a direct marker of HBV expression. Furthermore, because it is the only non-structural viral protein, the results will not be affected by the background of viral particles plus PLEG that might remain in the well even after cells washing and media replacement. Moreover, trypan blue was used to measure cell viability. Upon entry into the cell, trypan blue binds to intracellular proteins thereby rendering the cells a bluish colour. The trypan blue exclusion assay allows for a direct identification and enumeration of live (unstained) and dead (blue) cells in a given population.

## **RESULTS**

---

## 5. RESULTS

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

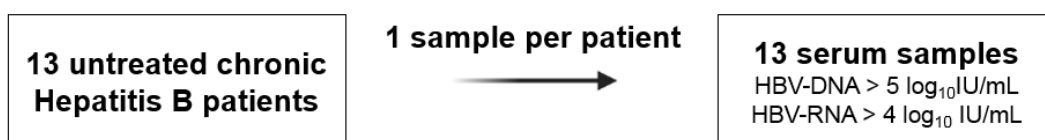
This first study was already published and is attached in **Appendix 1**.

#### 5.1.1. Aims

In this study, we aimed to establish a reliable methodology, based on NGS technology, to thoroughly analyse serum HBV-RNA quasispecies (QS) without interference from circulating HBV-DNA. With this methodology we analysed the region between nt 1255 to 1611, located at the hepatitis B X gene (*HBX*) 5' end, that we analysed in serum HBV-DNA QS in previous studies [2][3]. In these studies, we identified highly conserved regions that we speculated could be targets for targeted gene therapy. Therefore, with the methodology developed, we also aimed at describing similarities and differences between circulating HBV-RNA and HBV-DNA QS in *HBX* 5' end, in terms of complexity and conservation.

#### 5.1.2. Study design

It is a cross-sectional study that included a serum sample from 13 well-characterised untreated chronic hepatitis B (CHB) patients attending the outpatient clinic of Vall d'Hebron University Hospital (Barcelona, Spain) (**Figure 22**). All samples were selected considering their viral loads: HBV-DNA levels  $> 5 \log_{10}$ IU/mL and HBV-RNA levels  $> 4 \log_{10}$ copies/mL to ensure sufficient levels of both HBV-DNA and HBV-RNA to study their QS. 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 the differences between HBV-DNA and HBV-RNA QS.



**Figure 22. Outline of patients and samples included in the first study.**

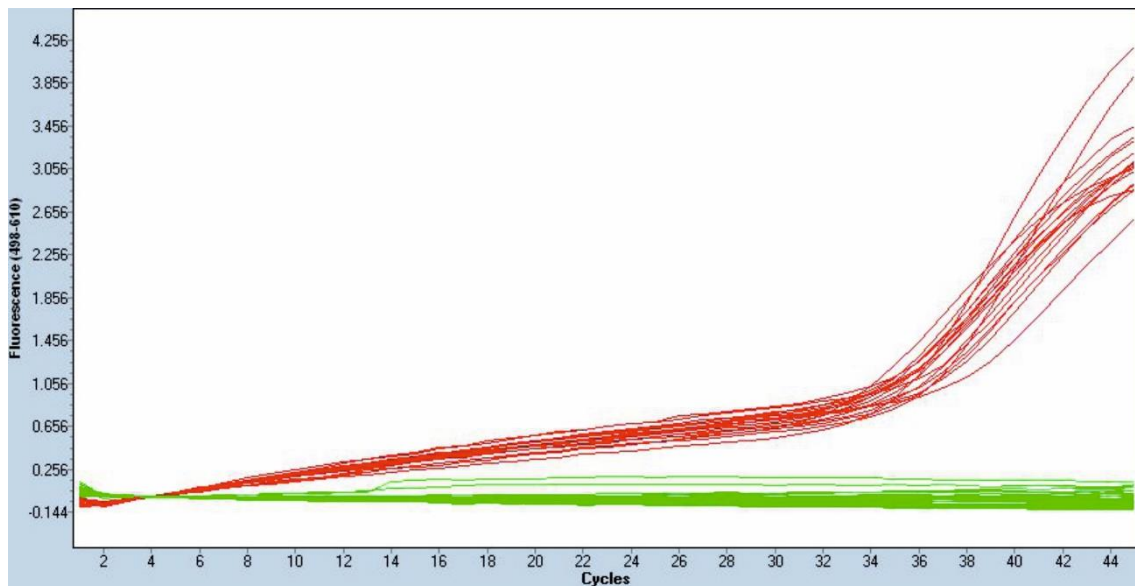
Clinical, virological, and serological parameters of the 13 patients included at the time of sample collection are summarised in **Table 13**.

<b>Total, n = 13</b>	
<b>Age, years, median (IQR)</b>	41.4 (31.7-47.2)
<b>Male, n (%)</b>	11 (84.6)
<b>Genotype, n (%)</b>	
A	3 (23.1)
C	3 (23.1)
D	2 (15.4)
E	2 (15.4)
F	3 (23.1)
<b>ALT, IU/L, median (IQR)</b>	124.0 (66.0-160.0)
<b>HBeAg +ve, n (%)</b>	11 (84.6)
<b>HBV-DNA, log<sub>10</sub>IU/mL, median (IQR)</b>	8.0 (8.0-8.4)
<b>HBV-RNA, log<sub>10</sub> copies/mL, median (IQR)</b>	5.2 (4.9-6.0)
<b>Fibrosis n (%)<sup>a</sup></b>	
Non-significant	8 (61.6)
Significant	3 (23.1)
<b>Cirrhosis n (%)</b>	2 (15.4)

**Table 13. Clinical, virological, and serological parameters from the 13 patients included in the study.** <sup>a</sup>Non-significant fibrosis by liver biopsy indicates Ishak fibrosis stage <3, significant fibrosis F≥3 and <5, and cirrhosis F5-6. Non-significant 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 WHO Guidelines for the prevention, care and treatment of persons with chronic hepatitis B infection criteria [185]. In this study, HBV-RNA was quantified using an in-house RT-qPCR assay (see section 4.3. *Hepatitis B virus DNA and RNA quantification*). Abbreviations: IQR, interquartile range; ALT, alanine aminotransferase; HBeAg +ve, hepatitis B e-antigen positive; HBsAg, hepatitis B virus surface antigen; HBV, hepatitis B virus; NS: Not specified.

### 5.1.3. Verification of residual hepatitis B virus DNA from RNA isolations

In this study, a protocol based on DNase I digestion (Life Technologies, Austin, United States) was used to eliminate residual HBV-DNA from HBV-RNA isolated in samples. The verification of residual DNA elimination by qPCR confirmed the absence of any traces of HBV-DNA from DNase I-treated isolated HBV-RNA; thus, no HBV-DNA contamination was present in cDNA resulting from HBV-RNA reverse transcription. An illustrative example is shown in **Figure 23**.



**Figure 23. The qPCR verification of residual hepatitis B virus DNA elimination RNA isolations after DNase I treatment.** Fluorescence through qPCR amplification samples is shown as green lines for DNase I-treated RNA samples, and as red lines for their respective cDNA reverse transcribed samples.

### 5.1.4. Next-generation sequencing results

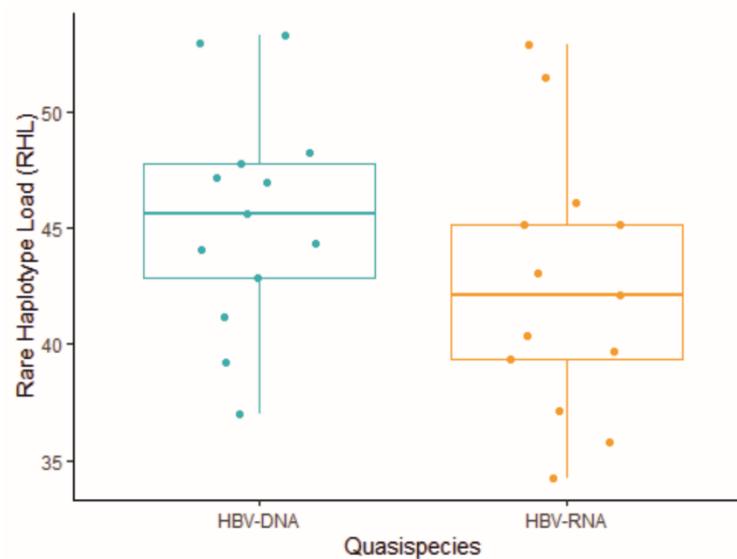
#### 5.1.4.1. Hepatitis B virus DNA and RNA quasispecies complexity analysis

In the present study, the complexity in the *HBX* 5' end of both HBV-DNA and HBV-RNA QS in the same sample was assessed by their Rare Haplotype Load (RHL). This index measures enrichment of each QS in minority viral populations, i.e., below a given threshold of frequency, which was set to 1% according to previous studies [212]. Therefore, this index allowed evaluating potential variations in the size of mutant spectrum between circulating HBV-DNA and RNA, probably due to the mutagenic effects of reverse transcription from HBV-RNA to DNA. To compute RHL, the set of haplotypes common to forward and reverse strands were taken before subsequent filtering by abundance. In this sense, a total of  $4.35 \times 10^6$  sequence reads with a median (IQR) of  $2.45 \times 10^5$  reads/sample ( $1.90 \times 10^5$ - $3.19 \times$

$10^5$ ) for HBV-RNA and  $8.27 \times 10^4$  reads/sample ( $6.03 \times 10^4 - 1.01 \times 10^5$ ) for HBV-DNA were analysed.

The RHL analysis showed a heterogeneous behaviour between patients. The following was observed:

- Mean RHL was greater in DNA ( $45.45 \pm 4.80$ ) than in RNA ( $42.50 \pm 5.63$ ), the difference was not significant ( $p = 0.1641$ ) (**Figure 24**).



**Figure 24. Comparison of mean hepatitis B virus quasispecies complexity.** Hepatitis B virus quasispecies complexity analyzed by rare haplotype load of all 13 patients in hepatitis B virus DNA quasispecies (blue-framed box) and RNA quasispecies (orange-framed box). Differences were statistically non-significant ( $p = 0.1641$ ). *Abbreviations: HBV, hepatitis B virus.*

- Similarly, no statistically significant differences were observed when comparing HBV-DNA and HBV-RNA QS separating patients according to the severity of liver disease (**Table 14**).

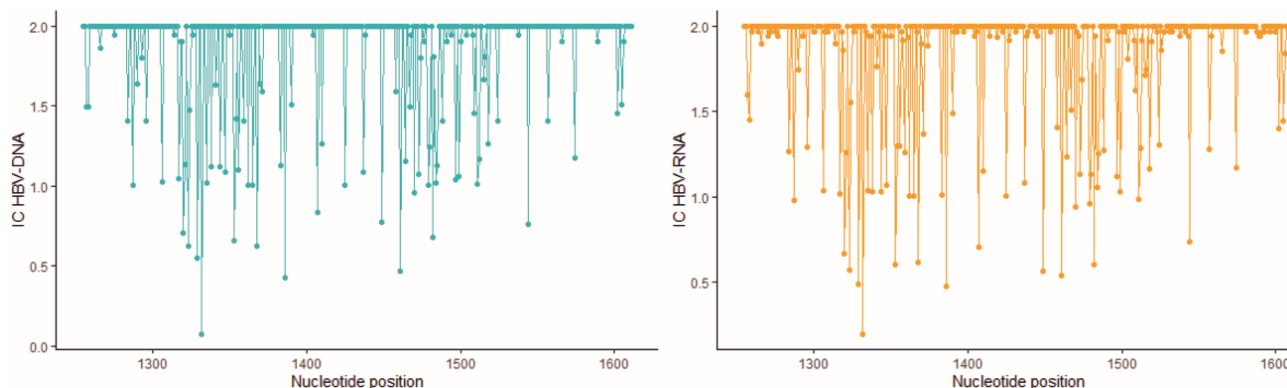


	Significant fibrosis or cirrhosis (n = 5)	Non-significant fibrosis (n = 8)	p-value*
<b>HBV-DNA QS</b>	44.09 (41.15-47.8)	46.31 (44-48.65)	0.4642
<b>HBV-RNA QS</b>	40.38 (37.12-43.06)	43.63 (39.6-47.44)	0.3055
p-value*	0.2506	0.2936	

**Table 14. Comparison of rare haplotype load between hepatitis B virus DNA and RNA quasispecies in patients separated by severity of liver disease.** Median (interquartile range) values are shown in each cell. A p-value  $\leq 0.05$  was considered statistically significant. *Abbreviations: HBV, hepatitis B virus; QS, quasispecies. \* Kruskal-Wallis test.*

#### 5.1.4.2. Hepatitis B virus DNA and RNA quasispecies conservation analysis

Conservation in the *HBX* 5' end of both HBV-DNA and HBV-RNA QS was analysed at the same sample by calculating the information content (IC) of each of 357 positions between nt 1255 – 1611, in a multiple alignment of all the different haplotypes [2]. To compute IC, all haplotypes with abundances below 0.25% were eliminated, obtaining  $2.91 \times 10^6$  sequence reads with a median (IQR) of  $1.58 \times 10^5$  reads/sample ( $1.29 \times 10^5 - 2.28 \times 10^5$ ) for HBV-RNA and  $5.66 \times 10^4$  reads/sample ( $3.2 \times 10^4 - 5.76 \times 10^4$ ) for HBV-DNA (**Figure 25**).

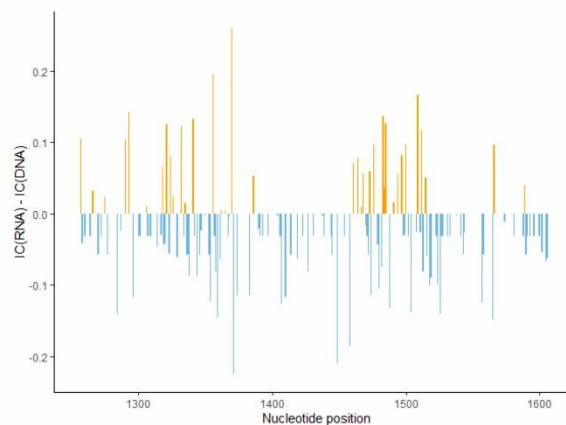


**Figure 25. Conservation and variability of 357 nucleotide positions analyzed.** IC of nt positions from 1255 to 1611 for HBV-DNA (blue lines) and HBV-RNA (orange lines) QS. IC ranges from 0 bits to 2 bits for a given position, where 2 bits represents the maximum IC value (no variability in that position, i.e., the same nt in all haplotypes in the alignment). *Abbreviations: IC, information content; nt, nucleotide; HBV, hepatitis B virus.*

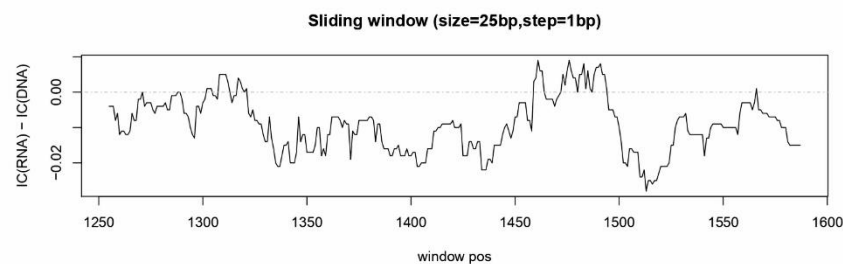
Differences in sequence conservation between both QS were determined by subtracting IC DNA values from IC RNA at each nt position analysed:

- In most of these positions, 218 (61.06%), IC values of both HBV-RNA and HBV-DNA QS were coincident. In the remaining 139 (38.93%) nt positions, differences between IC of both QS ranged from -0.26 to 0.23, and IC DNA > IC RNA in most of them (102/139, 73.38%) (**Figure 26; A**).
- 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) confirmed that the HBV-RNA QS was slightly less conserved than the DNA QS (**Figure 26; B**).
- In keeping with these results, HBV-RNA displayed more positions with some variability (IC < 2 bits), 135/357 (38%), than HBV-DNA, 85/357 (24%) ( $p < 0.01$ ).

**A**



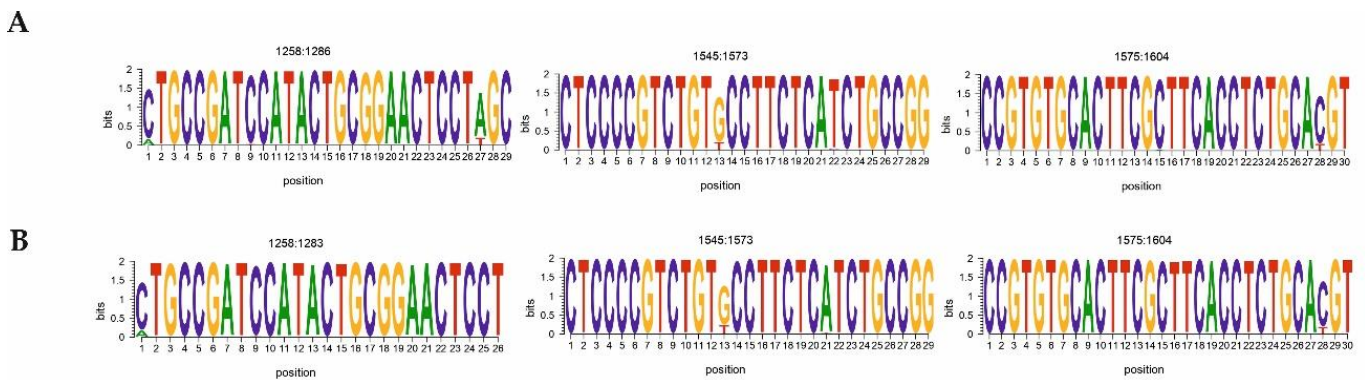
**B**



**Figure 26. Differences between information content of hepatitis B virus DNA and RNA quasispecies (A).** Nucleotide positions in which information content of hepatitis B virus RNA > DNA are depicted in orange lines while positions where information content of hepatitis B virus DNA > RNA in blue lines. **(B).** Sliding window analysis of the of the subtraction of mean information content RNA minus information content DNA values [IC(RNA)-IC(DNA)], in windows of 25 nt positions, displaced in steps of 1 position between them. *Abbreviations: IC, information content.*

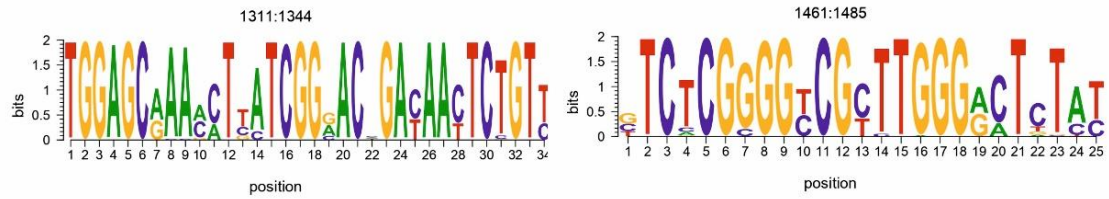
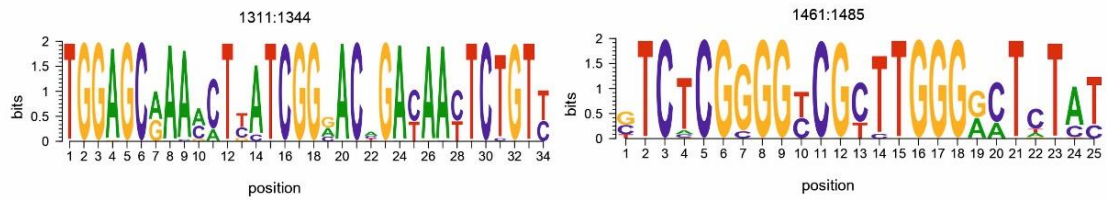
In order to confirm the highly conserved regions that we observed in previous studies in serum HBV-DNA QS [2][3], we performed a sliding window analysis of IC (computed in the same way as that of IC RNA-IC DNA values) in the multiple alignments of haplotypes obtained in both HBV-DNA and HBV-RNA QS. Then, concatenation of the 5% of windows with the highest and lowest mean IC values yielded seven different sequence logos:

- The most conserved windows (those with the highest IC values) gave rise to four sequence logos in both QS, three of which coincided between both QS: nts 1258-1286 (1283 in HBV-RNA QS), 1545-1573 and 1575-1604 (**Figure 27**). In addition, the sequence stretch 1519-1543 was found to be hyper-conserved only in HBV-DNA QS and the stretch 1559-1587 was found to be hyper-conserved only in HBV-RNA QS.



**Figure 27. Representation by sequence logos of the information content of the most conserved regions that coincided in hepatitis B virus DNA (A) and RNA (B) quasispecies.** The relative sizes of the letters in each stack, each of them representing a nucleotide position, 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).

- The most variable windows gave rise to two sequence logos, which coincided in both QS: nts 1311-1344 and 1461-1485 (**Figure 28**).

**A****B**

**Figure 28. Representation by sequence logos of the information content of the most variable regions in hepatitis B virus DNA (A) and RNA (B) quasispecies.** The relative sizes of the letters in each stack, each of them representing a nucleotide position, 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).

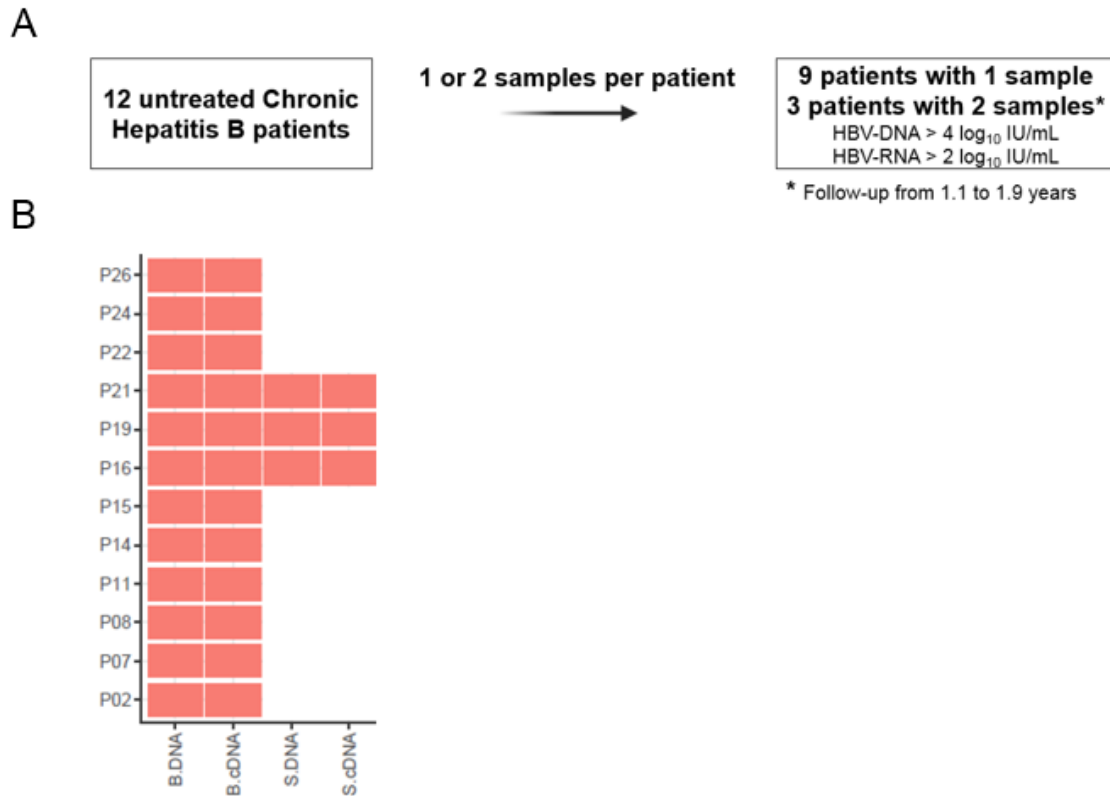
## 5.2. Second study: Comparison of circulating hepatitis B virus DNA and RNA quasispecies in the surface/polymerase overlapping region

### 5.2.1. Aims

In this study, we aimed to extend the study of similarities and differences between circulating HBV-DNA and RNA QS. Therefore, we first moved to a new region of the viral genome that is also relevant to HBV replication. In this sense, we studied HBV-DNA and HBV-RNA QS in the surface/polymerase overlapping region, analysing a fragment ranging from nt 615 to 972 of the HBV genome. This fragment covers most of the functional domain B and all of the functional domains C-E involved in reverse transcription [50]. Secondly, we used a set of genetic diversity and similarity indexes and genetic distances to improve our understanding of the similarities and differences between HBV-RNA and HBV-DNA QS, and their variation over time in two different samples.

### 5.2.2. Study design

It is a retrospective cross-sectional and longitudinal study in collaboration with the Research Group on Viral Hepatitis, Metabolic Liver Diseases and Autoimmune Hepatitis of the Hospital Clinic de Barcelona (Barcelona, Spain). From a group of 29 well-characterised CHB patients attending Vall d'Hebron and Hospital Clinic outpatient clinics, 12 untreated patients met the inclusion criteria shown in **Figure 29**, which allowed amplification of the surface/polymerase overlapping region of the viral genome in circulating HBV-DNA and RNA QS. Both QS were analysed using the NGS methodology established in the previous study.



**Figure 29. Diagram of the study design.** (A) Diagram of the patients included in the study, together with the inclusion criteria. (B) Diagram of 15 samples analysed from all 12 patients included: all had a baseline sample from which hepatitis B virus DNA and cDNA (the latter derived from HBV-RNA) were obtained. In addition, three of the 12 patients had follow-up samples from which DNA and RNA were also obtained. *Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; Patients are designated as PXX; B.DNA, baseline DNA quasispecies; B.cDNA, baseline cDNA quasispecies; S.DNA, follow-up DNA quasispecies; S.cDNA, follow-up cDNA quasispecies.*

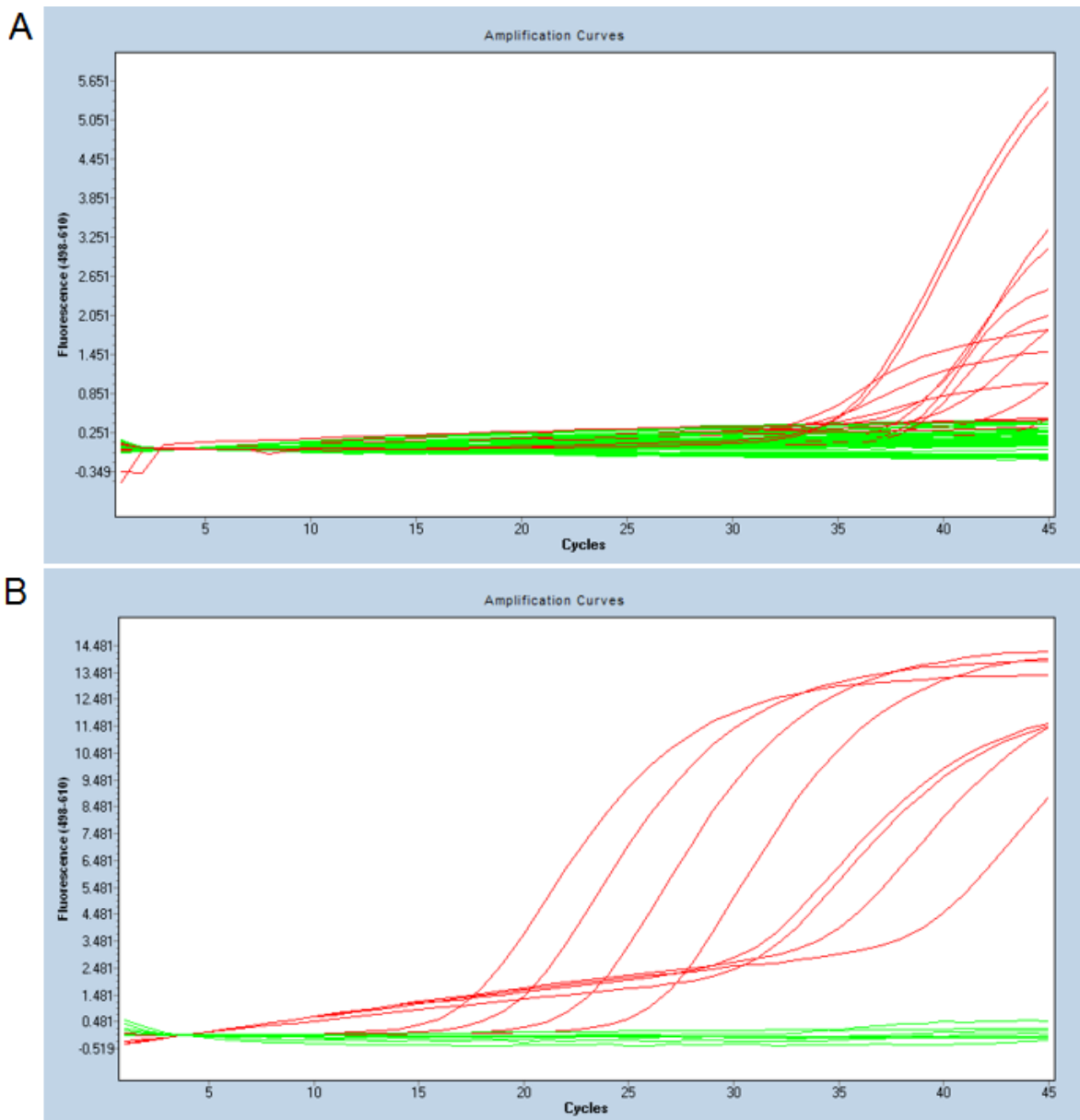
The two longitudinal serum samples obtained from three untreated patients were collected 1.1 to 1.9 years apart. Demographic and baseline, biochemical, virological, and serological parameters of the 12 patients at the time of sample collection are summarised **Table 15**.

<b>Total, n = 12</b>	
<b>Age, years, median (IQR)</b>	53.0 (46.6-55.1)
<b>Male, n (%)</b>	6 (50.0)
<b>Ethnic group, n (%)</b>	
Caucasian	6 (50.0)
African	1 (8.3)
Hispanic	1 (8.3)
Asian	4 (33.3)
<b>ALT, IU/L, median (IQR)</b>	77.0 (48.5-153.8)
<b>HBeAg +ve, n (%)</b>	9 (75.0)
<b>HBV-DNA, log<sub>10</sub>IU/mL, median (IQR)</b>	8.1 (7.4-8.6)
<b>HBV-RNA, log<sub>10</sub> copies/mL, median (IQR)</b>	6.0 (4.5-6.6)
<b>FIB-4, median (IQR)</b>	1.7 (1.1-3.1)

**Table 15. Demographical and baseline, biochemical, virological, and serological characteristics of chronic hepatitis B patients enrolled in the study.** In this study, hepatitis B virus RNA was quantified by RT-qPCR using an automated assay on the COBAS 6800 System (Roche Diagnostics, Mannheim, Germany) [190][191] (see section 4.3. Hepatitis B virus DNA and RNA quantification). Abbreviations: IQR, interquartile range; ALT, alanine aminotransferase; HBeAg +ve, hepatitis B e-antigen positive; HBV, hepatitis B virus; FIB-4, fibrosis-4 score.

### 5.2.3. Verification of residual hepatitis B virus DNA from RNA isolations

The DNase I protocol used in the first study to eliminate residual HBV-DNA from HBV-RNA isolated from samples was not fully effective in all samples included in this second study (an example of qPCR verification is shown in **Figure 30; A**). Therefore, a more rigorous DNase treatment was used in this study to remove residual HBV-DNA from the the isolated HBV-RNA. Verification of residual DNA removal by qPCR confirmed that there were no traces of contaminating HBV-DNA in HBV-RNA isolates treated with this DNase protocol. An illustrative example is shown in **Figure 30; B**. Thus, effective DNA digestion was achieved with this more rigorous DNase digestion.



**Figure 30. Example of qPCR verification of residual hepatitis B virus DNA elimination from RNA isolations after DNase treatment. (A)** Fluorescence through qPCR amplification is shown as green lines in all RNA isolates where DNA was successfully removed after DNase I treatment established in the first study, and in red lines all those DNase I-treated RNA isolations that continued showing DNA amplification. **(B)** The first four red lines starting from the left stand for the fluorescence for DNA isolations of four out of 15 samples selected with the highest viral load ( $> 8 \log_{10}$  IU/mL). The remaining four red lines stand for the fluorescence for the same isolations after the more rigorous DNase digestion used in the present study, and reverse transcription to cDNA. In green, fluorescence for RNA isolations successfully treated with the same rigorous DNase digestion protocol.



#### 5.2.4. Next-generation sequencing results

A total of  $2.23 \times 10^6$  sequence reads ( $1.08 \times 10^6$  and  $1.15 \times 10^6$  for HBV-DNA and HBV-RNA, respectively) with a median (IQR) of  $7.22 \times 10^4$  reads/sample ( $5.67 \times 10^4 - 8.43 \times 10^4$ ) for HBV-DNA and  $7.49 \times 10^4$  reads/sample ( $5.67 \times 10^4 - 8.47 \times 10^4$ ) for HBV-RNA were obtained. **Table 16** shows these observed values for all 15 samples included patients.

Patient	B.cDNA	B.DNA	S.cDNA	S.DNA	Genotype
	% of master frequency (total n° of reads)	% of master frequency (total n° of reads)	% of master frequency (total n° of reads)	% of master frequency (total n° of reads)	
P02	85.92 (32987)	14.05 (45280)			H
P07	57.58 (41737)	78.73 (57396)			C
P08	83.88 (101915)	84.62 (56710)			A
P11	86.57 (31780)	42.92 (21339*)			A
P14	9.04 (61828)	17.39 (56705)			B
P15	44.83 (87909)	30.01 (56522)			A
P16	66.73 (64748)	84.32 (103086)	62.42 (87246)	83.23 (87426)	A
P19	86.06 (86935)	81.76 (79589)	85.78 (78983)	79.82 (75349)	C
P21	62.06 (95955)	71.36 (84743)	86.00 (67675)	52.48 (90041)	D
P22	82.20 (97074)	82.53 (70673)			E
P24	81.26 (65715)	57.73 (73634)			A
P26	54.38 (79513)	46.65 (32606)			C

**Table 16. Number of reads obtained and percentage of master haplotypes for included samples.** Genotype was determined by DB-rule from both hepatitis B virus DNA and RNA masters. \*Rarefaction was performed considering the minimum coverage of 21339 reads obtained. Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; Patients are designated as PXX; B.DNA, baseline DNA quasispecies; B.cDNA, baseline cDNA quasispecies; S.DNA, follow-up DNA quasispecies; S.cDNA, follow-up cDNA quasispecies.

The HBV-DNA and HBV-RNA QS coverages were high in all samples (>10000 reads) [200], and no apparent bias towards HBV-DNA or HBV-RNA was observed (**Table 16**). This allowed the comparison of both HBV-DNA and HBV-RNA QS to proceed.

#### *5.2.4.1. Comparison of master haplotypes*

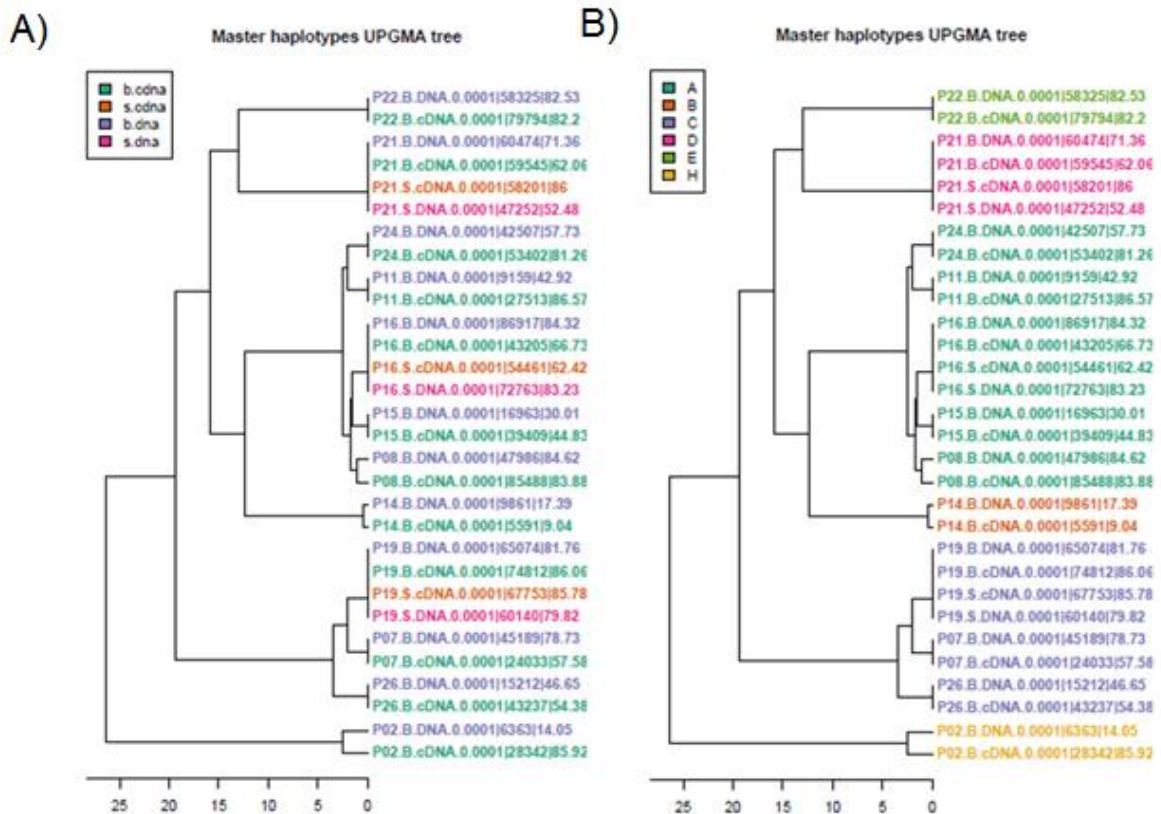
It is noteworthy the differences between the frequencies of the master haplotypes of both HBV-DNA and HBV-RNA QS, according to which we grouped patients into three patterns:

- A) High frequencies of master haplotypes, which were similar between HBV-DNA and HBV-RNA: P08 (84.62% vs. 83.88%, respectively), P19 (81.76% vs. 86.06% at baseline and 79.82% vs. 85.78% at follow-up) and P22 (82.53% vs. 82.2%, respectively).
- B) Master haplotype of HBV-DNA more frequent than that of HBV-RNA: P07 (78.73% vs. 57.58%), P14 (17.39% vs. 9.04%), P16 (84.32% vs. 66.73% at baseline and 83.23% vs. 62.42% at follow-up) and P21 (71.36% vs. 62.06% at baseline).
- C) Master haplotype of HBV-DNA less frequent than that of HBV-RNA: P02 (14.05% vs. 85.92%), P11 (42.92% vs. 86.57%), P15 (30.01% vs. 44.83%), P21 (52.48% vs. 86% at follow-up), P24 (57.73% vs. 81.26%) and P26 (46.65% vs. 54.38%).

Patients with longitudinal samples followed different trends during their follow-up. P19 showed similar frequencies of master haplotypes at baseline (81.76% vs. 86.06% for HBV-DNA and HBV-RNA QS, respectively) and follow-up (79.82% vs. 85.78% for HBV-DNA and HBV-RNA QS, respectively). In P16, the frequency balance of the master haplotypes remained almost identical, with a higher proportion of HBV-DNA master (84.32% vs. 83.23% at baseline and follow-up, respectively) than HBV-RNA master (66.73% vs. 62.42% at baseline and follow-up, respectively). In contrast to these two patients, in P21 while frequency of master haplotype was slightly higher in HBV-DNA than in HBV-RNA QS at baseline (71.36% vs. 62.06%, respectively), a clear decrease in HBV-DNA and increase in HBV-RNA master haplotype frequencies was observed at follow-up (52.48% vs. 86%, respectively), reversing the baseline situation since the HBV-RNA master haplotype became the most represented.

The UPGMA tree of the master haplotypes of all samples showed that most of patients (9/12, 75%) maintained the same master haplotypes in both HBV-DNA and HBV-RNA QS (**Figure 31; A**). Among patients with different master haplotypes in both QS, P02, P08 and P14, the P02 had the largest genetic distances between HBV-DNA and HBV-RNA master haplotypes, similar to genetic distances between the master haplotypes of different patients with the same genotype. Moreover, the same master haplotype was maintained during

longitudinal follow-up in patients P16, P19, P21. When analysing the genotype of these haplotypes (**Figure 31; B, Table 16**), it was observed in all patients that they coincided between HBV-DNA and HBV-RNA QS in both baseline and follow-up samples. No minor haplotypes with different genotypes were observed in any QS.



**Figure 31. Master haplotypes based on raw nucleotide distances. (A)** Master haplotypes are coloured according to the type of quasispecies (HBV-DNA or RNA [cDNA]) and time point (baseline [B] or follow-up [S]). **(B)** Master haplotypes are coloured according to the hepatitis B virus genotype assigned by DB-rule. *Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; Patients are designated as PXX; B.DNA, baseline DNA quasispecies; B.cDNA, baseline cDNA quasispecies; S.DNA, follow-up DNA quasispecies; S.cDNA, follow-up cDNA quasispecies.*

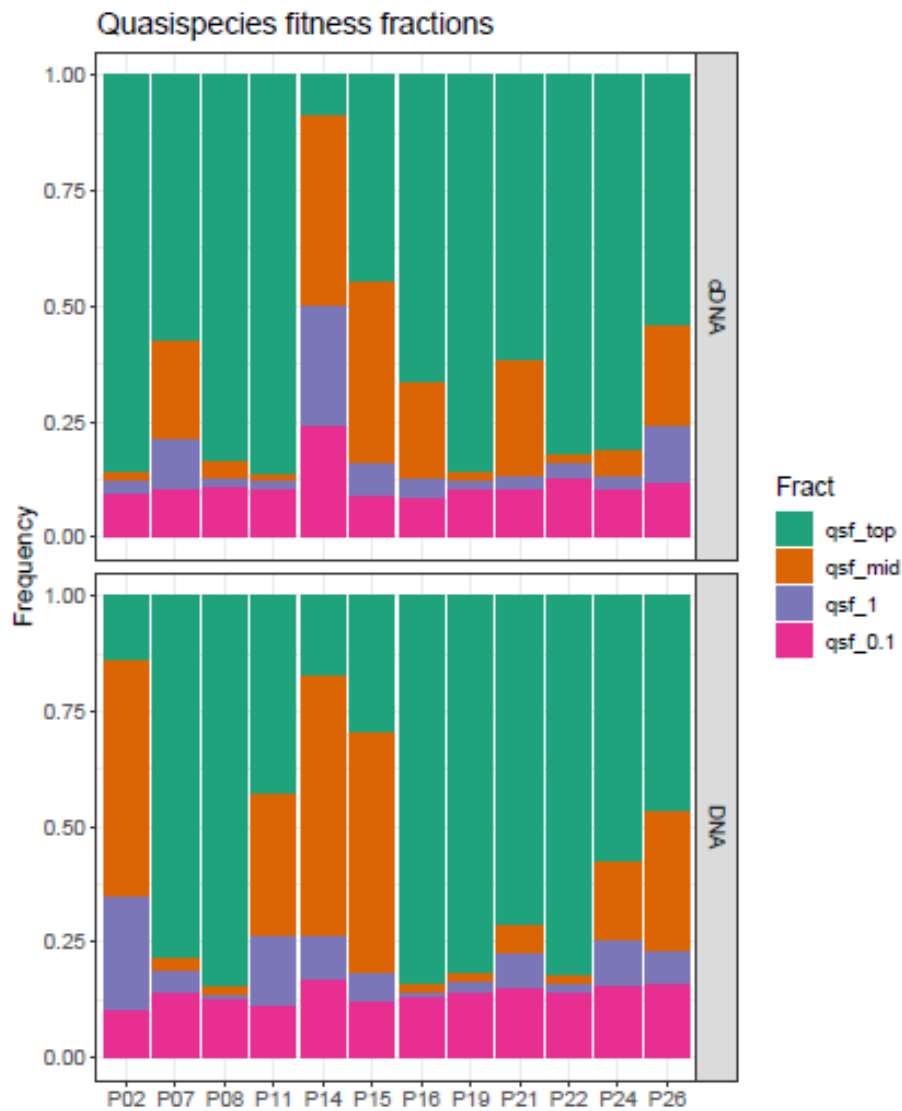
#### 5.2.4.2. Quasispecies complexity assessment with quasispecies fitness fractions and Hill number profiles

The composition of the QS at a given point in time and its evolution over a longitudinal follow-up of 1.1 to 1.9 years was also represented as the changes observed in the volume (fraction of molecules, i.e., sequence reads) of the four QS fitness fractions (QFF) combined with the Hill number profile (HNP). This combination allowed us to visualise and analyse the distribution of molecules within the QS into a dominant, emergent and a set of minority

haplotypes with low and very low fitness together with the distribution of relative abundances (i.e., frequencies) of the haplotypes (i.e., viral populations) within the QS. Due to the sensitivity of QFF and HNP to the different coverage obtained for each QS analysed, their calculations were performed with rarefaction, i.e., calculating the average of the QFF and HNP values obtained after 1000 cycles of resampling each QS to the lowest coverage obtained (21339 reads, **Table 16**) with repositioning.

#### 5.2.4.2.1. Baseline samples

With regard to the QFF distribution of haplotypes (**Figure 32**), three different patterns were observed in relation to the previous three in terms of differences between the frequencies of the master haplotype of both HBV-DNA and HBV-RNA QS:



**Figure 32. Partitioning of paired hepatitis B virus DNA and RNA (cDNA) quasispecies haplotypes into four fractions according to their fitness in baseline samples.** Four fractions are represented: the master haplotype, the emerging haplotypes (present at frequencies > 1% but less than that of the master haplotype), and the low and very low fitness haplotypes (present at 0.1-1% and at < 0.1%, respectively). The values used to generate this quasispecies fitness fraction plot, scaled from 0 to 1, are those given in **Appendix 3. Table 1.** *Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; qsf\_top, master fraction; qsf\_mid, emerging fraction; qsf\_1, low fitness fraction; qsf\_0.1, very low fitness fraction.*

- A) Both HBV-RNA and HBV-DNA QS showed similar distributions of haplotype fractions.
- a. P08, P19 and P22 (*pattern A*, see section 5.2.4.1. *Comparison of master haplotypes*) showed large fractions of master haplotypes that were dominant

in both HBV-DNA and HBV-RNA QS (84.60% vs. 83.89%, 81.75% vs. 86.06% and 82.53% vs. 82.19% in HBV-DNA and HBV-RNA QS for P08, P19 and P22, respectively) (**Figure 32** and **Appendix 3. Table 1**). In addition, all these patients showed emerging and low fitness fractions reduced in both QS, with most of the molecules not belonging to master haplotype fraction in the very low fitness haplotypes fraction.

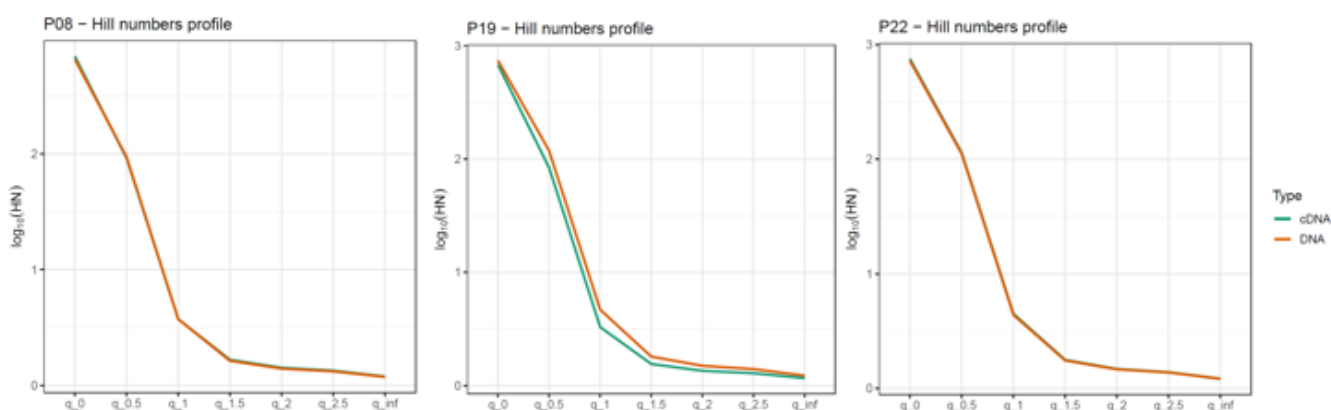
b. On the other hand, P14, P15 and P26 showed a reduced master haplotype fraction, in relation to the previous group of patients, and increased fractions of emerging and rare haplotypes in both HBV-DNA and HBV-RNA QS (frequencies of master haplotypes were 17.37% vs. 9.04%, 30.01% vs. 44.81% and 46.64% vs. 54.37% in P14, P15 and P26 for HBV-DNA and HBV-RNA QS, respectively) (**Figure 32** and **Appendix 3. Table 1**). Despite similarities between the QFF in HBV-DNA and HBV-RNA, the fraction of master haplotypes was slightly higher in HBV-DNA than in HBV-RNA for P14 and higher in HBV-RNA than in HBV-DNA for P15 and P26.

B) In P02, P11 and P24 the QFF of the HBV-DNA QS showed a shrinking master haplotype fraction in parallel with an increasing volume of molecules belonging to emerging haplotypes. In contrast, the HBV-RNA QS showed an opposite trend (frequencies of master haplotypes were 14.05% vs. 85.91%, 42.93% vs. 86.57%, 57.71% vs. 81.26% in P02, P11 and P24 for HBV-DNA and HBV-RNA QS, respectively) (**Figure 32** and **Appendix 3. Table 1**). All patients with this profile were also grouped as patients whose HBV-DNA master haplotype was less frequent than that of HBV-RNA (pattern C, see *section 5.2.4.1. Comparison of master haplotypes*).

C) On the other hand, P07, P16 and P21 showed a larger master haplotype and a shrinking fraction of emerging haplotypes in HBV-DNA compared to HBV-RNA QS (frequencies of master haplotypes were 78.72% vs. 57.58%, 84.32% vs. 66.72%, 71.35% vs. 62.06% in P07, P16 and P21 for HBV-DNA and HBV-RNA QS, respectively) (**Figure 32** and **Appendix 3. Table 1**). All the patients with this profile were also grouped as patients with an HBV-DNA master haplotype more frequent than that of HBV-RNA (pattern B, see *section 5.2.4.1. Comparison of master haplotypes*).

We decided to study the HNP grouping of patients in the three different patterns observed in relation to the master haplotype frequency (see *section 5.2.4.1. Comparison of master haplotypes*):

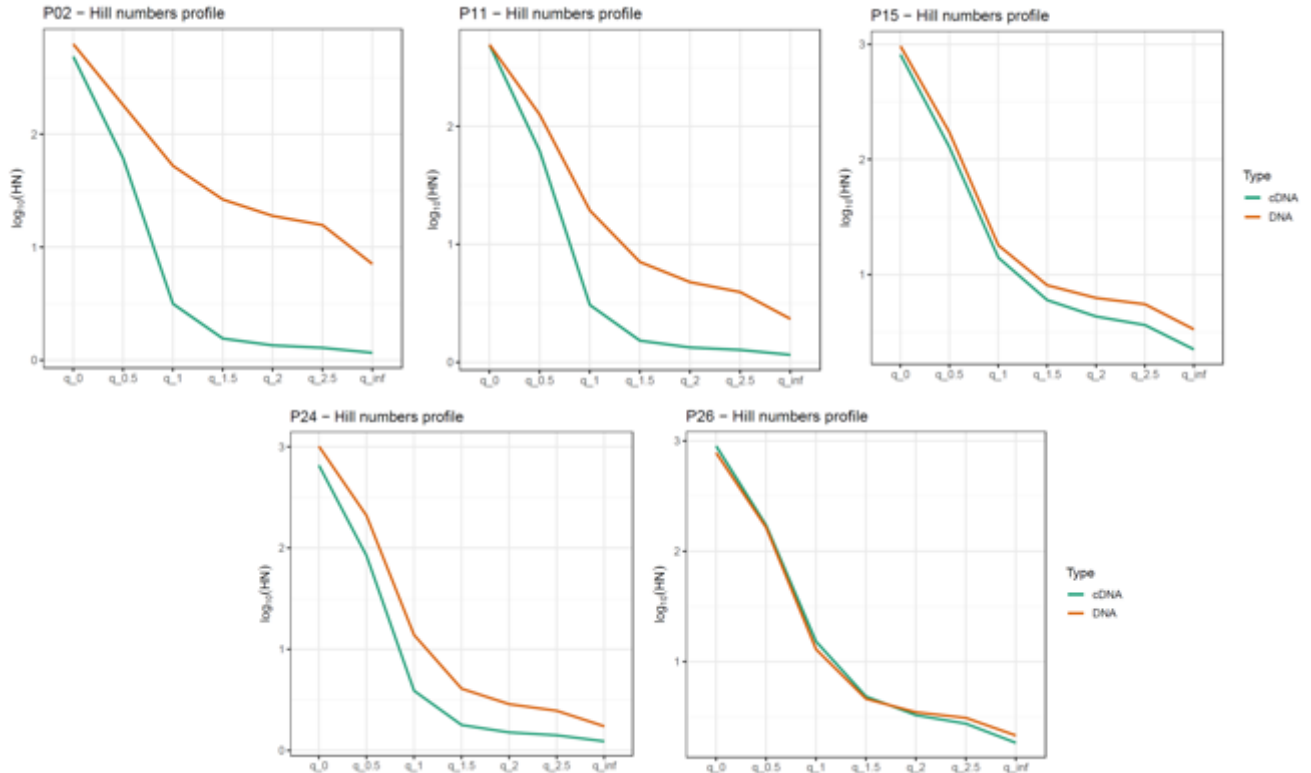
A) P08, P19 and P22 showed similar large master haplotype fractions (**Figure 33** and **Appendix 3. Table 2**). Their HNPs were characterised by a steeply sloping curve resulting in a minimal diversity, i.e., an uneven distribution of haplotypes due to dominance of the large master haplotype fraction over the others. Furthermore, these patients had almost identical frequency haplotypes between HBV-DNA and HBV-RNA QS, as shown by their overlapping or almost overlapping HNP curves. P08 is of particular interest because, although the HBV-DNA and HBV-RNA master haplotypes were not identical, the QS distribution in both sample types followed exactly the same trend.



**Figure 33. Hill numbers profile plots of baseline samples from P08, P19 and P22.** The x-axis is the order  $q$  in the Hill number plotted for values of  $q$  from 0 to  $\infty$  (inf). The y-axis is the decimal logarithm of the calculated Hill number values of each order  $q$ . For  $q = 0$ , the Hill number is the haplotype richness. Because the Hill number represents the equivalent number of equally abundant haplotypes, a perfectly even haplotype distribution would not change as  $q$  is increases (the Hill numbers profile would be a high, flat line). Larger values of  $q$  place progressively more weight on the master haplotype, so the equivalent number of equally abundant haplotypes would be much lower for the more uneven haplotype distribution than for the more even one. Hepatitis B virus DNA samples are shown in orange and RNA (cDNA) in green. *Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; HN, Hill numbers.*

B) P02, P11, P15, P24 and P26 showed lower master haplotype fractions in HBV-DNA than in HBV-RNA QS (**Figure 34** and **Appendix 3. Table 2**). Thus, they generally showed a higher curve in HBV-DNA, corresponding to higher diversity, than in HBV-RNA QS. Of note, P15 and P26 had more similar frequencies of HBV-DNA and HBV-RNA master haplotypes. They also showed closer HNP curves, even overlapping in P26. On the other hand, in P02, P11 and P24, the HBV-RNA displayed a much steeply sloping curve than HBV-DNA QS, indicating a more uneven distribution. The particular case of P02 showed the greatest haplotype

diversity in terms of HBV-DNA QS, whereas HBV-RNA QS showed a much steeper slope, indicating lower diversity.

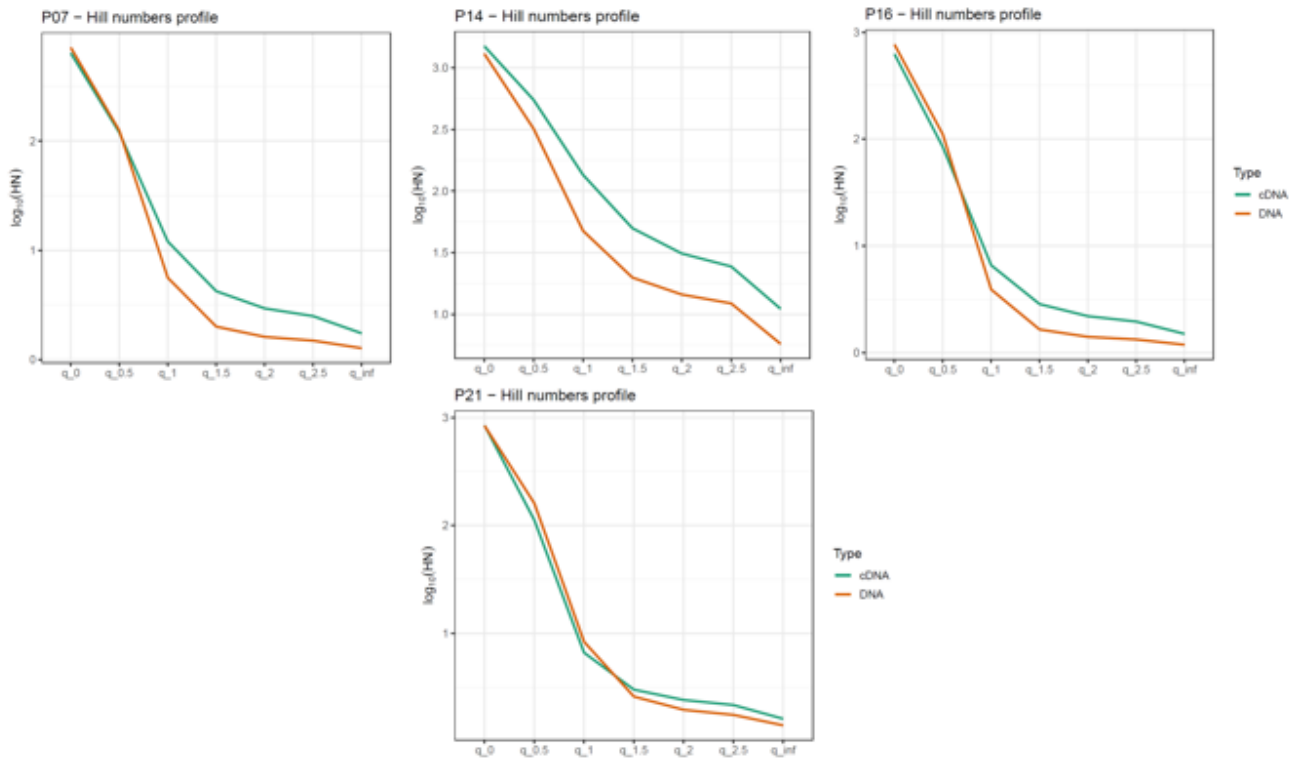


**Figure 34. Hill numbers profile plots from baseline samples of P02, P11, P15, P24 and P26.** The x-axis is the order  $q$  in the Hill number plotted for values of  $q$  from 0 to  $\infty$  (inf). The y-axis is the decimal logarithm of the calculated Hill number values of each order  $q$ . For  $q = 0$ , the Hill number is the haplotype richness. Because the Hill number represents the equivalent number of equally abundant haplotypes, a curve with a perfectly even haplotype distribution would not change as  $q$  increases (the Hill numbers profile would be a high, flat line). Larger values of  $q$  place progressively more weight on the master haplotype, so the equivalent number of equally abundant haplotypes would be much lower for the more uneven haplotype distribution than for the more even one. Hepatitis B virus DNA samples are shown in orange and RNA (cDNA) in green. *Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; HN, Hill numbers.*

C) P07, P14, P16 and P21 showed higher master haplotype fractions in HBV-DNA than in HBV-RNA QS (**Figure 35** and **Appendix 3. Table 2**). In general, these patients showed a similar haplotype abundance distribution between HBV-DNA and HBV-RNA QS, although with an inverse relationship to the previous pattern (B), since the higher curve was that of HBV-RNA QS. The HBV-DNA QS reported a minor diversity. Of note, the two HBV-DNA and HBV-RNA curves of P14 were higher than to the other patients, coinciding with high emerging and low fitness QFF. In the other



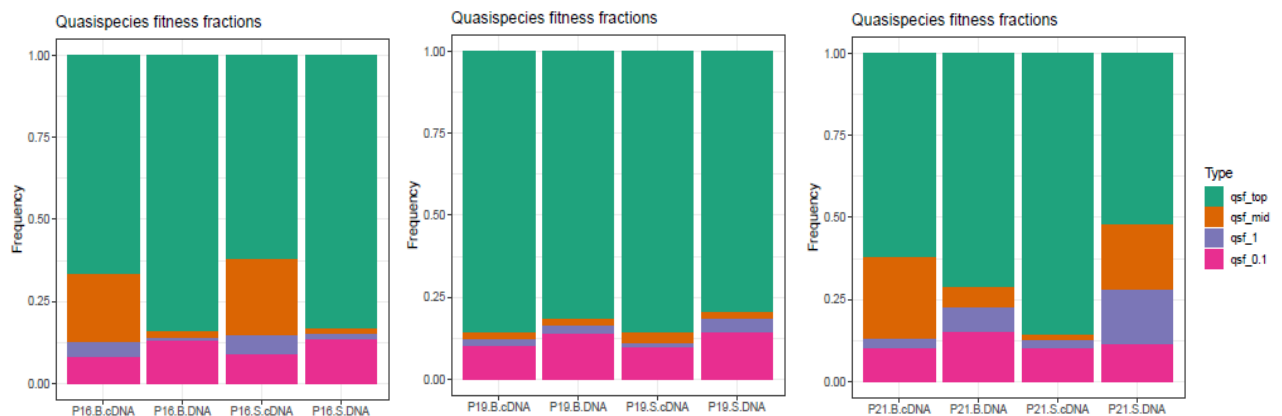
patients, the HBV-DNA HNP curve was higher than that of the HBV-RNA QS at  $q = 0$ , evidencing a higher number of haplotypes in DNA than in RNA. However, at higher orders ( $q$  values) the HBV-RNA curve crosses over to that of HBV-DNA due to the greater diversity of HBV-RNA QS, even with fewer haplotypes than HBV-DNA.



**Figure 35. Hill numbers profile plots from baseline samples of P07, P14, P16 and P21.** The x-axis is the order  $q$  in the Hill number plotted for values of  $q$  from 0 to  $\infty$  (inf). The y-axis is the decimal logarithm of the calculated Hill number values of each order  $q$ . For  $q = 0$ , the Hill number is the haplotype richness. Because the Hill number represents the equivalent number of equally abundant haplotypes, a curve with a perfectly even haplotype distribution would not change as  $q$  increases (the Hill numbers profile would be a high, flat line). Larger values of  $q$  place progressively more weight on the master haplotype, so the equivalent number of equally abundant haplotypes would be much lower for the more uneven haplotype distribution than for the more even one. Hepatitis B virus DNA samples are shown in orange and RNA (cDNA) in green. *Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; HN, Hill numbers.*

#### 5.2.4.2.2. Comparison of HBV-DNA and HBV-RNA quasispecies evolution

The changes occurring in the viral QS of three out of 12 patients over 1.1-1.9 years of follow-up are shown in **Figure 36**.



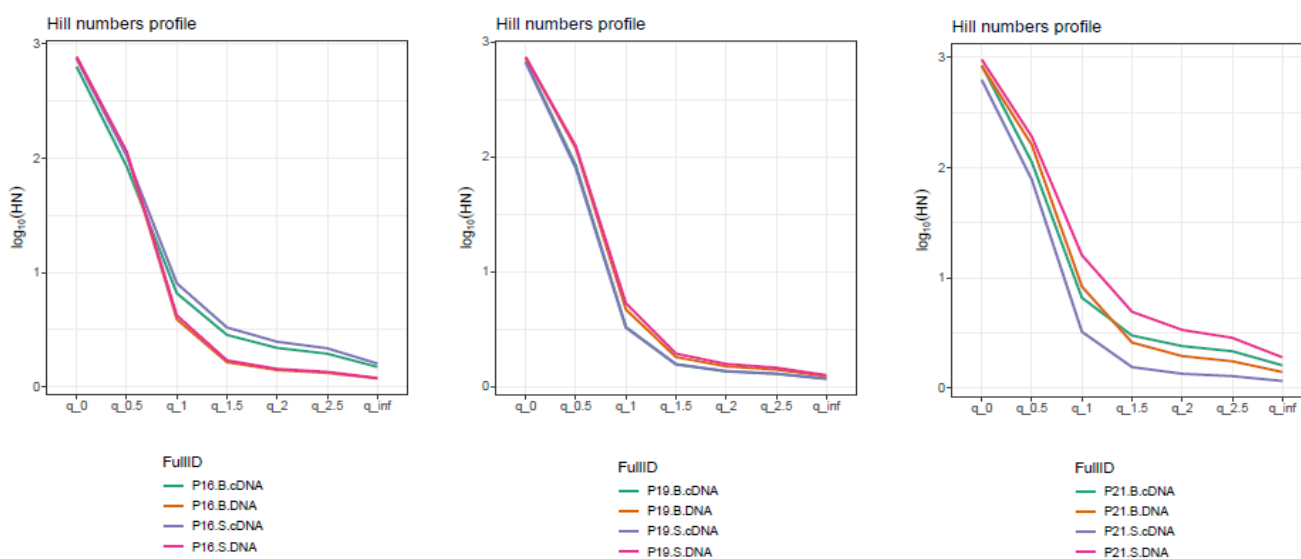
**Figure 36. Evolution of the distribution of hepatitis B virus DNA and RNA quasispecies haplotypes into four fractions according to their fitness for the three chronic hepatitis B patients with follow-up samples.** Four fractions are represented: the master haplotype, the emerging haplotypes (present at frequencies > 1% but less than that of the master haplotype), and the low and very low fitness haplotypes (present at 0.1-1% and at < 0.1%, respectively). The values used to generate this quasispecies fitness fraction plot, scaled from 0 to 1, are those given in **Appendix 3. Table 1**. *Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; Patients are designated as PXX; B.DNA, baseline DNA quasispecies; B.cDNA, baseline cDNA quasispecies; S.DNA, follow-up DNA quasispecies; S.cDNA, follow-up cDNA quasispecies; qsf\_top, master fraction; qsf\_mid, emerging fraction; qsf\_1, low fitness fraction; qsf\_0.1, very low fitness fraction.*

In terms of distributions of haplotype fractions according to their fitness (QFF), each patient follows distinct QS evolution:

- At baseline, HBV-DNA QS of the P16 (**Figure 36, left panel**) showed a larger master haplotype and shrinking emerging haplotypes fractions (2.10%) compared to HBV-RNA QS (20.77%). This picture was maintained at follow-up, with a higher burden of emerging haplotypes in HBV-RNA QS compared to HBV-DNA QS (23.24% vs 1.76%, respectively) (**Appendix 3. Table 1**).
- Like in the baseline sample, the P19 (**Figure 36, middle panel**) showed similar distributions of haplotype fractions in both HBV-DNA and HBV-RNA QS with large volumes of master haplotype during follow-up. A very slightly higher frequency of rare haplotypes with very low fitness was observed in both baseline and follow-up HBV-DNA QS (13.81% vs. 14.07%) compared to HBV-RNA QS (9.90% vs. 9.47%) (**Appendix 3. Table 1**).
- The P21 (**Figure 36, right panel**) showed a distinctive evolution between HBV-DNA and HBV-RNA QS. This patient displayed similar QFF profiles as P16 at baseline: HBV-RNA with a smaller master haplotype fraction and higher burden of emerging haplotypes fraction compared to HBV-DNA QS (62.06% vs. 71.35%, respectively for the master haplotype fraction and 25.10% vs. 6.24%, respectively for the

emerging haplotypes fraction). In addition, HBV-DNA QS showed larger fractions of haplotypes with low and very low fitness compared to HBV-RNA QS (22.41% vs. 12.84%, respectively). At follow-up, in HBV-RNA QS the master haplotype fraction increased from baseline and became the dominant in the QS (from 62.06% to 85.98%). On the other hand, HBV-DNA QS diversified to a higher level by increasing the emerging and low fitness haplotype fractions (from 6.24% at baseline to 19.87% at follow-up and from 7.52% at baseline to 16.28% at follow-up, respectively) and by reducing in parallel the master haplotype (from 71.35% at baseline to 52.47% at follow-up) (**Appendix 3. Table 1**).

Therefore, the QFF of HBV-DNA and HBV-RNA QS of P16 and P19 did not seem to undergo important changes during the follow-up period studied. On the other hand, in P21 the HBV-DNA and HBV-RNA QS followed opposite evolutionary trends during the follow-up, while HBV-DNA diversified HBV-RNA became more homogeneous. Notably, P16 and P21 showed similar QFF profiles at baseline, with the master haplotype of the HBV-DNA fraction being more frequent than that of the HBV-RNA. Nevertheless, they showed different evolutionary trends. Regarding the distribution of haplotype abundance based on HNP (**Figure 37**) and in accordance with what was observed by QFF:



**Figure 37. Comparison of Hill numbers profile plots of hepatitis B virus DNA and RNA quasiespecies for the three chronic hepatitis B patients with follow-up samples.** Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; Patients are designated as PXX; B.DNA, baseline DNA quasiespecies; B.cDNA, baseline cDNA quasiespecies; S.DNA, follow-up DNA quasiespecies; S.cDNA, follow-up cDNA quasiespecies.

- The P16 (**Figure 37, left panel**) and the P19 (**Figure 37, middle panel**) showed completely, or almost completely, overlapping baseline and follow-up HBV-DNA HNP curves, and a similar situation was observed for HBV-RNA from both samples, consistent with the absence of important changes in HBV-DNA and HBV-RNA QFFs between both samples:
  - a. In P16, the two HBV-DNA curves showed a steeper slope than those of HBV-RNAs, indicating a lower diversity, probably due to a greater influence of the master haplotype at both time points, as shown by QFF (**Figure 36, left panel**). In contrast to the HBV-DNA QS, the diversity of the HBV-RNA QS increased due to the higher presence of emerging haplotypes fraction.
  - b. P19 showed a very steep HNP curve for all four QS, resulting in a very uneven distribution of relative abundances due to the dominant master haplotype fraction in all the QS. In contrast to P16, the HBV-DNA QS showed a slightly more even distribution, i.e., higher HN curves, than the RNA. Thus, in P19 the HBV-DNA QS was slightly more diverse than the RNA.
- The P21 (**Figure 37, right panel**) showed a diversity profile with different evenness for baseline and follow-up samples, unlike P16 and P19:

- a. At baseline, the HNP curves of both HBV-DNA and HBV-RNA QS were very close between them, and the HBV-RNA curve crossed that of baseline HBV-DNA QS, consistent with its higher diversity (higher emerging haplotypes and lower master haplotype fractions than HBV-DNA).
- b. At follow-up in HBV-DNA QS, the master haplotype volume decreased and emerging haplotypes increased from baseline, resulting in the highest HNP curve of the QS studied in this patient (highly even distribution of haplotypes, thus high QS diversity). The opposite situation was observed for the HBV-RNA QS in the same sample. The increase of master haplotype volume resulted in sharp decrease of the HNP curve, which was in fact the lowest of the four QS of this patient (highly uneven distribution of haplotypes, with the master QS dominating, thus low QS diversity).

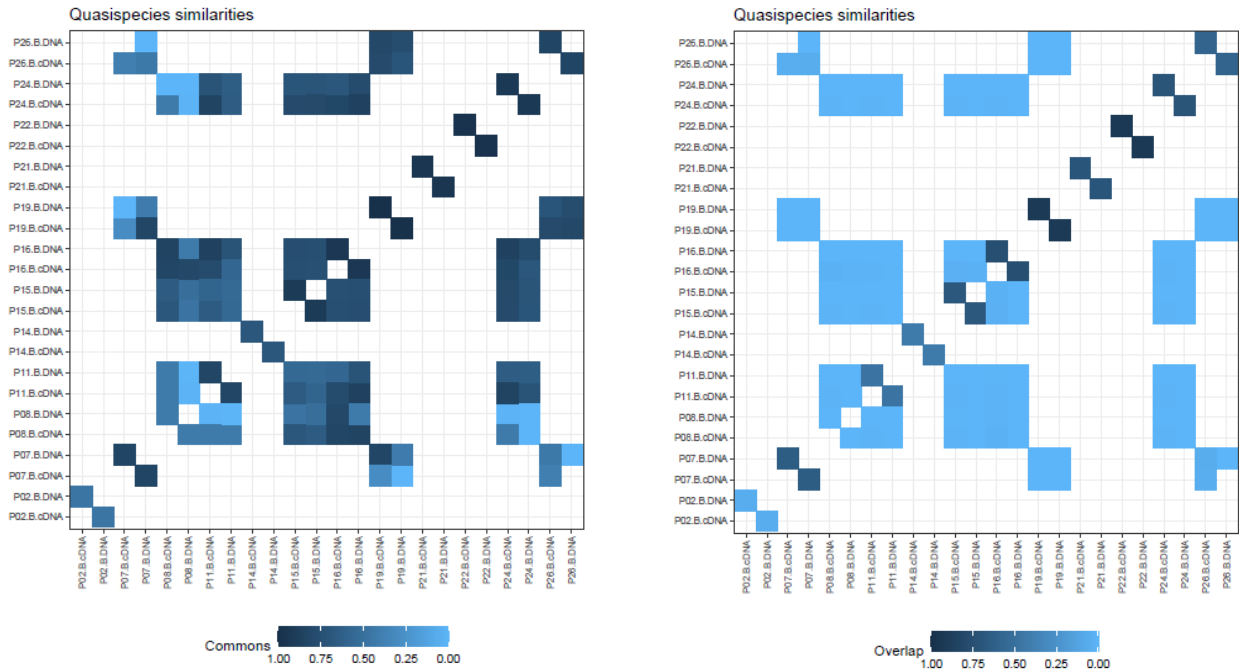
Thus, the HNPs also showed different evolutionary trends of HBV-DNA and HBV-RNA QS within each of the three patients. While HBV-DNA and HBV-RNA QS haplotype abundance distribution did not appear to experience important changes over time in P16 and P19, it did change in P21. Thus, the volume of the fraction of emerging haplotypes at baseline did not seem to be related to the evolution of the QS diversity.

#### *5.2.4.3. Haplotype distribution similarities: Indexes of Commons and Overlap*

Up to this point, the study of master haplotypes and QS complexity by QFF and HNP allowed to assess and compare the distribution of fitness fractions and of relative abundances of haplotypes in HBV-DNA and HBV-RNA QS. In addition, we also assessed the fraction of molecules in haplotypes common to both QS and compared the distribution of their abundance using two similarity indexes: **(a)** the index of Commons ( $C_m$ ), which is a strong indicator of QS relatedness, i.e., if two given QS have all their haplotypes identical, this index will result in a value of 1, even if their proportions are very dissimilar. **(b)** the Overlap index ( $O_v$ ), which is more related to the similarity between the frequencies of common haplotypes in two QS (the distribution of their abundances) and may result in low values even when all haplotypes of both QS are identical. It worth to note that, unlike QFF and HNP, the calculation of  $C_m$  and  $O_v$  was performed without rarefaction, considering the whole set of haplotypes from each QS.

##### *5.2.4.3.1. Baseline samples*

A summary of the values of the similarity indexes obtained by comparing the QS pairs in the baseline samples is given in **Appendix 3. Table 3** and **Appendix 3. Table 4**. These data are presented by means of heatmaps (**Figure 38**).



**Figure 38. Heatmaps showing Commons (left panel) and Overlap (right panel) values for comparisons between hepatitis B virus DNA and RNA quasispecies haplotypes.** Each square shows whether a common haplotype was identified between each pair of the quasispecies on each axis, and the values of  $C_m$  or  $O_v$  (from 0 to 1) are represented with a colour gradient: values closer to zero, i.e., lighter blue colours, mean that there is little similarity between the two quasispecies. Conversely, the closer to 1 (darker blue colours), the more similar the quasispecies. The diagonals show comparison between hepatitis B virus DNA and RNA quasispecies from each sample. *Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; Patients are designated as PXX; B.DNA, baseline DNA quasispecies; B.cDNA, baseline cDNA quasispecies; S.DNA, follow-up DNA quasispecies; S.cDNA, follow-up cDNA quasispecies.*

When comparing all HBV-DNA and HBV-RNA QS from the same sample, which is represented by the diagonal of both  $C_m$  and  $O_v$  heatmaps, it was observed that most of them, 9/12 (75%), showed a high  $C_m$  value (**Figure 38, left panel**), above 0.8. This indicated highly related QS with a high fraction of molecules belonging to common haplotypes present in high frequencies in both HBV-DNA and HBV-RNA QS. The remaining three patients, P02, P08 and P14 showed lower  $C_m$  values, thus their HBV-DNA and HBV-RNA QS shared a smaller fraction of molecules belonging to common haplotypes, due to the fractions of molecules included in their different master haplotypes (see section 5.2.4.1. *Comparison of master haplotypes*). These fractions were large in HBV-RNA QS from P02 (85.92%) and HBV-RNA and HBV-DNA QS from P08 (83.88% and 84.62%, respectively), while they were relatively small in both QS of P14 (HBV-RNA 9.04% and HBV-DNA 17.39%). Therefore, the dissimilarities between the both QS were higher in P02 and P08 ( $C_m = 0.464$  and  $0.422$ , respectively) than in P14 ( $C_m = 0.692$ ).

Regarding the  $O_v$  index, the range of values was more variable. For example, P19 and P22 showed most haplotypes displaying similar frequencies ( $O_v = 0.923$  and  $0.931$ , respectively). In addition, quite similar frequencies were observed in P07, P15, P16, P21, P24 and P26 ( $O_v$  ranging from  $0.592$  to  $0.752$ ) and more disparate frequencies in P11 and P14 ( $O_v = 0.472$  and  $0.421$ , respectively). In P02 and P08, common haplotypes in HBV-RNA and HBV-DNA QS showed very dissimilar frequencies, thus,  $O_v$  index also showed low values ( $O_v = 0.0513$  and  $0.00242$ , respectively).

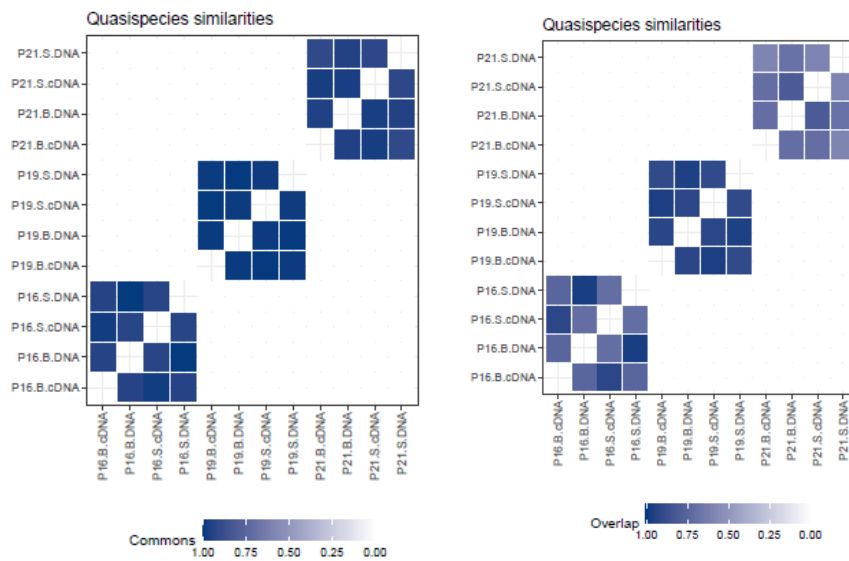
In addition to the HBV-DNA and HBV-RNA QS from the same sample, by comparing samples from different patients, common haplotypes were also observed to varying degrees between patients with the same genotype: P24, P11, P16, P15 and P08, with genotype A, and P19, P07 and P26, with genotype C, shared highly frequent common haplotypes between them, in many cases with  $C_m$  values close to 1. However, in all those samples sharing a high fraction of molecules belonging to common haplotypes,  $O_v$  values were near to 0 (**Figure 38, right panel**), meaning that master and other mid frequency haplotypes in both QS were different. It should be considered that QS from different patients sharing the same viral genotype may have some common haplotypes at high frequency in one of the QS, which may be present at low or very low frequency in the other. Considering all the patients who did not share genotype with the others (genotypes B, D, E and H in P14, P21, P22 and P02, respectively), they had no haplotype in common with any other patient.

#### 5.2.4.3.2. Changes in hepatitis B virus DNA and RNA quasispecies during longitudinal follow-up

Regarding the  $C_m$  index, the three patients with follow-up samples (during a period of 1.1-1.9 years) showed high  $C_m$  values (close to the maximum value of 1) when comparing HBV-DNA and HBV-RNA QS at baseline, at follow-up and when comparing both baseline and follow-up HBV-DNA or HBV-RNA QS longitudinally. None of the three patients showed evidence of genotype change during follow-up. Thus, when comparing the four QS of a patient, a great similarity among them was shown (**Figure 39, left panel**).

For the  $O_v$  index, they were close to the maximum value of 1 between all HBV-DNA and HBV-RNA QS analysed from P19 (**Figure 39, right panel**), indicating very similar QS structure. The lighter blue colours observed in **Figure 39 right panel** for comparisons of the different QS from P16 and from P21 indicated more dissimilar QS structure than that observed in P19. For example, in P16 the frequencies of haplotypes of HBV-RNA QS at baseline seemed to be more similar to the HBV-RNA QS at follow-up sample than to those of HBV-DNA QS in the same sample. The same situation was observed with HBV-DNA QS.

In P21, the  $O_v$  index was generally the lowest for all comparisons between baseline and follow-up HBV-DNA and HBV-RNA QS, evidencing a dissimilar structure between the four QS analysed. Of note, QFF and HNP assessment indicated a significant change in the distribution of haplotype frequencies in both HBV-RNA and HBV-DNA QS between baseline and follow-up samples (**Figures 36 and 37, right panel**), however the fraction of molecules in common haplotypes between those QS remained high (**Figure 39, left panel**). Thus, in this case, the  $O_v$  index reflected the changes in the distribution of the relative frequencies of the haplotypes shown by QFF and HNP.



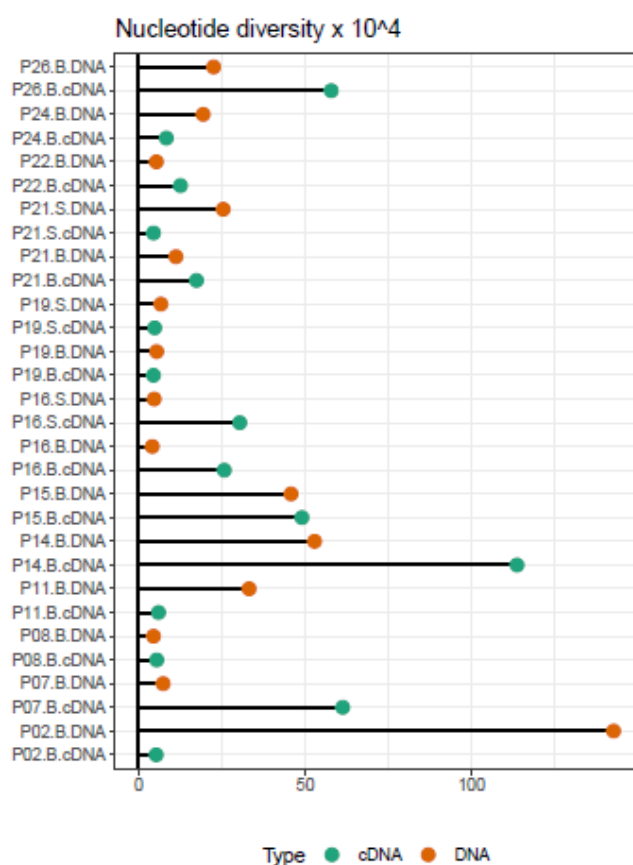
**Figure 39. Heatmaps showing Commons ( $C_m$ , left panel) and Overlap ( $O_v$ , right panel) values for comparisons between hepatitis B virus DNA and RNA quasispecies haplotypes for the three chronic hepatitis B patients with follow-up samples.** Each square shows whether a common haplotype was identified between each pair of the quasispecies on each axis, and the values of  $C_m$  or  $O_v$  (from 0 to 1) are represented with a colour gradient: values closer to zero, i.e., lighter blue colours, mean that there is little similarity between two quasispecies. Conversely, the closer to 1 (darker blue colours), the more similar the quasispecies. The diagonals show comparison between hepatitis B virus DNA and RNA quasispecies from each sample. *Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; Patients are designated as PXX; B.DNA, baseline DNA quasispecies; B.cDNA, baseline cDNA quasispecies; S.DNA, follow-up DNA quasispecies; S.cDNA, follow-up cDNA quasispecies.*

#### 5.2.4.4. Genetic distance between quasispecies

In addition to the similarities between the haplotype distributions of HBV-DNA and HBV-RNA QS in the same sample or at two separate time points represented by  $C_m$  and  $O_v$ , the extent of changes between these QS can also be quantified by their net genetic distance ( $D_A$ ).



To calculate the net genetic distances between circulating HBV-DNA and HBV-RNA QS, the sequence heterogeneity within each QS was first calculated to subtract their mean from the genetic distances between the two QS. Thus, the nucleotide diversity ( $\pi$ ), i.e., the mean number of mutations between each haplotype pair, was calculated for each QS. These calculations require the multiple alignment of all haplotypes in all samples, to make it feasible only the 100 most frequent haplotypes in each sample were considered (**Figure 40**).

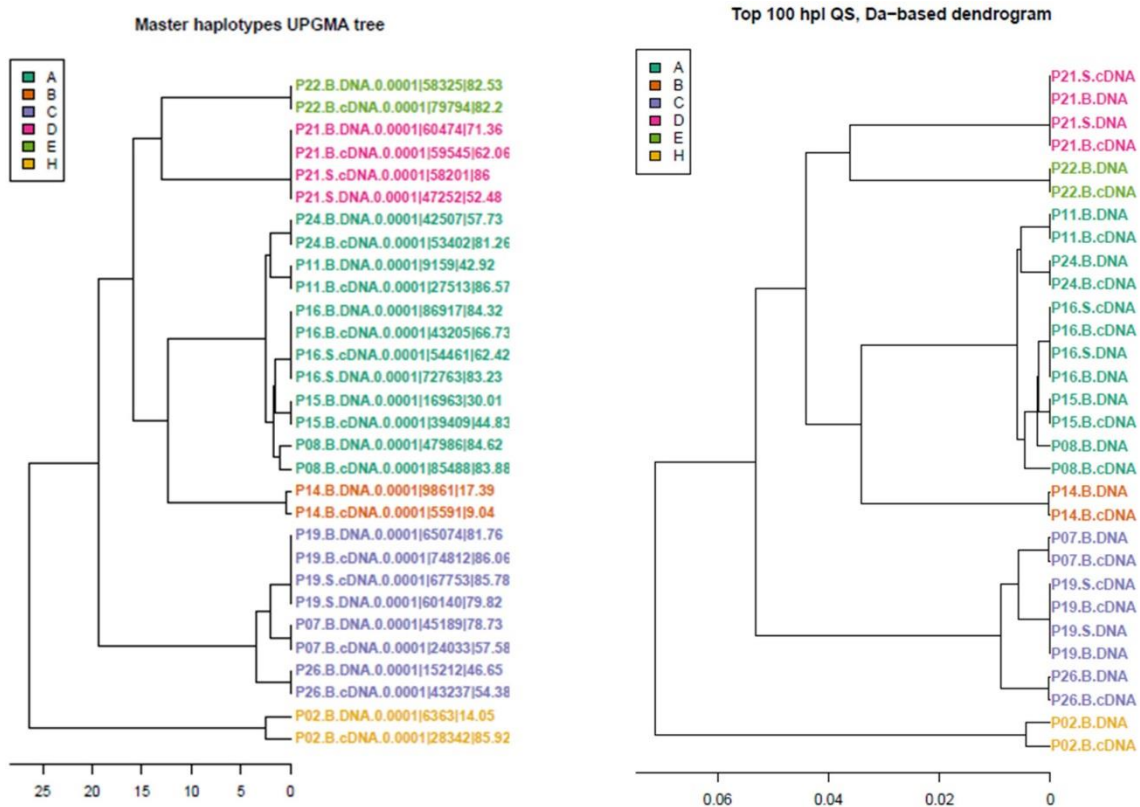


**Figure 40. Nucleotide diversity ( $\pi$ ) of each hepatitis B virus DNA and RNA quasispecies.** To facilitate visualisation and comparison of results, the nucleotide diversity values of each quasispecies (x-axis) have been multiplied by  $10^4$ . Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; Patients are designated as PXX; B.DNA, baseline DNA quasispecies; B.cDNA, baseline cDNA quasispecies; S.DNA, follow-up DNA quasispecies; S.cDNA, follow-up cDNA quasispecies.

Although  $\pi$  tended to be higher in HBV-RNA than in HBV-DNA QS (eight patients with  $\pi$  HBV-RNA >  $\pi$  HBV-DNA QS vs four patients with  $\pi$  HBV-RNA <  $\pi$  HBV-DNA QS at baseline), there were no clear differences between the genetic heterogeneities of the two QS. In the longitudinal study,  $\pi$  showed different trends in each of the three patients. P16 showed a  $\pi$  HBV-RNA >  $\pi$  HBV-DNA QS and P19 a  $\pi$  HBV-RNA <  $\pi$  HBV-DNA QS, at

both baseline and follow-up samples. In both patients,  $\pi$  HBV-RNA and  $\pi$  HBV-DNA QS slightly increased at follow-up. Regarding P21, a  $\pi$  HBV-RNA >  $\pi$  HBV-DNA QS was observed at baseline whereas this trend was reversed at follow-up, with a  $\pi$  HBV-RNA <  $\pi$  HBV-DNA QS. In keeping with this,  $\pi$  HBV-RNA QS decreased at follow up while  $\pi$  HBV-DNA QS increased.

The nucleotide distance (defined as the average number of nucleotide substitutions between the reads from both QS) between each pair of QS was then calculated. After subtracting the average  $\pi$  between the two QS, the  $D_A$  was calculated and represented as a UPGMA tree. Of note, the  $D_A$  UPGMA tree showed a similar structure to the UPGMA tree with the master haplotype representing the QS (**Figure 41**).



**Figure 41. Master haplotypes (left dendrogram) vs. net genetic distance ( $D_A$ ) based on the average genetic distances of the 100 most frequent haplotypes from each hepatitis B virus DNA and RNA quasispecies (right dendrogram).** Master haplotypes and labels with the name of each quasispecies are coloured according to the hepatitis B virus genotype assigned by the distance-based discriminant rule (DB-rule). Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; Patients are designated as PXX; B.DNA, baseline DNA quasispecies; B.cDNA, baseline cDNA quasispecies; S.DNA, follow-up DNA quasispecies; S.cDNA, follow-up cDNA quasispecies.

Regarding patients with different master sequences in HBV-DNA and HBV-RNA QS (P02, P08 and P14) they showed different trends:

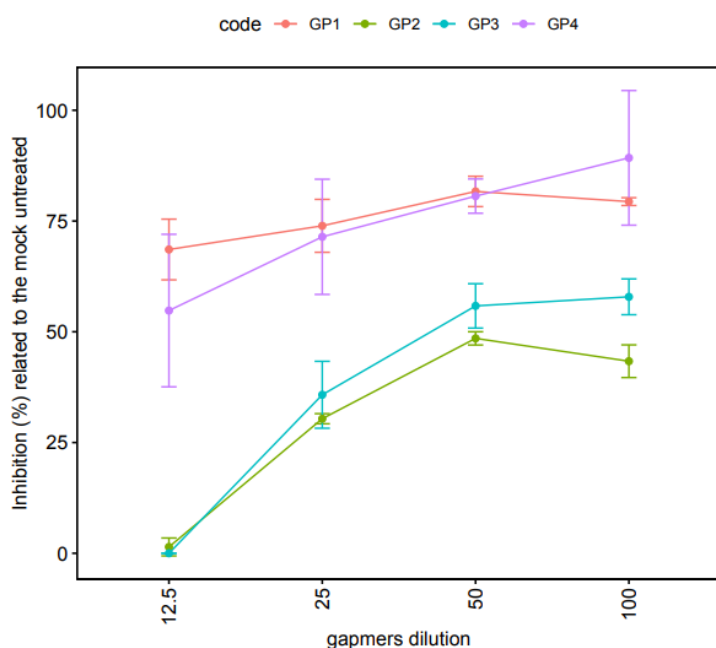
- HBV-DNA QS in P02 was very genetically heterogeneous while the contrary was observed for HBV-RNA ( $\pi$  HBV-RNA <  $\pi$  HBV-DNA). In addition, comparison between both HBV-DNA and HBV-RNA QS rendered small fractions of molecules in haplotypes common between both QS ( $C_m = 0.464$ ). In addition, frequencies of those common haplotypes showed different distributions in both QS ( $O_v = 0.0513$ ). This may be in keeping with the high volumes of emerging and low fitness QS fractions of HBV-DNA, in contrast to the high volume of master haplotype fraction in HBV-RNA. Probably due to the dominance of the master haplotype in HBV-RNA QS, the net genetic distance between HBV-DNA and HBV-RNA QS was similar than the distance between master haplotypes from both QS.
- On the contrary, P14 showed a much higher genetic heterogeneity in HBV-RNA QS than in HBV-DNA ( $\pi$  HBV-RNA >  $\pi$  HBV-DNA). In addition, the fraction of master haplotypes was very low in both QS (9.04% HBV-RNA and 17.39% HBV-DNA QS, **Table 16**), thus despite the reads included in these fractions in both QS showed differences, the relatively high  $C_m$  value between them (0.692), indicated that a relatively large fraction of QS molecules was included in haplotypes common between both HBV-DNA and RNA QS. Thus, the different master haplotypes had a limited weight in the HBV QS net genetic distances.
- P08 showed very similar and low  $\pi$  for HBV-RNA and HBV-DNA QS. Similar to P02, P08 showed few molecules in haplotypes common between both QS which showed different distributions in both QS ( $C_m = 0.422$  and  $O_v = 0.00242$ ). The differences between these masters probably had an important weight in HBV QS genetic distances between those QS with low genetic heterogeneity. Of note, as shown in **Figure 41**, P08 was the patient showing the most important differences between the HBV-DNA and HBV-RNA QS in the same sample, even being HBV-DNA QS closer to HBV-DNA and HBV-RNA QS of P15 and P16 than to HBV-RNA QS of the same sample. In addition, HBV-RNA QS of P08 was also closer to both QS of P15 and P16.

### 5.3. Third study: The combination of antisense locked nucleic acid Gapmers and/or siRNA as a potential hepatitis B virus gene therapy: preliminary in vitro results

#### 5.3.1. Antisense locked nucleic acid Gapmers entry into HepG2-hNTCP 2.5% DMSO-treated cells and dose-response curve

To ensure that GPs could efficiently entry by transfection, the cells treated with 5'-FAM scrambled GP, which generates a green colour when excited at 490nm, were observed by cytometry. The scrambled GP efficiently entered the DMSO-treated cells. However, the transfection efficiency observed was quite variable depending on cell viability and infection status.

In order to determine GPs working concentration, all the four oligonucleotides were serially diluted from 100 nM to 12.5 nM (**Figure 42**).

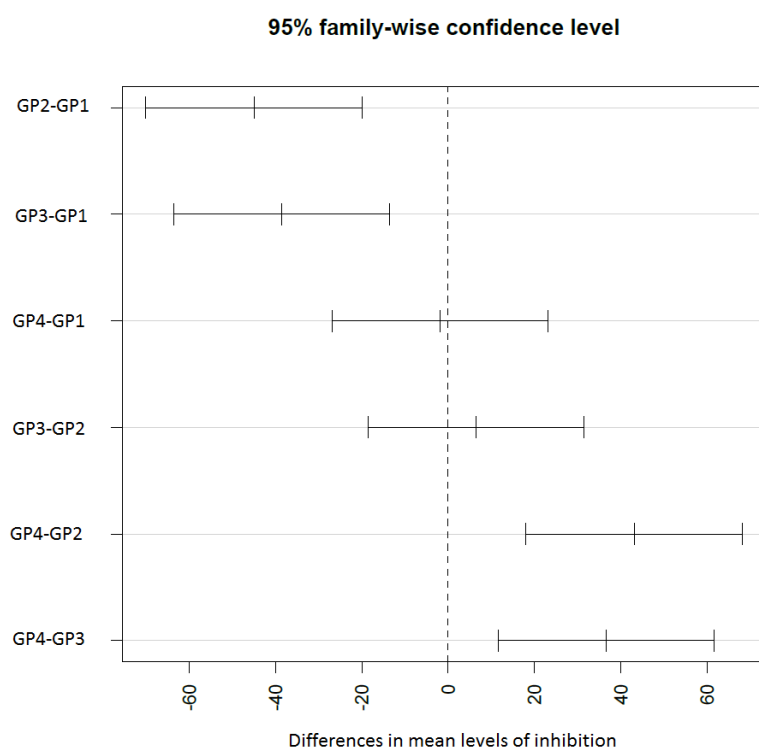


**Figure 42.** Dose-response data are graphed with the dose on the x-axis and the measured effect (response) on the y-axis. Abbreviations: GP, Gapmer.

All the GPs could induce an inhibition of the expression of HBeAg. GP1 and GP4 were more efficient in inhibiting the viral expression, with a percentage of inhibition of more than 50% at the lowest concentration ( $68.6 \pm 6.9$  and  $54.8 \pm 17.2$  for GP1 and GP4 at 12.5 nM, respectively). At the highest concentration, the reduction of the HBeAg in the supernatant was  $79.4 \pm 0.09$  and  $89.3 \pm 15.2$  for GP1 and GP4, respectively. The percentage of inhibition

was  $81.7 \pm 3.4$  and  $73.9 \pm 6$  for GP1 at 50 nM and 25 nM, respectively and  $80.6 \pm 3.9$  and  $71.4 \pm 13$  for GP4 at 50 nM and 25 nM, respectively. Differently, GP2 and GP3 induced a considerable inhibition at only higher concentration ( $43.4 \pm 3.7$ ,  $48.5 \pm 1.5$ ,  $30.4 \pm 1.1$  and  $1.2 \pm 2$  for GP2 at 100 nM, 50 nM, 25 nM and 12.5 nM, respectively and  $57.9 \pm 4$ ,  $55.8 \pm 5$ ,  $35.8 \pm 7.6$  and 0 for GP3 at 100 nM, 50 nM, 25 nM and 12.5 nM, respectively). Notably, for all the silencers, the silencing ability seemed to follow a sigmoid, and starting from 50nM a limited increased of the curve slope were observed, with a difference related to the 100nM of 81.7% vs 79.4%, 48.5% vs 43.4, 55.8% vs 57.9% and 80.6% vs 89.3% for GP1, GP2, GP3 and GP4 at 100 nM vs 50 nM, respectively. Particularly, cell viability was very similar when comparing Gapmers treatment at these both concentrations (100% vs 100%, 100% vs 80%, 85.7% vs 75% and 75% vs 88.89% for GP1, GP2, GP3, and GP4 at 100 nM vs 50 nM, respectively) and to the mock-untreated cells.

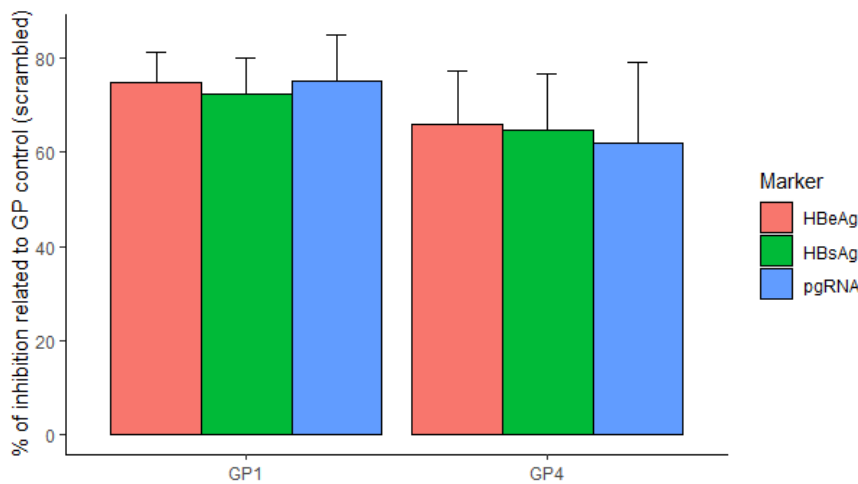
Moreover, confidence intervals and differences in terms of inhibition among the four GPs were also determined (**Figure 43**). The figure clearly shows that GP1 and GP4 had the best means levels of inhibition when compared to the other two GPs. For this reason, GP2 and GP3 were discarded from further studies.



**Figure 43. Confidence intervals and differences in terms of inhibition among the four Gapmers.** Abbreviations: GP, Gapmer.

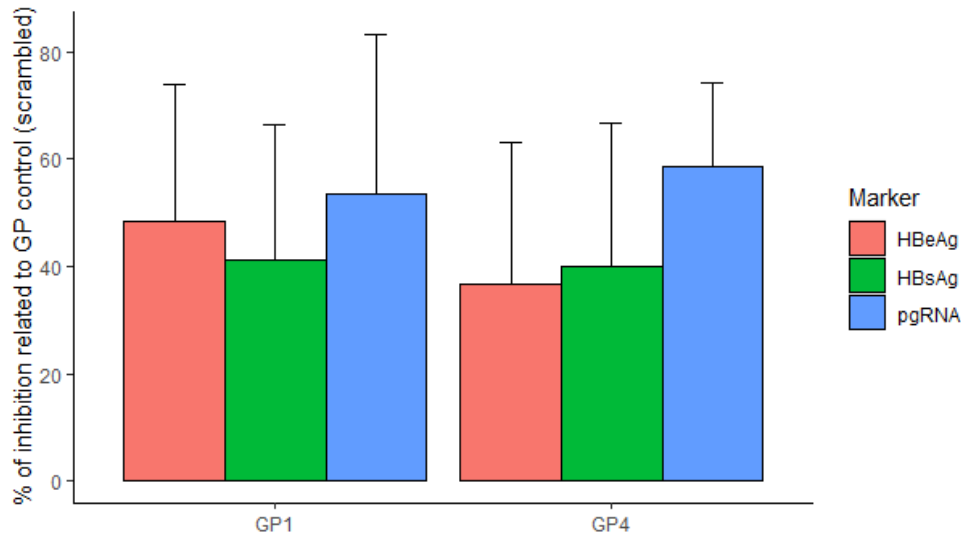
### 5.3.2. Antisense locked nucleic acid Gapmers inhibition of hepatitis B virus expression at early and late treatment

Early after infection, i.e., at 48 h pi, GP1 allowed an inhibition of more than 70% for both pgRNA, HBeAg and HBsAg (75.5±9.9%, 74.7±6.4%, and 72.4±7.7%, respectively) (**Figure 44**). Differently, GP4 showed an inhibition between 7.4 and 13 percentage points lower than GP1 (62±17%, 66.1±11.2% and 64.7±12% for pgRNA, HBeAg and HBsAg, respectively).



**Figure 44. Inhibition percentage of Gapmer 1 and 4 related to scrambled Gapmer at early time point, i.e., at 48 h post infection.** Abbreviations: GP, Gapmer, HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; pgRNA, pregenomic RNA.

When GPs were introduced at late time point, i.e., at 5 d pi, the inhibition of pgRNA decreased for both GP1 and GP4 (53.5±29% and 58.5±17%, respectively) (**Figure 45**). The efficiency was even lower if considering viral proteins expression (HBeAg: 48.5±25% and 36.7±26% for GP1 and GP4, respectively; HBsAg: 41±25% and 40±26% for GP1 and GP4, respectively).

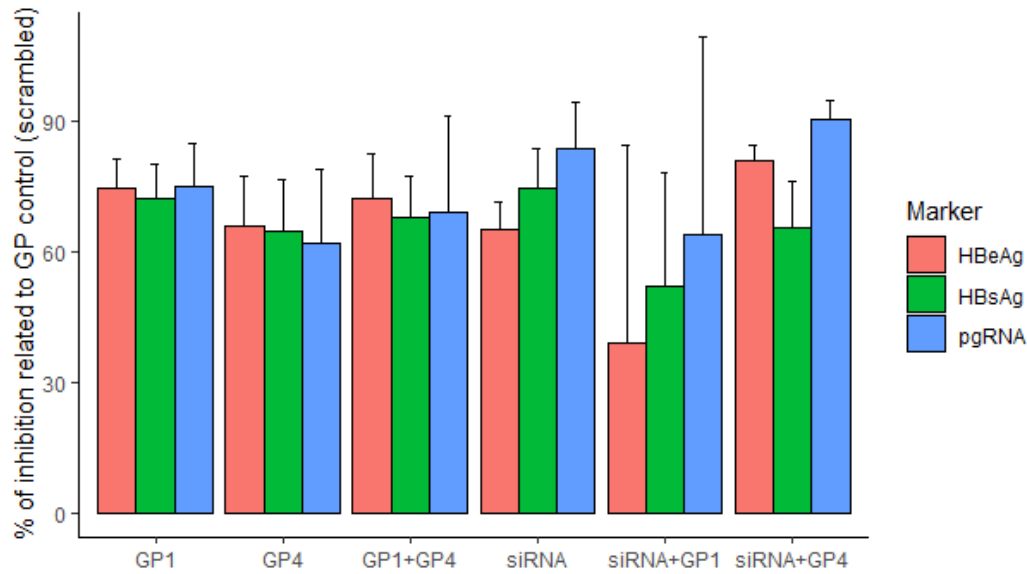


**Figure 45. Inhibition percentage of Gapmer 1 and 4 related to scrambled Gapmer at late time point, i.e., at 5 d post infection. Abbreviations: GP, gapmer, HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; pgRNA, pregenomic RNA.**

### 5.3.3. Increasing silencing efficiency by using an antisense locked nucleic acid Gapmers and siRNA combination strategy

Considering the results of GP1 and GP4, a combination strategy of both GPs together or in combination with a siRNA was tested.

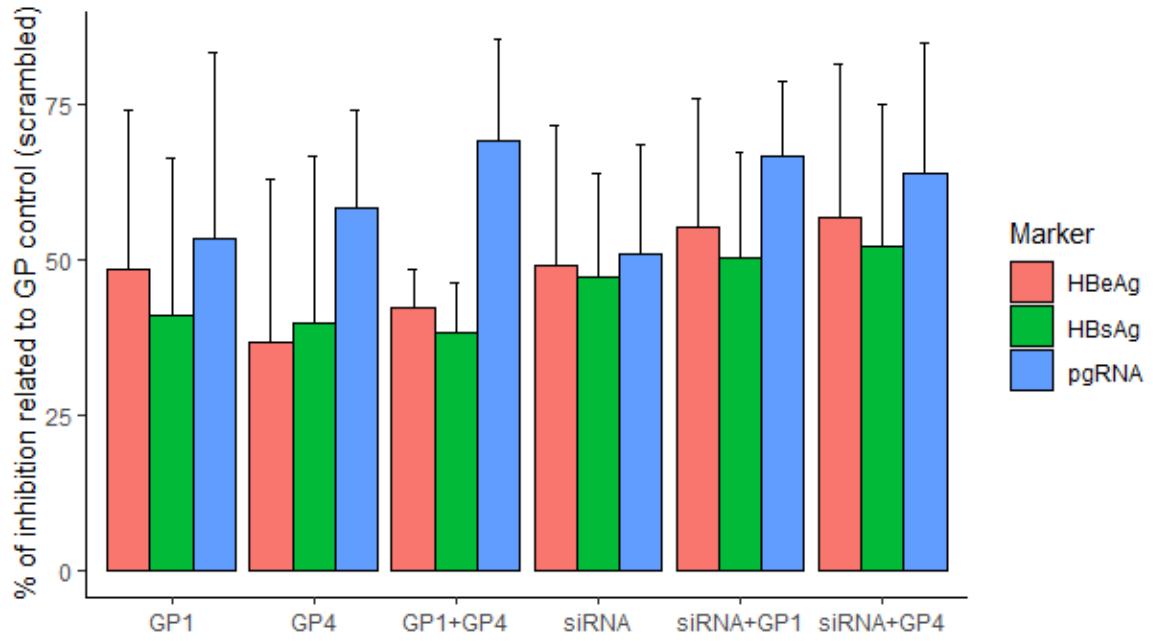
Early after infection, the GP1+GP4 combination showed no increased inhibition for any of the three viral markers compared to each GP individually (69±22%, 72.2±10%, 68.1±9% for pgRNA, HBeAg and HBsAg, respectively). The siRNA reached an inhibition percentage for pgRNA and HBsAg greater (83.8±10.73, 65.15±6.54 and 74.81±8.85 for pgRNA, HBeAg and HBsAg, respectively) than that obtained by each GP alone or in combination. Surprisingly, inhibition of the three viral markers decreased markedly for siRNA+GP1 combination (64.1±45%, 39.3±45% and 52.3±26% for pgRNA, HBeAg and HBsAg, respectively). Regarding siRNA+GP4 combination, inhibition was only increased for pgRNA and HBeAg markers (90.6±4% and 81±3%, respectively). HBsAg was shown to be not inhibited compared to siRNA individually (65.4±10%) (**Figure 46**).



**Figure 46. Inhibition percentage of combination of both Gappers or siRNA plus Gapper at early time point i.e., at 48 h post infection.** Abbreviations: *siRNA*, small interfering RNA; *GP*, gapper, *HBeAg*, hepatitis B e antigen; *HBsAg*, hepatitis B surface antigen; *pgRNA*, pregenomic RNA.

At late time point, the siRNA showed to increase the inhibition for both viral proteins ( $50.9 \pm 17.6$ ,  $49 \pm 22.6$ ,  $47.3 \pm 16.67$  for pgRNA, HBeAg and HBsAg, respectively) in comparison with GPs alone or in combination. Of note, the combination of both GPs or siRNA+GP at late time point, increased the inhibition efficiency of pgRNA ( $69.1 \pm 16.5\%$ ,  $66.7 \pm 12.2\%$  and  $63.8 \pm 21\%$  for GP1+GP4; siRNA+GP1; siRNA+GP4, respectively), just partially improving viral proteins inhibition (HBeAg:  $42.5 \pm 6.1\%$ ,  $55.4 \pm 20\%$  and  $57 \pm 24\%$ , for GP1+GP4, siRNA+GP1 and siRNA+GP4, respectively; HBsAg:  $38.2 \pm 8.2\%$ ,  $50.3 \pm 17\%$  and  $52.2 \pm 23$  for GP1+GP4, siRNA+GP1 and siRNA+GP4, respectively) (**Figure 47**).





**Figure 47. Inhibition percentage of combination of both Gappers or siRNA plus Gapper at late time point, i.e., at 5 d post infection. Abbreviations: siRNA, small interfering RNA; GP, gapper, HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; pgRNA, pregenomic RNA.**

## **DISCUSSION**

---

## 6. DISCUSSION

The World Health Organization (WHO) estimated that, in 2019, 296 million people worldwide were living with chronic hepatitis B virus (HBV) infection, with 1.5 million new infections each year despite the availability of a safe and effective vaccine [1]. It is estimated that around 15-40% of chronic hepatitis B (CHB) patients will develop liver cirrhosis, liver failure or hepatocellular carcinoma (HCC), with mortality rate of around 15-25% [223]. To date, there are two types of approved treatments for CHB infection. First, there is an immunomodulatory therapy based on pegIFN $\alpha$ , which is underused due to its severe side effects [224]. Second, the most widely used treatment is oral therapy with nucleos(t)ide analogues (NAs) that inhibit the action of viral polymerase in the reverse transcription of pregenomic RNA (pgRNA) to rcDNA. They are >99% effective in significantly reducing HBV-DNA levels to undetectable, but do not affect the expression of cccDNA or integrated DNA [127]. Although highly effective antiviral treatments are available, they do not cure the infection [128] or guarantee that the liver disease will not progress [129][130]. In fact, the cumulative risk of developing HCC in HBV infection (even in the absence of detectable viral load) at 5 and 10 years has been estimated at 1.63-5.9% respectively, higher in cirrhotic, male patients and those over 50 years of age [131]. Despite their high efficacy in reducing HBV replication and infection of new hepatocytes, NAs do not interfere with the expression of covalently closed DNA (cccDNA) or viral DNA integrated into the genome of the cells, which persists in the nucleus of the hepatocyte throughout its life. All this highlights the need to explore new intervention strategies for this infection, which is considered a global public health problem.

Given the important and wide role of HBx protein, a multifunctional transactivator protein encoded by the hepatitis B X gene (*HBX*), 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 [225]. Moreover, *HBX* is located near the 3' end of all the HBV mRNAs, which means that targeting its sequence may interfere with the synthesis of all the viral proteins [2]. This concept is in line with other strategies aiming for functional cure, which consist in a sustained loss of hepatitis B surface antigen (HBsAg), with or without seroconversion, and sustained suppression of viral replication leading to undetectable HBV-DNA in serum. Even when this kind of cure is achieved, the cccDNA still persists in infected hepatocytes, but a complete HBV cure, leading to both cccDNA and integrated HBV DNA elimination, is currently considered unattainable [226]. Moreover, HBV

infection rely on the proper pgRNA reverse transcription by the viral polymerase reverse transcriptase (RT). Theoretically, it would be expected to find highly conserved regions along the overlapping RT and surface (S) (S-RT) region, which includes part of the functional domains of the viral polymerase. Of note, some viral particles containing non-retrotranscribed pgRNA [98] and other HBV RNAs such as spliced pgRNA variants and *HBX* transcripts [59] have been described, although at a low levels compared to those containing HBV-DNA. In order to prevent not only viral replication but also viral protein translation and HBV-RNA accumulation, gene silencing could be a valuable strategy and *HBX* and S-RT regions optimal targets. Identifying highly conserved regions in these two genes may suggest valuable targets for pan-genotypic approaches. Several siRNAs are currently being tested against *HBX* and S ORFs. A study carried out in chimpanzees reported that administration of ARB-1467 (a combination of 3 interfering RNAs targeting both X and S ORFs [227]) triggered a 90% reduction in HBsAg levels and a 50% reduction in cccDNA within 28 days of treatment [228].

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 NAs cessation [225]. In addition, the slower decline of serum HBV RNA levels may also make it possible to study circulating viral quasispecies (QS), even when serum HBV-DNA levels are too low to do so. To date, few studies have analysed circulating HBV-RNA QS [119,120] and therefore there is little data on its similarities and differences to that of HBV-DNA QS. HBV is the most variable among DNA viruses because its life cycle includes the activity of error-prone RT, which catalyses the reverse transcription of encapsidated pgRNA into rcDNA, giving rise to circulating HBV-DNA QS with a very high virion production per day [55]. Since HBV-RNA QS has not been subjected to reverse transcription process, it is reasonable to assume that there may be significant differences between HBV-DNA and HBV-RNA QS.

With all this in mind, we analysed a region of the 5' end of *HBX* where we had previously identified hyper-conserved regions in the circulating HBV-DNA [2,3], which we intended to confirm in circulating HBV-RNA. In fact, serum HBV-RNA was found to be genetically homogenous with intrahepatic HBV-RNA [119] 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 HBV-RNA QS, when looking for targets for targeted gene therapy, verifying their conservation not only at serum HBV-DNA level but

also at HBV-RNA level would ensure their effect on their main target. By this reason, the first study of this doctoral thesis, which has been already published, aimed to compare both serum HBV-RNA and HBV-DNA QS in *HBX* in a group of 13 well-characterised untreated patients attending the outpatient clinic of Vall d'Hebron University Hospital (Barcelona, Spain). In this group of patients, we confirmed that most of the hyper-conserved regions identified by our group in previous studies at DNA level [2,3] coincided with those identified in serum HBV-RNA, and may thus also be conserved at intrahepatic HBV-RNA level. Moreover, the broad range of HBV genotypes and degrees of severity of liver disease from all patients included in that study enabled to get an overview of complexity and conservation of HBV-DNA and HBV-RNA QS.

In the second study, we aimed to compare similarities and differences between HBV-RNA and HBV-DNA QS in the S-RT region, using a set of genetic diversity, similarity indexes and genetic distances to provide detailed information. In this study we included a total of 12 well-characterized untreated patients attending the outpatient clinics of Vall d'Hebron University Hospital and Hospital Clinic de Barcelona (Barcelona, Spain). In addition, a follow-up sample from three out of 12 patients was also included. The possibility of analysing the relationship between HBV-DNA and HBV-RNA QS in a given time point but also changes emerging over time between both QS allowed us to get an idea of the independent evolutive trends of each viral QS, which may be different according to the patient. The results obtained demonstrate that the methodology established in both studies, based on next-generation sequencing (NGS) with the MiSeq platform, allowed detailed analysis of serum HBV-RNA QS, being a useful tool to perform these studies.

In the third study of this thesis, *in vitro* functional studies were performed in a cellular infection model to test the potential usefulness of gene therapy targeting the hyper-conserved sequences reported in the first study and even confirmed at HBV-RNA level. Taking all this into account, we tested several candidate oligonucleotides for gene therapy based on LNA Gapmer and siRNA.

## 6.1. Discussion first study: Cross-sectional evaluation of circulating hepatitis B virus RNA and DNA: Different quasispecies?

The aims of this study were to establish a reliable methodology for thorough analysis of serum HBV-RNA QS without interference from HBV-DNA QS, and to compare genetic

variability/conservation and complexity of circulating HBV-RNA QS with that of circulating HBV-DNA QS in a region including the 5' end of the *HBX*, which encodes the negative regulatory domain of the HBx protein.

Previous studies have found RNA levels between 0.8 and 2.8 logs lower than DNA levels [115,118,229]. However, even lower levels of circulating HBV-RNA have been observed in a recent study by our group [126], in which mean differences between circulating HBV-DNA (measured in IU/mL and converted to copies/mL using the conversion factor of 5.82 [230]) and HBV-RNA of  $3.6 \log_{10}$ copies/mL were observed in the absence of antiviral treatment, using an automated assay in 232 plasma samples from 50 CHB patients. Although we measured HBV-DNA and HBV-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. The low levels of HBV-RNA relative to HBV-DNA have made it difficult to recruit patients for whom HBV-RNA QS can be analysed by NGS. To solve this problem, we selected patients with high HBV-DNA viral loads (HBV-DNA levels  $> 5 \log_{10}$ IU/mL) in order to find patients with HBV-RNA levels  $> 4 \log_{10}$ copies/ mL, enough to amplify the *HBX* 5' end region. Even if HBV-DNA is present at low 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. The presence of DNA is an inherent problem during RNA extraction due, in part, to the similar physical and chemical properties of DNA and RNA [231]. To ensure that results obtained from HBV-RNA QS amplification procedure were free from HBV-DNA contamination, RNA isolates were pre-treated with DNase I, and verified the HBV-DNA elimination from these isolates by means of a qPCR assay. Thus, this verification step allowed us to confirm that no residual HBV-DNA was present in the HBV-RNA isolates.

#### Complexity comparison of HBV-DNA and HBV-RNA quasispecies using next-generation sequencing

Since errors during pgRNA reverse transcription to rcDNA are considered an important source of genetic variability for the HBV genome, we considered the rare haplotype load (RHL) to be a diversity index especially suitable for comparing complexities between HBV RNA and DNA QS. This is because this diversity index refers to the fraction of the QS with the lowest fitness, indicative of the intensity with which replication errors accumulate, in this case due to the effect of viral polymerase activity on HBV-RNA QS. High viremia and long infection times would tend to result in high RHL values. High values may also be given to mutagen treatments [212]. In our study, we considered as “rare” all QS minority viral populations below a threshold of frequency set to 1%, in keeping with a previous study [212]. To compute RHL, the set of haplotypes common to forward and reverse strands were

taken. Indeed, PCR or sequencing errors could be a source of artefactual “rare” haplotypes. However, it was assumed that technical noise such as pipetting, mistakes made by the polymerase, and instrument-derived variations would affect equally all samples in the experiment and that differences between HBV-DNA and RNA QS would be mainly due to reverse transcription or other biological sources of variability. HBV-DNA QS showed a higher mean RHL than HBV-RNA QS, suggesting an enrichment of low frequency (below 1%) HBV genomes in HBV-DNA QS, probably due to the error-prone reverse-transcription origin of HBV-DNA. However, these differences were not statistically significant.

Similarly, no statistically significant differences were observed when comparing both HBV-DNA and HBV-RNA QS between patients with significant fibrosis or cirrhosis and with non-significant fibrosis, or each HBV-DNA and HBV-RNA QS within themselves separating patients depending on their degree of fibrosis. The RHL shows high correlation with Shannon entropy, an abundance diversity index [212], which was used by Yu et al. [120] to compare HBV-DNA and HBV-RNA QS in sequential serum samples in a group of HBeAg positive patients receiving entecavir or pegIFN $\alpha$ . Similarly to our results, no significant difference was found between HBV-RNA and HBV-DNA QS complexity at baseline QS. 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 [120] (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 QS toward higher QS complexity, as shown by mean RHL values.

#### Conservation comparison of HBV-DNA and HBV-RNA quasispecies using next generation sequencing

Both HBV-DNA and HBV-RNA sequence conservation was studied to search for hyper-conserved nt regions regardless of clinical stage or viral genotype. Sequence conservation was calculated using the information content (IC) of each position in a multiple alignment of all the different haplotypes. Different from RHL, for IC calculation were only taken haplotypes obtained in frequencies > 0.25% in each sample to avoid underestimating sequence conservation due to technical noise (PCR or sequencing artefacts). Comparison of both HBV-DNA and HBV-RNA QS conservation yielded similar results, although HBV-DNA was slightly more conserved than HBV-RNA despite the effect of reverse-transcription over HBV-DNA QS. Therefore, although both HBV-DNA and HBV-RNA QS showed quite similar complexity (assessed by the RHL index that considers the haplotypes fraction < 1%

within a QS) and sequence conservation (assessed by the IC) at 5' *HBX*, HBV-DNA showed a trend towards higher complexity and conservation. The significance of this trend needs to be explored in further studies in larger and more homogeneous groups of patients. The higher complexity and sequence conservation showed by the HBV-DNA compared with HBV-RNA QS may be related with the indexes chosen for their respective study. In this study the QS complexity was studied with RHL at 1%, which refers to the fraction of molecules (i.e., reads) within the QS representing low- to very low-fitness haplotypes. This allowed us to reveal possible mutagenic effects due to the reverse transcription from HBV-RNA to HBV-DNA. On the other hand, conservation calculation took into account a different set of haplotypes than RHL, i.e., the haplotypes with high fitness and discard those with the lowest fitness (below 0.25% within the QS).

The slightly higher sequence variability showed in HBV-RNA QS may suggest that only a fraction of packaged pgRNA is reverse-transcribed and released. Thus, HBV-RNA would show a closer picture of the whole set of intracellular HBV transcripts produced, in keeping with previous report [119]. In addition it might exist different mechanisms, to be considered in future lines of research, responsible for generating RNA sequence diversity such as transcription-associated mutation mechanisms. With this regard, there are identified HBV pgRNA-interacting proteins directly related to the splicing machinery [232]. The variability due to such mechanisms would affect the fraction of circulating HBV-RNAs that cannot be subjected to reverse transcription, such as spliced pgRNA variants and *HBX* transcripts [59].

In addition, in this study, we identified three hyper-conserved regions which coincided in both HBV-RNA and HBV-DNA QS. The first hyper-conserved region identified was between nt 1258-1286, while the second hyper-conserved region consisted of two nt fragments (1545-1573 and 1575-1604) spanning a region between nt 1545-1604. Interestingly, these three conserved sequence fragments almost overlapped with hyper-conserved sequence fragments described in a previous study carried out by our group at circulating HBV-DNA QS level (1255-1286, 1545-1573 and 1575-1603)[2]. Likewise, another study by our group [3] also performed in HBV-DNA QS, reported some of these hyper-conserved regions (1258-1286 and 1575-1605). In addition, these two previous studies also reported as hyper-conserved the sequence stretch between positions 1519-1543, which we found among the 5% most conserved windows at HBV-DNA QS level but not at HBV-RNA QS level. Conversely, we identified the sequence stretch 1559-1587 as hyper-conserved only in HBV-RNA QS. This fragment was not identified in our previous studies, although it is included in the region between nt 1519 and 1603 defined by three conserved nt fragments (1519-1543,



1545-1573, and 1575-1603) described in one of them [2]. Thus, despite the high degree of similarity between sequence conservation in HBV-DNA and HBV-RNA QS, 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 HBV-DNA and HBV-RNA QS, similar to the most conserved stretches identified, which confirmed the similarities between both QS. The high degree of sequence conservation observed in the 5' HBX end may be indicative of the importance of the amino acid sequence encoded in the HBx protein function, positioning it as a good target for therapeutic strategies. It will be of particular interest to confirm whether these results are replicated in other regions of the viral genome, and in the whole genome, using third generation NGS technologies.

## 6.2. Discussion second study: Comparison of circulating hepatitis B virus DNA and RNA quasispecies in the surface/polymerase overlapping region

The aim of the second study was to extend the analysis of similarities and differences between circulating HBV-DNA and HBV-RNA QS. In this sense, we moved to a new region of the viral genome, that includes the overlapping surface (S) – reverse transcriptase (RT) which together with *HBX* is also very relevant to HBV replication. In this study, 12 untreated patients were included. Like the first study, all patients were also characterised by the absence of treatment pressure. Therefore, this study was the first step to verify if the similarities and differences observed in *HBX* in terms of both HBV-DNA and HBV-RNA QS complexity could be also reported in the S-RT region, and also analyse the similarity of their QS structures and their genetic distances. In addition, we analysed the evolution of HBV-RNA and HBV-DNA QS over time in two different samples, 1.1-1.9 years apart to each other, from three (25%) of the 12 patients included.

The region of interest ranged from nt 616 to 972 of the HBV genome, which covers most of the functional domain B and the entire functional domains C-E, being S-RT region one of the major targets of antiviral therapies. As the RT domain of the P ORF overlaps with the S domain of the preS/S ORF, mutations within the S gene could lead to structural and functional changes in the RT with potential impact on viral replication capacity and on the treatment effectiveness [233]. Thus, to understand differences between both QS in this region it was necessary to establish a baseline situation in which patients were not receiving

antiviral treatment, as the HBV-DNA QS could not be analysed during NAs treatment due to undetectable viral loads.

In the present study we faced the problem of co-extraction of HBV-DNA with HBV-RNA. Despite using the same protocol as in the previous study, it was not able to efficiently digest DNA traces in all the isolated RNA and DNA contamination by qPCR-based absolute quantification assay was detected. To solve this problem, a more rigorous DNase treatment was performed, based on a double digestion with the TURBO DNase enzyme (Thermo Fisher Scientific-Life Technologies, Austin, United States), an optimised variant of the wild-type DNase I with a greater DNA affinity and capable of digesting DNA into fragments even when the DNA concentration is in the nM range. TURBO DNase proved to be able to efficiently digest those small amounts of contaminating DNA. Furthermore, based on the results of the amplification of their respective cDNA retro transcribed samples, TURBO DNase treatment seemed to maintain a good cDNA concentration. Moreover, when comparing cDNA amplification derived from reverse transcription of DNase I-treated RNA isolations with those obtained with TURBO DNase treatment, it was observed no apparent degradation of the RNA. Therefore, to avoid bias in the NGS results obtained from HBV-RNA QS, we modified the DNase treatment protocol from the first study to this new digestion.

### Comparison of master haplotypes

The frequency of the master haplotype in a given QS provides a first measure of the homogeneity of that QS. The diversity of patterns observed when comparing these frequencies in HBV-DNA and HBV-RNA QS suggests that there is no systematic behaviour of these two QS. The HBV-RNA QS is an indirect measure of the intrahepatic transcription process without reverse transcription [119]. Although both HBV-DNA and HBV-RNA QS originate from transcription of cccDNA, in the previous study we observed small differences in QS conservation and complexity in both HBV-DNA and HBV-RNA QS. Viral RT are notoriously error-prone and expected to make an error every ~10,000 to 30,000 bp [234], whereas RNA polymerase II (RNAPII) is expected to make an error every 3,900,000 bp [235]. In fact, mRNA molecules synthesised by RNAPII contain the least amount of errors compared to all other cellular RNA polymerases [236]. Consistent with this, the pattern most commonly observed was that the master haplotype of HBV-RNA was more frequent than that of HBV-DNA identified in five patients (P02, P11, P15, P24 and P26) and in the follow-up sample of P21. It is noteworthy that in this pattern, the differences among these frequencies were greater than when the HBV-DNA frequency is the highest with the exception of one patient.

In addition, all but three patients retained the same master haplotypes in both HBV-DNA and HBV-RNA QS. Among these three patients, the UPGMA tree of master haplotypes of all samples showed that P02 showed the largest genetic distance between HBV-DNA and HBV-RNA master haplotypes, indeed similar to genetic distances between the master haplotypes of different patients with the same genotype. But if taking into account the master haplotypes frequencies according to the type of QS, i.e., HBV-DNA or HBV-RNA, they were very disparate from one another. Whereas in HBV-RNA it is the master haplotype that clearly includes most of molecules (i.e., reads representing single viral genomes) of the QS (85.92%), in HBV-DNA the QS population is not dominated by the master haplotype, on the contrary, it is underrepresented (14.05%). Therefore, while HBV-RNA QS was dominated by a sole haplotype, the HBV-DNA QS was dominated by haplotypes with different fitness than the master, probably evidencing different evolutive mechanisms in both QS which should be clarified. In addition to P02, the raw nucleotide distances of the master haplotypes represented in the UPGMA tree of these haplotypes for P08 and P14 were also surprising, as a considerable large distance was found when clustering the two HBV-DNA and HBV-RNA master haplotypes of each patient. However, unlike P02 different percentages of master haplotypes for both QS were observed. P14 showed low frequencies of master haplotype for both HBV-DNA (17.39%) and HBV-RNA QS (9.04%). The case of P08 is very surprising as it shows similar frequencies of master haplotypes for both QS (84.62% vs. 83.88%, for HBV-DNA and HBV-RNA QS respectively). This would mean that both QS were dominated by a different master haplotype. Among the patients with longitudinal samples, P21 showed a striking pattern, with the baseline master haplotype of HBV-DNA more frequent than that of HBV-RNA (71.36% vs. 62.06%), and the reverse situation at follow-up with the HBV-RNA master haplotype becoming the most represented (52.48% vs. 86% for HBV-DNA and HBV-RNA, respectively). The different patterns of master haplotypes frequencies and differences between HBV-DNA and HBV-RNA QS observed in P02, P08 and P14 did not seem to correspond to higher HBV-DNA or HBV-RNA viral loads nor alanine aminotransferase or liver damage levels. However, in general, this study included a small number of patients. Higher numbers showing different HBV-DNA and HBV-RNA master haplotype profiles should be assessed in order to draw conclusions about the clinical implications of these profiles. In addition, P21 showed that HBV-DNA and HBV-RNA QS structure is dynamic even in absence of antiviral treatment in S-RT region. This is supported by the changing profiles of the frequencies of the master haplotypes through longitudinal follow-up. In the case of P02, P08 and P14 we could only analyse a single serum sample which portrays their QS in a single moment. An exhaustive follow-up would have provided more information on clinical and QS evolution.

Differences between frequencies of HBV-DNA and HBV-RNA master haplotype could suggest that not all haplotypes of HBV-RNA QS are equally likely to be encapsidated for retrotranscription. At this point, even the possible different behaviour of the different HBV genotypes must also be taken into account [237]. Indeed, the *in vitro* replication phenotype of HBV genotypes A to D revealed marked differences in HBV-DNA, HBV-RNA and protein expression and secretion [238]. Concerning genotype H, it has a low replication phenotype, characterised by lower HBV-DNA and virion levels and altered expression of viral proteins, probably due to reduced promoter activity in the upper regulatory region encoding the basal core promoter, which is required for expression of pgRNA and preC RNA [24]. Despite the reported diversity of master haplotype frequencies, it was observed that all patients retained the same genotype between HBV-DNA and HBV-RNA QS. Regarding the three of 12 chronic patients with follow-up, the same genotype was maintained between baseline and follow-up samples and for HBV-DNA and HBV-RNA QS. This allowed the genotype to be tested and determined from both viral QS yielding the same results, which would be particularly important in situations where HBV-DNA levels are undetectable due to antiviral treatment. Therefore, if HBV-RNA can be detected in these situations, viral genotyping from an RNA sample would be possible. However, among our patients we only have a single case of genotypes B, D, E and H, thus further studies should be performed to assess this possibility. In terms of HBV genotype, P02 stood out as the only patient showing genotype H. This genotype is phylogenetically very distant from the others (A-E), which at least are distributed throughout the Afro-Eurasian-Asiatic multi continental mass that has been strongly communicated all over the existence of the humanity. On the other hand, human populations in the American continent have been strongly isolated since the ice age from populations in this multi continental mass, and therefore from the other genotypes. On the American continent, two genotypes have been described that are very distant from each other: genotype H, which is mainly found in Central America, and genotype F, which is found in the north of South America. Both regions are jungle with difficult access, which might justify a very specific evolution of a viral agent such as HBV among isolated populations in these environments.

However, an analysis based on the simple observation of the master haplotype frequency does not imply that one HBV-DNA and HBV-RNA QS is more homogeneous than the other. Then, more indices are required to deeply study the viral QS.

#### Quasispecies complexity by quasispecies fitness fraction and Hill numbers profile

In this study, an in-depth analysis about the differences between the two viral QS has been carried out. In the first study the QS analysis was only performed using the rare haplotype

load (RHL) as a biomarker of possible mutagenesis due to reverse transcription, just to compare accumulation of “rare” haplotypes in circulating HBV-DNA and HBV-RNA QS. Differently, in this second study, QS haplotypes were divided into four fractions according to their fitness (QFF) in which each fitness fraction was an aggregate of haplotypes according to their frequency. The first fraction included the master haplotype; the second included emerging haplotypes (present at frequencies > 1% but less than that of the master haplotype); and the third and the fourth the low and very low fitness haplotypes (present at 0.1-1% and at < 0.1%, respectively). Of note, RHL at 1% used in the first study is also an aggregation of haplotypes within the QS, but in this case it is restricted to the low and very low fitness fractions. Using the QFF we extended partitioning to whole QS according to the biological features of their viral populations (haplotypes), thus providing a whole picture of QS structure. Since QFF are aggregations of haplotypes, they have been used in combination with the Hill number profile (HNP), as recommended [217]. The HNP provide a complementary view of the QS structure since they summarize the distribution of relative frequencies of the haplotypes in QS, without aggregation. The HNP consists of different values or orders (q) that assign increasing weights to the haplotype frequencies within a QS as the order q increases.

With regard to the QFF distribution of haplotypes, also three patterns were identified at baseline samples in relation to the three previous patterns observed in the frequencies of master haplotypes. This made sense taking into account that the frequency of the master haplotype itself is one of the four QFF. However, P14, P15 and P26 showed similar distributions in QFF, with a fraction of master haplotypes slightly higher in HBV-DNA than in HBV-RNA for P14 (17.37% vs. 9.04% respectively) and higher in HBV-RNA than in HBV-DNA for P15 and P26 (44.81% vs. 30.01% and 54.37% vs 46.64%, respectively for P15 and P26). Despite these differences, in general, all those patients showed relatively low frequencies of their master haplotypes while relatively high frequencies of the other QFF. The different QFF patterns described in relation to both HBV-DNA and HBV-RNA QS showed that high values of the master haplotype fraction result in an uneven distribution of the fitness fractions in QS. Therefore, high values of the master haplotype fraction are generally associated with low QS diversity.

The structures of QS of the patients who showed different master haplotype P02, P08, and P14 were different in the three cases. The P02 showed large fractions of emerging and rare haplotypes dominating the HBV-DNA compared to the HBV-RNA QS in which rare haplotypes were very scarce. The HNP of this patient showed that the two curves corresponding to each of the HBV-DNA and HBV-RNA QS were far apart from each other.

In this case, HBV-DNA was plotted as higher and more horizontal curve than that of HBV-RNA QS, which presented a much steeper slope. This is in keeping with the high values of the three fractions that do not correspond to the master haplotype in HBV-DNA, in contrast with the high value of the master haplotype fraction in HBV-RNA QS. Altogether indicate a higher diversity in the QS structure of HBV-DNA QS than that of RNA in P02. Regarding P14, the clear dominance of the emerging haplotype and low/very low fitness fractions over the master haplotype fraction was demonstrated in both QS, showing HBV-RNA QS with more diversity translated into a lower frequency of master haplotype. This was also observed while studying the HNP of this patient. Differently from P02, P14 showed high HNP curves compared to the other patients, and HBV-RNA curve was higher than that of HBV-DNA, in keeping with the lower frequency of HBV-RNA QS master haplotype. The QFF also confirmed what was observed for master sequences of P08, both QS were found to be dominated by the master haplotype. The HNP curves of P08 followed exactly the same trend, mainly because they were overlapped, with a sharp decrease. This suggested an almost identical structure of the two QS dominated by the master haplotype fraction, which was different in both QS. These three patients illustrate different QFF and HNP profiles: P02 showed a very different QS structure between HBV-DNA (dominated by mid and low frequency haplotypes different from the master, thus high diversity) and HBV-RNA (dominated by the master fraction, thus low diversity). On the other hand, P08 and P14 showed similar structures in both QS, although these structures were completely opposite (high frequency of master, low diversity, and low frequency of master, high diversity for P08 and P14 respectively). Other patients, (P07, P16 and P21) showed a larger master haplotype and a shrinking fraction of emerging haplotypes in HBV-DNA compared to HBV-RNA QS, showing the last one a more diverse QS structure. Therefore, HBV-DNA and HBV-RNA QS evolve independently from each other in different patients, suggesting that differences between their structures might not only be explained by reverse transcription of pgRNA. In line with what was suggested in the first study, it seems possible that transcription-associated mutation mechanisms [232] may contribute to HBV-RNA QS diversity.

Regarding the longitudinal study in three of 12 patients, the results observed by QFF are also in accordance with those reported for the frequencies of the master haplotypes. The P16 and P19 showed a rather similar scenario in which both HBV-DNA and HBV-RNA QS distributions were maintained in relation to baseline time point. This was also evidenced by the HNPs, that showed completely or almost completely overlapping curves between baseline and follow-up HBV-DNA and HBV-RNA QS. In contrast, the P21 had a similar

master fraction in baseline samples, although higher in HBV-DNA QS. At follow-up, this trend appeared to be reversed and the master fraction of HBV-RNA QS became much higher. Consistent with this, HNP showed that both HBV-DNA and HBV-RNA QS were almost overlapped at baseline, thus showing similar diversity and structure, while the structures were completely different between both QS at follow-up. It seems possible that haplotypes in emerging fractions of HBV-DNA QS at baseline time point had the potential to proliferate, attain higher frequencies. However, they did not overtake the current master sequence, which was the same in all four samples of P21. So, while at the DNA level other haplotypes would be selected, at the RNA level the opposite would happen; the master would be selected. These results confirm what was suggested solely taking into account the frequencies of both master haplotypes. Both HBV-DNA and HBV-RNA QS structure are dynamic even in absence of antiviral treatment. Overall, in patients with a different evolutionary profile during follow-up, it could be suggested that the resulting QS should be better adapted to its current environment.

As mentioned above, QS complexity was assessed in combination with the HNP. The different QFF patterns described in relation to both HBV-DNA and HBV-RNA QS showed that high values of the master haplotype fraction result in an uneven distribution of the fitness fractions in QS. Therefore, high values of the master haplotype fraction are generally associated with low QS diversity. This is observed in both QS of P08, whose HNP curves follow exactly the same trend, suggesting an almost identical composition of the two QS dominated by the master haplotype fraction, which is also observed by QFF. P02 showed that the two HNP curves corresponding to each of the HBV-DNA and HBV-RNA QS were far apart from each other. In this case, HBV-DNA was plotted as higher and more horizontal curve than that of HBV-RNA QS, which presented a much steeper slope, similar to those of P08. The high values in the other three fractions, which do not correspond to the master haplotype fraction, indicate a high diversity in the HBV-DNA QS of P02, which is also consistent with the study of master haplotype and QS complexity by QFF.

### Similarity between common haplotypes distributions

In this study, we also assessed the fraction of molecules belonging to common haplotypes between both QS, and the distribution of their frequencies using the similarity indexes of Commons ( $C_m$ ) and Overlap ( $O_v$ ) were compared [106]. Therefore,  $C_m$  and  $O_v$  allowed to study the extent of changes between the HBV-DNA and HBV-RNA QS regardless haplotypes fitness, simply considering common haplotypes frequencies between both QS. At baseline, most of patients showed related QS with a high fraction of molecules belonging to common haplotypes in both HBV-DNA and HBV-RNA QS, translated into a high  $C_m$  value.

Of note, P19 and P22 showed a dominance of HBV-DNA and HBV-RNA QS by the same master haplotype, which included most of QS molecules showing similar frequencies, all of them around 80% or higher. On the other hand, P08 showed a very similar distribution of reads into different QS haplotypes as P19 and P22. However, master haplotypes were different between both QS and not common to each other, with a low  $C_m$  value (0.422). In addition, the common haplotypes different from the master, in particular those haplotypes common at lower frequencies, showed quite similar frequency distributions in P19 and P22 yielding extremely high  $O_v$  values (0.923 and 0.931, respectively) whereas in P08 these distributions showed to be quite different, yielding a very low  $O_v$  value (0.00242). This example illustrates the usefulness of the  $C_m$  and  $O_v$  indexes, because while P08, P19 and P22 were very similar in terms of HBV-DNA and HBV-RNA QS structure as assessed by QFF and HNP, the distribution of molecules in common haplotypes and the similarities between their frequencies as assessed by  $C_m$  and  $O_v$  were completely different. This suggests that different evolutionary mechanisms are acting on the three patients, but these remain unexplained. No different evolutionary pressures were known between them and, as previously explained, no apparent differences in viral loads, ALT or liver damage levels were observed among these patients. In addition, it should be borne in mind that the similar QS structure in these three patients was a picture of a given moment, and a longitudinal follow-up may reveal different evolutionary trends between them. It is worth to note that, HBV-DNA and HBV-RNA QS from different samples containing haplotypes showing the same genotypes could also be highly related as observed in the five genotype A and 3 genotype C patients included. The closer phylogenetic relationship between sequences from the same genotype [16][17] facilitates that common haplotypes may be present in QS from different patients. However, the  $O_v$  values were near to 0 in all cases, outlining the different distribution of frequencies in different QS, evidencing different evolution of QS in different environments.

Together with P08, P02 and P14 were the three patients who showed the lowest  $C_m$  values, indicating that their HBV-DNA and HBV-RNA QS shared a smaller fraction of molecules belonging to common haplotypes than the rest of patients included. These results were congruent with the ones obtained when comparing their QS structures. In the case of P02, the low  $C_m$  value (0.464) could be explained because the large differences observed between the fractions of master haplotypes of the two QS. The master haplotype fraction of HBV-RNA was much larger than that of HBV-DNA QS (85.92% vs. 14.05% respectively). Thus, most of the molecules from HBV-RNA were probably not included in common haplotypes with HBV-DNA QS. In addition, the HBV-DNA QS, was dominated by emerging



and low fitness fractions. Therefore, common haplotypes are likely to be represented at different frequencies in both QS, resulting in a low  $O_v$  value (0.0513). On the other hand, comparison of the fraction molecules included in common haplotypes between HBV-DNA and HBV-RNA QS in P14 showed a different scenario. The different master haplotypes represented very reduced fractions in both QS (9.04% and 17.39% for HBV-RNA and HBV-DNA QS, respectively). Thus, higher fractions of molecules may have accumulated in other haplotypes included in the other fractions common to HBV-DNA and RNA QS, resulting in a higher  $C_m$  value (0.692) than P02 and P08. Regarding the  $O_v$ , although P14 showed lower values (0.421) than most of patients, these values were much higher than those reported for P02 and P08. Therefore, despite having different master haplotypes, the HBV-DNA and HBV-RNA QS haplotype distributions appear to be more homogenous between them in P14 than in P02 and P08. Although these patients showed the lowest  $O_v$  values, a certain variability in the values of this index was observed in all the other patients, despite the high degree of relatedness between HBV-DNA and HBV-RNA QS, as evidenced by the high  $C_m$  values in most of them. The  $O_v$  values ranged from 0.592 to 0.752 in half of patients and were even lower in the three patients discussed above, including P11. This again demonstrates the independent evolutionary trends of HBV-DNA and HBV-RNA QS.

Regarding the three patients with longitudinal follow-up, all QS comparisons at baseline, at follow-up and between baseline and follow-up showed very high  $C_m$  values (near 1 between all QS from the same patient), indicating a great relatedness among them. For the  $O_v$  index, a very similar haplotype frequency distribution was especially observed in P19 for all the four QS analysed. Moreover, similar haplotype frequency distributions were shown between baseline and follow-up HBV-RNA and baseline and follow-up HBV-DNA QS in P16. The P21 was the one with the lowest  $O_v$  values, with HBV-RNA QS at baseline sharing a closer similarity with HBV-DNA QS at follow-up. Therefore,  $C_m$  and  $O_v$  proved to be useful to follow QS evolution between two time points, as was previously reported in a patient chronically infected by hepatitis E virus who underwent an off-label treatment with Ribavirin for three years [106]. In this sense, it would be interesting to apply these indexes in the same S-RT region analysed in the present study, in patients under antiviral treatment or with an active immune response to which this region of the viral genome could be exposed [239].

#### Nucleotide diversity in hepatitis B virus DNA and RNA quasispecies and genetic distances between them

The  $C_m$  and  $O_v$  indexes we used to assess the degree to which two haplotype distributions were similar between HBV-DNA and HBV-RNA QS. However, both QS could also be compared by the degree of differentiation between them. In this sense, we measured their

genetic distance, which takes into account the genetic differences between two QS. The distance between two QS was measured as the average number of nucleotide substitutions between their reads. From this calculation we subtracted the mean between the nucleotide diversity ( $\pi$ ) of each QS, in order to obtain the net genetic distance ( $D_A$ ) [106]. Of note  $\pi$  is defined as the average number of nucleotide substitutions for a couple of arbitrary reads in each QS [218]. Assessment of  $\pi$  in baseline QS, showed a trend of HBV-RNA to be higher than HBV-DNA QS in more than half of the patients at baseline. In the study reported by Yu *et al.* [240], evolution and comparison between serum HBV-RNA and HBV-DNA QS was performed in a group of HBeAg positive CHB patients receiving entecavir or pegIFN $\alpha$  with a mean genetic distance assessed in 122 paired HBV-RNA and HBV-DNA samples. The authors observed that the mean genetic distance of baseline HBV-RNA was significantly greater than that of HBV-DNA. These results seem to support our observations; however, our group of patients was more heterogeneous, and the region of viral genome analysed in that study (the terminal protein domain of the viral polymerase) is different from the region analysed in the present study. Therefore, further studies in the S-RT region are needed to confirm the trend to a higher  $\pi$  of HBV-RNA QS. During the follow-up,  $\pi$  tended to increase in both QS in our three patients, only P21 showed a decrease from baseline to follow-up in HBV-RNA QS. Interestingly, the structure of HBV-RNA QS became more uneven at follow-up due to the increase of the master haplotype fraction as observed by QFF and HNP analysis. The same reasoning may serve to explain why  $\pi$  increased in the HBV-DNA QS.

Once again, HBV-RNA demonstrated that it is not necessarily less diverse or heterogeneous than that of HBV-DNA QS, despite reflecting the direct transcription process of cccDNA without the mutagenic or evolutionary effects of reverse transcription. The origin of this variability, or nucleotide diversity, is not likely to be due to errors during transcription, since as mentioned above, mRNA molecules synthesised by RNAPII contain the least amount of errors compared to all other cellular RNA polymerases [236]. As mentioned above, a reason for the reduced  $\pi$  of HBV-DNA compared to HBV-RNA QS could be speculated that only a fraction of the packaged pgRNA would be reverse transcribed and released, thus HBV-RNA would show an accurate picture of the full set of intracellular HBV transcripts while HBV-DNA diversity would be limited. Another possibility that could render a higher HBV-RNA QS could be, the post-transcriptional modifications although these errors could be incompatible with the viral reverse transcription. So, there would be haplotypes that will not be retrotranscribed due to these changes, leading to a decrease in HBV-DNA complexity [232]. These possibilities, among others, should be considered in future lines of

research aimed at elucidating the surprisingly high complexity and genetic diversity of circulating HBV-RNA QS.

Finally, calculation of  $D_A$  allowed representing the genetic distances between the QS as a UPGMA tree. The resulting tree was very similar to the tree showing the genetic distances between master haplotypes, with only P02 and P08 showing considerable genetic distances between their QS. As for the other patients, including P14 who also showed different master haplotypes in HBV-DNA and HBV-RNA QS, distances between both QS were 0 or near 0. In addition, the distances between HBV-DNA and HBV-RNA QS of the three patients followed for 1.1-1.9 years were close to 0 when the two baseline samples or the two follow-up samples were compared with each other or when the same QS was compared between baseline and follow-up samples. This situation was the same for P21, who changed the HBV-DNA and HBV-RNA QS structure through the follow-up. The  $D_A$  values correlates with the distribution of common haplotypes between the assessed by  $C_m$  and  $O_v$  [106]. In this sense, P02 and P08 showed the lowest  $C_m$  and  $O_v$  values, indicating a very low fraction of molecules sharing common haplotypes with different frequencies between their HBV-DNA and RNA QS. In this situation, most of the haplotypes would be different between both QS, so  $D_A$  took high values. Of note, HBV-DNA QS of P08 was closer to HBV-DNA and HBV-RNA QS from P15 and P16 than to the HBV-RNA QS of the same sample. Both P08, P15 and P16 share the HBV genotype A, and it should be remembered that QS composed of haplotypes with the same genotype could be related by showing high fractions of reads in common haplotypes. In fact,  $C_m$  values of HBV-DNA and HBV-RNA QS of P08 compared to those of P15 and P16 were closer to 1 than between HBV-DNA and HBV-RNA QS of P08. This explains why the  $D_A$  of both HBV-DNA and HBV-RNA QS from P08 was longer between them than to viral QS from P15 and P16. We observed similar low  $C_m$  and  $O_v$  values between the HBV-DNA and HBV-RNA QS in P02. However, as this patient was the only one with genotype H in our group of patients, the  $D_A$  to QS from any other patient was longer than between both QS from P02. Finally, P14 was the only patient whose differences were diluted when considering the whole QS and not only the master haplotype. In this case,  $C_m$  and  $O_v$  values were higher than in P02, P08, so the distributions of the haplotypes were more related, and their frequencies more similar between HBV-DNA and HBV-RNA QS. In this situation, and taking into account the low weight of different master haplotypes in the structure of both QS, it seems reasonable the short  $D_A$  between HBV-DNA and HBV-RNA QS.

### 6.3. Discussion third study: The combination of antisense locked nucleic acid Gapmers and/or siRNA as a potential hepatitis B virus gene therapy: preliminary *in vitro* results

The aim of the third study was to design and test in an *in vitro* cellular infection model a new therapeutic strategy based on gene silencing with Gapmers (GPs) and siRNA targeting hyper-conserved *HBX* regions previously identified by our group and confirmed on HBV-RNA QS.

Due to the persistence of the cccDNA, HBV flares and relapses are common after cessation of the antiviral therapy [241][242]. To limit the continued expression of the viral 'minichromosome', new therapeutic strategies aimed at silencing or reducing the HBV reservoir (cccDNA) are needed to control HBV expression, interfere with disease progression and potentially promote the functional cure of the chronic HBV infection (2).

The strategy of silencing the expression of viral transcripts appears to be the most promising. The most advanced RNA therapeutic is the siRNA ARC-520/521, which successfully reduces HBsAg in HBeAg-positive patients. However, its effect was limited in HBeAg-negative patients or those on long-term NAs treatment, highlighting the importance of circulating HBsAg derived from integrated HBV, as this molecule was not designed to take into account the sequence of this integrated DNA or cccDNA [243]. Preliminary results with a second generation of siRNAs (JNJ3989 and AB-729) suggest that HBsAg reduction was comparable in HBeAg-positive and HBeAg-negative patients [244][245].

The aim of the third study was to design and test in an *in vitro* cellular infection model a new therapeutic strategy based on gene silencing with GPs and siRNA targeting hyper-conserved regions in the *HBX* previously identified by our group and confirmed on HBV-RNA QS (first study of this thesis).

All the silencers designed by our group, both GPs and siRNA, were selected based on several criteria such as targeting the 3' end common to all HBV transcripts and targeting a sequence in the viral genome that is highly conserved between major HBV genotypes.

GPs are antisense oligonucleotides that were modified by the addition of locked nucleic acids at their ends. This modification improves their affinity for the target and increases their structural stability and their resistance to cellular nucleases. As a result, their potential clinical application is extremely high, even for the production of therapeutic strategies with

different routes of administration, such as inhalable drugs [246]. Due to their small size and stability, they can easily enter the nucleus of cells, even by non-assisted cellular uptake. For this reason, we chose this type of silencing strategy against HBV infection. In order to guarantee a strategy that could be suitable in different clinical stages and in the presence of different genotypes, we selected as targets the hyper-conserved regions previously described by our group [2] and further confirmed in circulating RNA in the first study included in this thesis.

Four different types of GPs and one siRNA were designed. To ensure that the observed inhibitory effect was sequence specific, a 5'-FAM-scrambled GP was also used as a control in each experiment. Although all GPs were able to induce inhibition of HBeAg expression, GP1 and GP4 were the most efficient in inhibiting the viral expression. In particular, when treated early after infection (48 h pi), GP1 showed the best percentage inhibition at both pgRNA and protein levels (HBeAg and HBsAg), while GP4 was slightly less efficient at inhibition. Although both GPs were designed to target hyper-conserved regions of *HBX*, each has its own specific sequence and each target sequence can be exposed to a greater or lesser extent. This may probably explain the difference in the inhibition efficiency. Nevertheless, the GPs designed by our group showed a good inhibition percentage for both pgRNA and HBeAg and HBsAg proteins.

However, late after infection (5 d pi), treatment with GPs showed an overall decrease in the percentage of inhibition. At this time point, cccDNA is well-established with active transcriptional activity [247], which could determine an accumulation of viral transcripts in the hepatocyte cytoplasm. Consequently, the single action of GPs at both nuclear and even cytoplasmic levels would not be sufficient to effectively inhibit all of this accumulated RNA, thus requiring an adjustment in the concentration of GPs. This led to the idea of combining the two silencers. The combination should double the antiviral efficacy under later conditions, especially in combination with siRNA, which would act as a second "line of defence" against those mRNAs that escape intranuclear inhibition. This was confirmed by the fact that, although the GPs were able to detect more than 50% inhibition of pgRNA expression, the GPs alone slightly silenced the expression of viral proteins at late expression. Moreover, the combination of both GPs increased the inhibition of pgRNA but had no effect on HBsAg or HBeAg. In contrast, the combination with a siRNA was able to increase the inhibition of both viral proteins and pgRNA, suggesting a kind of additive effect. This effect is widely used in pharmacology as it allows the antiviral effect to be enhanced without increasing the concentration of the individual drug. It is well known, for example, in the treatment of HIV infection, where the administration of antiretroviral drugs belonging to

different classes of compounds targeting different steps of the viral life cycle is often used to achieve an additional effect on viral load suppression and better long-term control of viral replication compared to monotherapy [248].

Notably, the sequence of the GPs and siRNA is included in all viral transcripts due to the co-terminal localisation of the *HBX*. Of the designed GPs and siRNA, GP1 and the siRNA target a hyper-conserved region within the *HBX* promoter. This is very important because they can potentially silence viral transcripts produced from integrated HBV-DNA. By analysing the integrated HBV-DNA from 14 HBV-associated HCC, it was observed that most cases had deletions in the 3' end of the *HBX*, which could produce C-terminal truncated HBx proteins [249]. Of note, the silencing strategy presented in this dissertation targets the other end of the gene (the 5' *HBX*) and includes the gene promoter. This means that our GPs and siRNA can also inhibit the expression of transcripts produced from the integrated HBV-DNA, which is particularly important in preventing liver disease progression and the development of HCC.

Despite the promising results obtained on the antiviral efficacy of GPs alone or in combination, the *in vitro* infection system used here represents the main limitation of the study. The HepG2-hNTCP cells are a good approximation of the HBV infection as they allow the establishment of cccDNA and support the expression of all viral proteins and transcripts. However, they do not support cellular re-infection, which limits its spread *in vitro*. Therefore, the study of HBV infection should be improved. Three-dimensional culture strategies have been reported to provide results with far greater reliability two-dimensional cell culture models, as they allow the cell culture to grow in the same way as in the human body, retaining single-cell characteristics and analogously reproducing complex structures and physiological functions [250].

The study of HBV infection requires a continuous effort to develop new molecules, treatment combinations and novel therapeutic strategies to achieve the goal of HBV elimination [251]. Although the silencing strategy is very promising in the context of the HBV infection, the major challenge that hinders the therapeutic use of oligonucleotides and interference RNAs is their delivery. Consistent with our results, despite the promising inhibitory capacity, the transfection efficiency of the fluorescent scrambled GP was quite variable depending on cell status and permissiveness.

Based on that, a new delivery system should be tested. In recent years, a significant progress has been made in overcoming some of the obstacles associated with *in vivo* delivery of siRNA, and nanoparticles (NPs) are very promising tools. In nanomedicine, a NP

refers to an organic and biodegradable nanometric structure that contains molecules inside to be released once they reach their destination (into the target tissue). This idea has allowed to effectively treat human diseases, such as glioblastoma, colorectal cancer, and breast cancer [252][253]. This delivery system is very important in cancer research to protect hydrophobic drugs from premature degradation and to promote a targeted delivery [254]. Encapsulation would therefore solve these problems by preventing an early elimination of the drug by mononuclear phagocyte system and by supporting a prolonged retention period, which would allow NPs to accumulate in target tissue to an optimal concentration [255]. It has been reported that the most common strategy to avoid phagocytic uptake by the liver is to coat the NP with polyethylene glycol (PEG) to increase surface hydrophilicity and improve circulation half-life [256]. There are two types of NPs worth mentioning depending on their composition: the lipid NPs (LNPs) and the polymeric NPs. LNPs have been specifically tested for the delivery of siRNA molecules to hepatocytes since the liver is a well-perfused organ with a fenestrated endothelium. Moreover, LNPs are known to interact with serum proteins, exchanging components and acquiring proteins in the circulation that can potentially target LNPs to specific cell types [257]. It has been reported that siRNA-loaded LNP absorb apolipoprotein E (ApoE) on their surface, enhancing uptake into hepatoma cells and primary hepatocytes [257][258]. ApoE binds to the low-density lipoprotein receptor, which is highly expressed on the outer membrane of hepatocytes, thereby favouring liver-specific delivery. This mechanism of uptake by hepatocytes is the basis of the first FDA-approved siRNA-based drug, Patisiran [259], a siRNA that targets hepatocytes and prevents the production of misfolded proteins that accumulate in different parts of the body [260].

On the other hand, among the polymeric NPS, those formed by polyurethanes (PUs) have received increasing interest due to their synthetic versatility, excellent mechanical properties, and good biocompatibility [261]. PUs have been widely used in the clinic as prosthetic heart valves, catheters, blood pumps, and wound dressings [261]. Their release mechanism is based on redox/pH-response, allowing controlled drug release according to the intracellular protein concentration and pH [262], which is particularly important in tumour diseases.

Liver-specific targeting may take advantage of the asialoglycoprotein receptor (ASGPR), which is specifically expressed on the surface of hepatocytes [176]. The ASGPR ligand is the N-acetylgalactosamine (GalNAc), which can be conjugate to ASOs [263][264], siRNAs [265], and siRNA-contained GalNAc-NPs [257][266][267][268], the latter with an uptake efficiency of up to 96-98% [268]. Roche has developed an LNA-ASO-GalNAc conjugate

(the RO7062931) that has successfully passed phase I clinical trial [269], while another ASO-GalNAc conjugate, the GSK3389404, is in phase II clinical trial in NAs-treated patients [270].



## **CONCLUSIONS**

---

## 7. CONCLUSIONS

Regarding the first study:

1. A methodology for in-depth analysis of serum HBV-RNA QS, based on next-generation sequencing, was established. This methodology allowed a reliable elimination of serum HBV-DNA QS, avoiding its interference.
2. The analysis of the complexity in the *HBX* 5' region of HBV genome using the rare haplotype load index showed no statistically significant differences between HBV-DNA and HBV-RNA QS analysing all patients together or separating them according to the severity of liver disease. However, in general, an increased presence of low frequency HBV genomes was observed in HBV-DNA QS, which may be due to reverse transcription errors.
3. Interestingly, although HBV-DNA and HBV-RNA QS showed a similar degree of conservation in the same region of viral genome, HBV-RNA QS tended to show higher sequence variability than HBV-DNA.
4. Most of the hyper-conserved and hyper-variable regions identified were almost identical between HBV-DNA and HBV-RNA QS and were highly consistent with our previous findings.

Regarding the second study:

5. The methodology for in-depth analysis of HBV-RNA QS established in the previous study was successfully adapted to eliminate the HBV-DNA interference in all included samples.
6. A comprehensive comparative analysis of HBV-DNA and HBV-RNA QS using specific indexes to analyse their structure and compare their haplotype distributions and genetic distances, allowed us to explore the differences between both QS in the S-RT HBV genome region.
7. The master haplotype frequency of the HBV-DNA and HBV-RNA QS was very informative about their structure. However, the genetic complexity indexes complemented this information, showing that HBV-DNA and HBV-RNA QS structure and genetic diversity evolve independently and follow different patterns in different patients.

8. Regarding the longitudinal study, HBV-DNA and HBV-RNA QS structure and genetic diversity were dynamic even in the absence of antiviral treatment and could evolve differently in different patients. It would be necessary to include more patients to determine specific evolutionary patterns in relation to any virological or clinical characteristics of the patients.

Regarding the third study:

9. The study of the viral QS by next-generation sequencing is a valuable preliminary step for the design of HBV-specific antisense oligonucleotides (ASOs) and small interfering RNA (siRNA) targeting a highly-conserved region in the 5' *HBX*, the latter thanks to the confirmation of the conservation observed in serum HBV-DNA also at HBV-RNA level.
10. Two of the four designed GPs were highly efficient in inhibiting the HBV expression at an early stage of infection (48 h pi), although this efficacy appeared to be reduced when treated at a late time point (5 d pi). The combination of both GPs or one GP with a siRNA may overcome this limitation.
11. Improvement of the infection model is required to complete our preclinical evaluation, followed by a suitable encapsulation system to enhance delivery and support *in vivo* administration of the molecules.

## LIMITATIONS

---

## 8. LIMITATIONS

On the one hand, the main limitation of the first and second studies were the high differences between circulating HBV-DNA and HBV-RNA in the absence of treatment. These differences were found to be greater than previously published [118][229][115]. In the absence of antiviral treatment, circulating HBV-RNA was 3.6 log<sub>10</sub>copies/mL lower than HBV-DNA in average, as observed in a recent study by our group [126] This made it very difficult to obtain samples with sufficient levels of circulating HBV-RNA to analyse its quasispecies by next-generation sequencing, limiting the number of patients in whom this could be analysed. As a consequence of this, we compared HBV-DNA and HBV-RNA QS in small and heterogeneous groups of patients, which just enabled to take a preliminary picture of its comparison in 5' *HBX* and S-RT regions of HBV genome.

On the other hand, the main limitation of the third study was the rather variable transfection efficiency observed in all the experiments, which was strongly dependent on cell viability after DMSO treatment and infection status.

## **FUTURE PERSPECTIVES**

---

## 9. FUTURE PERSPECTIVES

The inclusion of a large number of patients with available follow-up samples and with different clinical and virological characteristics is necessary to draw firm conclusions regarding the circulating HBV-RNA QS and its comparison with the structure, haplotype distribution and genetic diversity of the HBV-DNA QS. Therefore, it would be necessary a comprehensive search of well-characterized patients under different immune and antiviral treatment pressures. This would allow to draw conclusions about the clinical implications of the HBV-RNA QS structure and genetic diversity, and its potential utility as a prognostic factor. In addition, it should be taken in mind that analyzing a QS in a single sample is like taking a picture of its structure and genetic diversity in a given moment. In this sense, a longitudinal study involving not only more patients but also longer follow-up period (long enough to continue to detect circulating DNA levels) would allow for an optimal and more precise determination of what has happened to the viral QS over time.

Among the most notable results presented in the first and second studies of the present thesis, there is the independent evolution of HBV-DNA and HBV-RNA QS structure and genetic diversity. This evolution seems to depend on the patient but is not unusual for the HBV-RNA QS to show a more diverse structure and higher genetic diversity than for the HBV-DNA QS. In addition, in the first study it was observed that the HBV-RNA QS tended to show higher sequence variability than HBV-DNA QS. In future studies, it would be necessary to explore mechanisms, in addition to the reverse transcription from HBV pgRNA to rcDNA, to generate diversity in both QS.

Regarding gene therapy studies, further experiments are required to confirm these results and other delivery systems should be tested to improve treatment efficiency. Nanotechnology-based delivery represents a significant step for gene therapy. Our group is currently working with lipid and polymeric (polyurethane) nanoparticles (NPs). In addition, NP surfaces can be functionalised with multiple moieties that allow them to be carriers of gene therapeutics. This versatility allows nanoparticles to be used as a strategy aiming for functional cure. However, as for a complete HBV cure both cccDNA and integrated HBV-DNA need to be targeted. Since HBsAg expression from integrated HBV-DNA is of importance for maintaining high levels of the protein itself, gene therapy is also prompted to be performed by studying the QS conservation of HBV-RNA in the S-RT region. Thus, the discovery of other conserved regions may be revealed, providing additional potential

targets for gene therapy and valuables for pan-genotypic approaches. This strategy would provide a therapeutic tool that could be used regardless of the patient's viral genotype or clinical stage.

Finally, our group is now focused on overcoming the expansion limitation of HepG2-hNTCP cells by generating a 3D-culture model, called liver organoids, from human induced pluripotent stem cells (iPSCs), which are able to self-renew, self-organise, and recapitulate the functionality of the tissue of origin [271].



## REFERENCES

---

## 10. REFERENCES

1. World Health Organization (WHO) Hepatitis B. *Fact sheet* **2022**, N°204.
2. González, C.; Tabernero, D.; Cortese, M.F.; Gregori, J.; Casillas, R.; Riveiro-Barciela, M.; Godoy, C.; Sopena, S.; Rando, A.; Yll, M.; et al. Detection of hyper-conserved regions in hepatitis B virus X gene potentially useful for gene therapy. *World J. Gastroenterol.* **2018**, *24*, 2095–2107, doi:10.3748/wjg.v24.i19.2095.
3. Cortese, M.F.; González, C.; Gregori, J.; Casillas, R.; Carioti, L.; Guerrero-Murillo, M.; Riveiro-Barciela, M.; Godoy, C.; Sopena, S.; Yll, M.; et al. 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**, 1–11, doi:10.1038/s41598-021-83762-4.
4. Graber-Stiehl, I. The silent epidemic killing more people than HIV, malaria or TB. *Nature* **2018**, *564*, 24–26, doi:10.1038/d41586-018-07592-7.
5. Blumberg, B.S. A “New” Antigen in Leukemia Sera. *JAMA J. Am. Med. Assoc.* **1965**, *191*, 541, doi:10.1001/jama.1965.03080070025007.
6. Blumberg, B.S.; Millman, I.; Sutnick, A.I.; London, W.T. The nature of australia antigen and its relation to antigen-antibody complex formation. *J. Exp. Med.* **1971**, *134*, 320–9.
7. BAYER, M.E.; BLUMBERG, B.S.; WERNER, B. Particles associated with Australia Antigen in the Sera of Patients with Leukaemia, Down’s Syndrome and Hepatitis. *Nature* **1968**, *218*, 1057–1059, doi:10.1038/2181057a0.
8. Alter, H.J.; Blumberg, B.S. Further studies on a “new” human isoprecipitin system (Australia antigen). *Blood* **1966**, *27*, 297–309.
9. Lin, W.-L.; Hung, J.-H.; Huang, W. Association of the Hepatitis B Virus Large Surface Protein with Viral Infectivity and Endoplasmic Reticulum Stress-mediated Liver Carcinogenesis. *Cells* **2020**, *9*, 2052, doi:10.3390/cells9092052.
10. Dane, D.S.; Cameron, C.H.; Briggs, M. VIRUS-LIKE PARTICLES IN SERUM OF PATIENTS WITH AUSTRALIA-ANTIGEN-ASSOCIATED HEPATITIS. *Lancet* **1970**, *295*, 695–698, doi:10.1016/S0140-6736(70)90926-8.
11. Blumberg, B.S. Australia Antigen and the Biology of Hepatitis B. *Science (80- )*. **1977**, *197*, 17–25, doi:10.1126/science.325649.
12. Galibert, F.; Mandart, E.; Fitoussi, F.; Tiollais, P.; Charnay, P. Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in *E. coli*. *Nature* **1979**, *281*, 646–650, doi:10.1038/281646a0.
13. Gerlich, W.H. Medical Virology of Hepatitis B: how it began and where we are now. *Viol. J.* **2013**, *10*, 239, doi:10.1186/1743-422X-10-239.
14. Herrscher, C.; Roingeard, P.; Blanchard, E. Hepatitis B Virus Entry into Cells. *Cells* **2020**, *9*, 1486, doi:10.3390/cells9061486.
15. Magnius, L.; Mason, W.S.; Taylor, J.; Kann, M.; Glebe, D.; Dény, P.; Sureau, C.; Norder, H. ICTV Virus Taxonomy Profile: Hepadnaviridae. *J. Gen. Virol.* **2020**, *101*,

- 571–572, doi:10.1099/jgv.0.001415.
16. Karlsen, A.A.; Kyuregyan, K.K.; Isaeva, O. V.; Kichatova, V.S.; Asadi Mobarkhan, F.A.; Bezuglova, L. V.; Netesova, I.G.; Manuylov, V.A.; Pochtovyi, A.A.; Gushchin, V.A.; et al. Different evolutionary dynamics of hepatitis B virus genotypes A and D, and hepatitis D virus genotypes 1 and 2 in an endemic area of Yakutia, Russia. *BMC Infect. Dis.* **2022**, *22*, 452, doi:10.1186/s12879-022-07444-w.
  17. Kramvis, A. Genotypes and Genetic Variability of Hepatitis B Virus. *Intervirology* **2014**, *57*, 141–150, doi:10.1159/000360947.
  18. Rajoriya, N.; Combet, C.; Zoulim, F.; Janssen, H.L.A. How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis B. Time for an individualised approach? *J. Hepatol.* **2017**, *67*, 1281–1297, doi:10.1016/j.jhep.2017.07.011.
  19. Liu, Z.; Zhang, Y.; Xu, M.; Li, X.; Zhang, Z. Distribution of hepatitis B virus genotypes and subgenotypes. *Medicine (Baltimore)*. **2021**, *100*, e27941, doi:10.1097/MD.00000000000027941.
  20. Mayerat, C.; Mantegani, A.; Frei, P.C. Does hepatitis B virus (HBV) genotype influence the clinical outcome of HBV infection? *J. Viral Hepat.* **1999**, *6*, 299–304, doi:10.1046/j.1365-2893.1999.00174.x.
  21. Lin, C.-L.; Kao, J.-H. Hepatitis B Virus Genotypes and Variants. *Cold Spring Harb. Perspect. Med.* **2015**, *5*, a021436–a021436, doi:10.1101/cshperspect.a021436.
  22. Kao, J.-H.; Chen, P.-J.; Lai, M.-Y.; Chen, D.-S. Hepatitis B virus genotypes and spontaneous hepatitis B e antigen seroconversion in Taiwanese hepatitis B carriers. *J. Med. Virol.* **2004**, *72*, 363–369, doi:10.1002/jmv.10534.
  23. Wong, G.L.-H.; Chan, H.L.-Y.; Yiu, K.K.-L.; Lai, J.W.-Y.; Chan, V.K.-K.; Cheung, K.K.-C.; Wong, E.W.-N.; Wong, V.W.-S. Meta-analysis: the association of hepatitis B virus genotypes and hepatocellular carcinoma. *Aliment. Pharmacol. Ther.* **2013**, *37*, 517–526, doi:10.1111/apt.12207.
  24. Sozzi, V.; Shen, F.; Chen, J.; Colledge, D.; Jackson, K.; Locarnini, S.; Yuan, Z.; Revill, P.A. In vitro studies identify a low replication phenotype for hepatitis B virus genotype H generally associated with occult HBV and less severe liver disease. *Virology* **2018**, *519*, 190–196, doi:10.1016/j.virol.2018.04.015.
  25. Sunbul, M. Hepatitis B virus genotypes: Global distribution and clinical importance. *World J. Gastroenterol.* **2014**, *20*, 5427, doi:10.3748/wjg.v20.i18.5427.
  26. Garcia-Garcia, S.; Cortese, M.F.; Rodríguez-Algarra, F.; Tabernero, D.; Rando-Segura, A.; Quer, J.; Buti, M.; Rodríguez-Frías, F. Next-generation sequencing for the diagnosis of hepatitis B: current status and future prospects. *Expert Rev. Mol. Diagn.* **2021**, *21*, 381–396, doi:10.1080/14737159.2021.1913055.
  27. Wei, L.; Ploss, A. Mechanism of Hepatitis B Virus cccDNA Formation. *Viruses* **2021**, *13*, 1463, doi:10.3390/v13081463.
  28. Liang, T.J. Hepatitis B: The virus and Disease. *Hepatology* **2010**, *49*, 1–17, doi:10.1002/hep.22881.
  29. Bousali, M.; Papatheodoridis, G.; Paraskevis, D.; Karamitros, T. Hepatitis B Virus DNA Integration, Chronic Infections and Hepatocellular Carcinoma. *Microorganisms*

- 2021**, 9, 1787, doi:10.3390/microorganisms9081787.
30. Rodriguez-Frias, F.; Buti, M.; Tabernero, D.; Homs, M. Quasispecies structure, cornerstone of hepatitis B virus infection: Mass sequencing approach. *World J. Gastroenterol.* **2013**, *19*, 6995–7023, doi:10.3748/wjg.v19.i41.6995.
  31. Miller, R.H.; Kaneko, S.; Chung, C.T.; Girones, R.; Purcell, R.H. Compact organization of the hepatitis B virus genome. *Hepatology* **1989**, *9*, 322–327, doi:10.1002/hep.1840090226.
  32. Salpini, R.; D’Anna, S.; Benedetti, L.; Piermatteo, L.; Gill, U.; Svicher, V.; Kennedy, P.T.F. Hepatitis B virus DNA integration as a novel biomarker of hepatitis B virus-mediated pathogenetic properties and a barrier to the current strategies for hepatitis B virus cure. *Front. Microbiol.* **2022**, *13*, 1–18, doi:10.3389/fmicb.2022.972687.
  33. García-García, S.; Caballero-Garralda, A.; Tabernero, D.; Cortese, M.F.; Gregori, J.; Rodriguez-Algarra, F.; Quer, J.; Riveiro-Barciela, M.; Homs, M.; Rando-Segura, A.; et al. Hepatitis B Virus Variants with Multiple Insertions and/or Deletions in the X Open Reading Frame 3’ End: Common Members of Viral Quasispecies in Chronic Hepatitis B Patients. *Biomedicines* **2022**, *10*, doi:10.3390/biomedicines10051194.
  34. Inoue, J.; Sato, K.; Ninomiya, M.; Masamune, A. Envelope Proteins of Hepatitis B Virus: Molecular Biology and Involvement in Carcinogenesis. *Viruses* **2021**, *13*, 1124, doi:10.3390/v13061124.
  35. Tu, T.; Budzinska, M.; Shackel, N.; Urban, S. HBV DNA Integration: Molecular Mechanisms and Clinical Implications. *Viruses* **2017**, *9*, 75, doi:10.3390/v9040075.
  36. Podlaha, O.; Wu, G.; Downie, B.; Ramamurthy, R.; Gaggar, A.; Subramanian, M.; Ye, Z.; Jiang, Z. Genomic modeling of hepatitis B virus integration frequency in the human genome. *PLoS One* **2019**, *14*, e0220376, doi:10.1371/journal.pone.0220376.
  37. Salpini, R.; Surdo, M.; Warner, N.; Cortese, M.F.; Colledge, D.; Soppe, S.; Bellocchi, M.C.; Armenia, D.; Carioti, L.; Continenza, F.; et al. Novel HBsAg mutations correlate with hepatocellular carcinoma, hamper HBsAg secretion and promote cell proliferation in vitro. *Oncotarget* **2017**, *8*, 15704–15715, doi:10.18632/oncotarget.14944.
  38. Lazarevic, I. Clinical implications of hepatitis B virus mutations: Recent advances. *World J. Gastroenterol.* **2014**, *20*, 7653, doi:10.3748/wjg.v20.i24.7653.
  39. Thi Cam Huong, N.; Trung, N.Q.; Luong, B.A.; Tram, D.B.; Vu, H.A.; Bui, H.H.; Pham Thi Le, H. Mutations in the HBV PreS/S gene related to hepatocellular carcinoma in Vietnamese chronic HBV-infected patients. *PLoS One* **2022**, *17*, e0266134, doi:10.1371/journal.pone.0266134.
  40. Heermann, K.H.; Goldmann, U.; Schwartz, W.; Seyffarth, T.; Baumgarten, H.; Gerlich, W.H. Large surface proteins of hepatitis B virus containing the pre-s sequence. *J. Virol.* **1984**, *52*, 396–402, doi:10.1128/jvi.52.2.396-402.1984.
  41. Montali, I.; Vecchi, A.; Rossi, M.; Tiezzi, C.; Penna, A.; Reverberi, V.; Laccabue, D.; Missale, G.; Boni, C.; Fisicaro, P. Antigen Load and T Cell Function: A Challenging Interaction in HBV Infection. *Biomedicines* **2022**, *10*, 1224, doi:10.3390/biomedicines10061224.
  42. Chai, N.; Chang, H.E.; Nicolas, E.; Han, Z.; Jarnik, M.; Taylor, J. Properties of

- Subviral Particles of Hepatitis B Virus. *J. Virol.* **2008**, *82*, 7812–7817, doi:10.1128/JVI.00561-08.
43. Sayan, M.; Şentürk, Ö.; Akhan, S.Ç.; Hülagü, S.; Çekmen, M.B. Monitoring of hepatitis B virus surface antigen escape mutations and concomitantly nucleos(t)ide analog resistance mutations in Turkish patients with chronic hepatitis B. *Int. J. Infect. Dis.* **2010**, *14*, e136–e141, doi:10.1016/j.ijid.2009.11.039.
  44. Arbuthnot, P.; Kew, M. Hepatitis B virus and hepatocellular carcinoma. *Int. J. Exp. Pathol.* **2001**, *82*, 77–100, doi:10.1111/j.1365-2613.2001.iep0082-0077-x.
  45. Chen, B.-F. Hepatitis B virus pre-S/S variants in liver diseases. *World J. Gastroenterol.* **2018**, *24*, 1507–1520, doi:10.3748/wjg.v24.i14.1507.
  46. Pollicino, T.; Cacciola, I.; Saffiotti, F.; Raimondo, G. Hepatitis B virus PreS/S gene variants: Pathobiology and clinical implications. *J. Hepatol.* **2014**, *61*, 408–417, doi:10.1016/j.jhep.2014.04.041.
  47. Li, Y.W.; Yang, F.C.; Lu, H.Q.; Zhang, J.S. Hepatocellular carcinoma and hepatitis B surface protein. *World J. Gastroenterol.* **2016**, *22*, 1943–1952, doi:10.3748/wjg.v22.i6.1943.
  48. Padarath, K.; Deroubaix, A.; Kramvis, A. The Complex Role of HBeAg and Its Precursors in the Pathway to Hepatocellular Carcinoma. *Viruses* **2023**, *15*, 857, doi:10.3390/v15040857.
  49. Feng, H.; Hu, K. Structural characteristics and molecular mechanism of hepatitis B virus reverse transcriptase. *Virol. Sin.* **2009**, *24*, 509–517, doi:10.1007/s12250-009-3076-6.
  50. Ghany, M.; Liang, T.J. Drug Targets and Molecular Mechanisms of Drug Resistance in Chronic Hepatitis B. *Gastroenterology* **2007**, *132*, 1574–1585, doi:10.1053/j.gastro.2007.02.039.
  51. Das, K.; Xiong, X.; Yang, H.; Westland, C.E.; Gibbs, C.S.; Sarafianos, S.G.; Arnold, E. Molecular Modeling and Biochemical Characterization Reveal the Mechanism of Hepatitis B Virus Polymerase Resistance to Lamivudine (3TC) and Emtricitabine (FTC). *J. Virol.* **2001**, *75*, 4771–4779, doi:10.1128/JVI.75.10.4771-4779.2001.
  52. Tavis, J.E.; Cheng, X.; Hu, Y.; Totten, M.; Cao, F.; Michailidis, E.; Aurora, R.; Meyers, M.J.; Jacobsen, E.J.; Parniak, M.A.; et al. The hepatitis B virus ribonuclease H is sensitive to inhibitors of the human immunodeficiency virus ribonuclease H and integrase enzymes. *PLoS Pathog.* **2013**, *9*, e1003125, doi:10.1371/journal.ppat.1003125.
  53. Ono-Nita, S.K.; Kato, N.; Shiratori, Y.; Masaki, T.; Lan, K.-H.; Carrilho, F.J.; Omata, M. YMDD motif in hepatitis B virus DNA polymerase influences on replication and lamivudine resistance: A study by in vitro full-length viral DNA transfection. *Hepatology* **1999**, *29*, 939–945, doi:10.1002/hep.510290340.
  54. Park, S.G.; Kim, Y.; Park, E.; Ryu, H.M.; Jung, G. Fidelity of hepatitis B virus polymerase. *Eur. J. Biochem.* **2003**, *270*, 2929–2936, doi:10.1046/j.1432-1033.2003.03650.x.
  55. Nowak, M.A.; Bonhoeffer, S.; Hill, A.M.; Boehme, R.; Thomas, H.C.; McDade, H. Viral dynamics in hepatitis B virus infection. *Proc. Natl. Acad. Sci.* **1996**, *93*, 4398–

- 4402, doi:10.1073/pnas.93.9.4398.
56. Whalley, S.A.; Murray, J.M.; Brown, D.; Webster, G.J.M.; Emery, V.C.; Dusheiko, G.M.; Perelson, A.S. Kinetics of Acute Hepatitis B Virus Infection in Humans. *J. Exp. Med.* **2001**, *193*, 847–854, doi:10.1084/jem.193.7.847.
  57. Levrero, M.; Zucman-Rossi, J. Mechanisms of HBV-induced hepatocellular carcinoma. *J. Hepatol.* **2016**, *64*, S84–S101, doi:10.1016/j.jhep.2016.02.021.
  58. Slagle, B.L.; Andrisani, O.M.; Bouchard, M.J.; Lee, C.G.L.; Ou, J. - H. J.; Siddiqui, A. Technical standards for hepatitis B virus X protein (HBx) research. *Hepatology* **2015**, *61*, 1416–1424, doi:10.1002/hep.27360.
  59. Stadelmayer, B.; Diederichs, A.; Chapus, F.; Rivoire, M.; Neveu, G.; Alam, A.; Fraisse, 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, doi:10.1016/j.jhep.2020.01.028.
  60. Xie, N.; Chen, X.; Zhang, T.; Liu, B.; Huang, C. Using proteomics to identify the HBx interactome in hepatitis B virus: how can this inform the clinic? *Expert Rev. Proteomics* **2014**, *11*, 59–74, doi:10.1586/14789450.2014.861745.
  61. Nassal, M. HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. *Gut* **2015**, *64*, 1972–84, doi:10.1136/gutjnl-2015-309809.
  62. Niu, C.; Livingston, C.M.; Li, L.; Beran, R.K.; Daffis, S.; Ramakrishnan, D.; Burdette, D.; Peiser, L.; Salas, E.; Ramos, H.; et al. The Smc5/6 complex restricts HBV when localized to ND10 without inducing an innate immune response and is counteracted by the HBV X protein shortly after infection. *PLoS One* **2017**, *12*, 1–32, doi:10.1371/journal.pone.0169648.
  63. Murphy, C.M.; Xu, Y.; Li, F.; Nio, K.; Reszka-Blanco, N.; Li, X.; Wu, Y.; Yu, Y.; Xiong, Y.; Su, L. Hepatitis B Virus X Protein Promotes Degradation of SMC5/6 to Enhance HBV Replication. *Cell Rep.* **2016**, *16*, 2846–2854, doi:10.1016/j.celrep.2016.08.026.
  64. Murphy, C.M.C.; Xu, Y.; Li, F.; Nio, K.; Reszka-Blanco, N.; Li, X.; Wu, Y.; Yu, Y.; Xiong, Y.; Su, L. Hepatitis B Virus X Protein Promotes Degradation of SMC5/6 to Enhance HBV Replication. *Cell Rep.* **2016**, *16*, 2846–2854, doi:10.1016/j.celrep.2016.08.026.
  65. Altinel, K.; Hashimoto, K.; Wei, Y.; Neuveut, C.; Gupta, I.; Suzuki, A.M.; Dos Santos, A.; Moreau, P.; Xia, T.; Kojima, S.; et al. Single-Nucleotide Resolution Mapping of Hepatitis B Virus Promoters in Infected Human Livers and Hepatocellular Carcinoma. *J. Virol.* **2016**, *90*, 10811–10822, doi:10.1128/jvi.01625-16.
  66. Doria, M.; Klein, N.; Lucito, R.; Schneider, R.J. The hepatitis B virus HBx protein is a dual specificity cytoplasmic activator of Ras and nuclear activator of transcription factors. *EMBO J.* **1995**, *14*, 4747–4757, doi:10.1002/j.1460-2075.1995.tb00156.x.
  67. Henkler, F.; Hoare, J.; Waseem, N.; Goldin, R.D.; McGarvey, M.J.; Koshy, R.; King, I.A. Intracellular localization of the hepatitis B virus HBx protein. *J. Gen. Virol.* **2001**, *82*, 871–882, doi:10.1099/0022-1317-82-4-871.
  68. Su, Q.; Schröder, C.H.; Hofmann, W.J.; Otto, G.; Pichlmayr, R.; Bannasch, P. Expression of hepatitis B virus X protein in HBV-infected human livers and hepatocellular carcinomas. *Hepatology* **1998**, *27*, 1109–1120,

doi:10.1002/hep.510270428.

69. Martin-Vilchez, S.; Lara-Pezzi, E.; Trapero-Marugán, M.; Moreno-Otero, R.; Sanz-Cameno, P. The molecular and pathophysiological implications of hepatitis B X antigen in chronic hepatitis B virus infection. *Rev. Med. Virol.* **2011**, *21*, 315–329, doi:10.1002/rmv.699.
70. Cha, M.-Y.; Ryu, D.-K.; Jung, H.-S.; Chang, H.-E.; Ryu, W.-S. Stimulation of hepatitis B virus genome replication by HBx is linked to both nuclear and cytoplasmic HBx expression. *J. Gen. Virol.* **2009**, *90*, 978–986, doi:10.1099/vir.0.009928-0.
71. Klein, N.P.; Bouchard, M.J.; Wang, L.H.; Kobarg, C.; Schneider, R.J. Src kinases involved in hepatitis B virus replication. *EMBO J.* **1999**, *18*, 5019–27, doi:10.1093/emboj/18.18.5019.
72. Tang, H.; Delgermaa, L.; Huang, F.; Oishi, N.; Liu, L.; He, F.; Zhao, L.; Murakami, S. The Transcriptional Transactivation Function of HBx Protein Is Important for Its Augmentation Role in Hepatitis B Virus Replication. *J. Virol.* **2005**, *79*, 5548–5556, doi:10.1128/JVI.79.9.5548-5556.2005.
73. Leupin, O.; Bontron, S.; Schaeffer, C.; Strubin, M. Hepatitis B Virus X Protein Stimulates Viral Genome Replication via a DDB1-Dependent Pathway Distinct from That Leading to Cell Death. *J. Virol.* **2005**, *79*, 4238–4245, doi:10.1128/JVI.79.7.4238-4245.2005.
74. Hwang, K.B.; Kyaw, Y.Y.; Kang, H.R.; Seong, M.S.; Cheong, J. Mitochondrial dysfunction stimulates HBV gene expression through lipogenic transcription factor activation. *Virus Res.* **2020**, *277*, 197842, doi:10.1016/j.virusres.2019.197842.
75. Lucifora, J.; Arzberger, S.; Durantel, D.; Belloni, L.; Strubin, M.; Levrero, M.; Zoulim, F.; Hantz, O.; Protzer, U. Hepatitis B virus X protein is essential to initiate and maintain virus replication after infection. *J. Hepatol.* **2011**, *55*, 996–1003, doi:10.1016/j.jhep.2011.02.015.
76. Chen, H.S.; Kaneko, S.; Girones, R.; Anderson, R.W.; Hornbuckle, W.E.; Tennant, B.C.; Cote, P.J.; Gerin, J.L.; Purcell, R.H.; Miller, R.H. The woodchuck hepatitis virus X gene is important for establishment of virus infection in woodchucks. *J. Virol.* **1993**, *67*, 1218–1226, doi:10.1128/jvi.67.3.1218-1226.1993.
77. Zoulim, F.; Saputelli, J.; Seeger, C. Woodchuck hepatitis virus X protein is required for viral infection in vivo. *J. Virol.* **1994**, *68*, 2026–2030, doi:10.1128/jvi.68.3.2026-2030.1994.
78. Tang, H.; Oishi, N.; Kaneko, S.; Murakami, S. Molecular functions and biological roles of hepatitis B virus x protein. *Cancer Sci.* **2006**, *97*, 977–983, doi:10.1111/j.1349-7006.2006.00299.x.
79. Xu, Z.; Yen, T.S.B.; Wu, L.; Madden, C.R.; Tan, W.; Slagle, B.L.; Ou, J. Enhancement of Hepatitis B Virus Replication by Its X Protein in Transgenic Mice. *J. Virol.* **2002**, *76*, 2579–2584, doi:10.1128/jvi.76.5.2579-2584.2002.
80. Reifenberg, K.; Nusser, P.; Löhler, J.; Spindler, G.; Kuhn, C.; von Weizsäcker, F.; Köck, J. Virus replication and virion export in X-deficient hepatitis B virus transgenic mice. *J. Gen. Virol.* **2002**, *83*, 991–996, doi:10.1099/0022-1317-83-5-991.
81. Melegari, M.; Wolf, S.K.; Schneider, R.J. Hepatitis B Virus DNA Replication Is

- Coordinated by Core Protein Serine Phosphorylation and HBx Expression. *J. Virol.* **2005**, *79*, 9810–9820, doi:10.1128/JVI.79.15.9810-9820.2005.
82. Song, H.; Xu, F.; Xiao, Q.; Tan, G. Hepatitis B virus X protein and its host partners. *Cell. Mol. Immunol.* **2021**, *18*, 1345–1346, doi:10.1038/s41423-021-00674-z.
  83. Kumar, V.; Jayasuryan, N.; Kumar, R. A truncated mutant (residues 58-140) of the hepatitis B virus X protein retains transactivation function. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 5647–52.
  84. Zhang, A.Y.; Lai, C.L.; Poon, R.T.P.; Huang, F.Y.; Seto, W.K.; Fung, J.; Wong, D.K.H.; Yuen, M.F. Hepatitis B virus full-length genomic mutations and quasispecies in hepatocellular carcinoma. *J. Gastroenterol. Hepatol.* **2016**, *31*, 1638–1645, doi:10.1111/jgh.13316.
  85. Jiang, B.; Hildt, E. Intracellular Trafficking of HBV Particles. *Cells* **2020**, *9*, 2023, doi:10.3390/cells9092023.
  86. Königer, C.; Wingert, I.; Marsmann, M.; Rösler, C.; Beck, J.; Nassal, M. Involvement of the host DNA-repair enzyme TDP2 in formation of the covalently closed circular DNA persistence reservoir of hepatitis B viruses. *Proc. Natl. Acad. Sci.* **2014**, *111*, doi:10.1073/pnas.1409986111.
  87. Xia, Y.; Guo, H. Hepatitis B virus cccDNA: Formation, regulation and therapeutic potential. *Antiviral Res.* **2020**, *180*, 104824, doi:10.1016/j.antiviral.2020.104824.
  88. Li, H.-C.; Huang, E.-Y.; Su, P.-Y.; Wu, S.-Y.; Yang, C.-C.; Lin, Y.-S.; Chang, W.-C.; Shih, C. Nuclear Export and Import of Human Hepatitis B Virus Capsid Protein and Particles. *PLoS Pathog.* **2010**, *6*, e1001162, doi:10.1371/journal.ppat.1001162.
  89. Eckhardt, S.G.; Milich, D.R.; McLachlan, A. Hepatitis B virus core antigen has two nuclear localization sequences in the arginine-rich carboxyl terminus. *J. Virol.* **1991**, *65*, 575–582, doi:10.1128/jvi.65.2.575-582.1991.
  90. Peng, B.; Jing, Z.; Zhou, Z.; Sun, Y.; Guo, G.; Tan, Z.; Diao, Y.; Yao, Q.; Ping, Y.; Li, X.; et al. Nonproductive Hepatitis B Virus Covalently Closed Circular DNA Generates HBx-Related Transcripts from the HBx/Enhancer I Region and Acquires Reactivation by Superinfection in Single Cells. *J. Virol.* **2023**, *97*, doi:10.1128/jvi.01717-22.
  91. Nassal, M. Hepatitis B viruses: Reverse transcription a different way. *Virus Res.* **2008**, *134*, 235–249, doi:10.1016/j.virusres.2007.12.024.
  92. Pollicino, T.; Caminiti, G. HBV-integration studies in the clinic: Role in the natural history of infection. *Viruses* **2021**, *13*, 368, doi:10.3390/v13030368.
  93. Heger-Stevic, J.; Zimmermann, P.; Lecoq, L.; Böttcher, B.; Nassal, M. Hepatitis B virus core protein phosphorylation: Identification of the SRPK1 target sites and impact of their occupancy on RNA binding and capsid structure. *PLOS Pathog.* **2018**, *14*, e1007488, doi:10.1371/journal.ppat.1007488.
  94. Zhao, K.; Liu, A.; Xia, Y. Insights into Hepatitis B Virus DNA Integration-55 Years after Virus Discovery. *Innov.* **2020**, *1*, 100034, doi:10.1016/j.xinn.2020.100034.
  95. Tu, T.; Budzinska, M.A.; Vondran, F.W.R.; Shackel, N.A.; Urban, S. Hepatitis B Virus DNA Integration Occurs Early in the Viral Life Cycle in an In Vitro Infection Model via Sodium Taurocholate Cotransporting Polypeptide-Dependent Uptake of Enveloped Virus Particles. *J. Virol.* **2018**, *92*, doi:10.1128/JVI.02007-17.



96. Lampertico, P.; Agarwal, K.; Berg, T.; Buti, M.; Janssen, H.L.A.; Papatheodoridis, G.; Zoulim, F.; Tacke, F. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J. Hepatol.* **2017**, *67*, 370–398, doi:10.1016/j.jhep.2017.03.021.
97. Sung, W.-K.; Zheng, H.; Li, S.; Chen, R.; Liu, X.; Li, Y.; Lee, N.P.; Lee, W.H.; Ariyaratne, P.N.; Tennakoon, C.; et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* **2012**, *44*, 765–769, doi:10.1038/ng.2295.
98. Hu, J.; Liu, K. Complete and Incomplete Hepatitis B Virus Particles: Formation, Function, and Application. *Viruses* **2017**, *9*, 56, doi:10.3390/v9030056.
99. Possehl, C.; Repp, R.; Heermann, K.-H.; Korec, E.; Uy, A.; Gerlich, W.H. Absence of free core antigen in anti-HBc negative viremic hepatitis B carriers. In: 1992; pp. 39–41.
100. Selzer, L.; Zlotnick, A. Assembly and Release of Hepatitis B Virus. *Cold Spring Harb. Perspect. Med.* **2015**, a021394, doi:10.1101/cshperspect.a021394.
101. Tang, H.; McLachlan, A. A Pregenomic RNA Sequence Adjacent to DR1 and Complementary to Epsilon Influences Hepatitis B Virus Replication Efficiency. *Virology* **2002**, *303*, 199–210, doi:10.1006/viro.2002.1645.
102. Oropeza, C.E.; McLachlan, A. Complementarity between epsilon and phi sequences in pregenomic RNA influences hepatitis B virus replication efficiency. *Virology* **2007**, *359*, 371–381, doi:10.1016/j.virol.2006.08.036.
103. Christofi, T.; Zaravinos, A. RNA editing in the forefront of epitranscriptomics and human health. *J. Transl. Med.* **2019**, *17*, 319, doi:10.1186/s12967-019-2071-4.
104. Di Giorgio, S.; Martignano, F.; Torcia, M.G.; Mattiuz, G.; Conticello, S.G. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci. Adv.* **2020**, *6*, eabb5813, doi:10.1126/sciadv.abb5813.
105. Domingo, E.; Sheldon, J.; Perales, C. Viral Quasispecies Evolution. *Microbiol. Mol. Biol. Rev.* **2012**, *76*, 159–216, doi:10.1128/MMBR.05023-11.
106. Gregori, J.; Ibañez-Lligoña, M.; Quer, J. Quantifying In-Host Quasispecies Evolution. *Int. J. Mol. Sci.* **2023**, *24*, 1301, doi:10.3390/ijms24021301.
107. 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, doi:10.1016/j.antiviral.2018.07.019.
108. Hepatitis B Foundation What Is Hepatitis B? Available online: <https://www.hepb.org/what-is-hepatitis-b/what-is-hepb/acute-vs-chronic/#acute> (accessed on Feb 12, 2023).
109. Fanning, G.C.; Zoulim, F.; Hou, J.; Bertoletti, A. Therapeutic strategies for hepatitis B virus infection: towards a cure. *Nat. Rev. Drug Discov.* **2019**, *18*, 827–844, doi:10.1038/s41573-019-0037-0.
110. Kamatani, Y.; Wattanapokayakit, S.; Ochi, H.; Kawaguchi, T.; Takahashi, A.; Hosono, N.; Kubo, M.; Tsunoda, T.; Kamatani, N.; Kumada, H.; et al. A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. *Nat. Genet.* **2009**, *41*, 591–595, doi:10.1038/ng.348.

111. Bourne, E.J.; Dienstag, J.L.; Lopez, V.A.; Sander, T.J.; Longlet, J.M.; Hall, J.G.; Kwiatkowski, R.W.; Wright, T.; Lai, C.L.; Condey, L.D. Quantitative analysis of HBV cccDNA from clinical specimens: correlation with clinical and virological response during antiviral therapy. *J. Viral Hepat.* **2007**, *14*, 55–63, doi:10.1111/j.1365-2893.2006.00775.x.
112. Takkenberg, B.; Terpstra, V.; Zaaier, H.; Weegink, C.; Dijkgraaf, M.; Jansen, P.; Beld, M.; Reesink, H. Intrahepatic response markers in chronic hepatitis B patients treated with peginterferon alpha-2a and adefovir. *J. Gastroenterol. Hepatol.* **2011**, *26*, 1527–1535, doi:10.1111/j.1440-1746.2011.06766.x.
113. Bai, F.; Yano, Y.; Fukumoto, T.; Takebe, A.; Tanaka, M.; Kuramitsu, K.; Anggorowati, N.; Rinonce, H.T.; Widasari, D.I.; Saito, M.; et al. Quantification of Pregenomic RNA and Covalently Closed Circular DNA in Hepatitis B Virus-Related Hepatocellular Carcinoma. *Int. J. Hepatol.* **2013**, *2013*, 1–9, doi:10.1155/2013/849290.
114. Kramvis, A.; Chang, K.-M.; Dandri, M.; Farci, P.; Glebe, D.; Hu, J.; Janssen, H.L.A.; Lau, D.T.Y.; Penicaud, C.; Pollicino, T.; et al. A roadmap for serum biomarkers for hepatitis B virus: current status and future outlook. *Nat. Rev. Gastroenterol. Hepatol.* **2022**, *19*, 727–745, doi:10.1038/s41575-022-00649-z.
115. Wang, J.; Shen, T.; Huang, X.; Kumar, G.R.; Chen, X.; Zeng, Z.; Zhang, R.; Chen, R.; Li, T.; Zhang, T.; et al. 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, doi:10.1016/j.jhep.2016.05.029.
116. Giersch, K.; Allweiss, L.; Volz, T.; Dandri, M.; Lütgehetmann, M. Serum HBV pgRNA as a clinical marker for cccDNA activity. *J. Hepatol.* **2017**, *66*, 460–462, doi:10.1016/j.jhep.2016.09.028.
117. Su, Q.; Wang, S.F.; Chang, T.E.; Breikreutz, R.; Hennig, H.; Takegoshi, K.; Edler, L.; Schröder, C.H. Circulating hepatitis B virus nucleic acids in chronic infection: representation of differently polyadenylated viral transcripts during progression to nonreplicative stages. *Clin. Cancer Res.* **2001**, *7*, 2005–15.
118. van Bömmel, F.; Bartens, A.; Mysickova, A.; Hofmann, J.; Krüger, D.H.; Berg, T.; Edelmann, A.; H.Krüger, D.; 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, doi:10.1002/hep.27381.
119. Wang, J.; Yu, Y.; Li, G.; Shen, C.; Meng, Z.; Zheng, J.J.; Jia, Y.; Chen, S.; Zhang, X.; Zhu, M.; et al. Relationship between serum HBV-RNA levels and intrahepatic viral as well as histologic activity markers in entecavir-treated patients. *J. Hepatol.* **2018**, *68*, 16–24, doi:10.1016/j.jhep.2017.08.021.
120. Yu, X.; Wang, M.; Yu, D.; Chen, P.; Zhu, M.; Huang, W.; Han, Y.; Gong, Q.; Zhang, X. 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*, e00075-20, doi:10.1128/JCM.00075-20.
121. Ji, X.; Xia, M.; Zhou, B.; Liu, S.; Liao, G.C.; 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, doi:10.2147/IDR.S252994.

122. van Bömmel, F.; van Bömmel, A.; Krauel, A.; Wat, C.; Pavlovic, V.; Yang, L.; Deichsel, D.; Berg, T.; Böhm, S. Serum HBV RNA as a Predictor of Peginterferon Alfa-2a Response in Patients With HBeAg-Positive Chronic Hepatitis B. *J. Infect. Dis.* **2018**, *218*, 1066–1074, doi:10.1093/infdis/jiy270.
123. Kaewdech, A.; Tangkijvanich, P.; Sripongpun, P.; Witeerungrot, T.; Jandee, S.; Tanaka, Y.; Piratvisuth, T. Hepatitis B surface antigen, core- related antigen and HBV RNA: Predicting clinical relapse after NA therapy discontinuation. *Liver Int.* **2020**, *40*, 2961–2971, doi:10.1111/liv.14606.
124. Butler, E.K.; Gersch, J.; McNamara, A.; Luk, K.C.; Holzmayer, V.; de Medina, M.; Schiff, E.; Kuhns, M.; Cloherty, G.A. Hepatitis B Virus Serum DNA and RNA Levels in Nucleos(t)ide Analog-Treated or Untreated Patients During Chronic and Acute Infection. *Hepatology* **2018**, *68*, 2106–2117, doi:10.1002/hep.30082.
125. Seto, W.-K.K.; Liu, K.S.; Mak, L.-Y.Y.; Cloherty, G.; Wong, D.K.-H.H.; Gersch, J.; Lam, Y.-F.F.; Cheung, K.-S.S.; Chow, N.; Ko, K.-L.L.; et al. Role of serum HBV RNA and hepatitis B surface antigen levels in identifying Asian patients with chronic hepatitis B suitable for entecavir cessation. *Gut* **2021**, *70*, 775–783, doi:10.1136/gutjnl-2020-321116.
126. Cortese, M.F.; Riveiro-Barciela, M.; Tabernero, D.; Rodriguez-Algarra, F.; Palom, A.; Sopena, S.; Rando-Segura, A.; Roade, L.; Kuchta, A.; Ferrer-Costa, R.; et al. Standardized Hepatitis B Virus RNA Quantification in Untreated and Treated Chronic Patients: a Promising Marker of Infection Follow-Up. *Microbiol. Spectr.* **2022**, *10*, doi:10.1128/spectrum.02149-21.
127. Degasperis, E.; Anolli, M.P.; Lampertico, P. Towards a Functional Cure for Hepatitis B Virus: A 2022 Update on New Antiviral Strategies. *Viruses* **2022**, *14*, 2404, doi:10.3390/v14112404.
128. Testoni, B.; Levrero, M.; Zoulim, F. Challenges to a Cure for HBV Infection. *Semin. Liver Dis.* **2017**, *37*, 231–242, doi:10.1055/s-0037-1606212.
129. Gounder, P.P.; Bulkow, L.R.; Snowball, M.; Negus, S.; Spradling, P.R.; Simons, B.C.; McMahon, B.J. Nested case-control study: hepatocellular carcinoma risk after hepatitis B surface antigen seroclearance. *Aliment. Pharmacol. Ther.* **2016**, *43*, 1197–1207, doi:10.1111/apt.13621.
130. Cheung, K.-S.; Seto, W.-K.; Wong, D.K.-H.; Lai, C.-L.; Yuen, M.-F. Relationship between HBsAg, HBcrAg and hepatocellular carcinoma in patients with undetectable HBV DNA under nucleos(t)ide therapy. *J. Viral Hepat.* **2017**, *24*, 654–661, doi:10.1111/jvh.12688.
131. Kim, G.-A.; Lee, H.C.; Kim, M.-J.; Ha, Y.; Park, E.J.; An, J.; Lee, D.; Shim, J.H.; Kim, K.M.; Lim, Y.-S. Incidence of hepatocellular carcinoma after HBsAg seroclearance in chronic hepatitis B patients: A need for surveillance. *J. Hepatol.* **2015**, *62*, 1092–1099, doi:10.1016/j.jhep.2014.11.031.
132. Papatheodoridis, G. V.; Chan, H.L.-Y.; Hansen, B.E.; Janssen, H.L.A.; Lampertico, P. Risk of hepatocellular carcinoma in chronic hepatitis B: Assessment and modification with current antiviral therapy. *J. Hepatol.* **2015**, *62*, 956–967, doi:10.1016/j.jhep.2015.01.002.
133. Terrault, N.A.; Lok, A.S.F.; McMahon, B.J.; Chang, K.; Hwang, J.P.; Jonas, M.M.; Brown, R.S.; Bzowej, N.H.; Wong, J.B. Update on prevention, diagnosis, and

- treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. *Hepatology* **2018**, *67*, 1560–1599, doi:10.1002/hep.29800.
134. Buti, M.; Tsai, N.; Petersen, J.; Flisiak, R.; Gurel, S.; Krastev, Z.; Aguilar Schall, R.; Flaherty, J.F.; Martins, E.B.; Charuwarn, P.; et al. Seven-Year Efficacy and Safety of Treatment with Tenofovir Disoproxil Fumarate for Chronic Hepatitis B Virus Infection. *Dig. Dis. Sci.* **2015**, *60*, 1457–1464, doi:10.1007/s10620-014-3486-7.
  135. Sheldon, J.; Camino, N.; Rodés, B.; Bartholomeusz, A.; Kuiper, M.; Tacke, F.; Núñez, M.; Mauss, S.; Lutz, T.; Klausen, G.; et al. Selection of hepatitis B virus polymerase mutations in HIV-coinfected patients treated with tenofovir. *Antivir. Ther.* **2005**, *10*, 727–34.
  136. Amini-Bavil-Olyaei, S.; Herbers, U.; Sheldon, J.; Luedde, T.; Trautwein, C.; Tacke, F. The rtA194T polymerase mutation impacts viral replication and susceptibility to tenofovir in hepatitis B e antigen-positive and hepatitis B e antigen-negative hepatitis B virus strains. *Hepatology* **2009**, *49*, 1158–1165, doi:10.1002/hep.22790.
  137. Fung, J.; Lai, C.-L.; Yuen, M.-F. LB80380: a promising new drug for the treatment of chronic hepatitis B. *Expert Opin. Investig. Drugs* **2008**, *17*, 1581–1588, doi:10.1517/13543784.17.10.1581.
  138. Korean Association for the Study of the Liver (KASL) KASL clinical practice guidelines for management of chronic hepatitis B. *Clin. Mol. Hepatol.* **2019**, *25*, 93–159, doi:10.3350/cmh.2019.1002.
  139. Song, D.S.; Kim, W.; Ahn, S.H.; Yim, H.J.; Jang, J.Y.; Kweon, Y.O.; Cho, Y.K.; Kim, Y.J.; Hong, G.Y.; Kim, D.J.; et al. Continuing besifovir dipivoxil maleate versus switching from tenofovir disoproxil fumarate for treatment of chronic hepatitis B: Results of 192-week phase 3 trial. *Clin. Mol. Hepatol.* **2021**, *27*, 346–359, doi:10.3350/cmh.2020.0307.
  140. Kim, J.C.; Lee, H.Y.; Lee, A.R.; Dezhbord, M.; Lee, D.R.; Kim, S.G.S.H.; Won, J.; Park, S.; Kim, N.Y.; Shin, J.J.; et al. Identification and Characterization of Besifovir-Resistant Hepatitis B Virus Isolated from a Chronic Hepatitis B Patient. *Biomedicines* **2022**, *10*, 282, doi:10.3390/biomedicines10020282.
  141. Won, J.; Lee, A.R.; Dezhbord, M.; Lee, D.R.; Kim, S.H.; Kim, J.C.; Park, S.; Kim, N.; Jae, B.; Kim, K.-H. Susceptibility of Drug Resistant Hepatitis B Virus Mutants to Besifovir. *Biomedicines* **2022**, *10*, 1637, doi:10.3390/biomedicines10071637.
  142. Boyd, A.; Lacombe, K.; Lavocat, F.; Maylin, S.; Mialhes, P.; Lascoux-Combe, C.; Delaugerre, C.; Girard, P.-M.; Zoulim, F. Decay of ccc-DNA marks persistence of intrahepatic viral DNA synthesis under tenofovir in HIV-HBV co-infected patients. *J. Hepatol.* **2016**, *65*, 683–691, doi:10.1016/j.jhep.2016.05.014.
  143. Terrault, N.A.; Bzowej, N.H.; Chang, K.M.; Hwang, J.P.; Jonas, M.M.; Murad, M.H. AASLD guidelines for treatment of chronic hepatitis B. *Hepatology* **2016**, *63*, 261–283, doi:10.1002/hep.28156.
  144. Wu, J.; Mao, W. Review of Serum Biomarkers and Models Derived from Them in HBV-Related Liver Diseases. *Dis. Markers* **2020**, *2020*, 1–12, doi:10.1155/2020/2471252.
  145. Parikh, P.; Ryan, J.D.; Tsochatzis, E.A. Fibrosis assessment in patients with chronic hepatitis B virus (HBV) infection. *Ann. Transl. Med.* **2017**, *5*, 40–40,

doi:10.21037/atm.2017.01.28.

146. Liu, F.; Wang, X.-W.; Chen, L.; Hu, P.; Ren, H.; Hu, H.-D. Systematic review with meta-analysis: development of hepatocellular carcinoma in chronic hepatitis B patients with hepatitis B surface antigen seroclearance. *Aliment. Pharmacol. Ther.* **2016**, *43*, 1253–1261, doi:10.1111/apt.13634.
147. Coffin, C.S.; Zhou, K.; Terrault, N.A. New and Old Biomarkers for Diagnosis and Management of Chronic Hepatitis B Virus Infection. *Gastroenterology* **2019**, *156*, 355-368.e3, doi:10.1053/j.gastro.2018.11.037.
148. Testoni, B.; Lebossé, F.; Scholtes, C.; Berby, F.; Miaglia, C.; Subic, M.; Loglio, A.; Facchetti, F.; Lampertico, P.; Levrero, M.; et al. Serum hepatitis B core-related antigen (HBcrAg) correlates with covalently closed circular DNA transcriptional activity in chronic hepatitis B patients. *J. Hepatol.* **2019**, *70*, 615–625, doi:10.1016/j.jhep.2018.11.030.
149. Inoue, T.; Tanaka, Y. The Role of Hepatitis B Core-Related Antigen. *Genes (Basel)*. **2019**, *10*, 357, doi:10.3390/genes10050357.
150. Riveiro-Barciela, M.; Bes, M.; Rodríguez-Frías, F.; Tabernero, D.; Ruiz, A.; Casillas, R.; Vidal-González, J.; Homs, M.; Nieto, L.; Sauleda, S.; et al. 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, doi:10.1016/j.cmi.2017.03.003.
151. Adraneda, C.; Tan, Y.C.; Yeo, E.J.; Kew, G. Sen; Khakpoor, A.; Lim, S.G. A critique and systematic review of the clinical utility of hepatitis B core-related antigen. *J. Hepatol.* **2023**, *78*, 731–741, doi:10.1016/j.jhep.2022.12.017.
152. Tada, T.; Kumada, T.; Toyoda, H.; Kiriya, S.; Tanikawa, M.; Hisanaga, Y.; Kanamori, A.; Kitabatake, S.; Yama, T.; Tanaka, J. HBcrAg predicts hepatocellular carcinoma development: An analysis using time-dependent receiver operating characteristics. *J. Hepatol.* **2016**, *65*, 48–56, doi:10.1016/j.jhep.2016.03.013.
153. Kimura, T.; Ohno, N.; Terada, N.; Rokuhara, A.; Matsumoto, A.; Yagi, S.; Tanaka, E.; Kiyosawa, K.; Ohno, S.; Maki, N. Hepatitis B Virus DNA-negative Dane Particles Lack Core Protein but Contain a 22-kDa Precore Protein without C-terminal Arginine-rich Domain. *J. Biol. Chem.* **2005**, *280*, 21713–21719, doi:10.1074/jbc.M501564200.
154. Gural, N.; Mancio-Silva, L.; He, J.; Bhatia, S.N. Engineered Livers for Infectious Diseases. *Cell. Mol. Gastroenterol. Hepatol.* **2018**, *5*, 131–144, doi:10.1016/j.jcmgh.2017.11.005.
155. Verrier, E.; Colpitts, C.; Schuster, C.; Zeisel, M.; Baumert, T. Cell Culture Models for the Investigation of Hepatitis B and D Virus Infection. *Viruses* **2016**, *8*, 261, doi:10.3390/v8090261.
156. Lempp, F.A.; Qu, B.; Wang, Y.-X.; Urban, S. Hepatitis B Virus Infection of a Mouse Hepatic Cell Line Reconstituted with Human Sodium Taurocholate Cotransporting Polypeptide. *J. Virol.* **2016**, *90*, 4827–4831, doi:10.1128/JVI.02832-15.
157. Gripon, P.; Rumin, S.; Urban, S.; Le Seyec, J.; Glaise, D.; Cannie, I.; Guyomard, C.; Lucas, J.; Trepo, C.; Guguen-Guillouzo, C. Infection of a human hepatoma cell line by hepatitis B virus. *Proc. Natl. Acad. Sci.* **2002**, *99*, 15655–15660, doi:10.1073/pnas.232137699.

158. Zhou, M.; Zhao, K.; Yao, Y.; Yuan, Y.; Pei, R.; Wang, Y.; Chen, J.; Hu, X.; Zhou, Y.; Chen, X.; et al. Productive HBV infection of well-differentiated, hNTCP-expressing human hepatoma-derived (Huh7) cells. *Viol. Sin.* **2017**, *32*, 465–475, doi:10.1007/s12250-017-3983-x.
159. Ni, Y.; Lempp, F.A.; Mehrle, S.; Nkongolo, S.; Kaufman, C.; Fälth, M.; Stindt, J.; Königer, C.; Nassal, M.; Kubitz, R.; et al. Hepatitis B and D Viruses Exploit Sodium Taurocholate Co-transporting Polypeptide for Species-Specific Entry into Hepatocytes. *Gastroenterology* **2014**, *146*, 1070-1083.e6, doi:10.1053/j.gastro.2013.12.024.
160. Sumida, K.; Igarashi, Y.; Toritsuka, N.; Matsushita, T.; Abe-Tomizawa, K.; Aoki, M.; Urushidani, T.; Yamada, H.; Ohno, Y. Effects of DMSO on gene expression in human and rat hepatocytes. *Hum. Exp. Toxicol.* **2011**, *30*, 1701–1709, doi:10.1177/0960327111399325.
161. Sells, M.A.; Chen, M.L.; Acs, G. Production of hepatitis B virus particles in Hep G2 cells transfected with cloned hepatitis B virus DNA. *Proc. Natl. Acad. Sci.* **1987**, *84*, 1005–1009, doi:10.1073/pnas.84.4.1005.
162. van Bömmel, F.; Berg, T. Risks and Benefits of Discontinuation of Nucleos(t)ide Analogue Treatment: A Treatment Concept for Patients With HBeAg-Negative Chronic Hepatitis B. *Hepatology Commun.* **2021**, *5*, 1632–1648, doi:10.1002/hep4.1708.
163. World Health Organization Combating hepatitis B and C to reach elimination by 2030. *World Heal. Organ.* 2016, 1–16.
164. Allweiss, L.; Volz, T.; Giersch, K.; Kah, J.; Raffa, G.; Petersen, J.; Lohse, A.W.; Beninati, C.; Pollicino, T.; Urban, S.; et al. Proliferation of primary human hepatocytes and prevention of hepatitis B virus reinfection efficiently deplete nuclear cccDNA in vivo. *Gut* **2018**, *67*, 542–552, doi:10.1136/gutjnl-2016-312162.
165. Levrero, M.; Pollicino, T.; Petersen, J.; Belloni, L.; Raimondo, G.; Dandri, M. Control of cccDNA function in hepatitis B virus infection. *J. Hepatology*. **2009**, *51*, 581–592, doi:10.1016/j.jhep.2009.05.022.
166. Martinez, M.G.; Smekalova, E.; Combe, E.; Gregoire, F.; Zoulim, F.; Testoni, B. Gene Editing Technologies to Target HBV cccDNA. *Viruses* **2022**, *14*, 2654, doi:10.3390/v14122654.
167. Vaillant, A. Oligonucleotide-Based Therapies for Chronic HBV Infection: A Primer on Biochemistry, Mechanisms and Antiviral Effects. *Viruses* **2022**, *14*, 2052, doi:10.3390/v14092052.
168. Gebbing, M. Gene therapeutic approaches to inhibit hepatitis B virus replication. *World J. Hepatology*. **2014**, *7*, 150, doi:10.4254/wjh.v7.i2.150.
169. Tarn, W.-Y.; Cheng, Y.; Ko, S.-H.; Huang, L.-M. Antisense Oligonucleotide-Based Therapy of Viral Infections. *Pharmaceutics* **2021**, *13*, 2015, doi:10.3390/pharmaceutics13122015.
170. Zhang, Y. RNA-induced Silencing Complex (RISC). In *Encyclopedia of Systems Biology*; Springer New York: New York, NY, 2013; pp. 1876–1876.
171. Yuen, M.-F.; Locarnini, S.; Lim, T.H.; Strasser, S.I.; Sievert, W.; Cheng, W.;

- Thompson, A.J.; Given, B.D.; Schlupe, T.; Hamilton, J.; et al. Combination treatments including the small-interfering RNA JNJ-3989 induce rapid and sometimes prolonged viral responses in patients with CHB. *J. Hepatol.* **2022**, *77*, 1287–1298, doi:10.1016/j.jhep.2022.07.010.
172. Yuen, M.; Schiefke, I.; Yoon, J.; Ahn, S.H.; Heo, J.; Kim, J.H.; Lik Yuen Chan, H.; Yoon, K.T.; Klinker, H.; Manns, M.; et al. RNA Interference Therapy With ARC-520 Results in Prolonged Hepatitis B Surface Antigen Response in Patients With Chronic Hepatitis B Infection. *Hepatology* **2020**, *72*, 19–31, doi:10.1002/hep.31008.
  173. Liang, X.-H.; Sun, H.; Nichols, J.G.; Crooke, S.T. RNase H1-Dependent Antisense Oligonucleotides Are Robustly Active in Directing RNA Cleavage in Both the Cytoplasm and the Nucleus. *Mol. Ther.* **2017**, *25*, 2075–2092, doi:10.1016/j.ymthe.2017.06.002.
  174. Yuen, M.-F.; Lim, S.-G.; Plesniak, R.; Tsuji, K.; Janssen, H.L.A.; Pojoga, C.; Gadano, A.; Popescu, C.P.; Stepanova, T.; Asselah, T.; et al. Efficacy and Safety of Bepirovirsen in Chronic Hepatitis B Infection. *N. Engl. J. Med.* **2022**, *387*, 1957–1968, doi:10.1056/NEJMoa2210027.
  175. Kupryushkin, M.S.; Filatov, A. V.; Mironova, N.L.; Patutina, O.A.; Chernikov, I. V.; Chernolovskaya, E.L.; Zenkova, M.A.; Pyshnyi, D. V.; Stetsenko, D.A.; Altman, S.; et al. Antisense oligonucleotide gapmers containing phosphoryl guanidine groups reverse MDR1-mediated multiple drug resistance of tumor cells. *Mol. Ther. - Nucleic Acids* **2022**, *27*, 211–226, doi:10.1016/j.omtn.2021.11.025.
  176. Debacker, A.J.; Voutila, J.; Catley, M.; Blakey, D.; Habib, N. Delivery of Oligonucleotides to the Liver with GalNAc: From Research to Registered Therapeutic Drug. *Mol. Ther.* **2020**, *28*, 1759–1771, doi:10.1016/j.ymthe.2020.06.015.
  177. Thomas, G.S.; Cromwell, W.C.; Ali, S.; Chin, W.; Flaim, J.D.; Davidson, M. Mipomersen, an Apolipoprotein B Synthesis Inhibitor, Reduces Atherogenic Lipoproteins in Patients With Severe Hypercholesterolemia at High Cardiovascular Risk. *J. Am. Coll. Cardiol.* **2013**, *62*, 2178–2184, doi:10.1016/j.jacc.2013.07.081.
  178. Benson, M.D.; Waddington-Cruz, M.; Berk, J.L.; Polydefkis, M.; Dyck, P.J.; Wang, A.K.; Planté-Bordeneuve, V.; Barroso, F.A.; Merlini, G.; Obici, L.; et al. Inotersen Treatment for Patients with Hereditary Transthyretin Amyloidosis. *N. Engl. J. Med.* **2018**, *379*, 22–31, doi:10.1056/NEJMoa1716793.
  179. Witztum, J.L.; Gaudet, D.; Freedman, S.D.; Alexander, V.J.; Digenio, A.; Williams, K.R.; Yang, Q.; Hughes, S.G.; Geary, R.S.; Arca, M.; et al. Volanesorsen and Triglyceride Levels in Familial Chylomicronemia Syndrome. *N. Engl. J. Med.* **2019**, *381*, 531–542, doi:10.1056/NEJMoa1715944.
  180. Pervez, M.T.; Hasnain, M.J. ul; Abbas, S.H.; Moustafa, M.F.; Aslam, N.; Shah, S.S.M. A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. *Biomed Res. Int.* **2022**, *2022*, 1–12, doi:10.1155/2022/3457806.
  181. Slatko, B.E.; Gardner, A.F.; Ausubel, F.M. Overview of Next-Generation Sequencing Technologies. *Curr. Protoc. Mol. Biol.* **2018**, *122*, doi:10.1002/cpmb.59.
  182. Illumina Inc. Illumina Sequencing Overview Available online: [https://www.ogc.ox.ac.uk/wp-content/uploads/2017/09/Illumina\\_Sequencing\\_Overview\\_15045845\\_D.pdf](https://www.ogc.ox.ac.uk/wp-content/uploads/2017/09/Illumina_Sequencing_Overview_15045845_D.pdf) (accessed on Jan 21, 2023).

183. Bentley, D.R.; Balasubramanian, S.; Swerdlow, H.P.; Smith, G.P.; Milton, J.; Brown, C.G.; Hall, K.P.; Evers, D.J.; Barnes, C.L.; Bignell, H.R.; et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **2008**, *456*, 53–59, doi:10.1038/nature07517.
184. Illumina Inc. Optimizing Cluster Density on Illumina Sequencing Systems. Understanding cluster density limitations and strategies for preventing under- and overclustering Available online: <https://www.illumina.com/content/dam/illumina-marketing/documents/products/other/miseq-overclustering-primer-770-2014-038.pdf> (accessed on Jan 22, 2023).
185. World Health Organization (WHO) *Guidelines for the prevention, care and treatment of persons with chronic hepatitis B infection*; World Health Organization (WHO), Ed.; 2015; ISBN 978-92-4-154905-9.
186. Pauly, M.D.; Kamili, S.; Hayden, T.M. Impact of nucleic acid extraction platforms on hepatitis virus genome detection. *J. Virol. Methods* **2019**, *273*, 113715, doi:10.1016/j.jviromet.2019.113715.
187. ROCHE S.A. MagNA Pure 24 Total NA Isolation Kit. **2021**, 28.
188. QIAGEN QIAamp DNA Mini and Blood Mini Handbook. *Qiagen* **2016**, 1–72.
189. Durantel, D.; Carrouée-Durantel, S.; Werle-Lapostolle, B.; Brunelle, M.-N.; 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, doi:10.1002/hep.20388.
190. Maasoumy, B.; Geretti, A.M.; Frontzek, A.; Austin, H.; Aretzweiler, G.; Garcia-Álvarez, M.; Leuchter, S.; Simon, C.O.; Marins, E.G.; Canchola, J.A.; et al. HBV-RNA Co-amplification May Influence HBV DNA Viral Load Determination. *Hepatol. Commun.* **2020**, *4*, 983–997, doi:10.1002/hep4.1520.
191. Scholtès, C.; Hamilton, A.T.; Plissonnier, M.-L.; Charre, C.; Scott, B.; Wang, L.; Berby, F.; French, J.; Testoni, B.; Blair, A.; et al. Performance of the cobas® HBV RNA automated investigational assay for the detection and quantification of circulating HBV RNA in chronic HBV patients. *J. Clin. Virol.* **2022**, *150–151*, 105150, doi:10.1016/j.jcv.2022.105150.
192. Invitrogen TURBO DNA-free kit. *Computer (Long Beach, Calif.)* **2020**, 169–232.
193. Godoy, C.; Taberner, D.; Sopena, S.; Gregori, J.; Cortese, M.F.; González, C.; Casillas, R.; Yll, M.; Rando, A.; López-Martínez, R.; et al. Characterization of hepatitis B virus X gene quasispecies complexity in mono-infection and hepatitis delta virus superinfection. *World J. Gastroenterol.* **2019**, *25*, 1566–1579, doi:10.3748/wjg.v25.i13.1566.
194. Sonagra, A.; Dholariya, S. Electrophoresis [Updated 2022 Aug 8] Available online: <https://www.ncbi.nlm.nih.gov/books/NBK585057/> (accessed on Jan 8, 2023).
195. Illumina Inc. How short inserts affect sequencing performance Available online: <https://support.illumina.com/bulletins/2020/12/how-short-inserts-affect-sequencing-performance.html> (accessed on Jan 22, 2023).
196. Quail, M.A.; Swerdlow, H.; Turner, D.J. Improved Protocols for the Illumina Genome Analyzer Sequencing System. *Curr. Protoc. Hum. Genet.* **2009**, *62*,



doi:10.1002/0471142905.hg1802s62.

197. Illumina Inc. PhiX Control v3 Available online: <https://www.illumina.com/products/by-type/sequencing-kits/cluster-gen-sequencing-reagents/phix-control-v3.html> (accessed on Jan 20, 2023).
198. Gregori, J.; Esteban, J.I.; Cubero, M.; Garcia-Cehic, D.; Perales, C.; Casillas, R.; Alvarez-Tejado, M.; Rodríguez-Frías, F.; Guardia, J.; Domingo, E.; et al. Ultra-Deep Pyrosequencing (UDPS) Data Treatment to Study Amplicon HCV Minor Variants. *PLoS One* **2013**, *8*, e83361, doi:10.1371/journal.pone.0083361.
199. Illumina Inc. FASTQ files explained Available online: <https://emea.support.illumina.com/bulletins/2016/04/fastq-files-explained.html> (accessed on Jan 22, 2023).
200. Soria, M.E.; Gregori, J.; Chen, Q.; García-Cehic, D.; Llorens, M.; de Ávila, A.I.; Beach, N.M.; Domingo, E.; Rodríguez-Frías, F.; Buti, M.; et al. Pipeline for specific subtype amplification and drug resistance detection in hepatitis C virus. *BMC Infect. Dis.* **2018**, *18*, 446, doi:10.1186/s12879-018-3356-6.
201. R Development Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Viena, Austria. **2020**.
202. Pages, H.; Aboyoun, P.; Gentleman, R.; DebRoy, S. Memory efficient string containers, string matching algorithms, and other utilities, for fast manipulation of large biological sequences or sets of sequences. [*Internet*] **2022**, R package, doi:10.18129/B9.bioc.Biostrings.
203. Morgan, M.; Anders, S.; Lawrence, M.; Aboyoun, P.; Pagès, H.; Gentleman, R. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* **2009**, *25*, 2607–2608, doi:10.1093/bioinformatics/btp450.
204. Paradis, E.; Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2019**, *35*, 526–528, doi:10.1093/bioinformatics/bty633.
205. Magoc, T.; Salzberg, S.L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **2011**, *27*, 2957–2963, doi:10.1093/bioinformatics/btr507.
206. Ewing, B.; Hillier, L.; Wendl, M.C.; Green, P. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Res.* **1998**, *8*, 175–185, doi:10.1101/gr.8.3.175.
207. Hall, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids Symp. Ser* **199AD**, *41*, 95–98.
208. 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, doi:10.1007/BF01731581.
209. Schneider, T.D. Information Content of Individual Genetic Sequences. *J. Theor. Biol.* **1997**, *189*, 427–441, doi:10.1006/jtbi.1997.0540.
210. Ou, J.; Wolfe, S.A.; Brodsky, M.H.; Zhu, L.J. motifStack for the analysis of transcription factor binding site evolution. *Nat. Methods* **2018**, *15*, 8–9,

doi:10.1038/nmeth.4555.

211. Gregori, J.; Perales, C.; Rodriguez-Frias, F.; Esteban, J.I.; Quer, J.; Domingo, E. Viral quasispecies complexity measures. *Virology* **2016**, *493*, 227–237, doi:10.1016/j.virol.2016.03.017.
212. Gregori, J.; Soria, M.E.; Gallego, I.; Guerrero-Murillo, M.; Esteban, J.I.; Quer, J.; Perales, C.; Domingo, E. Rare haplotype load as marker for lethal mutagenesis. *PLoS One* **2018**, *13*, 1–17, doi:10.1371/journal.pone.0204877.
213. Hill, M.O. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* **1973**, *54*, 427–432, doi:10.2307/1934352.
214. Chao, A.; Chiu, C.-H.; Jost, L. Phylogenetic diversity measures based on Hill numbers. *Philos. Trans. R. Soc. B Biol. Sci.* **2010**, *365*, 3599–3609, doi:10.1098/rstb.2010.0272.
215. Chao, A.; Chiu, C.-H.; Jost, L. Unifying Species Diversity, Phylogenetic Diversity, Functional Diversity, and Related Similarity and Differentiation Measures Through Hill Numbers. *Annu. Rev. Ecol. Evol. Syst.* **2014**, *45*, 297–324, doi:10.1146/annurev-ecolsys-120213-091540.
216. Gotelli, N.J.; Chao, A. Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data. In *Encyclopedia of Biodiversity*; Elsevier, 2013; pp. 195–211.
217. Gregori, J.; Colomer-Castell, S.; Campos, C.; Ibañez-Lligoña, M.; Garcia-Cehic, D.; Rando-Segura, A.; Adombie, C.M.; Pintó, R.; Guix, S.; Bosch, A.; et al. Quasispecies Fitness Partition to Characterize the Molecular Status of a Viral Population. Negative Effect of Early Ribavirin Discontinuation in a Chronically Infected HEV Patient. *Int. J. Mol. Sci.* **2022**, *23*, doi:10.3390/ijms232314654.
218. Li, W.H.; Sadler, L.A. Low nucleotide diversity in man. *Genetics* **1991**, *129*, 513–523, doi:10.1093/genetics/129.2.513.
219. Gregori, J.; Salicrú, M.; Domingo, E.; Sanchez, A.; Esteban, J.I.; Rodríguez-Frías, F.; Quer, J. Inference with viral quasispecies diversity indices: clonal and NGS approaches. *Bioinformatics* **2014**, *30*, 1104–1111, doi:10.1093/bioinformatics/btt768.
220. Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press, 1987; ISBN 9780231886710.
221. SPEARMAN, C. THE METHOD OF 'RIGHT AND WRONG CASES' ('CONSTANT STIMULI') WITHOUT GAUSS'S FORMULAE. *Br. J. Psychol. 1904-1920* **1908**, *2*, 227–242, doi:10.1111/j.2044-8295.1908.tb00176.x.
222. Reissinger, A.; Volkers, P.; Scheiblaue, H.; Nick, S. Paul-Ehrlich-Institut.
223. Lavanchy, D.; Kane, M. Global Epidemiology of Hepatitis B Virus Infection. In; 2016; pp. 187–203.
224. Zoulim, F.; Durantel, D. Antiviral Therapies and Prospects for a Cure of Chronic Hepatitis B. *Cold Spring Harb. Perspect. Med.* **2015**, *5*, a021501–a021501, doi:10.1101/cshperspect.a021501.
225. Lok, A.S.; Zoulim, F.; Dusheiko, G.; Ghany, M.G. Hepatitis B cure: From discovery to regulatory approval. *J. Hepatol.* **2017**, *67*, 847–861,

doi:10.1016/j.jhep.2017.05.008.

226. Martinez, M.G.; Testoni, B.; Zoulim, F. Biological basis for functional cure of chronic hepatitis B. *J. Viral Hepat.* **2019**, *26*, 786–794, doi:10.1111/jvh.13090.
227. Flisiak, R.; Jaroszewicz, J.; Łucejko, M. siRNA drug development against hepatitis B virus infection. *Expert Opin. Biol. Ther.* **2018**, *18*, 609–617, doi:10.1080/14712598.2018.1472231.
228. Durantel, D. New treatments to reach functional cure: Virological approaches. *Best Pract. Res. Clin. Gastroenterol.* **2017**, *31*, 329–336, doi:10.1016/j.bpg.2017.05.002.
229. Jansen, L.; Kootstra, N.A.; Van Dort, K.A.; Takkenberg, R.B.; Reesink, H.W.; Zaaijer, H.L. 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, doi:10.1093/infdis/jiv397.
230. Palom, A.; Sopena, S.; Riveiro-Barciela, M.; Carvalho-Gomes, A.; Madejón, A.; Rodríguez-Tajes, S.; Roade, L.; García-Eliz, M.; García-Samaniego, J.; Lens, S.; et al. One-quarter of chronic hepatitis D patients reach HDV-RNA decline or undetectability during the natural course of the disease. *Aliment. Pharmacol. Ther.* **2021**, *54*, 462–469, doi:10.1111/apt.16485.
231. Hashemipetroudi, S.H.; Nematzadeh, G.; Ahmadian, G.; Yamchi, A.; Kuhlmann, M. Assessment of DNA Contamination in RNA Samples Based on Ribosomal DNA. *J. Vis. Exp.* **2018**, doi:10.3791/55451.
232. Duriez, M.; Mandouri, Y.; Lekbaby, B.; Wang, H.; Schnuriger, A.; Redelsperger, F.; Guerrero, C.I.; Lefevre, M.; Fauveau, V.; Ahodantin, J.; et al. Alternative splicing of hepatitis B virus: A novel virus/host interaction altering liver immunity. *J. Hepatol.* **2017**, *67*, 687–699, doi:10.1016/j.jhep.2017.05.025.
233. Ding, F.; Miao, X.-L.; Li, Y.-X.; Dai, J.-F.; Yu, H.-G. Mutations in the S gene and in the overlapping reverse transcriptase region in chronic hepatitis B Chinese patients with coexistence of HBsAg and anti-HBs. *Brazilian J. Infect. Dis.* **2016**, *20*, 1–7, doi:10.1016/j.bjid.2015.08.014.
234. Acevedo, A.; Brodsky, L.; Andino, R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **2014**, *505*, 686–690, doi:10.1038/nature12861.
235. Gout, J.-F.; Thomas, W.K.; Smith, Z.; Okamoto, K.; Lynch, M. Large-scale detection of in vivo transcription errors. *Proc. Natl. Acad. Sci.* **2013**, *110*, 18584–18589, doi:10.1073/pnas.1309843110.
236. Gout, J.-F.; Li, W.; Fritsch, C.; Li, A.; Haroon, S.; Singh, L.; Hua, D.; Fazelinia, H.; Smith, Z.; Seeholzer, S.; et al. The landscape of transcription errors in eukaryotic cells. *Sci. Adv.* **2017**, *3*, doi:10.1126/sciadv.1701484.
237. Elizalde, M.M.; Tadey, L.; Mammana, L.; Quarleri, J.F.; Campos, R.H.; Flichman, D.M. Biological Characterization of Hepatitis B virus Genotypes: Their Role in Viral Replication and Antigen Expression. *Front. Microbiol.* **2021**, *12*, doi:10.3389/fmicb.2021.758613.
238. Sozzi, V.; Walsh, R.; Littlejohn, M.; Colledge, D.; Jackson, K.; Warner, N.; Yuen, L.; Locarnini, S.A.; Revill, P.A. In Vitro Studies Show that Sequence Variability

- Contributes to Marked Variation in Hepatitis B Virus Replication, Protein Expression, and Function Observed across Genotypes. *J. Virol.* **2016**, *90*, 10054–10064, doi:10.1128/JVI.01293-16.
239. Cento, V.; Mirabelli, C.; Dimonte, S.; Salpini, R.; Han, Y.; Trimoulet, P.; Bertoli, A.; Micheli, V.; Gubertini, G.; Cappiello, G.; et al. Overlapping structure of hepatitis B virus (HBV) genome and immune selection pressure are critical forces modulating HBV evolution. *J. Gen. Virol.* **2013**, *94*, 143–149, doi:10.1099/vir.0.046524-0.
  240. Yu, X.; Wang, M.; Yu, D.; Chen, P.; Zhu, M.; Huang, W.; Han, Y.; Gong, Q.; Zhang, X. 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*, e00075-20, doi:10.1128/JCM.00075-20.
  241. Papatheodoridis, G.; Vlachogiannakos, I.; Cholongitas, E.; Wursthorn, K.; Thomadakis, C.; Touloumi, G.; Petersen, J. Discontinuation of oral antivirals in chronic hepatitis B: A systematic review. *Hepatology* **2016**, *63*, 1481–1492, doi:10.1002/hep.28438.
  242. Kao, J.-H.; Jeng, W.-J.; Ning, Q.; Su, T.-H.; Tseng, T.-C.; Ueno, Y.; Yuen, M.-F. APASL guidance on stopping nucleos(t)ide analogues in chronic hepatitis B patients. *Hepatol. Int.* **2021**, *15*, 833–851, doi:10.1007/s12072-021-10223-5.
  243. Wooddell, C.I.; Yuen, M.-F.; Chan, H.L.-Y.; Gish, R.G.; Locarnini, S.A.; Chavez, D.; Ferrari, C.; Given, B.D.; Hamilton, J.; Kanner, S.B.; et al. RNAi-based treatment of chronically infected patients and chimpanzees reveals that integrated hepatitis B virus DNA is a source of HBsAg. *Sci. Transl. Med.* **2017**, *9*, doi:10.1126/scitranslmed.aan0241.
  244. Yuen, M.-F.; Locarnini, S.; Lim, T.H.; STRASSER, S.; Sievert, W.; Cheng, W.; Thompson, A.; Given, B.; Schlupe, T.; Hamilton, J.; et al. PS-080-Short term RNA interference therapy in chronic hepatitis B using JNJ-3989 brings majority of patients to HBsAg < 100 IU/ml threshold. *J. Hepatol.* **2019**, *70*, e51–e52, doi:10.1016/S0618-8278(19)30092-1.
  245. Lee, A.C.H.; Thi, E.P.; Cuconati, A.; Ardzinski, A.; Holland, R.; Huang, H.; Kondratowicz, A.S.; Kowalski, R.; Palmer, L.; Pasetka, C.; et al. FRI-184-Function and drug combination studies in cell culture models for AB-729, a subcutaneously administered siRNA investigational agent for chronic hepatitis B infection. *J. Hepatol.* **2019**, *70*, e471, doi:10.1016/S0618-8278(19)30929-6.
  246. Verma, N.K.; Fazil, M.H.U.T.; Duggan, S.P.; Kelleher, D. Combination Therapy Using Inhalable GapmeR and Recombinant ACE2 for COVID-19. *Front. Mol. Biosci.* **2020**, *7*, doi:10.3389/fmolb.2020.00197.
  247. Qi, Y.; Gao, Z.; Xu, G.; Peng, B.; Liu, C.; Yan, H.; Yao, Q.; Sun, G.; Liu, Y.; Tang, D.; et al. DNA Polymerase  $\kappa$  Is a Key Cellular Factor for the Formation of Covalently Closed Circular DNA of Hepatitis B Virus. *PLOS Pathog.* **2016**, *12*, e1005893, doi:10.1371/journal.ppat.1005893.
  248. Carey, I.; Harrison, P.M. Monotherapy versus combination therapy for the treatment of chronic hepatitis B. *Expert Opin. Investig. Drugs* **2009**, *18*, 1655–1666, doi:10.1517/13543780903241599.
  249. Wang, Y.; Lau, S.H.; Sham, J.S.-T.; Wu, M.-C.; Wang, T.; Guan, X.-Y.

- Characterization of HBV integrants in 14 hepatocellular carcinomas: association of truncated X gene and hepatocellular carcinogenesis. *Oncogene* **2004**, *23*, 142–148, doi:10.1038/sj.onc.1206889.
250. Centeno, E.G.Z.; Cimarosti, H.; Bithell, A. 2D versus 3D human induced pluripotent stem cell-derived cultures for neurodegenerative disease modelling. *Mol. Neurodegener.* **2018**, *13*, 27, doi:10.1186/s13024-018-0258-4.
  251. Roca Suarez, A.A.; Testoni, B.; Zoulim, F. HBV 2021: New therapeutic strategies against an old foe. *Liver Int.* **2021**, *41*, 15–23, doi:10.1111/liv.14851.
  252. Sobczak, M.; Kędra, K. Biomedical Polyurethanes for Anti-Cancer Drug Delivery Systems: A Brief, Comprehensive Review. *Int. J. Mol. Sci.* **2022**, *23*, 8181, doi:10.3390/ijms23158181.
  253. Pathak, N.; Singh, P.; Singh, P.K.; Sharma, S.; Singh, R.P.; Gupta, A.; Mishra, R.; Mishra, V.K.; Tripathi, M. Biopolymeric nanoparticles based effective delivery of bioactive compounds toward the sustainable development of anticancerous therapeutics. *Front. Nutr.* **2022**, *9*, doi:10.3389/fnut.2022.963413.
  254. Tran, S.; DeGiovanni, P.; Piel, B.; Rai, P. Cancer nanomedicine: a review of recent success in drug delivery. *Clin. Transl. Med.* **2017**, *6*, doi:10.1186/s40169-017-0175-0.
  255. Li, Q.; Li, X.; Zhao, C. Strategies to Obtain Encapsulation and Controlled Release of Small Hydrophilic Molecules. *Front. Bioeng. Biotechnol.* **2020**, *8*, doi:10.3389/fbioe.2020.00437.
  256. Suk, J.S.; Xu, Q.; Kim, N.; Hanes, J.; Ensign, L.M. PEGylation as a strategy for improving nanoparticle-based drug and gene delivery. *Adv. Drug Deliv. Rev.* **2016**, *99*, 28–51, doi:10.1016/j.addr.2015.09.012.
  257. Akinc, A.; Querbes, W.; De, S.; Qin, J.; Frank-Kamenetsky, M.; Jayaprakash, K.N.; Jayaraman, M.; Rajeev, K.G.; Cantley, W.L.; Dorkin, J.R.; et al. Targeted Delivery of RNAi Therapeutics With Endogenous and Exogenous Ligand-Based Mechanisms. *Mol. Ther.* **2010**, *18*, 1357–1364, doi:10.1038/mt.2010.85.
  258. Bisgaier, C.L.; Siebenkas, M. V; Williams, K.J. Effects of apolipoproteins A-IV and A-I on the uptake of phospholipid liposomes by hepatocytes. *J. Biol. Chem.* **1989**, *264*, 862–6.
  259. Suhr, O.B.; Coelho, T.; Buades, J.; Pouget, J.; Conceicao, I.; Berk, J.; Schmidt, H.; Waddington-Cruz, M.; Campistol, J.M.; Bettencourt, B.R.; et al. Efficacy and safety of patisiran for familial amyloidotic polyneuropathy: a phase II multi-dose study. *Orphanet J. Rare Dis.* **2015**, *10*, 109, doi:10.1186/s13023-015-0326-6.
  260. Adams, D.; Koike, H.; Slama, M.; Coelho, T. Hereditary transthyretin amyloidosis: a model of medical progress for a fatal disease. *Nat. Rev. Neurol.* **2019**, *15*, 387–404, doi:10.1038/s41582-019-0210-4.
  261. Morral-Ruiz, G.; Melgar-Lesmes, P.; Solans, C.; García-Celma, M.J. Polyurethane nanoparticles, a new tool for biomedical applications? In *Advances in Polyurethane Biomaterials*; Elsevier, 2016; pp. 195–216.
  262. Argenziano, M.; Lombardi, C.; Ferrara, B.; Trotta, F.; Caldera, F.; Blangetti, M.; Koltai, H.; Kapulnik, Y.; Yarden, R.; Gigliotti, L.; et al. Glutathione/pH-responsive

- nanosponges enhance strigolactone delivery to prostate cancer cells. *Oncotarget* **2018**, *9*, 35813–35829, doi:10.18632/oncotarget.26287.
263. Prakash, T.P.; Graham, M.J.; Yu, J.; Carty, R.; Low, A.; Chappell, A.; Schmidt, K.; Zhao, C.; Aghajan, M.; Murray, H.F.; et al. Targeted delivery of antisense oligonucleotides to hepatocytes using triantennary N -acetyl galactosamine improves potency 10-fold in mice. *Nucleic Acids Res.* **2014**, *42*, 8796–8807, doi:10.1093/nar/gku531.
  264. BIESSEN, E.A.L.; VIETSCH, H.; RUMP, E.T.; FLUITER, K.; KUIPER, J.; BIJSTERBOSCH, M.K.; VAN BERKEL, T.J.C. Targeted delivery of oligodeoxynucleotides to parenchymal liver cells in vivo. *Biochem. J.* **1999**, *340*, 783–792, doi:10.1042/bj3400783.
  265. Nair, J.K.; Willoughby, J.L.S.; Chan, A.; Charisse, K.; Alam, M.R.; Wang, Q.; Hoekstra, M.; Kandasamy, P.; Kel'in, A. V.; Milstein, S.; et al. Multivalent N -Acetylgalactosamine-Conjugated siRNA Localizes in Hepatocytes and Elicits Robust RNAi-Mediated Gene Silencing. *J. Am. Chem. Soc.* **2014**, *136*, 16958–16961, doi:10.1021/ja505986a.
  266. Sato, Y.; Matsui, H.; Yamamoto, N.; Sato, R.; Munakata, T.; Kohara, M.; Harashima, H. Highly specific delivery of siRNA to hepatocytes circumvents endothelial cell-mediated lipid nanoparticle-associated toxicity leading to the safe and efficacious decrease in the hepatitis B virus. *J. Control. Release* **2017**, *266*, 216–225, doi:10.1016/j.jconrel.2017.09.044.
  267. Kasiewicz, L.N.; Biswas, S.; Beach, A.; Ren, H.; Dutta, C.; Mazzola, A.M.; Rohde, E.; Chadwick, A.; Cheng, C.; Musunuru, K.; et al. Lipid nanoparticles incorporating a GalNAc ligand enable in vivo liver ANGPTL3 editing in wild-type and somatic LDLR knockout non-human primates. *bioRxiv* **2021**.
  268. Korin, E.; Bejerano, T.; Cohen, S. GalNAc bio-functionalization of nanoparticles assembled by electrostatic interactions improves siRNA targeting to the liver. *J. Control. Release* **2017**, *266*, 310–320, doi:10.1016/j.jconrel.2017.10.001.
  269. Gane, E.; Yuen, M.; Kim, D.J.; Chan, H.L.; Surujbally, B.; Pavlovic, V.; Das, S.; Triyatni, M.; Kazma, R.; Grippio, J.F.; et al. Clinical Study of Single-Stranded Oligonucleotide RO7062931 in Healthy Volunteers and Patients With Chronic Hepatitis B. *Hepatology* **2021**, *74*, 1795–1808, doi:10.1002/hep.31920.
  270. Yuen, M.-F.; Heo, J.; Kumada, H.; Suzuki, F.; Suzuki, Y.; Xie, Q.; Jia, J.; Karino, Y.; Hou, J.; Chayama, K.; et al. Phase IIa, randomised, double-blind study of GSK3389404 in patients with chronic hepatitis B on stable nucleos(t)ide therapy. *J. Hepatol.* **2022**, *77*, 967–977, doi:10.1016/j.jhep.2022.05.031.
  271. Rao, S.; Hossain, T.; Mahmoudi, T. 3D human liver organoids: An in vitro platform to investigate HBV infection, replication and liver tumorigenesis. *Cancer Lett.* **2021**, *506*, 35–44, doi:10.1016/j.canlet.2021.02.024.

## **APPENDIX**

---

# 11. APPENDIXES

## Appendix 1



Basic Study

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

Selene Garcia-Garcia, Maria Francesca Cortese, David Taberero, Josep Gregori, Marta Vila, Beatriz Pacin, 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 Taberero 0000-0002-1146-4084; Josep Gregori 0000-0002-4253-8015; Marta Vila 0000-0001-9303-5189; Beatriz Pacin 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 Taberero D equally coordinated the research; Garcia-Garcia S and Quer J designed the experiments; Garcia-Garcia S, Vila M, Pacin B and Casillas R and Castillo-Ribelles L performed the experiments; Cortese MF, Taberero D, Garcia-Garcia S and Gregori J analyzed data acquired during the experiments and

Selene Garcia-Garcia, Maria Francesca Cortese, Marta Vila, Beatriz Pacin, 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 Taberero, Josep Gregori, Beatriz Pacin, 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 Taberero, 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.taberero@ciberehd.org](mailto:david.taberero@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, Pacin 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 MEGAclean 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 Cycler 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 HBV 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>18</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 re-purified 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  $2.45 \times 10^3$  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 (61.54)
Significant	3 (23.08)
Cirrhosis, n (%)	2 (15.39)

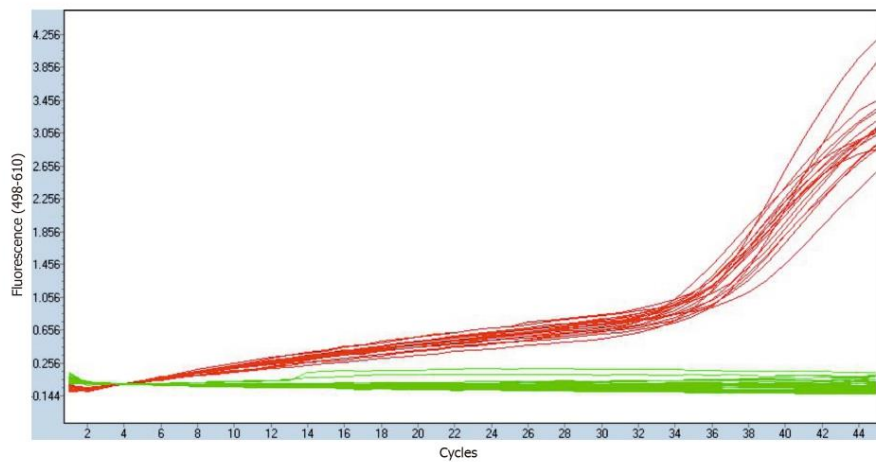
<sup>1</sup>Nonsignificant fibrosis indicates Ishak fibrosis stage < 3, significant fibrosis F ≥ 3 and < 5, and cirrhosis F5-6. Nonsignificant fibrosis by noninvasive markers indicates liver stiffness < 7-8.5 kPa, significant fibrosis > 7-8.5 and < 11-14 kPa and cirrhosis > 11-14 kPa by transient elastography 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.

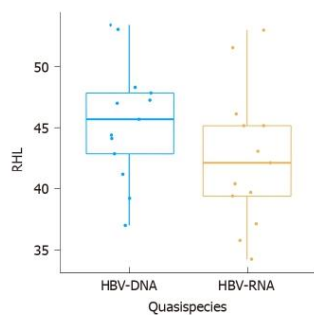
(IQR,  $1.90 \times 10^3$ - $3.19 \times 10^3$ ) for HBV-RNA and  $8.27 \times 10^4$  reads/sample (IQR,  $6.03 \times 10^4$ - $1.01 \times 10^5$ ) for HBV-DNA and were included in the RHL analysis. By setting the threshold of abundance as equal to or less than 1% HBV quasispecies complexity showed a 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 DNaseI 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.



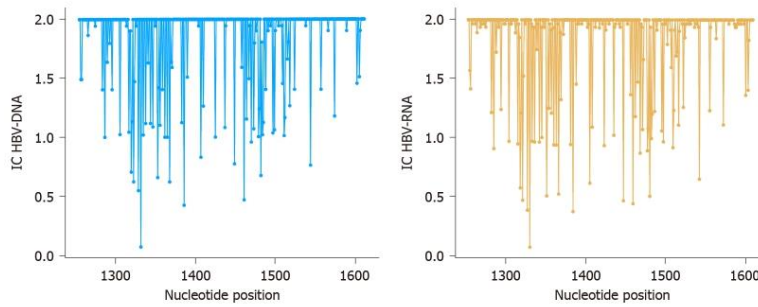
**Figure 2** Comparison of mean hepatitis B virus quasispecies complexity. Hepatitis B virus (HBV) quasispecies complexity analyzed by Rare Haplotype Load of all 13 patients in HBV-DNA quasispecies (blue-framed boxes) and HBV-RNA quasispecies (yellow-framed boxes). Differences were statistically non-significant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

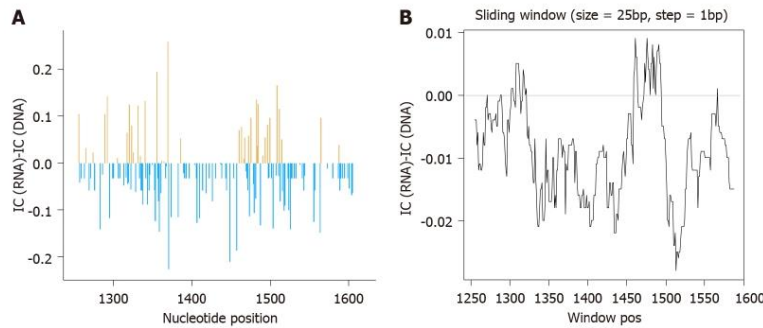
## DISCUSSION

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





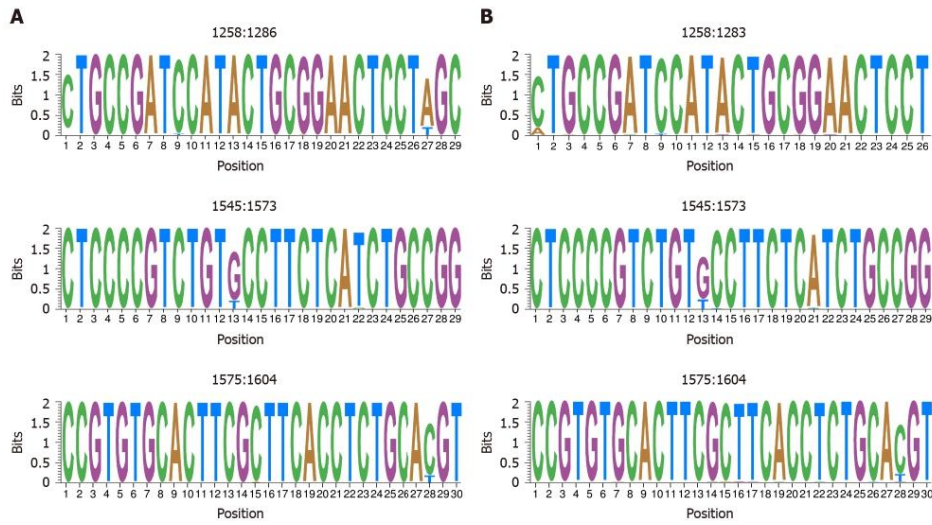
**Figure 3 Conservation and variability of 357 nucleotide positions analyzed.** Information content of nucleotide positions from 1255 to 1611 for hepatitis B virus (HBV)-DNA (blue lines) and HBV-RNA (orange lines) quasispecies. IC: Information content; HBV: Hepatitis B virus.



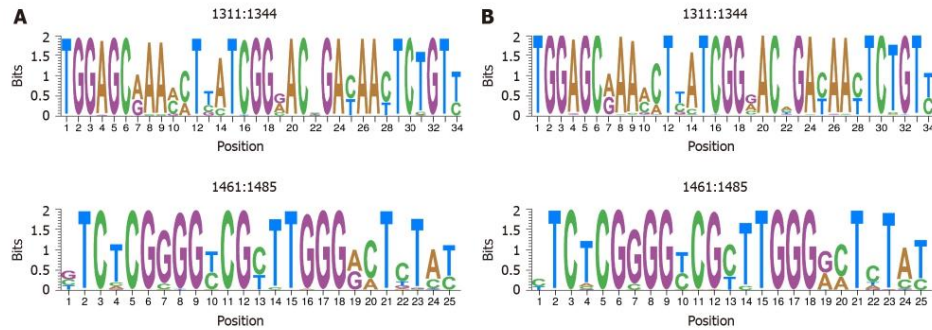
**Figure 4 Differences between information content of hepatitis B virus-DNA and RNA quasispecies.** A: Nucleotide positions in which information content (IC) hepatitis B virus (HBV)-RNA > HBV-DNA are depicted in orange while positions where IC HBV-DNA > HBV-RNA in blue. B: Sliding window analysis of 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  $3.6$ - $15 \times 10^6$  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, Tabemero 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 **Stadelmayer 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, Rodriguez-Frias F, Gregori J, Tabemero D, Soria ME, Garcia-Celici 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, Locamini 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**, Taberner 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, Rodriguez-Frias 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, Taberner D, Buti M, Rodriguez-Frias F. Sophisticated viral quasiespecies 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 **Durantel D**, Carroué-Durantel 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**, Taberner D, Sopena S, Gregori J, Cortese MF, González C, Casillas R, Yll M, Rando A, López-Martinez R, Quer J, González-Aseguinolaza G, Esteban R, Riveiro-Barciela M, Buti M, Rodriguez-Frias F. Characterization of hepatitis B virus X gene quasiespecies 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 JI, 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, Colomo 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**, García-García S, Casillas R, Lopez-Martinez R, Pacin Ruiz B, Sopena S, Taberner D, Ferrer-Costa RM, Riveiro-Barciela M, Buti M, Rodriguez-Frias 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**, García-García S, Pacin Ruiz B, Casillas R, Taberner D, Riveiro-Barciela M, Buti M, Rodriguez-Frias 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

## Appendix 2

Appendix 2. Table 1

<b>HBV Genotype</b> (n sequences)	<b>Accession numbers</b> <b>on NCBI GenBank</b>
<b>A (19)</b>	A2_AJ309371, A2_AM282986, A2_X02763, A2_X51970, A1_AY233278, A1_AB241115, A3_AB194952, A3_AB194951, A4_AY934764, A6_GQ331047, A6_GQ331048, A__FJ692557, A__FM199979, A__DQ020002, A__AY233274, A__AB453985, A__AP007263, A__EU594391, A__AM184126,
<b>B (21)</b>	B1_D00329, B2_AP011084, B3_M54923, B3_AP011085, B5_AB219427, B6_AB287316, B6_DQ463787, B7_EF473977, B8_AP011093, B1_AB073858, B1_AB362933, B2_GQ924653, B2_GU815751, B__KJ410516, B__KJ803816, B__JN792901, B__JX661483, B__GQ358151, B__KC774390, B__AB900106, B__FJ562316
<b>C (19)</b>	C2_X52939, C2_D23681, C3_X75656, C3_X75665, C1_AB112066, C2_AB033553, C2_AF533983, C6_AP011102, C6_AP011103, C7_EU670263, C8_AP011107, C9_AP011108, C__KJ410521, C__AB971714, C__GQ358157, C__KP784761, C__KF214667, C__EU305541, C__DQ246215
<b>D (20)</b>	D1_X80926, D3_V01460, D2_Z35716, D3_AY233291, D4_AB048702, D4_AB033559, D5_DQ315779, D5_AB033558, D7_AM494716, D3_X65258, D2_X97848, D__KJ803771, D__KR905423, D__JN792907, D__KM577667, D__HQ833465, D__KC875335, D__AB555496, D__JN664919, D__JX096955
<b>E (20)</b>	E4_HM363569, E4_FJ349226, E2_X75657, E5_DQ060828, E__KT192626, E__KM606738, E__KF922438, E__JQ000009, E__HM363566, E__FJ349238, E__GQ161834, E__EU239222, E__DQ060825, E__AM494714, E__AB205189, E__HM363591, E__GQ161787, E__EU239217, E__DQ060822, E__AB194948
<b>F (19)</b>	F3_X75663, F3_AB036910, F2_AY311369, F2_X69798, F4_HE974368, F4_EU366116, F1_HM590471, F2_AY090455, F3_AB036915, F4_DQ823090, F4_AB166850, F__KJ843197,



	F__KC494404, F__KC494394, F__EF576808, F__JN811656, F__HQ378247, F__HM585200, F__AB116654
<b>G (7)</b>	G__KF414679, G__JQ707679, G__JQ707678, G__JQ707657, G__GU565217, G__AF405706, G__AP007264
<b>H (9)</b>	H__AB818694, H__HM117851, H__HM117850, H__HM066946, H__EU498228, H__AB266536, H__AB298362, H__EF157291, H__AB205010

**Appendix 2. Table 1.** Accession numbers of 134 reference sequences from HBV genotypes A to H on NCBI GenBank used in the second study.

## Appendix 3

Appendix 3. Table 1

	Patient	B.cDNA	B.DNA	S.cDNA	S.DNA
<b>Master fraction</b> (present at the highest frequency)	P02	85.91	14.05		
	P07	57.58	78.72		
	P08	83.89	84.60		
	P11	86.57	42.93		
	P14	9.04	17.37		
	P15	44.81	30.01		
	P16	66.72	84.32	62.43	83.23
	P19	86.06	81.75	85.77	79.79
	P21	62.06	71.35	85.98	52.47
	P22	82.19	82.53		
	P24	81.26	57.71		
	P26	54.37	46.64		
	<b>Emerging fraction</b> (below the master value and above 1%)	P02	1.85	51.16	
P07		21.55	2.58		
P08		3.83	2.20		
P11		1.61	30.89		
P14		41.26	56.69		
P15		39.39	52.08		

	P16	20.77	2.10	23.24	1.76
	P19	1.75	2.19	3.37	1.80
	P21	25.10	6.24	1.55	19.87
	P22	1.87	1.89		
	P24	5.80	17.11		
	P26	21.96	30.71		
	<hr/>				
	P02	3.29	24.99		
	P07	10.9	4.83		
	P08	1.71	0.94		
	P11	1.89	15.1		
	P14	25.95	9.13		
	P15	7.08	5.88		
<b>Low fitness fraction</b> (from 1% to 0.1%)	P16	4.47	0.81	5.51	1.69
	P19	2.29	2.25	1.39	4.34
	P21	2.84	7.52	2.42	16.28
	P22	3.46	1.95		
	P24	2.86	9.79		
	P26	12.14	7.19		
	<hr/>				
	P02	8.95	9.79		
<b>Very low fitness fraction</b> ( $< 0.1\%$ )	P07	9.96	13.87		
	P08	10.57	12.26		
	P11	9.93	11.08		

P14	23.75	16.81		
P15	8.72	12.03		
P16	8.04	12.76	8.82	13.33
P19	9.90	13.81	9.47	14.07
P21	10.00	14.89	10.04	11.38
P22	12.48	13.63		
P24	10.08	15.39		
P26	11.54	15.46		

**Appendix 3. Table 1. Values of the partitions in four fractions of paired HBV-DNA and HBV-RNA quasispecies haplotypes according to their fitness.** *Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; Patients are designated as PXX; B.DNA, baseline DNA quasispecies; B.cDNA, baseline cDNA quasispecies; S.DNA, follow-up DNA quasispecies; S.cDNA, follow-up cDNA quasispecies.*

Appendix 3. Table 2

	<b>q_0</b>	<b>q_0.5</b>	<b>q_1</b>	<b>q_1.5</b>	<b>q_2</b>	<b>q_2.5</b>	<b>q_inf</b>
<b>P02.B.cDNA</b>	488,10	61,27	3,16	1,56	1,35	1,29	1,16
<b>P02.B.DNA</b>	629,53	181,23	52,61	26,49	18,96	15,73	7,12
<b>P07.B.cDNA</b>	639,03	119,75	12,04	4,22	2,94	2,50	1,74
<b>P07.B.DNA</b>	717,30	124,74	5,61	2,00	1,61	1,49	1,27
<b>P08.B.cDNA</b>	698,99	91,58	3,73	1,66	1,42	1,34	1,19
<b>P08.B.DNA</b>	658,60	94,20	3,74	1,63	1,40	1,32	1,18
<b>P11.B.cDNA</b>	491,04	61,92	3,06	1,52	1,33	1,27	1,16
<b>P11.B.DNA</b>	493,68	126,02	19,38	7,08	4,78	3,93	2,33
<b>P14.B.cDNA</b>	1498,39	548,27	135,24	49,91	31,09	24,42	11,06
<b>P14.B.DNA</b>	1302,25	320,17	47,62	19,88	14,47	12,27	5,76
<b>P15.B.cDNA</b>	815,71	127,63	14,00	5,98	4,31	3,63	2,23
<b>P15.B.DNA</b>	967,94	172,30	17,97	8,06	6,25	5,51	3,33
<b>P16.B.cDNA</b>	626,42	84,28	6,58	2,85	2,20	1,95	1,50
<b>P16.B.DNA</b>	768,81	110,51	3,90	1,65	1,41	1,33	1,19
<b>P16.S.cDNA</b>	735,93	103,88	8,04	3,31	2,49	2,18	1,60

<b>P16.S.DNA</b>	755,35	114,16	4,22	1,71	1,44	1,36	1,20
<b>P19.B.cDNA</b>	680,54	84,07	3,28	1,55	1,35	1,28	1,16
<b>P19.B.DNA</b>	740,52	118,41	4,68	1,80	1,50	1,40	1,22
<b>P19.S.cDNA</b>	661,83	79,45	3,25	1,56	1,36	1,29	1,17
<b>P19.S.DNA</b>	730,53	124,11	5,31	1,93	1,57	1,46	1,25
<b>P21.B.cDNA</b>	844,52	112,01	6,61	3,00	2,41	2,16	1,61
<b>P21.B.DNA</b>	840,32	160,73	8,35	2,59	1,95	1,75	1,40
<b>P21.S.cDNA</b>	626,94	77,86	3,25	1,55	1,35	1,29	1,16
<b>P21.S.DNA</b>	957,76	191,15	16,06	4,93	3,38	2,86	1,91
<b>P22.B.cDNA</b>	749,53	113,60	4,48	1,77	1,48	1,39	1,22
<b>P22.B.DNA</b>	721,07	111,34	4,38	1,75	1,47	1,38	1,21
<b>P24.B.cDNA</b>	658,19	84,31	3,90	1,78	1,51	1,41	1,23
<b>P24.B.DNA</b>	1015,93	210,19	13,89	4,08	2,87	2,47	1,73
<b>P26.B.cDNA</b>	896,51	171,94	15,20	4,84	3,27	2,74	1,84
<b>P26.B.DNA</b>	778,05	163,91	12,93	4,60	3,48	3,09	2,14

**Appendix 3. Table 2. Values of the orders ( $q$  values) of Hill numbers.**  
*Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; B.DNA, baseline DNA quasispecies; B.cDNA, baseline cDNA quasispecies; S.DNA, follow-up DNA quasispecies; S.cDNA, follow-up cDNA quasispecies.*

Appendix 3. Table 3

<b>ID 1</b>	<b>ID 2</b>	<b>C<sub>m</sub> index</b>
P02.B.DNA	P02.B.cDNA	0,464
P07.B.DNA	P19.B.DNA	0,410
P07.B.DNA	P26.B.DNA	0,001
P07.B.DNA	P07.B.cDNA	0,830
P07.B.DNA	P19.B.cDNA	0,825
P07.B.DNA	P26.B.cDNA	0,438
P08.B.DNA	P11.B.DNA	0,003
P08.B.DNA	P15.B.DNA	0,517
P08.B.DNA	P16.B.DNA	0,424
P08.B.DNA	P24.B.DNA	0,001
P08.B.DNA	P08.B.cDNA	0,422
P08.B.DNA	P11.B.cDNA	0,004
P08.B.DNA	P16.B.cDNA	0,809
P08.B.DNA	P15.B.cDNA	0,472
P08.B.DNA	P24.B.cDNA	0,004
P11.B.DNA	P15.B.DNA	0,544
P11.B.DNA	P16.B.DNA	0,720
P11.B.DNA	P24.B.DNA	0,637
P11.B.DNA	P08.B.cDNA	0,420
P11.B.DNA	P11.B.cDNA	0,820
P11.B.DNA	P16.B.cDNA	0,573
P11.B.DNA	P15.B.cDNA	0,556

P11.B.DNA	P24.B.cDNA	0,651
P14.B.DNA	P14.B.cDNA	0,692
P15.B.DNA	P16.B.DNA	0,750
P15.B.DNA	P24.B.DNA	0,708
P15.B.DNA	P08.B.cDNA	0,662
P15.B.DNA	P11.B.cDNA	0,587
P15.B.DNA	P16.B.cDNA	0,742
P15.B.DNA	P15.B.cDNA	0,918
P15.B.DNA	P24.B.cDNA	0,795
P16.B.DNA	P24.B.DNA	0,782
P16.B.DNA	P08.B.cDNA	0,842
P16.B.DNA	P11.B.cDNA	0,858
P16.B.DNA	P16.B.cDNA	0,940
P16.B.DNA	P15.B.cDNA	0,762
P16.B.DNA	P24.B.cDNA	0,857
P19.B.DNA	P26.B.DNA	0,774
P19.B.DNA	P07.B.cDNA	0,001
P19.B.DNA	P19.B.cDNA	0,990
P19.B.DNA	P26.B.cDNA	0,703
P21.B.DNA	P21.B.cDNA	0,951
P22.B.DNA	P22.B.cDNA	0,971
P24.B.DNA	P08.B.cDNA	0,001
P24.B.DNA	P11.B.cDNA	0,723
P24.B.DNA	P16.B.cDNA	0,694
P24.B.DNA	P15.B.cDNA	0,703
P24.B.DNA	P24.B.cDNA	0,933



P26.B.DNA	P19.B.cDNA	0,806
P26.B.DNA	P26.B.cDNA	0,832
P07.B.cDNA	P19.B.cDNA	0,288
P07.B.cDNA	P26.B.cDNA	0,392
P08.B.cDNA	P11.B.cDNA	0,422
P08.B.cDNA	P16.B.cDNA	0,827
P08.B.cDNA	P15.B.cDNA	0,691
P08.B.cDNA	P24.B.cDNA	0,422
P11.B.cDNA	P16.B.cDNA	0,782
P11.B.cDNA	P15.B.cDNA	0,662
P11.B.cDNA	P24.B.cDNA	0,845
P16.B.cDNA	P15.B.cDNA	0,751
P16.B.cDNA	P24.B.cDNA	0,813
P15.B.cDNA	P24.B.cDNA	0,792
P19.B.cDNA	P26.B.cDNA	0,791

**Appendix 3. Table 3. Values of Commons index for comparisons between HBV-DNA and HBV-RNA quasispecies haplotypes.** *Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; Patients are designated as PXX; B.DNA, baseline DNA quasispecies; B.cDNA, baseline cDNA quasispecies; S.DNA, follow-up DNA quasispecies; S.cDNA, follow-up cDNA quasispecies; C<sub>m</sub>, Commons.*

Appendix 3. Table 4

<b>ID 1</b>	<b>ID 2</b>	<b>O<sub>v</sub> index</b>
P02.B.DNA	P02.B.cDNA	0,05
P07.B.DNA	P19.B.DNA	0,00
P07.B.DNA	P26.B.DNA	0,00
P07.B.DNA	P07.B.cDNA	0,65
P07.B.DNA	P19.B.cDNA	0,00
P07.B.DNA	P26.B.cDNA	0,05
P08.B.DNA	P11.B.DNA	0,00
P08.B.DNA	P15.B.DNA	0,00
P08.B.DNA	P16.B.DNA	0,00
P08.B.DNA	P24.B.DNA	0,00
P08.B.DNA	P08.B.cDNA	0,00
P08.B.DNA	P11.B.cDNA	0,00
P08.B.DNA	P16.B.cDNA	0,00
P08.B.DNA	P15.B.cDNA	0,00
P08.B.DNA	P24.B.cDNA	0,00
P11.B.DNA	P15.B.DNA	0,00
P11.B.DNA	P16.B.DNA	0,00
P11.B.DNA	P24.B.DNA	0,00
P11.B.DNA	P08.B.cDNA	0,00
P11.B.DNA	P11.B.cDNA	0,47
P11.B.DNA	P16.B.cDNA	0,00
P11.B.DNA	P15.B.cDNA	0,00

P11.B.DNA	P24.B.cDNA	0,00
P14.B.DNA	P14.B.cDNA	0,42
P15.B.DNA	P16.B.DNA	0,00
P15.B.DNA	P24.B.DNA	0,00
P15.B.DNA	P08.B.cDNA	0,00
P15.B.DNA	P11.B.cDNA	0,00
P15.B.DNA	P16.B.cDNA	0,03
P15.B.DNA	P15.B.cDNA	0,69
P15.B.DNA	P24.B.cDNA	0,00
P16.B.DNA	P24.B.DNA	0,00
P16.B.DNA	P08.B.cDNA	0,00
P16.B.DNA	P11.B.cDNA	0,00
P16.B.DNA	P16.B.cDNA	0,75
P16.B.DNA	P15.B.cDNA	0,00
P16.B.DNA	P24.B.cDNA	0,00
P19.B.DNA	P26.B.DNA	0,00
P19.B.DNA	P07.B.cDNA	0,00
P19.B.DNA	P19.B.cDNA	0,92
P19.B.DNA	P26.B.cDNA	0,00
P21.B.DNA	P21.B.cDNA	0,71
P22.B.DNA	P22.B.cDNA	0,93
P24.B.DNA	P08.B.cDNA	0,00
P24.B.DNA	P11.B.cDNA	0,00
P24.B.DNA	P16.B.cDNA	0,00
P24.B.DNA	P15.B.cDNA	0,00
P24.B.DNA	P24.B.cDNA	0,71

P26.B.DNA	P19.B.cDNA	0,00
P26.B.DNA	P26.B.cDNA	0,59
P07.B.cDNA	P19.B.cDNA	0,00
P07.B.cDNA	P26.B.cDNA	0,05
P08.B.cDNA	P11.B.cDNA	0,01
P08.B.cDNA	P16.B.cDNA	0,02
P08.B.cDNA	P15.B.cDNA	0,00
P08.B.cDNA	P24.B.cDNA	0,00
P11.B.cDNA	P16.B.cDNA	0,01
P11.B.cDNA	P15.B.cDNA	0,00
P11.B.cDNA	P24.B.cDNA	0,01
P16.B.cDNA	P15.B.cDNA	0,03
P16.B.cDNA	P24.B.cDNA	0,01
P15.B.cDNA	P24.B.cDNA	0,00
P19.B.cDNA	P26.B.cDNA	0,00

**Appendix 3. Table 4. Values of Overlap index for comparisons between HBV-DNA and HBV-RNA quasispecies haplotypes.** Abbreviations: cDNA stands for complementary DNA and corresponds to reads obtained from hepatitis B virus RNA reverse transcription libraries; Patients are designated as PXX; B.DNA, baseline DNA quasispecies; B.cDNA, baseline cDNA quasispecies; S.DNA, follow-up DNA quasispecies; S.cDNA, follow-up cDNA quasispecies; O<sub>v</sub>, Overlap.

## Appendix 4

### **Publications during the course of the doctoral thesis:**

**Scientific article.** Garcia-Garcia, Selene; Caballero-Garralda, Andrea; Tabernero, David; et al; Rodriguez-Frias, Francisco. (1/16). 2022. Hepatitis B Virus Variants with Multiple Insertions and/or Deletions in the X Open Reading Frame 3' End: Common Members of Viral Quasispecies in Chronic Hepatitis B Patients BIOMEDICINES. MDPI. 10-5. ISSN 2227-9059. Impact Factor = 4.757, Quartile 1.

**Scientific article.** Cortese, Maria Francesca; Riveiro-Barciela, Mar; Tabernero, David; et al; Garcia-Garcia, Selene; Rodriguez-Frias, Francisco. (15/19). 2022. Standardized Hepatitis B Virus RNA Quantification in Untreated and Treated Chronic Patients: a Promising Marker of Infection Follow-Up. Microbiology spectrum. 10-2, pp.e0214921. ISSN 2165-0497. Impact Factor = 9.043, Quartile 1.

**Scientific article.** Pacin-Ruiz, Beatriz; Cortese, Maria Francesca; Tabernero, David; et al; Garcia-Garcia, Selene; Rodriguez-Frias, Francisco. (6/18). 2022. Inspecting the Ribozyme Region of Hepatitis Delta Virus Genotype 1: Conservation and Variability VIRUSES-BASEL. MDPI. 14-2. ISSN 1999-4915. Impact Factor = 5.818, Quartile 2.

**Review.** Rodriguez-Frias, Francisco; Quer, Josep; Tabernero, David; Cortese, Maria Francesca; Garcia-Garcia, Selene; Rando-Segura, Ariadna; Pumarola, Tomas. (5/7). 2021. Microorganisms as Shapers of Human Civilization, from Pandemics to Even Our Genomes: Villains or Friends? A Historical Approach MICROORGANISMS. MDPI. 9-12. ISSN 2076-2607. Impact Factor = 4.926, Quartile 2.

**Scientific article.** Garcia-Garcia, Selene; Cortese, Maria Francesca; Tabernero, David; et al; Rodriguez-Frias, Francisco. (1/18). 2021. Cross-sectional evaluation of circulating hepatitis B virus RNA and DNA: Different quasispecies? WORLD JOURNAL OF GASTROENTEROLOGY. BAISHIDENG PUBLISHING GROUP INC. 27-41, pp.7144-7158. ISSN 2219-2840. Impact Factor = 5.374, Quartile 2.

**Scientific article.** Francesca Cortese, Maria; Gonzalez, Carolina; Gregori, Josep; et al; Garcia-Garcia, Selene; Rodriguez-Frias, Francisco. (15/19). 2021. Sophisticated viral quasispecies with a genotype-related pattern of mutations in the hepatitis B X gene of HBeAg-ve chronically infected patients SCIENTIFIC REPORTS. NATURE RESEARCH. 11-1. ISSN 2045-2322. Impact Factor = 4.997, Quartile 1.

**Review.** Garcia-Garcia, Selene; Cortese, Maria Francesca; Rodriguez-Algarra, Francisco; Taberero, David; Rando-Segura, Ariadna; Quer, Josep; Buti, Maria; Rodriguez-Frias, Francisco. (1/8). 2021. Next-generation sequencing for the diagnosis of hepatitis B: current status and future prospects EXPERT REVIEW OF MOLECULAR DIAGNOSTICS. TAYLOR & FRANCIS AS. 21-4, pp.381-396. ISSN 1744-8352. Impact Factor = 5.670, Quartile 1.