



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. The access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.



Universitat Autònoma de Barcelona

Departamento de Bioquímica y Biología Molecular

Programa de doctorado en Bioquímica, Biología Molecular y Biomedicina

Tesis doctoral

Complejidad de la cuasiespecie del virus de la hepatitis B
en la región X/preCore: asociación con la evolución de la
infección con y sin tratamiento antiviral

Autora de la tesis

Andrea Caballero Garralda

Directores de la tesis

Dr. Francisco Rodríguez Frías

Dra. María Buti Ferret

Tutor de la tesis

Dr. Francisco Rodríguez Frías

Universidad Autònoma de Barcelona. Barcelona, 2016

ÍNDICE

ÍNDICE	3
AGRADECIMIENTOS	7
1. ABREVIATURAS	9
2. INTRODUCCIÓN	13
2.1 VIROLOGÍA DEL VIRUS DE LA HEPATITIS B	15
2.1.1 HISTORIA	15
2.1.2 CLASIFICACIÓN	15
2.1.3 LA PARTÍCULA VIRAL	15
2.1.4 ORGANIZACIÓN DEL GENOMA VIRAL Y PROTEÍNAS CODIFICADAS	17
2.1.4.1 ORFs	18
2.1.4.1.1 ORF preS/S	18
2.1.4.1.2 ORF P	21
2.1.4.1.3 ORF PC/C	24
2.1.4.1.4 ORF X	27
2.1.4.2 Elementos reguladores	37
2.1.4.2.1 Promotores y enhancers	37
2.1.4.2.2 La señal de encapsidación épsilon	39
2.1.4.2.3 Otros elementos reguladores	41
2.1.5 CICLO REPLICATIVO	41
2.1.5.1 Entrada del virus	44
2.1.5.2 Transporte de las cápsides desde el citoplasma al núcleo y liberación del ADNrc en el núcleo	44
2.1.5.3 Formación del ADNccc	44
2.1.5.4 Transcripción viral del ADNccc y traducción proteica	45
2.1.5.5 Formación de las partículas core y transcripción reversa	45
2.1.5.5.1 Unión polimerasa-ε: encapsidación y cebado proteico	45
2.1.5.5.2 Translocación de 4NT y la polimerasa a DR1*	46
2.1.5.5.3 Síntesis de la hebra(-) por la actividad retrotranscriptasa	46
2.1.5.5.4 Translocación del complejo P+ARN e inicio síntesis de la hebra(+)	46
2.1.5.5.5 Elongación hebra(+)	47
2.1.5.6 Maduración y secreción de viriones y partículas subvirales	47
2.1.5.6.1 Liberación de partículas virales nuevas	47
2.1.5.6.2 Reciclaje de las cápsides	47
2.2 IMPLICACIONES CLÍNICAS DE LA INFECCIÓN POR EL VHB	48
2.2.1 EPIDEMIOLOGÍA	48
2.2.1.1 Prevalencia global	48
2.2.1.2 Transmisión	49
2.2.2 HISTORIA NATURAL	49
2.2.2.1 Hepatitis aguda	50
2.2.2.2 Hepatitis B crónica (HBC)	50
2.2.2.2.1 Fase de tolerancia inmune	51
2.2.2.2.2 Fase de actividad inmune	51
2.2.2.2.3 Fase de portador inactivo	51
2.2.2.2.4 Hepatitis crónica HBeAg(-)	51
2.2.2.2.5 Fase de remisión	52
2.2.3 INMUNOPATOGENÉESIS DEL VHB	53
2.2.3.1 Respuesta innata	53
2.2.3.2 Respuesta adaptativa	53
2.2.3.3 Establecimiento de la infección crónica	54

2.2.3.3.1 Rol de HBeAg en la inmunopatogénesis	55
2.2.3.3.2 Rol de HBcAg en la inmunopatogénesis	55
2.2.4 HEPATOCARCINOMA RELACIONADO CON EL VHB (HCC-VHB)	57
2.2.4.1 Mecanismos moleculares de la hepatocarcinogénesis asociada al VHB	58
2.2.4.1.1 Integración del ADN-VHB y hepatocarcinogénesis	58
2.2.4.1.2 Proteínas virales y HCC	59
2.2.5 TRATAMIENTOS ANTIVIRALES	60
2.2.5.1 Fármacos aprobados y estrategias de tratamiento	60
2.2.5.1.1 IFN α /PEG-IFN	60
2.2.5.1.2 NUCs	61
2.2.5.2 Estrategias de tratamiento y objetivo	61
2.2.5.3 Valoración del tratamiento	63
2.2.5.4 Resistencias a NUCs	63
2.2.5.4.1 Tasas de resistencia	64
2.2.5.4.2 Patrones de mutación	65
2.2.5.5 Nuevas estrategias para curar la infección por el VHB	66
2.2.5.5.1 Agentes antivirales directos (DAAs)	66
2.2.5.5.2 Agentes contra el hospedador (HTAs)	67
2.3 VARIABILIDAD GENÉTICA	69
2.3.1 GENOTIPOS Y SUBGENOTIPOS	69
2.3.1.1 Clasificación	69
2.3.1.2 Distribución geográfica	70
2.3.1.3 Recombinación intergenotípica	71
2.3.1.3.1 Mecanismo de recombinación	72
2.3.1.3.2 Puntos de recombinación (<i>breakpoints</i>)	72
2.3.1.3.3 Descripción detallada de los genomas recombinantes	73
2.3.1.4 Implicación del genotipo en la virología	74
2.3.1.4.1 Tamaño genoma	74
2.3.1.4.2 Modo de transmisión	75
2.3.1.4.3 Características genéticas particulares de los genotipos	75
2.3.1.5 Implicación del genotipo en la patogénesis	76
2.3.1.5.1 Genotipos B y C	76
2.3.1.5.2 Genotipos A, D, F Y G	76
2.3.1.6 Implicación del genotipo en el tratamiento	77
2.3.1.6.1 IFN	77
2.3.1.6.2 NUCs	77
2.3.1.7 Impacto de los genotipos en las medidas profilácticas	78
2.3.2 PRINCIPALES VARIANTES GENÉTICAS DE INTERÉS CLÍNICO EN EL VHB	78
2.3.2.1 Variabilidad en el ORF preS/S	79
2.3.2.2 Variabilidad en el ORF P	79
2.3.2.3 Variabilidad en el core	80
2.3.2.4 Variabilidad en la región X	81
2.3.2.4.1 Variantes en el PC	81
2.3.2.4.2 Variantes en el PBC	81
2.3.2.4.3 Variantes en la región X/Cp/PC	82
2.3.2.4.4 Variantes en el ORF X	82
2.3.3 ESTRUCTURA EN QUASIESPECIES DE LAS POBLACIONES DEL VHB	85
2.3.3.1 Definición de cuasiespecie (QS)	85
2.3.3.2 Origen de la cuasiespecie viral	85
2.3.3.3 Importancia del estudio de la QS	86
2.3.4 COMPLEJIDAD DE QS	88
2.3.4.1 Definición y significado clínico	88
2.3.4.2 Evolución de la complejidad de QS en la historia natural de la Infección	88
2.3.4.3 Evolución de la complejidad de QS durante el tratamiento	88

2.3.5	ÍNDICES PARA EL ESTUDIO DE LA DIVERSIDAD DE CUASIESPECIE	89
2.3.5.1	Incidencia (total)	89
2.3.5.2	Abundancia (frecuencias)	89
2.3.5.3	Función (diferencias entre haplotipos)	90
2.3.6	MÉTODOS PARA EL ESTUDIO DE LAS VARIANTES GENÉTICAS DEL VHB	91
2.3.6.1	Métodos para el genotipaje	91
2.3.6.1.1	Métodos moleculares para el genotipaje	91
2.3.6.1.2	Métodos on line para genotipar	92
2.3.6.2	Herramientas para el estudio de la QS	93
2.3.6.2.1	Técnicas moleculares indirectas	93
2.3.6.2.2	Técnicas de secuenciación	93
2.3.6.3	Descripción de las plataformas de NGS	96
2.3.6.4	Comparación de las plataformas	103
2.4	HIPÓTESIS Y OBJETIVOS	107
2.4.1	HIPÓTESIS	107
2.4.2	OBJETIVOS	109
3.	PRIMER ESTUDIO: Clinical Application of Estimating Hepatitis B Virus Quasispecies Complexity by Massive Sequencing: Correlation between Natural Evolution and On-Treatment Evolution	113
3.1	Introducción	115
3.2	Hipótesis y objetivos	116
3.2.1	Hipótesis	116
3.2.2	Objetivos	116
3.3	Resumen del estudio	116
3.2.1.	Complejidad de la QS y su relación con el período de toma de la muestra, genotipo, ADN-VHB, niveles ALT y estatus HBeAg	117
3.2.2	Complejidad de QS y estatus HBeAg: estudio longitudinal	118
3.2.3.	Distribución de los cambios en preCore y core	120
3.2.4	Evolución de la complejidad de la QS sin tratamiento y bajo la presión de NUCs	121
3.2.5	Conclusiones	122
3.4	Manuscrito completo	123
4.	SEGUNDO ESTUDIO: Complex Genotype Mixtures Analyzed by Deep Sequencing in Two Different Regions of Hepatitis B Virus	137
4.1	Introducción	139
4.2	Hipótesis y objetivos	139
4.1.1	Hipótesis	139
4.2.2	Objetivos	140
4.3	Resumen del estudio	140
4.3.1.	Validación de las regiones P/S y X/PC para el genotipaje	141
4.3.2	Resultados del genotipaje por UDPS	141
4.3.2.1	Genotipaje en la región P/S	142
4.3.2.2	Genotipaje en la región X/PC	142
4.3.2.3	Comparación de la región P/S vs. X/PC para el genotipaje	142
4.3.3	Cambios de genotipo durante los períodos de tiempo estudiados	145
4.3.4	Complejidad de la QS	146
4.3.5	Conclusiones	147
4.4	Manuscrito completo	148
5.	DISCUSIÓN	171
5.1	Técnicas para el estudio de la variabilidad	175
5.2	Complejidad de la QS y seroconversión anti-HBe	178
5.3	Complejidad de la QS y actividad viral	179

5.4 Descenso de la complejidad asociada al fallo del tratamiento	180
5.5 Utilidad clínica de los índices de diversidad	183
5.6 Complejidad del VHB: Mezclas de genotipos	184
5.7 Posible recombinación inter-genotípica	186
5.8 Importancia de una correcta clasificación genotípica	189
5.9 Perspectivas de futuro	191
6. CONCLUSIONES	193
7. BIBLIOGRAFÍA	199

AGRADECIMIENTOS

Si puedo escribir estos agradecimientos es gracias a dos personas, el Dr. Francisco Rodríguez-Frías y la Dra. María Homs. Agradecer al Dr. Rodríguez-Frías el haber creado en mí la inquietud por conocer más y más de esos seres, no diré que vivos, quizás bioquímicos, que son los virus, por su infinito elevado a infinito conocimiento, por sus ideas geniales, por hacerme pensar, por ver el mundo de manera diferente, por su saber estar siempre presente, por evitar una y otra vez que tirara la toalla, por enseñarme a trabajar duro predicando con el ejemplo y, en definitiva, por haber confiado en mí. A la Dra. Homs por haberme ayudado a reflejar los estudios en papel como si no costara esfuerzo, todo parece fácil con ella, quizás porque las personas brillantes tienen el don de hacerlo. Quisiera agradecer a la Dra. María Buti su aportación fundamental del punto de vista clínico, su entusiasmo y su capacidad de trabajo incombustibles, al Dr. Josep Gregori su invaluable procesamiento de los datos y al Dr. Ernesto Casis y la Dra. Dolors Pelegrí el haberme puesto todas las facilidades para la redacción de esta tesis.

No puedo dejar de agradecer a Roche Diagnostics, su soporte tecnológico y personal, apostando desde el principio por nuestros proyectos, y por este en particular, en línea con la política de búsqueda de soluciones innovadoras de la compañía. El desarrollo de esta tesis también ha sido posible gracias al Instituto Carlos III y la financiación del proyecto FIS PI12/1983, cuyo IP es el Dr. Rodríguez-Frías.

También quisiera dar las gracias al Dr. Josep Quer, a la Dra. Anna Messeguer y a la Dra. María Cubero por sus consejos en las comisiones de seguimiento, en especial dar las gracias al Dr. Quer por haber puesto siempre a mi disposición su laboratorio de Enfermedades Hepáticas del VHIR y por su cariño.

A continuación quisiera poder reflejar mi profundo agradecimiento a mis compañeros del "Hepatits Team", agradecer al Dr. David Tabernero su inmenso conocimiento del VHB reflejado en su maravillosa tesis, por la revisión de esta y por la información administrativa, a María Blasi por enseñarme cómo se trabaja aspirando a la perfección en el laboratorio, a Chari por explicarme por primera vez el protocolo de pirosecuenciación, a Irene y a Leo por transmitirme su optimismo ¡Ánimo chicos! ya os queda menos. Agradecer a todos por haberme animado, preguntado qué tal y haberme ofrecido vuestra ayuda una y otra vez, pero sobre todo por haberme hecho sentir parte de vuestro equipo, un equipo de GRANDES PERSONAS.

Creo que esta tesis tampoco existiría sin mis compañeras de preanalítica, Luzma y Teresa no tengo palabras para agradecer vuestro esfuerzo para liberarme de la asistencia y vuestro apoyo cuando tenía un “día malo”, Luzma te debo un favor ultra-mega-gigante, la mitad de esta tesis está escrita gracias a ti. Gracias a mis técnicos de la UCRE, administrativ@s e informáticos y a los chicos del Aqua, por haber aguantado mis cambios de humor y sus palabras de ánimo constantes. A Adam por las conversaciones a las 15 h. A todos los Facultativos por preocuparse por mí, sobre todo a la Dra. Roser Ferrer por insistir una y otra vez “¡Hazla!”, y a las futuras Dra. Imma Comas, Dra. Silvia Camós, Dra. Clara Carnicer y Dra. Paula Fernández ¡Hacerla! Al Dr. Caballero por su sabiduría y hacerme ser mejor persona. A los “resis” y “ex-resis” (Andrea, Vicky, Laura, Pilar, Silvia, Guille, Alicia y Carol) por haber hecho las guardias solos. Gracias en especial a la Dra. Nuria Llecha por haberme hecho ver la importancia de desarrollar una tesis doctoral, por enseñarme cómo se organiza un laboratorio, por hacerme ver mis errores y cómo solucionarlos, mis debilidades y cómo superarlas, por inspirarme a evolucionar y no ponerme límites, por su paciencia y su afecto y por demostrarme que si uno quiere puede.

A mi familia, sin ellos sí que no estaría escribiendo esto, papá, mamá, Lidia, gracias por protegerme y cuidarme cada día, gracias Marina y Alba por sacarme siempre una sonrisa y gracias Pablo por cuidar de mis tres chicas. A Rocío por su energía desbordante. A las Amparitos por no abandonarme nunca, y por resumirme chats de 200 WhatsApp 🤖, a Marina por enviarme su tesis desde Nueva York, a Lucía y Gloria por organizar una despedida y una boda, a María por su llamada, a Ale por darme techo y no dejarme sola en un momento difícil, no sé qué hubiera hecho sin ti, y a Sheila por sus “Buenos días”, por hacerme pasar muy buenos momentos pero sobre todo por estar ahí en los no tan buenos. Gracias a todos los que no os he nombrado pero que sabéis que en algún momento me habéis hecho ver un rayo de luz. Gracias Cristián por tu ánimo, por haberme impulsado a seguir adelante, por tu paciencia extrema, por conocerme mejor que nadie y saber lo que pensaba sin hablar.

La inspiración existe, pero tiene que encontrarte trabajando.
Pablo Picasso

1. ABREVIATURAS

ϵ	Señal de encapsidación ϵ
3'- ϵ	Extremo 3' de la señal de encapsidación ϵ
4NT	Primer de cuatro nucleótidos TGAA
5'- ϵ	Extremo 5' de la señal de encapsidación ϵ

A

A1-A7	Subgenotipos del genotipo A del virus de la hepatitis B
aa	Aminoácidos
ADN	Ácido desoxirribonucleico
ADNbc	ADN bicatenario
ADNccc	ADN del virus de la hepatitis B circular covalentemente cerrado
ADNmc	ADN monocatenario
ADNpol/RT	Actividad ADN polimerasa y retrotranscriptasa de la proteína P del virus de la hepatitis B
ADNrc	ADN del virus de la hepatitis B relajado circular
ADN-VHB	Genoma del VHB
ADV	Adefovir
AG	Aparato de Golgi
ALT	Alanina aminotransferasa
ARNpc	ARN preCore
ARNpg	ARN pregenómico
Ag	Antígeno
anti-HBc	Anticuerpos contra la proteína core del virus de la hepatitis B
anti-HBe	Anticuerpos contra el antígeno e del virus de la hepatitis B
anti-HBs	Anticuerpos contra las proteínas de superficie del virus de la hepatitis B
ARNm	ARN mensajero
AS	Sitio aceptor para el primer 4NT

B

B1-B9	Subgenotipos del genotipo B del virus de la hepatitis B
BER	<i>Base excision repair</i>

C

C1-C16	Subgenotipos del genotipo C del virus de la hepatitis B
C	Región core del genoma del virus de la hepatitis B
cB	Células B del Sistema Inmune
cccADN	ADN circular cerrado covalentemente
CD	Célula dendrítica
Cp	Promotor del preCore/pregenómico
cT	Células T del Sistema Inmune
Ct	Extremo carboxi terminal
CURS	<i>Core upstream regulatory sequence</i>

D

D1-D9	Subgenotipos del genotipo D del virus de la hepatitis B
DAAs	<i>Direct acting antivirals</i>
del	deleciones
DDB1	<i>DNA damage-binding protein 1</i>
DNMTs	ADN Metiltransferasas

D

dNTPs	desoxinucleósidos trifosfato
DR1/DR2	Repeticiones directas 1 y 2
dsIADN	ADN lineal de doble cadena
DVHB	Virus de la hepatitis B del pato (del inglés <i>Duck</i>)

E

emPCR	PCR en emulsión
ERH	Elementos de respuesta a hormonas
ESLD	<i>End-stage liver disease</i>
ETV	Entecavir

F

F1-F4	Subgenotipos del genotipo F del virus de la hepatitis B
FT	Factor de transcripción

G

Gen/A	Genotipo A del virus de la hepatitis B
Gen/B	Genotipo B del virus de la hepatitis B
Gen/C	Genotipo C del virus de la hepatitis B
Gen/D	Genotipo D del virus de la hepatitis B
Gen/E	Genotipo E del virus de la hepatitis B
Gen/F	Genotipo F del virus de la hepatitis B
Gen/G	Genotipo G del virus de la hepatitis B
Gen/H	Genotipo H del virus de la hepatitis B
Gen/I	Genotipo I del virus de la hepatitis B
Gen/J	Genotipo J del virus de la hepatitis B
GRE	Elementos de respuesta a glucocorticoides

H

HBC	Hepatitis B Crónica
HBcAg	Antígeno core del virus de la hepatitis B
HBeAg	Antígeno e del virus de la hepatitis B
HBsAg	Antígeno de superficie del virus de la hepatitis B
HBSP	<i>Hepatitis B spliced protein</i>
HBx	Proteína X del virus de la hepatitis B
HBwx	Proteína X completa del virus del virus de la hepatitis B
HCC	Hepatocarcinoma
HCC-VHB	Hepatocarcinoma asociado a la infección por el virus de la hepatitis B
HR	Recombinación homóloga
HSC	<i>Hepatic Stellate Cells</i>
HTAs	<i>Host targeting agents</i>

I

I1-I2	Subgenotipos del genotipo I del virus de la hepatitis B
IFN α	Interferón α
indel	variantes con inserciones y/o deleciones
Inr	Elementos iniciadores de la transcripción
ins	inserciones
IOB	Infección oculta por el virus de la hepatitis B

K

Kb	Quilobases
kDa	Kilodalton

L

LAM	Lamivudina
LdT	Telbivudina
LTC	Linfocitos T citotóxicos
LHBs	Proteína de superficie del virus de la hepatitis B grande

M

MEC	Matriz extracelular
Mf	Frecuencia de mutación
MHBs	Proteína de superficie del virus de la hepatitis B mediana
MHR	<i>Major hydrophilic region</i>
miARN	Micro ARNs
MIR	<i>Major Immunodominant Region</i>
MMPs	Metaloproteasas de matriz
MVB	Cuerpos multivesiculares

N

NGS	<i>Next generation sequencing</i>
NNGS	<i>Next-next generation sequencing</i>
NK	Células <i>Natural Killer</i>
NRE	<i>Negative regulatory element</i>
NRREs	<i>Nuclear receptor-responding elements</i>
nt	Nucleótido
Nt	Extremo amino terminal
NTPs	Nucleósidos trifosfato
NUCs	Antivirales orales inhibidores análogos de nucleós(t)idos

O

OMS	Organización Mundial de la Salud
ORF	Marco de lectura abierto

P

P	Proteína codificada por el ORF P
PCB	Promotor básico del core
PC	Región preCore del genoma del virus de la hepatitis B
PC/C	Región preCore/Core del genoma del virus de la hepatitis B
PCR-RFLP	Análisis de fragmentos de restricción de longitud polimórfica
PEG-IFN	Interferón pegilado
Pi	Diversidad nucleotídica
PKC	Proteín kinasa C
PPI	Pirofosfato inorgánico
PRE	Elemento regulador post-transcripcional
preS/S	Región preS/superficie del genoma del virus de la hepatitis B
PT	Dominio que codifica la proteína terminal de la polimerasa del virus de la hepatitis B
PTP	<i>PicoTiterPlate</i>

Q

QS Cuasiespecies del virus de la hepatitis B

R

r Redundancia terminal de la hebra(-) del ADNrc
RE Retículo endoplasmático
RH Dominio Ribonucleasa H de la polimerasa del virus de la hepatitis B
RN Superfamilia de los receptores nucleares
RP Ribonucleoproteína
ROS Especies reactivas de oxígeno
RT Retrotranscriptasa
RT pol Dominio retrotranscriptasa de la polimerasa del virus de la hepatitis B
RVS Respuesta virológica sostenida

S

SHBs Proteína de superficie del virus de la hepatitis B pequeña
SI Sistema inmune
SLN Señal de Localización Nuclear
Sn Entropía de Shannon normalizada
spVHB Variantes de splicing del VHB
SV Partículas subvirales

T

TDF Tenofovir disoproxil fumarato
TDP Tirosil-ADN fosfodiesterasas
TEM Transición epitelio-mesénquima
Th Linfocitos T *helper*
TLR *Toll like receptor*
TM Transmembrana
TOP Topoisomerasas

U

UDPS *Ultra-deep pyrosequencing*
URR *Upper regulatory region*

V

V Partículas virales completas del virus de la hepatitis B
VBK *Virological Breakthrough*
VIH Virus de la Inmunodeficiencia Humana
VHB Virus de la Hepatitis B
VHC Virus de la Hepatitis C
VHD Virus de la Hepatitis D

X

X Región X del genoma del virus de la hepatitis B

Y

Y Residuo de tirosina en el dominio PT del virus de la hepatitis B

Introducción

2. INTRODUCCIÓN

2.1 VIROLOGÍA DEL VIRUS DE LA HEPATITIS B

2.1.1 HISTORIA

El virus de la hepatitis B (VHB) fue descrito por primera vez por Alter y Blumberg en 1965 con la detección de un nuevo antígeno en sangre de un aborigen australiano y que fue llamado antígeno Australia [1], posteriormente se supo que ese antígeno correspondía al antígeno de superficie (HBsAg) del VHB [2]. En 1970, Dane *et al.* detectaron las partículas virales mediante microscopía electrónica [3] y en 1979 el genoma viral fue secuenciado en su totalidad [4].

2.1.2 CLASIFICACIÓN

El VHB es un miembro de la Familia *Hepadnaviridae*, género *Orthohepadnavirus* (*Hepa*; infectan al hígado y *dna*; virus de ácido desoxirribonucleico), cuyo hospedador natural es el ser humano aunque se han aislado virus similares en simio, marmota (woodchuck hepatitis virus [WHV]), ardillas (ground squirrel hepatitis virus [GSHV]), garzas (heron hepatitis B virus), patos (duck hepatitis B virus [DVHB]), gansos (goose hepatitis B virus) y grullas (crane hepatitis B virus) [5].

2.1.3 LA PARTÍCULA VIRAL

Las partículas virales circulantes en sangre, de entre 40-42 nm de diámetro, están compuestas por una envuelta externa lipídica y una cápside interna proteica de entre 30-32 nm, dentro de la cual se encuentra el ADN viral (ADN-VHB), la polimerasa viral y otras proteínas del hospedador que participan en el ciclo replicativo (Figura 1). La envuelta, derivada de la célula a la que infecta, está formada por una bicapa de fosfolípidos y contiene el HBsAg, que engloba tres tipos diferentes de glicoproteínas de superficie; grande (LHBs), mediana (MHBs) y pequeña (SHBs) [6]. La nucleocápside está compuesta por 180-240 monómeros de una única cadena polipeptídica de 20 kDa, el antígeno core (HBcAg) del VHB y tiene una conformación icosaédrica [7].

Además el HBsAg se puede ensamblar en partículas lipoproteicas sin nucleocápside, llamadas partículas subvirales (SV), no infecciosas y que pueden tener dos morfologías; esféricas o filamentosas (Figura 2). Una partícula esférica de 22 nm puede contener hasta 100 monómeros de HBsAg, que se ensamblan en las membranas del compartimento post-Retículo

endoplasmático (RE), pre-Aparato de Golgi (AG)[8] y que en el suero de una persona infectada pueden ser entre 10^4 - 10^6 veces más abundantes que las propias partículas virales. Las partículas filamentosas, de unos 22 nm de diámetro y longitud variable, son menos abundantes que las esféricas pero aun así hay 100 veces más que de las virales [9].

Figura 1: Viriones de VHB circulantes en sangre. Imagen modificada de © 2002 James A. Perkins, Medical and Scientific illustrations, <http://people.rit.edu/japfaa/infectious.html>.

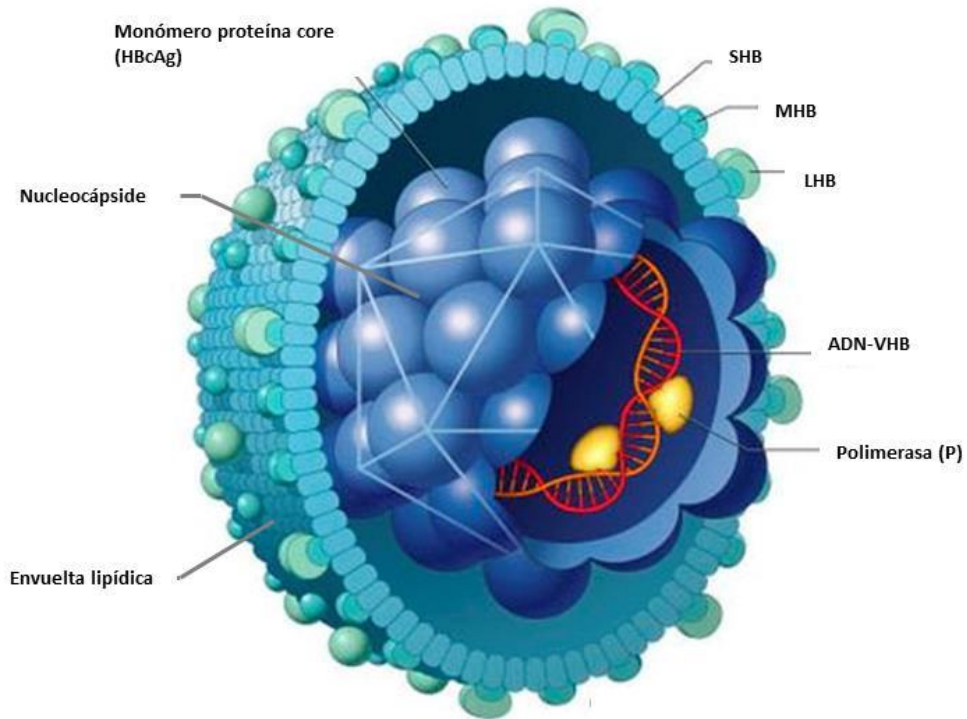
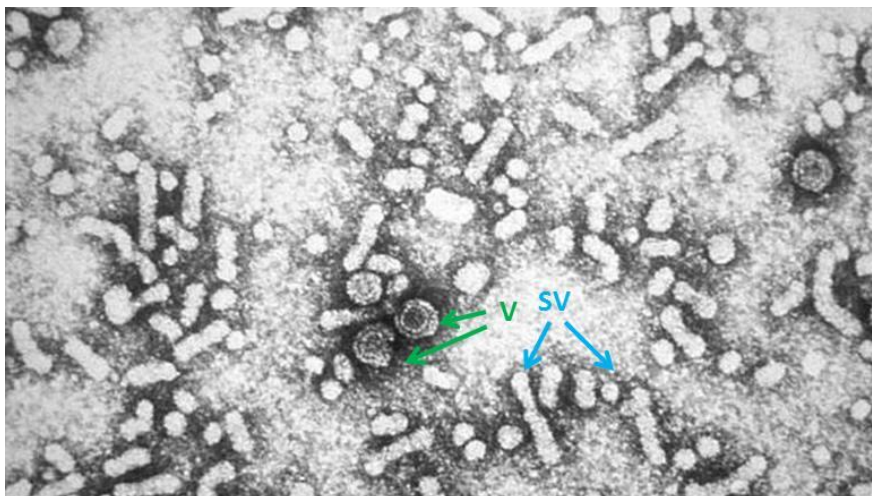


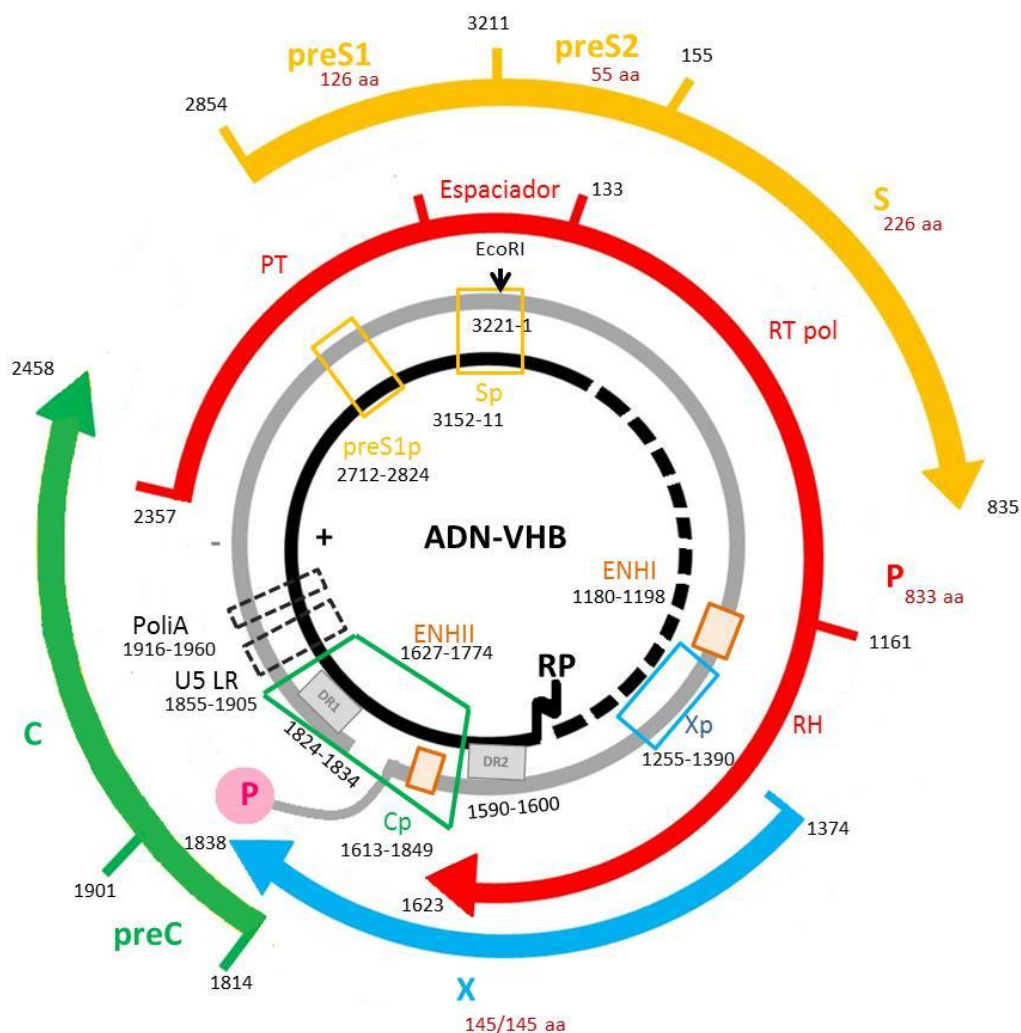
Figura 2: Partículas subvirales (SV) y viriones (V) completos del VHB con microscopía electrónica. Imagen modificada de http://www.abc.es/salud/enfermedades/abci-virus-hepatitis-dificil-erradicar-organismo-201603161651_noticia.html.



2.1.4 ORGANIZACIÓN DEL GENOMA VIRAL Y PROTEÍNAS CODIFICADAS

El genoma viral consiste en una única molécula de ADN circular bicatenario parcial, constituido por una hebra larga negativa de 3,2 kb y una corta positiva incompleta de longitud variable (Figura 3). Dentro de la nucleocápside adopta una conformación relajada circular (ADNrc) gracias a una región cohesiva en los extremos 5' de ambas cadenas, las repeticiones directas 1 (DR1) y 2 (DR2). Además la polimerasa viral (P) permanece unida covalentemente al extremo 5' de la hebra(-) y un pequeño ribonucleótido (RP) al 5' de la hebra(+). Dentro del núcleo del hepatocito el ADN viral adopta una conformación circular covalentemente cerrada (ADNccc). El ADNccc se pliega como un minicromosoma junto con histonas y proteínas no histonas celulares, así como proteínas virales, y permanece durante toda la vida de la célula actuando como molde para la síntesis de los ARN virales [10].

Figura 3: Organización del genoma del VHB (ADN-VHB) dentro de la partícula viral. Se representa la hebra(+) en negro con el ribonucleótido (RP) unido al extremo 5' y la hebra(-) en gris con la polimerasa (P) unida al extremo 5'. Aparecen representados los cuatro marcos de lectura abiertos (ORF), preS/superficie (preS/S) en naranja, polimerasa (P) en rojo, preCore/Core (PC/C) en verde y X en azul. También aparecen diferentes elementos reguladores.



2.1.4.1 ORFs

El genoma del VHB posee cuatro marcos de lectura abiertos (ORF) altamente solapados; preS/Superficie (preS/S), Polimerasa (P), preCore/Core (PC/C) y X, cuya transcripción por la ARN polimerasa II del hospedador da lugar a cinco ARNm; preS1, preS2/S, pregenómico (ARNpg), preCore (ARNpc) y X, bajo el control de cuatro promotores virales (preS1p, Sp, Cp y Xp), que se traducen en siete proteínas virales; tres tipos de proteínas de superficie (SHBs, MHBs y LHBs), la polimerasa viral (P) y los antígenos/proteínas e (HBeAg), core (HBcAg), y X (HBx) (Tabla 1).

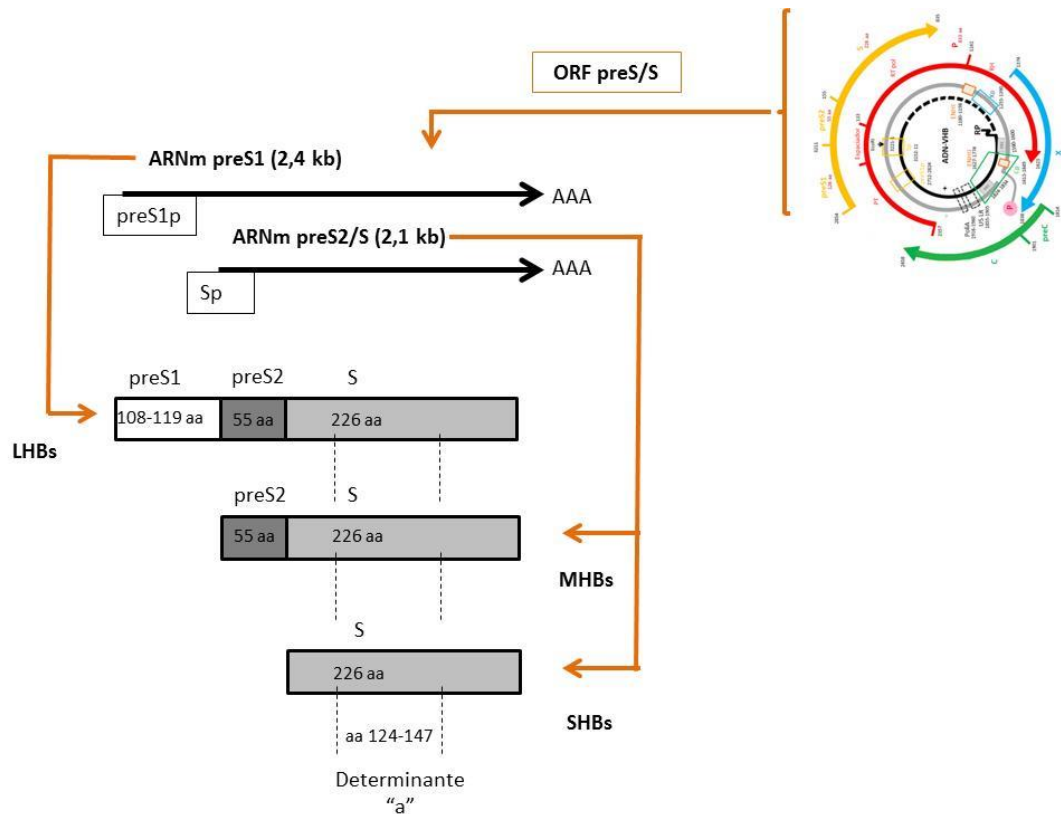
Tabla 1: ORFs, promotores, transcritos y proteínas del VHB. nt: nucleótido en el genoma del VHB.

ORF	Posición	Promotor	Posición	ARNm	Longitud (kb)	Proteína	PM (kDa)
preS/S	nt 2854-835	preS1p	nt 2712-2824	preS1	2,4	LHBs	39
		Sp	nt 3152-11	preS2/S	2,1	MHBs	33
						SHBs	24
P	nt 2357-1623	preCore/ pregenómico	nt 1613-1849	ARNpg	3,5	Polimerasa	90
preCore/Core	nt 1814-2458			preCore	3,5	HBeAg	16
				X	nt 1374-1838	X	0,9

2.1.4.1.1 ORF preS/S

El marco de lectura preS/S codifica tres proteínas de superficie diferentes estructuralmente relacionadas, que son sintetizadas desde tres codones de inicio alternativos en fase; LHBs desde la posición nt 2848, MHBs desde la posición nt 3205 y SHBs desde la posición nt 155 así pues, las tres comparten el mismo extremo carboxi terminal (Ct) pero tienen diferente longitud en el extremo amino terminal (Nt) (Figura 4). Las proteínas son traducidas desde dos ARNm subgenómicos, el más largo (ARNm preS1) codifica la LHBs, mientras que el ARNm preS2/S codifica las proteínas MHBs y SHBs desde codones de inicio separados (Figura 4). La síntesis de MHBs y SHBs está regulada por un potente promotor, no específico del hígado, Sp, mientras que la transcripción de la LHBs está controlada por un promotor más débil, hepato-específico, preS1p [11]. Todas tienen la región principal hidrofílica MHR (*Major hydrophilic region*) entre los aminoácidos (aa) 99-169, donde se localiza el determinante “a” (aa 124-147), principal dominio antigénico diana para la respuesta neutralizante de anticuerpos de células B (cB) tras una infección o la vacunación [21].

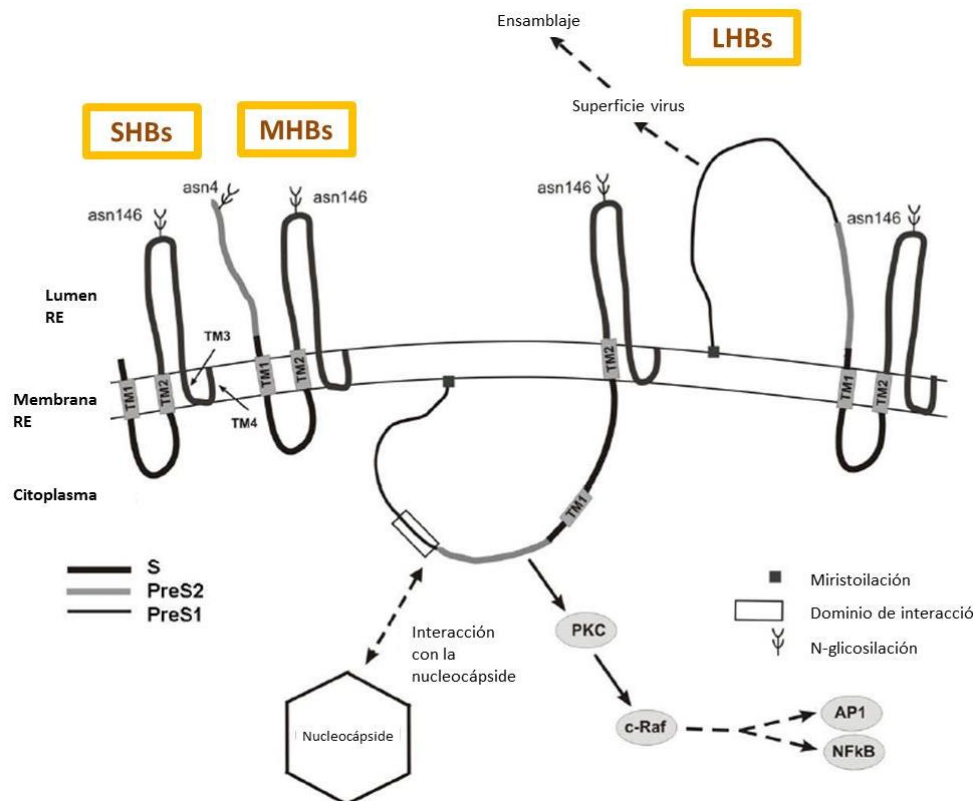
Figura 4: Representación de las proteínas de la envuelta del VHB. El ADN-VHB representado en la parte superior derecha se transcribe a dos ARNm: ARNm preS1 desde el promotor preS1 (preS1p) y ARNm preS2/S desde el promotor S (Sp). La traducción del ARNm preS1 da lugar a la proteína LHBs, la traducción del ARNm preS2/S da lugar a las proteínas MHBs y SHBs. Se muestra el determinante “a” común a todas ellas.



Las proteínas de superficie llevan a cabo diferentes acciones durante la morfogénesis y liberación de las partículas virales y sus funciones específicas parecen estar estrictamente relacionadas con su topología transmembrana (Figura 5). Son proteínas integrales de membrana ancladas en la superficie de la envuelta de las partículas virales a través del dominio S, común a todas ellas. La inserción en la membrana de este dominio se produce durante su traducción en la membrana del RE. Se inicia por una secuencia señal Nt (aa 8-22) que no se escinde y que constituye el primer dominio transmembrana (TM1), una segunda señal, que forma el segundo dominio transmembrana (TM2) (aa 80-90), dirige la translocación de la cadena creciente a través de la membrana del RE hacia el lumen, hasta que comienza, alrededor del aa 170, el extremo Ct hidrofóbico del dominio S, localizado dentro de la membrana del RE. La región Ct (aa 170-226) se considera como una o dos regiones transmembrana (TM3/4) (Figura 5). La SHBs es la proteína más abundante tanto en las partículas subvirales como en los viriones, LHBs sólo represente el 2% de las proteínas esféricas subvirales y MHBs junto con LHBs se encuentren en mayor proporción en viriones y partículas filamentosas subvirales, pero aun y todo sólo representen aproximadamente el 20% del total de las proteínas de la envuelta [11].

La LHBs muestra una biosíntesis inusual, resultando en una topología de membrana dual para las LHBs maduras. Las secuencias preS1 y preS2 no se translocan co-traduccionalmente al lumen del RE, y por tanto quedan inicialmente en la cara citosólica del RE [12]. La señal que determina la inserción co-traduccional de esta proteína es el TM2, por tanto el dominio TM1 no se inserta en la membrana del RE. Después de la traducción, aproximadamente el 50% de las LHBs sufren una reorientación topológica post-traduccional por la que su extremo Nt se transloca al lumen del RE [12]. En estas proteínas LHBs el dominio TM1 es integrado en la membrana del RE y el dominio preS1-preS2 mira al lumen del RE, y por lo tanto hacia el exterior de las partículas virales (Figura 5). Dada la orientación de su dominio preS, esta forma de la LHBs es responsable del reconocimiento de la célula huésped y de la infectividad del virus, a través de su receptor específico identificado recientemente [13], el péptido co-transportador de sodio-taurocolato (NTCP). Cuando el dominio preS1-preS2 es proyectado en el citoplasma, LHBs juega un papel esencial en la morfogénesis del virus, interaccionando con la nucleocápside, y está asociado con la función adicional de preS2 como proteína reguladora. En esta conformación citosólica, el dominio preS2 se une y activa la protein quinasa C (PKC), resultando en la activación de la cascada de señalización c-Raf/MEK que controla la expresión de una variedad de promotores celulares y virales.

Figura 5: Esquema de la estructura de las proteínas de superficie del VHB. Se indican los residuos Asn que pueden ser glicosilados en cada una de ellas [14].



2.1.4.1.2 ORF P

El gen de la polimerasa (P) es el más largo del VHB, cubre casi el 80% del genoma, (nt 2357-1623) y tiene una capacidad de codificación de alrededor de 830 aa, solapa con la totalidad del gen preS/S, la porción Ct del preCore/Core y la porción Nt del X (Figura 3). Codifica una única proteína de 90 KDa multifuncional implicada en la encapsidación, retrotranscripción y posterior degradación del ARNpg y en la síntesis de ADN (ADN pol/RT) (Ver apartado 2.1.5 *Ciclo replicativo*). Está compuesto por cuatro dominios que incluyen un dominio Nt proteína terminal (PT), una región espaciadora, de longitud variable según el genotipo, un dominio ADN polimerasa/retrotranscriptasa (RT pol) y un dominio Ct ARNasa H (RH) (Figura 3) [15,16]. Pero su estructura tridimensional no ha sido dilucidada, debido a la dificultad de aislar suficiente cantidad de proteína P pura y soluble, no obstante mediante análisis genéticos y bioinformáticos se ha demostrado que el dominio RT de Hepadnavirus es homólogo a los correspondientes dominios RT de retrovirus. Por eso los modelos propuestos se basan en esta homología con la secuencia de la RT del Virus de la Inmunodeficiencia Humana (VIH) (Figura 6A) [10,17]. En el dominio RH la homología de secuencia con el dominio ARNasa H de transcriptasas reversas de retrovirus y otros retroelementos también es significativa, aunque baja [18,19]. Por otro lado los motivos importantes y residuos críticos para varias actividades han sido identificados por estudios genéticos y bioquímicos (Figura 6B) [20,21].

Figura 6: Proteína P del VHB. **A:** Modelo tridimensional teórico propuesto para la proteína P del VHB. Los colores muestran el grado de conservación, los aa conservados se agrupan en el core catalítico, se indica el motivo YMDD [22]. **B:** Se muestra el dominio PT, espaciador, RT pol y RH, los motivos importantes (T3 y RT-1) y residuos críticos. Los dominios A-F están conservados en las diferentes RT virales, las regiones para la unión al ARN se muestran en verde [20]. *Función verificada en el DVHB pero no en VHB. **El sitio activo YMDD es requerido para el cebado proteico y la síntesis de ADN.

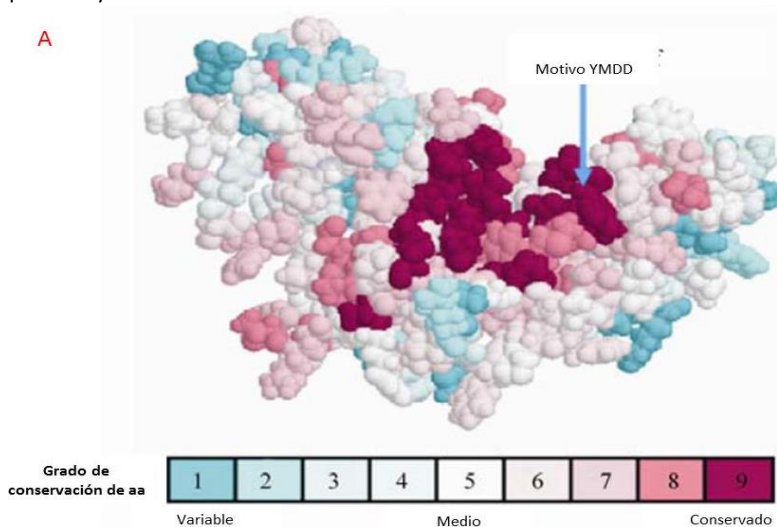
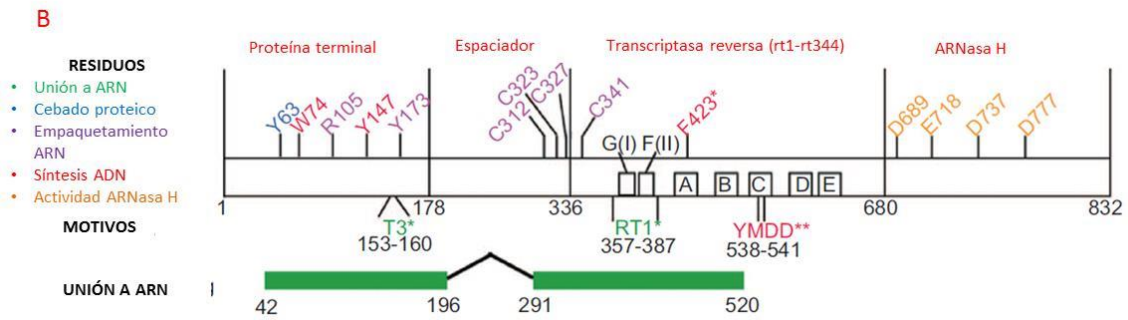


Figura 6: Continuación.



Dominio PT (nt 2307-2862)

Identificado originalmente por su unión al extremo 5' de la hebra(-) del ADN viral (Figura 3), se requiere para la unión a la señal de encapsidación épsilon (ϵ) (Ver apartado 2.1.4.2.2 *La señal de encapsidación ϵ*), el empaquetamiento del ARN y el cebado proteico (Ver apartado 2.1.5 *Ciclo replicativo*) y es una característica exclusiva de las RT de Hepadnavirus [20]. Localizado en el extremo Nt de la proteína P, proporciona el residuo de tirosina (Y63) al que se une covalentemente el primer nt de la hebra(-) durante el inicio de la replicación, el grupo OH fenólico de la tirosina cumple la función que de normal realiza el extremo 3' hidroxil de un primer de ácidos nucleicos [17,21].

El screening genético en el DVHB ha identificado además un motivo T3, una secuencia corta cerca del extremo Ct que forma parte de una región más larga, expuesta transitoriamente en la superficie de la proteína tras la activación por chaperonas dependiente de ATP, y que se cree contribuye de manera directa a la unión a ϵ . Como el dominio RT pol también es importante para la unión a ϵ , se cree que el dominio T3 interacciona con el dominio RT pol en una región llamada RT-1 (Figura 6B), pero no hay evidencia directa de esta interacción. En el VHB se ha comprobado que T3 y RT-1 pueden unirse al ARN y que ambos contribuyen a la unión proteína P-ARNpg y a la encapsidación [23].

Región espaciadora (nt 2863-129)

Situada entre el dominio PT y RT pol, gran parte de su secuencia puede sufrir mutagénesis sin que se altere la función de la proteína P, sin embargo tres residuos de cisteína en la región Ct y otro en la región Nt del dominio RT pol se requieren para el empaquetamiento del ARN (Figura 6B) [20,21].

La función exacta es desconocida pero, por un lado estas cisteínas podrían ser parte de un dedo de Zn que coordina la unión a ϵ , por otro podría tener un importante papel

estructural en las funciones de P, como la formación de puentes disulfuro [20], o podría hacer el dominio PT más flexible, facilitando el acercamiento de la PT a ϵ [21].

Dominio RT pol (nt 130-1161)

Su estructura ha sido deducida por su alta homología con la RT del VIH-1 y la información funcional se basa en los datos de resistencias a fármacos, el dominio RT es la diana de los antivirales orales inhibidores análogos de nucleos(t)idos (NUCs) [17,20].

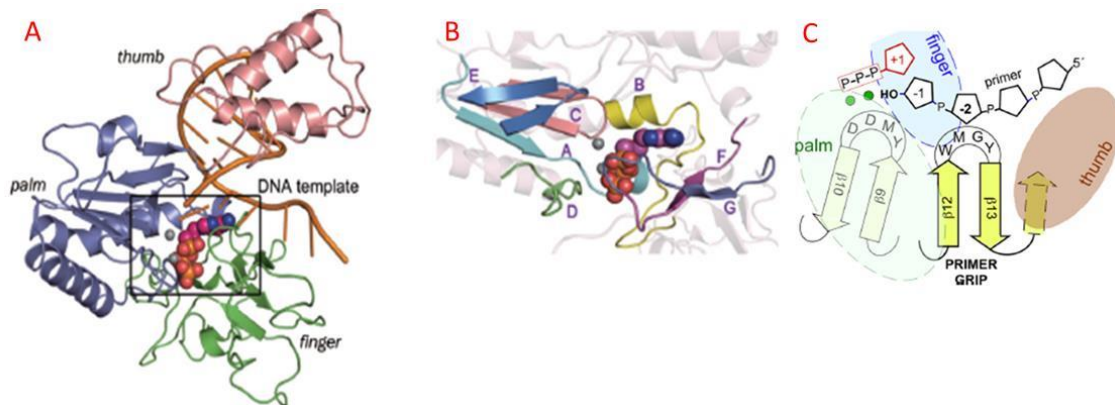
Al igual que la RT del HIV-1, la retrotranscriptasa del VHB tiene una configuración de mano derecha con tres subdominios típicos palma (*palm*), índice (*finger*) y pulgar (*thumb*) (Figura 7A) [20,24]. La posición de los subdominios *finger* y *thumb* crea una ranura en el subdominio *palm* a través del cual pasa la cadena 5' en extensión, este subdominio además contiene el motivo catalítico YMDD, común a todas las RT y que es el sitio activo que cataliza la adición de nt al DNA en el proceso de síntesis. Los nt y NUCs se unen adyacentes al extremo 3' de la hebra naciente. Los subdominios *finger* facilitan la interacción con los nt entrantes y el molde al que están unidos. El subdominio *thumb* podría tener una función en el posicionamiento, procesado y translocación del dúplex de ADN [25]. Existe una nomenclatura consenso (rt) de los aa del dominio, el inicio está en el motivo muy conservado EDWGPCDEHG y contiene 344 aa en total, así el dominio catalítico está en las posiciones rt203-rt206 (Y₂₀₃M₂₀₄D₂₀₅D₂₀₆) [26].

El dominio RT pol tiene siete dominios funcionales: A-E, más el G (o subdominio I) y el F (o II) (Figura 6B y 7B).

- Dominio A (rt75-rt90): Forma parte del bolsillo de unión a los dNTPs [27].
- Dominio B (rt159-rt182): Formado por una hélice alfa, tiene un bucle que interviene en el correcto posicionamiento de la cadena molde dentro del core catalítico [28].
- Dominio C (rt200-rt210): Junto con A participa en la interacción con los dNTPs y la formación del enlace fosfodiéster en la cadena de ADN en el proceso de síntesis. Incluye el centro activo YMDD [22].
- Dominio D (rt230-rt241): Está fuera de la región de unión a dNTPs pero interacciona con residuos del dominio A [29].
- Dominio E (rt247-rt257): Contiene la región de interacción con el primer de ADN y es importante para la replicación, específicamente actúa en la interacción de P con ϵ y en la polimerización del ADN, interviniendo en el posicionamiento correcto de la cadena molde en la región de unión a dNTPs [20]. Su bucle también contacta con el motivo YMDD y con el segundo residuo desde el extremo 3' del primer (Figura 7C) [17].

- Dominio F (rt37-rt47) y G (rt24-rt36): Análogos a los motivos 2 y 1 de la RT del VIH-1, F no forma parte del bolsillo de unión a dNTPs pero G sí [30].

Figura 7: Modelo estructural del dominio RT pol por homología con la RT de VIH-1. **A:** Configuración similar a una mano derecha con tres subdominios *finger* (verde), *palm* (azul) y *thumb* (salmón). El dúplex molde de ADN-primer en naranja y cCTP (átomos de P, N y O en naranja, azul y rosa, respectivamente). **B:** Dominios catalíticos estructuralmente conservados representados en cintas: Dominio A (cian), B (amarillo), C (rosa), D (verde), E (azul), F (magenta) y G (gris oscuro). **C:** La secuencia del dominio E forma una horquilla ($\beta 12$ - $\beta 13$) que hace de núcleo conector para otros subdominios, coordina la posición relativa de *finger*, *palm* y *thumb*, su bucle también contacta con el motivo YMDD y con el segundo residuo desde el extremo 3' del primer y esto ayuda al posicionamiento del grupo 3' OH terminal cerca del sitio catalítico para el ataque nucleofílico del dNTP entrante (*primer grip*) [17].



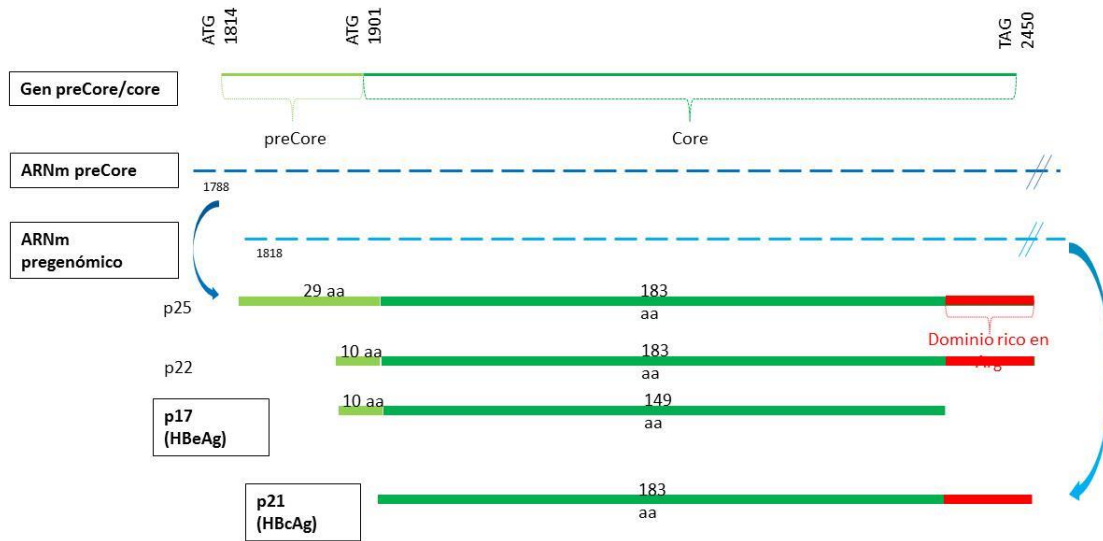
Dominio ARNasa H (nt 1162-1623)

Está localizado en el extremo Ct del gen P y degrada el ARNpg durante la replicación viral. En los 100 aa Nt se localiza el motivo DEDD (Figura 6B), la carga negativa de estos aa es importante para su actividad, coordinando la unión a iones metálicos. Además de la actividad ARNasa H, este dominio también es importante para el empaquetamiento del ARNpg [17].

2.1.4.1.3 ORF PC/C

El gen preCore/Core (nt 1814-2452) tiene dos codones de inicio ATG en fase, nt 1814 y nt 1901, la secuencia entre ambos codones se denomina región preCore (PC) y entre las posiciones 1901 y 2452 región core (Figura 8). Si la traducción comienza en el primer codón (nt 1814) se sintetiza un péptido de 25 kDa (p25) que contiene ambas regiones preCore y core. La región preCore contiene una secuencia señal que hace que p25 sea dirigido a la vía secretora y que tras proteólisis de los 19 primeros aa Nt (en el lumen del RE) y 34 últimos aa Ct (en el AG) da lugar al HBeAg, de 16-17 kDa. Si la traducción comienza en el segundo codón (nt 1901) se sintetiza el HBcAg, de 21 kDa, que forma las cápsides virales [31–34].

Figura 8: El ORF preCore/Core tiene dos codones de inicio en fase que sintetizan dos proteínas co-terminales, la proteína preCore/Core que es la precursora del HBeAg, usando el primer codón, y la proteína core desde el segundo. Las dos proteínas son sintetizadas desde diferentes transcritos el ARNpc y el ARNpg, respectivamente. Se indica el dominio rico en argininas.



HBeAg

HBeAg es secretado como una proteína soluble no estructural, y aunque no se requiere para el ensamblaje viral o la replicación, está conservado en todos los ortohepadnavirus y es importante para la infección natural *in vivo* [35].

La función exacta de HBeAg no se conoce pero es una proteína inmunoreguladora e inmunotolerógena (Ver apartado 2.2.3.3.1 *Rol de HBeAg en la inmunopatogénesis*). Clínicamente HBeAg es un índice de replicación viral, infectividad, inflamación, severidad de la enfermedad y respuesta al tratamiento y la seroconversión de HBeAg(+) a HBeAg(-)/anti-HBe(+) normalmente indica el control de la infección por el sistema inmune. Sin embargo, la aparición de mutaciones en el promotor básico del core (PBC) y PC pueden llevar a infecciones HBeAg(-) donde la replicación y la inflamación son altas (Ver apartado 2.3.2.4.1 *Variantes en el PC* y 2.3.2.4.2 *Variantes en el PBC*) [35].

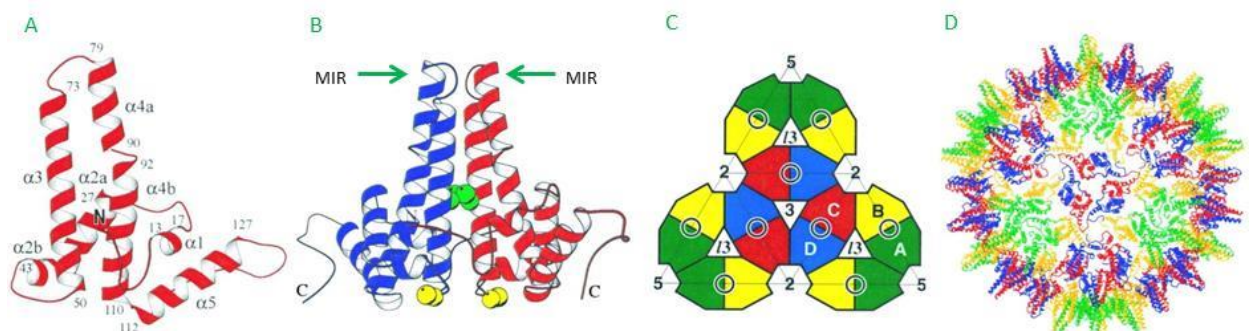
HBcAg

Aunque la secuencia de aa de HBcAg es idéntica a la de la proteína HBeAg, ambas se pliegan y tienen funciones biológicas totalmente diferentes. Las proteínas core se ensamblan formando partículas virales icosaédricas para que tenga lugar la encapsidación del ARNpg y la replicación del genoma (Ver apartado 2.1.5 *Ciclo replicativo*).

La estructura de los monómeros core (Hbc) está dominada por una larga horquilla α -helicoidal (Figura 9A), la asociación de dos horquillas anfipáticas α -helicoidales resulta en la formación de un dímero, unido a través de puentes disulfuro intermoleculares entre cisteínas (Cys 61) (Figura 9B). Estos dímeros se pueden ensamblar en dos tipos de partículas, grandes y pequeñas, con un diámetro de 34 y 30 nm y una simetría con número de triangulación de empaquetamiento T=4 y T=3, conteniendo 240 y 180 moléculas de Hbc, respectivamente. Se forma entonces un haz de cuatro hélices, dos de cada monómero, con una protusión que forma una espículas de aproximadamente 25 Å de largo y 20 Å de ancho (Figura 9C y 9D), el extremo Ct de cada Hbc, rico en argininas (Figura 8), queda orientado hacia el interior de la cápside e interacciona con las cargas negativas del ARNpg, mientras que el extremo Nt queda en el exterior, la secuencia de aa 76-81 localizado en la punta de las espículas constituye la mayor parte de la región inmunodominante principal MIR (*Major Immunodominant Region*) de la partícula core. La cápside está fenestrada con poros de diámetro 12-15 Å, que podrían ser esenciales para el acceso de los nt al interior durante la síntesis de ADN [36–40].

La proteína Hbc también puede interactuar con el ADNccc, formando parte de la estructura de los minicromosomas virales [41], puede encontrarse tanto en el citoplasma como en el núcleo celular o en ambos, en parte según la fase de la hepatitis B crónica (HBC): se encuentra preferentemente en el citoplasma durante la inflamación hepática en la fase de actividad inmune, mientras que incrementa su concentración nuclear en la fase de tolerancia inmune (Ver apartado 2.2.2 *Historia natural*) [42].

Figura 9: Representación de la estructura de la cápside y sus componentes [7]. **A:** Plegamiento de los monómeros de la cápside. Estructura dominada por una larga horquilla α -helicoidal constituida por aa 50-73 ($\alpha 3$) y aa 79-110 ($\alpha 4$), la hélice $\alpha 4$ está torcida entre aa 90 y 92, con lo que la parte inferior ($\alpha 4b$) se coloca lejos de $\alpha 3$. La región Nt (aa 1-12) adopta una estructura irregular seguida por una hélice $\alpha 1$ corta (aa 13-17) y $\alpha 2$ más larga (aa 27-43), torcida después del residuo 37, que conjuntamente rodean tres lados de la horquilla. La región Ct consiste en una hélice $\alpha 5$ (aa 112-127) seguida de un bucle irregular rico en prolinas (aa 128-136) y una hebra desplegada (aa 137-142/3). Los aa de $\alpha 1$, el bucle que conecta $\alpha 1$ y $\alpha 2$, $\alpha 4b$ y $\alpha 5$ forman un core hidrofóbico conservado responsable de la estabilizar el plegamiento. **B:** Formación de los dímeros. Dos monómeros se asocian para dar lugar a dímeros compactos en los que la horquilla α -helicoidal se empaqueta para formar un haz de 4 hélices y que forman las espículas características en la superficie de la cápside. En la interfaz del dímero se forma un puente disulfuro entre las Cys61 (en verde). **C:** Representación esquemática de 4 subunidades independientes en cápsides T=4, desde una cara icosaédrica. La posición de un haz de 4 hélices aparece en círculos negros. Uno de los dímeros en azul-tojo, otro en verde-amarillo. **D:** Visión espacial del empaquetamiento de los dímeros en cápsides T=4.



2.1.4.1.4 ORF X

El ORF X es el más pequeño con una longitud de 465 pb (nt 1374-1838) y codifica para una proteína de 16,5 kDa y 154 aa. Su nombre (gen/proteína X) hace referencia a que originalmente no se conocía su función y a su falta de homología con otras proteínas. La proteína X (HBx) está conservada a lo largo de todos los Hepadnavirus de mamíferos, pero ausente en los aviares, el hecho de que estos últimos no desarrollen hepatocarcinoma (HCC) pero sí el resto sugirió para esta proteína un papel oncogénico en la infección por el VHB que se ha ido confirmando en estudios posteriores [43]. El gen X se localiza cerca del extremo cohesivo del genoma viral donde se solapa con el gen P en el extremo Nt y preCore en el Ct (Figura 3 y 10). Su secuencia nucleotídica también contiene varios elementos reguladores incluyendo el promotor del preCore/pregenómico (Cp) (nt 1613-1849), el enhancer II (ENHII) (nt 1627-1774), las DR1 (nt 1824-134) y DR2 (nt 1590-1600) y las cuatro TATA boxes-like, además contiene una región de unión a miARN (nt 1362-1383), un motivo tipo BH3 (nt 1719-1769) de interacción con proteínas anti-apoptóticas Bcl-2 y Bcl-XL y regiones epitópicas para cB (nt 1458-1517) y células T (cT) (nt 1719-1754) (Figura 10) [44–47].

Estructura proteica

La proteína HBx no ha sido cristalizada, con la excepción del dominio de interacción a DDB1 (*DNA damage-binding protein 1*) que ha sido co-cristalizado con la molécula [22], con lo cual la estructura ha sido caracterizada por ensayos menos precisos como dicroísmo circular, fluorescencia, resonancia magnética nuclear o predicción *in silico* con aproximaciones bioinformáticas (Figura 11A). Estos estudios sugieren que los 30 aa Nt son desestructurados, mientras que la porción Ct es estructurada y que hay numerosas estructuras α -helicoidales en la porción central [22]. Mediante la búsqueda en la base de datos PDB (*Protein Data Bank*) se ha visto que HBx tiene una homología estructural considerable con los miembros de la familia MUG de ADN glicosilasas (*mismatch-specific uracil DNA glycosylases*) (Figura 11B), enzimas clave en la reparación por escisión de bases BER (*Base excision repair*) y se ha confirmado mediante la modelización *in silico* de la porción de HBx que co-cristalizada con DDB1 (Figura 11C), no obstante HBx no se une a la doble hélice de ADN *in vivo* [22,43,46,48–50]. HBx contiene 9 residuos de Cys conservados (Figura 12), más del doble del número esperado para una proteína de mamíferos de su tamaño, que forman cuatro puentes disulfuro y que parecen ser importantes para su función, ya que la mutagénesis dirigida que eliminan específicamente estos residuos de Cys lleva a la pérdida de las funciones de HBx [45]. También presenta 6 Lys conservadas que podrían ser diana para la ubiquitinación [46]. No se han establecido órdenes superiores de estructura pero existen algunas evidencias de que podría formar dímeros [46].

Figura 10: ORF X completo y región X del VHB, que incluye el extremo 3' del ORF X, el extremo 5' del ORF preCore/Core, el Cp y los elementos reguladores enhancers, DRs y TATA boxes-*like* (TA1 nt 1750-1755, TA2 nt 1758-1762, TA3 nt 1771-1775 y TA4 nt 1788-1795). Se detalla también la estructura del Cp.

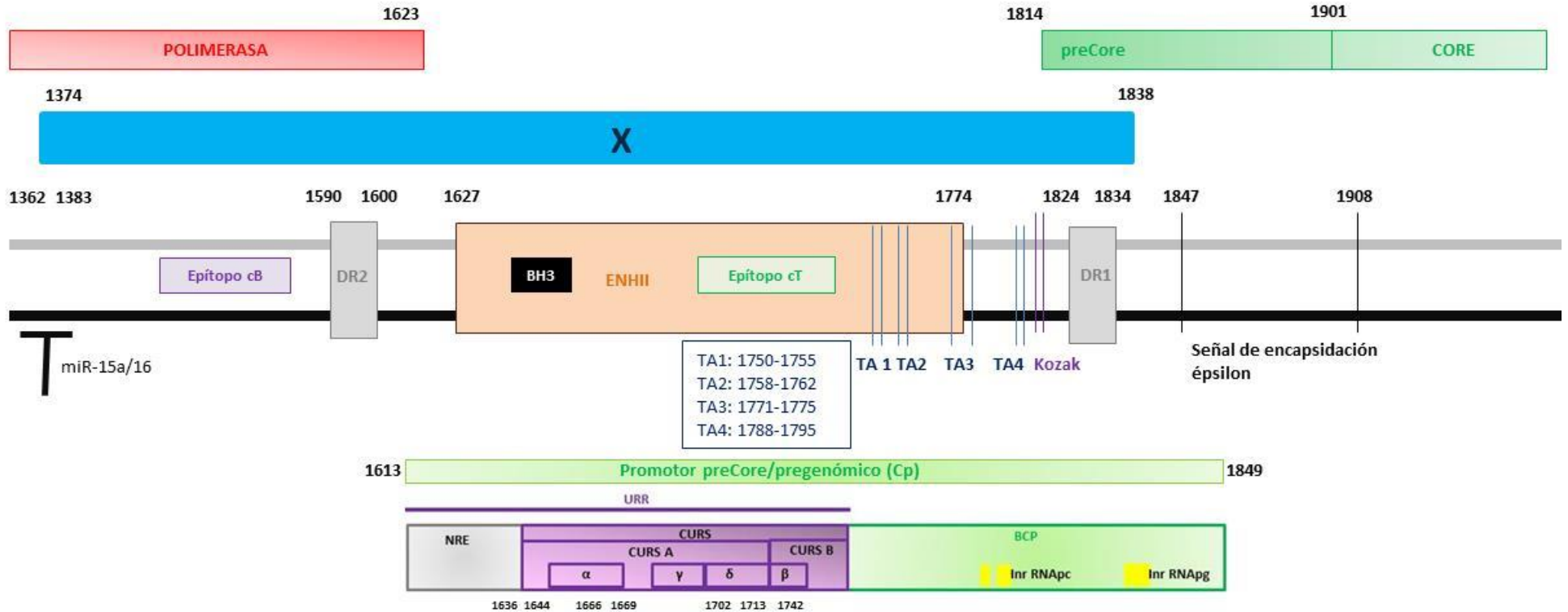
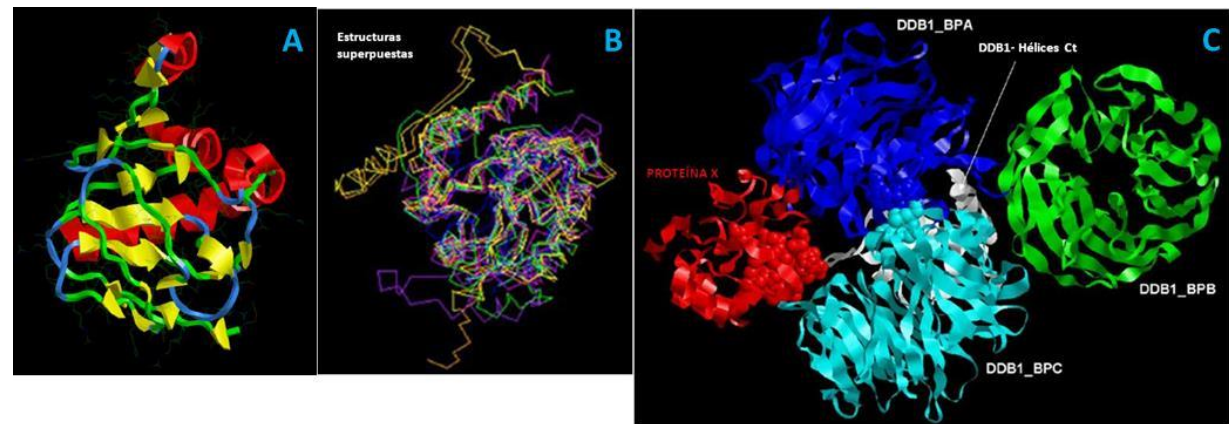


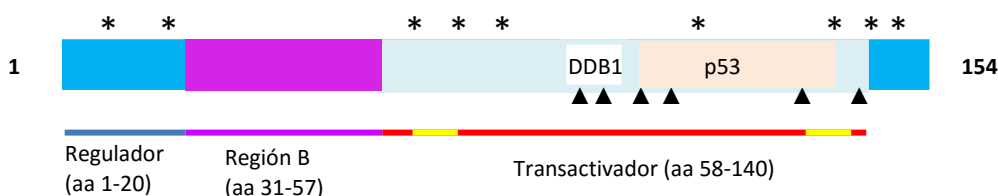
Figura 11: Modelo de la estructura 3D de HBx [22]. **A:** Modelo de la secuencia consenso del genotipo D. Rojo, amarillo, azul y verde para las hélices, láminas, giros y otros, respectivamente. **B:** Similitud 3D entre HBx y las ADN glicosilasas. La similitud estructural está restringida al dominio central conservado, los motivos oligopeptídicos flanqueantes, incluyendo el residuo Asn catalítico crítico para la actividad glicosilasa están ausentes. **C:** Modelo de interacción HBx-DDB1. HBx (aa 88-100) interacciona con DDB1 en un bolsillo creado por dos dominios hélices- β en DDB1 (BPC y BPA) con el extremo Ct de HBx saliendo del complejo. HBx en rojo, los dominios A, B y C de DDB1 en azul, verde y magenta, respectivamente y el dominio α -helicoidal Ct de DDB1 en blanco. El nonapéptido 91KVLHKRTLGG99 participa en la superficie de interacción, excepto 92V y 99G.



Dominios funcionales

En líneas generales, HBx tiene un dominio Nt regulador negativo (aa 1-20), separado por una región rica en Pro y Ser (Región B aa 31-57) y un dominio Ct transactivador (aa 58-140) (Figura 12), dentro de este, existen tres regiones altamente conservadas a lo largo de los Hepadnavirus (aa 58-84, 61-69 y 132-138), que son esenciales para la transactivación, dos de esas secuencias (aa 61-69 y 132-138) son homólogas al dominio Kunitz, una secuencia funcional esencial de inhibidores de serín-proteasas tipo Kunitz. El extremo Ct (aa 143-154) no parece ser esencial para la transactivación, ya que proteínas HBx que pierden una pequeña región Ct mantienen esta actividad [43,49,50].

Figura 12: Representación esquemática de los dominios funcionales de HBx. Se representan el dominio regulador (aa 1-30), dominio transactivador (aa 58-140), con los dos dominios tipo Kunitz en amarillo, dominio de unión al proteosoma (aa 82-154) en azul claro, que abarca la región de unión a DDB1 (aa 88-100) en blanco, y el dominio de unión a p53 (aa 102-136) en naranja. Se muestran los 9 residuos de Cys conservados (*aa 7, 17, 61, 69, 78, 115, 137, 143 y 148) y seis de Lys (▲ aa 91, 95, 113, 118, 130 y 140) [45].



Localización HBx

HBx se localiza mayoritariamente en el citoplasma pero hay una cierta cantidad detectable en el núcleo, se transporta entre ambos compartimentos a través de la vía de exportación nuclear dependiente de Crm-1 (*chromosome region maintenance 1*) gracias a una señal de exportación nuclear [44,46]. En el citoplasma HBx estimula múltiples vías de señalización, en el núcleo, HBx asociada con el ADNccc, regula la replicación y la expresión de genes del VHB e interactúa con componentes de la maquinaria basal transcripcional y factores de transcripción (FT) específicos para regular los genes celulares [49,51].

Funciones moleculares de HBx

HBx es una proteína multifuncional que está implicada en la transactivación de promotores celulares y virales, la activación de vías de señalización, el control del ciclo celular, la inducción o prevención de la apoptosis, la inhibición de la reparación del ADN celular, la proliferación celular, la autofagia/mitofagia, la degradación proteica, el metabolismo, la estabilidad genética y la adhesión celular [49,52–54] (Figura 13). A continuación se detallan sus funciones.

A. FUNCIÓN TRANSACTIVADORA

HBx es un transactivador promiscuo que puede activar una amplia variedad de promotores y enhancers celulares y virales. No se une directamente al ADN, con lo que la activación de la transcripción está mediada por interacciones proteína-proteína. Puede ejercer su función transactivadora tanto en el citoplasma, a través de vías de señalización, como en el núcleo, a través de proteínas de unión al ADN (FT y maquinaria basal de la transcripción). La activación transcripcional desencadenada por HBx es esencial para la replicación viral y está implicada en la regulación de la expresión de ciertos genes celulares cruciales para la proliferación celular, la progresión del ciclo celular, la respuesta inmune, la degradación de proteínas, la estabilidad genética, la apoptosis celular y la organización de tejidos [52].

B. FUNCIÓN EN LA REPLICACIÓN DEL VHB

HBx tiene un papel fundamental en mantener la replicación del virus a través de varios mecanismos, incluyendo la transactivación de enhancers o promotores del VHB, la inhibición del proteosoma, la activación de polimerasas por una señalización dependiente de Ca^{2+} , la inducción de autofagia y las modificaciones epigenéticas [47].

B.1 Implicación del proteosoma y DDB1 en el papel estimulante de la transcripción y replicación viral de HBx

Existen experimentos que demuestran que el aumento de la transcripción y replicación llevada a cabo por HBx es por una vía dependiente del proteosoma y que DDB1 podría estar implicado, de manera dependiente de HBx, en la transcripción y replicación del VHB.

B.1.1 HBx interacciona con el proteosoma a través de DDB1

DDB1 es una proteína multifuncional descubierta originalmente por su implicación en el reconocimiento de ADN dañado por luz UV y el reclutamiento de la maquinaria NER (*Nuclear Excision Repair*) en el sitio del daño, funciona como una proteína adaptadora para el complejo CRL4 del proteosoma, reclutando y uniendo proteínas denominadas DCAFs (*DDB1 Cullin Accesory Factor*) que a su vez reclutan sustratos a CRL4 para la ubiquitinación y la degradación por el proteosoma. DDB1 tiene un papel importante en diversos procesos celulares como la síntesis de ADN, la expresión génica, la división celular y la apoptosis [46,54]. HBx (aa 88-100) interacciona con DDB1 en la misma región a la que se unen los DCAFs (Figura 11C), con el extremo Ct protruyendo del complejo, se une a diferentes FT necesarios para la replicación viral. La interacción HBx-DDB1 está conservada a lo largo de todos los Hepadnavirus de

mamíferos lo que sugiere el importante papel de esta interacción en la replicación viral [46,54].

B.1.2 HBx interacciona con subunidades del proteosoma

Además HBx interacciona por su parte Ct con las subunidades del proteosoma PSMA7, una subunidad α del proteosoma 20S y PSMC1, una subunidad *ATPase-like* en el componente regulador 19S del proteosoma 26S, y esta interacción es necesaria para la transactivación de HBx y la replicación viral [55].

B.2 Vías de señalización dependientes de calcio

HBx estimula la replicación viral a través de la activación de vías de señalización dependientes de Ca^{2+} . HBx desencadena la liberación de Ca^{2+} desde la mitocondrial al citoplasma al interactuar con un componente del poro mitocondrial. Los niveles elevados de Ca^{2+} citoplasmático activan Pyk2, que une y activa las quinasas Src, Src activa la polimerasa del VHB estimulando la replicación viral a nivel de la transcripción reversa [51].

B.3 Autofagia

HBx también puede estimular la replicación desencadenando la autofagia, proceso de digestión celular de macromoléculas y orgánulos durante el ayuno para reciclar componentes en vías biosintéticas y producir energía. Pero para el VHB, la autofagia no lleva a un aumento en la tasa de degradación de proteínas, en su lugar, HBx se une y activa la fosfatidilinositol 3-quinasa clase III y esto *up*-regula la replicación del VHB [51].

B.4 Modificaciones epigenéticas

HBx regula parcialmente la replicación del VHB a través de modificaciones epigenéticas del ADNccc. Se ha observado que cuando el ADNccc se forma en ausencia de HBx es transcripcionalmente silente y contiene histonas no acetiladas, histona desacetilasas HDAC1 y hSiert1 y H4 metilada, por el contrario cuando se forma en presencia de HBx, el ADNccc está enriquecido en H3/H4 acetiladas, acetiltransferasa p300 y H4 hipometilada [49].

C. FUNCIÓN EN LA APOPTOSIS

Se ha descrito que HBx media en la apoptosis de diferentes maneras [46,49,52].

Existen estudios que demuestran que:

- HBx bloquea la apoptosis mediada por $TNF\alpha$, Fas, p53 o $TGF\beta$ en algunas líneas celulares.
- HBx sensibiliza a la apoptosis inducida por $TNF\alpha$ en otras líneas celulares, esto podría implicar la activación de MAPK 1 (*mitogen-activated-protein kinase 1*) y la proteína Myc.

- HBx inhibe la apoptosis inducida por p53 por asociación directa con este.
- HBx forma complejos con la proteína anti-apoptótica survivina que suprime la acción de las caspasas y la apoptosis.
- HBx induce agregación mitocondrial y liberación de citocromo c, que es indicativo de inducción de apoptosis a través de la disfunción mitocondrial.
- HBx es anti-apoptótica en cultivos primarios de hepatocitos de rata y este efecto está unido a la activación de NF- κ B.

D. FUNCIÓN EN EL CICLO CELULAR

HBx puede regular la proliferación celular aunque su impacto varía en función del sistema experimental usado. HBx puede inducir a células en quiescencia a salir de la fase G₀ y entrar en el ciclo celular, actuando a través de la activación de factores de vías de señalización Src, MAPK y JNK [46,49].

E. FUNCIÓN EN LA REPARACIÓN DEL ADN

HBx también puede regular la respuesta a ADN dañado y la reparación del ADN (NER) porque se une directamente a DDB1 y a la helicasa XBP (Xeroderma pigmentosa B) o XPD (Xeroderma pigmentosa D) en el FT general IIIH (TFHII) [49].

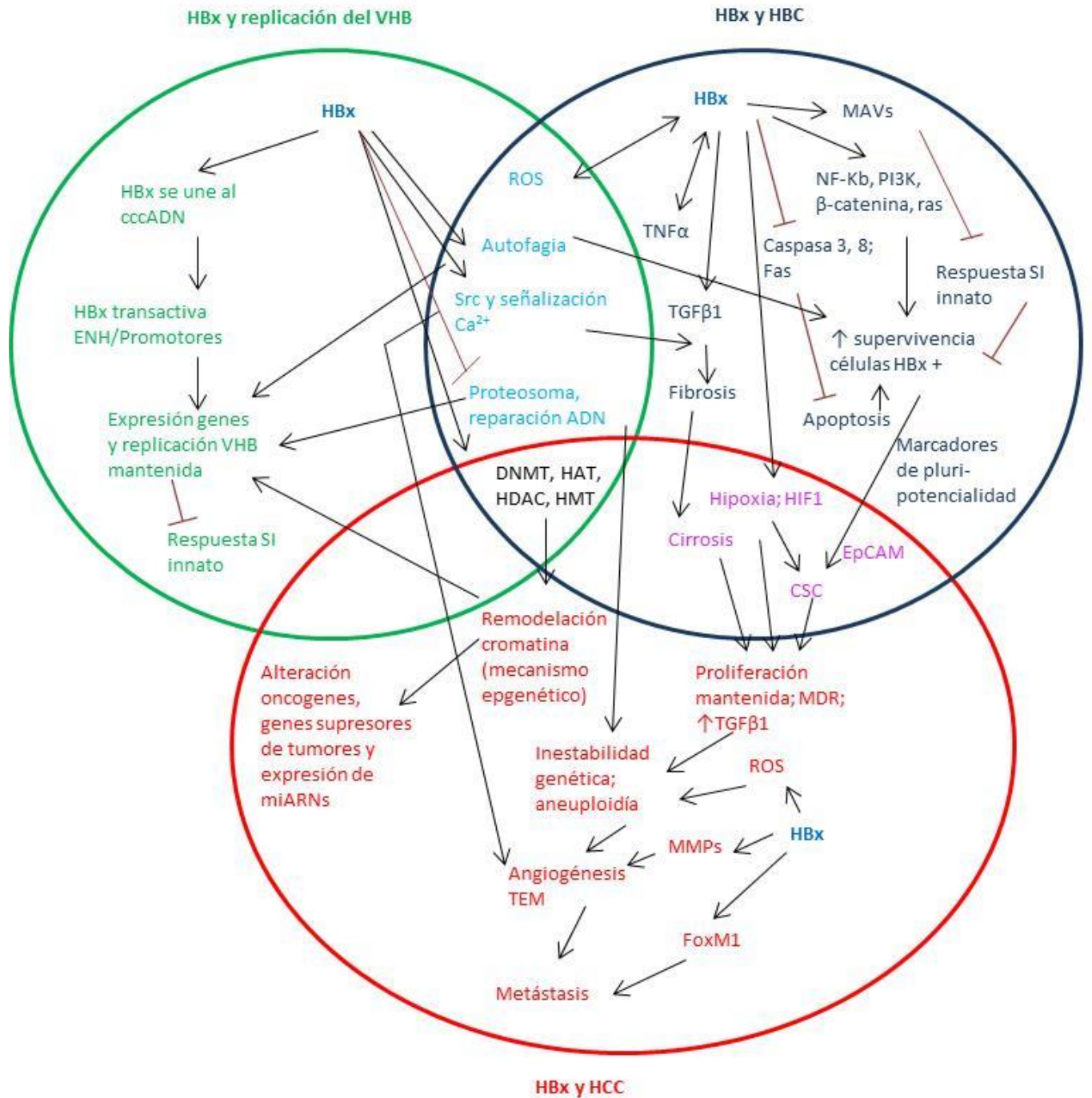
F. FUNCIÓN EN EL METABOLISMO CELULAR

HBx altera el metabolismo lipídico interactuando con PPAR- γ (*peroxisome proliferator-activated receptor- γ*) y SREBP1 (*sterol regulatory element binding protein 1*), que puede llevar a una acumulación lipídica en los hepatocitos, y aumenta la β -oxidación bajo condiciones de glucosa limitantes. También se ha relacionado con un metabolismo alterado de la glucosa, aa y ácidos nucleicos [56,57].

HBx en la infección crónica y HCC

Como se puede deducir del punto anterior HBx contribuye a la biología del VHB mediante la estimulación de la replicación, pero también a la patogénesis modulando un amplio número de procesos celulares e interactuando con una gran variedad de moléculas (Figura 13).

Figura 13: Las funciones de HBx en la replicación del VHB y en la patogénesis de la infección crónica y HCC están representadas en tres círculos solapantes, ya que muchos de los mecanismos por los que HBx estimula la replicación viral y protege las células infectadas de la destrucción mediada por el sistema inmune, también contribuyen a la patogénesis de la HBC y al desarrollo/progresión del HCC [51]. MAVs (*mitochondrial antiviral-signaling proteins*), SI: sistema inmune, ROS: especies reactivas de oxígeno.



A. HBx Y LA INFECCIÓN CRÓNICA

A.1 Inflamación

HBx promueve la inflamación crónica a través de un feed-back positivo con TNF α , las células inflamatorias expresan TNF α , que estimulan la secreción de HBx y la estabiliza y HBx *up*-regula a su vez TNF α [47].

A.2 Fibrosis

HBx promueve la fibrogénesis a través de varios mecanismos:

- Aumentando la actividad transcripcional de TGF β 1 (fuertemente pro-fibrogénico en el contexto de enfermedad hepática crónica) [51].
- Mediante la activación y proliferación paracrina y migración de células hepáticas estrelladas HSC (*Hepatic Stellate Cells*) [58].
- Induciendo la transición epitelio-mesénquima (TEM) a través de la *down*-regulación de E-cadherina (*Cdh1*) [51].

A.3 Hipoxia y angiogénesis

En los nódulos cirróticos se observa una amplia fracción de células en proliferación y no se observan triadas portales ni vena central, lo que resulta en áreas con una baja tensión de oxígeno. HBx *trans*-activa fuertemente la expresión de HDAC1, que físicamente forma complejos y estabiliza HIF1 (*hypoxia inducible factor 1*), un potente FT que promueve la angiogénesis, a través de desacetilación, de hecho en los nódulos cirróticos se han detectado grandes cantidades de HBx [51].

A.4 Células progenitoras/madre hepáticas (CPHs/CMHs)

Aunque no está claro si el VHB infecta a las células progenitoras hepáticas (CPHs), la introducción de HBx resulta en la *down*-regulación de marcadores pro-apoptóticos y la *up*-regulación constitutiva de marcadores anti-apoptóticos. Además HBx induce la aparición de factores de pluripotencialidad [51].

A.5. Desregulación respuesta innata

HBx inhibe moléculas importantes en la señalización de la inmunidad innata y *up*-regula moléculas proinflamatorias como IL-23 (que promueve la expansión de linfocitos T helper Th₁₇), ciclo-oxigenasa 2 (COX2), prostaglandina E2, IL-32 e IL-28 (que *up*-regula FasL). Esta desregulación de la respuesta inmune promueve la persistencia de los virus en el hígado dañado [51].

A.6 Apoptosis

HBx protege a células infectadas de la apoptosis mediada por TNF α a través de la activación de la señalización NF-kB, PI3K, SAPK/JNK y β -catenina (todos inhiben la apoptosis) y por inhibición de la caspasa 3 y 8, que bloquea la eliminación de hepatocitos infectados por linfocitos T citotóxicos (LTC) a través de la vía Fas [51].

B. HBx Y HCC

Existen evidencias tanto *in vitro* como *in vivo* del papel de HBx en el proceso de hepatocarcinogénesis, HBx es el antígeno viral más prevalente en tejidos de HCC de portadores del VHB, transforma líneas celulares de hepatocitos, ratones transgénicos para HBx con niveles elevados mantenidos de la proteína desarrollan HCC y en Hepadnavirus que infectan a aves donde el ORF X no está conservado no existen evidencias de HCC [44,51]. Pero su papel oncogénico es controvertido, existen evidencias de que HBx promueve el crecimiento tumoral alterando la expresión de genes celulares, modulando la apoptosis, la actividad telomerasa, la reparación del ADN dañado, el ciclo celular, la adhesión celular y las modificaciones epigenéticas [59]. Para más detalle sobre el hepatocarcinoma asociada al VHB (HCC-VHB) ver el apartado 2.2.4 *Hepatocarcinoma relacionada con el VHB (HCC-VHB)*.

B.1 HBx en la expresión de los genes celulares

HBx altera la expresión de genes celulares que contribuyen al desarrollo de HCC por una variedad de mecanismos.

B.1.1 Integración en el ADN celular

En la infección crónica por el VHB fragmentos del genoma del virus se integran en el celular, la mayoría de estas secuencias integradas sintetizan ARNm X que es traducido a HBx [51]. El gen X con frecuencia es parcialmente deletado durante el proceso de integración, causando el truncamiento Ct de HBx y estas HBx con Ct truncado juegan un papel pro-oncogénico y pro-metastásico crítico en la hepatocarcinogénesis [43,59–61].

B.1.2 Transactivación

Como se ha visto en los puntos anteriores HBx afecta a promotores y enhancers de una variedad de genes celulares reguladores del ciclo celular y de la proliferación, promoviendo la invasión celular, la angiogénesis y la metástasis [51].

B.1.3 Mecanismos epigenéticos

En el núcleo HBx puede estimular o bloquear la expresión de genes virales a través de la modulación de mecanismos epigenéticos [62].

B.1.3.1 HBx y metilación de promotores

HBx induce la *up*-regulación de la actividad de las ADN metiltransferasas 1 y 3A (DNMT1 y DNMT3A) e induce la hipermetilación del promotor de algunos genes supresores de tumores como p16^{INK4A}, E-caderina (CDH1), IGFBP-3 (*insulin-like growth factor binding protein 3*) y GSTP1 (*glutathione S-transferase P1*). Por otro lado, HBx aumenta la expresión de oncogenes a través de la hipometilación de sus promotores como: ASPP2 (*ankyrin-repeat-containing, SH3-domain-containing, and proline-rich-region-containing protein 2*), MT1F (*metallothionein 1F*), IL-4R, ALDH1 (*retinal dehydrogenase 1*), RBP (*plasma retinol-binding protein precursor*), CRBP1 (*cellular retinol-binding protein I*) y CDH6 (*cadherin 6*) [62].

B.1.3.2 HBx y acetilación/desacetilación de histonas

HBx puede ocupar físicamente el promotor de sus genes diana y aumentar el reclutamiento de los complejos reguladores de transcripción CREB-binding protein (CBP)/p300 (con actividad acetiltransferasa intrínseca), así induce la *up*-regulación de la acetilación de histonas y la expresión génica de IL-8 y PCNA. Por otro lado, HBx puede inducir la desacetilación de histonas de genes supresores de tumores [62].

B.1.3.3 HBx y miARNs

HBx puede controlar multitud de miARNs de diferentes maneras para promover la progresión del HCC (Tabla 2) [62].

Tabla 2: miARN regulados por HBx y las dianas de estos [62].

EFEECTO DE HBx	miARN	DIANA
Up-regulación	miR-602	RASSF1A
	miR-143	FNDC3B (fibronectin type III domain-containing 3B)
	miRNA-29a/miR-148a	PTEN (phosphatase and tensin homolog)
Down-regulación	miR-661	MTA1
	miR-373	CDH1
	let-7a	STAT3 (signal transducer and activator of transcription 3)
	miR-101	DNMT3A
	miR-152	DNMT1
	miR-148a	HPIP (hematopoietic pre-B cell leukemia transcription factor interacting protein)
	miR-16	CCND1 (ciclina D1)

B.2 Inestabilidad genética

HBx causa una formación excesiva de centrosomas y la aparición de husos mitóticos multipolares llevando a defectos en la segregación cromosómica y a la inestabilidad genética. Esto se ha asociado con la unión de HBx a HBXIP (*HBx-interacting protein*), proteína que es requerida para el crecimiento y supervivencia de los hepatocitos. HBx también contribuye a la

inestabilidad genética alterando varias vías de reparación de ADN, HBx se une y secuestra p53 en el citoplasma, inhibiendo su función como supresor de tumores, bloqueando su acción en NER y evitando su acción apoptótica [51].

B.3 MMPs, TEM y metástasis

HBx contribuye a los estadios tardíos de la tumorigénesis induciendo metaloproteasas de matriz (MMPs), promoviendo la transición epitelio-mesénquima (TEM) y facilitando la metástasis [51].

B.4 Senescencia y telómeros

HBx *up*-regulan la expresión de hTERT (*human telomerase reverse transcriptase*), a pesar de su activación, los telómeros permanecen muy cortos en células de HCC, lo que predispone a la inestabilidad de telómeros, inestabilidad cromosómica y poliploidía.

HBwx (*whole HBx*)

En 2016 se ha identificado un ORF adicional 168 nt *upstream* del codón de inicio normal del gen X con actividad transcripcional y que puede ser traducido en pauta con el gen X para formar un gen X completo (wX: *whole-X gene*) de 630 pb que codifica para una proteína X completa de 210 aa (HBwx). Se ha encontrado principalmente en pacientes asiáticos con HCC y genotipo C, con lo que podría estar asociado con la hepatocarcinogénesis, de hecho HBwx favorece la tumorigénesis y crecimiento *in vivo* e *in vitro* así como migración celular e invasión [63].

2.1.4.2 ELEMENTOS REGULADORES

2.1.4.2.1 PROMOTORES Y ENHANCERS

El ADNccc sirve como molde para la síntesis de cinco ARNm; preS1, preS2/S, pregenómico (ARNpg), preCore (ARNpc) y X, bajo el control de cuatro promotores virales (preS1p, Sp, Cp y Xp) que se traducen en las proteínas virales anteriores. Todos los ARNs tienen un casquete-5' o cap, resultado de la adición de 7-metilguanósina trifosfato, y una cola poliA-3' (nt 1916-1960) [64]. Además de los promotores, regulan la transcripción los dos enhancers: ENHI (nt 1180-1198) y ENHII (nt 1627-1774) y los FT que se unen a los promotores y enhancers (Tabla 3) [49].

Tabla 3: FT que se unen a promotores y enhancers. El ENHI contiene secuencias de unión para FT que se expresan en muchos tipos celulares mientras que el ENHII contiene secuencias reconocidas por FT hepato-específicos [49].

Elemento regulador	FT
preS1p	OCT1, Sp1, HNF1 y 3 (<i>hepatocyte nuclear factor</i>), TBP (<i>TATA-binding protein</i>), NF1
Sp	Sp1, NF1
ENHII, Cp	Sp1, C/EBP (<i>CCAAT-enhancer-binding protein</i>), HNF3 y 4, PPAR α -RXR α
ENHI, Xp	C/EBP, C/ENB, HNF1, 3 y 4, OCT2, PBF, RFX1, NF1, CREB (<i>cAMP response element-binding protein</i>)/ATF2 (<i>activating transcription factor</i>), AP1, RXR α , PPAR (<i>peroxisome proliferator-activated receptor</i>), COUP-TF

A. REGULACIÓN DE LA TRANSCRIPCIÓN DEL ARN_{pg} Y ARN_{pc}

El promotor preCore/pregenómico (Cp, nt 1613-1849) dirige la transcripción de dos ARNs: el ARN_{pc}, que codifica la proteína preCore (PC) que es la precursora del HBeAg y el ARN_{pg}, que es bicistrónico y codifica para la proteína core y la polimerasa γ , además, actúa como molde para la transcripción reversa que da lugar al ADN_{rc} [65]. Estos ARNs son más largos que el propio genoma, contienen una secuencia redundante 3' (r), que corresponde al sitio de inicio de la transcripción de estos ARNm (nt 1789) hasta el sitio de poliadenilación (nt 1916-1960), y coincide con la señal de encapsidación ϵ [66].

El promotor Cp está formado por el promotor básico del core (PBC, nt 1742-1849), que es suficiente para iniciar la transcripción, y una secuencia URR (*upper regulatory region*, nt 1613-1742) que regula su actividad (Figura 10). Esta secuencia URR contiene elementos reguladores positivos CURS (*core upstream regulatory sequence*, nt 1636-1742) y negativos NRE (*negative regulatory element*, nt 1613-1636) de la actividad del PBC. A su vez el elemento CURS está formado por CURS A (nt 1636-1702), que contiene tres dominios reguladores positivos: *box* α , γ y δ , y CURS B (nt 1702-1742), que contiene un dominio regulador negativo: *box* β . Este promotor es muy débil y necesita la activación por parte del ENHI (nt 1180-1198). El ENHII (nt 1627-1774), solapado con el promotor, también puede estimular su expresión. Además existen cuatro regiones TATA *boxes-like* (TA1 nt 1750-1755, TA2 nt 1758-1762, TA3 nt 1771-1775 y TA4 nt 1788-1795), las tres primeras controlan la síntesis del ARN_{pc} y la cuarta la del ARN_{pg}. De manera homóloga a los promotores celulares de la ARN_{pol} II, el ARN_{pc} y ARN_{pg} tienen sus respectivos elementos iniciadores de la transcripción (Inr), que son elementos promotores mínimos y que potencian el efecto de las TATA *boxes-like* [65,66].

B. REGULACIÓN DE LA TRANSCRIPCIÓN DEL ARN_{preS1} Y ARN_{preS2/S}

Sp es un potente promotor, no específico del hígado, que dirige la transcripción de las proteínas MHBs y SHBs, mientras que preS1p es un promotor más débil, hepato-específico, que dirige la transcripción de LHBs, es el único con una TATA-box clásica, donde se une TBP

(*TATA-binding protein*), que junto con el sitio de unión a HNF1 *upstream*, son los dos únicos motivos necesarios para la actividad transcripcional [11]. El ENHI tiene poco efecto en los promotores Sp y preS1, pero el ENHII puede estimular ambos [67].

C. REGULACIÓN DE LA TRANSCRIPCIÓN DEL ARNx

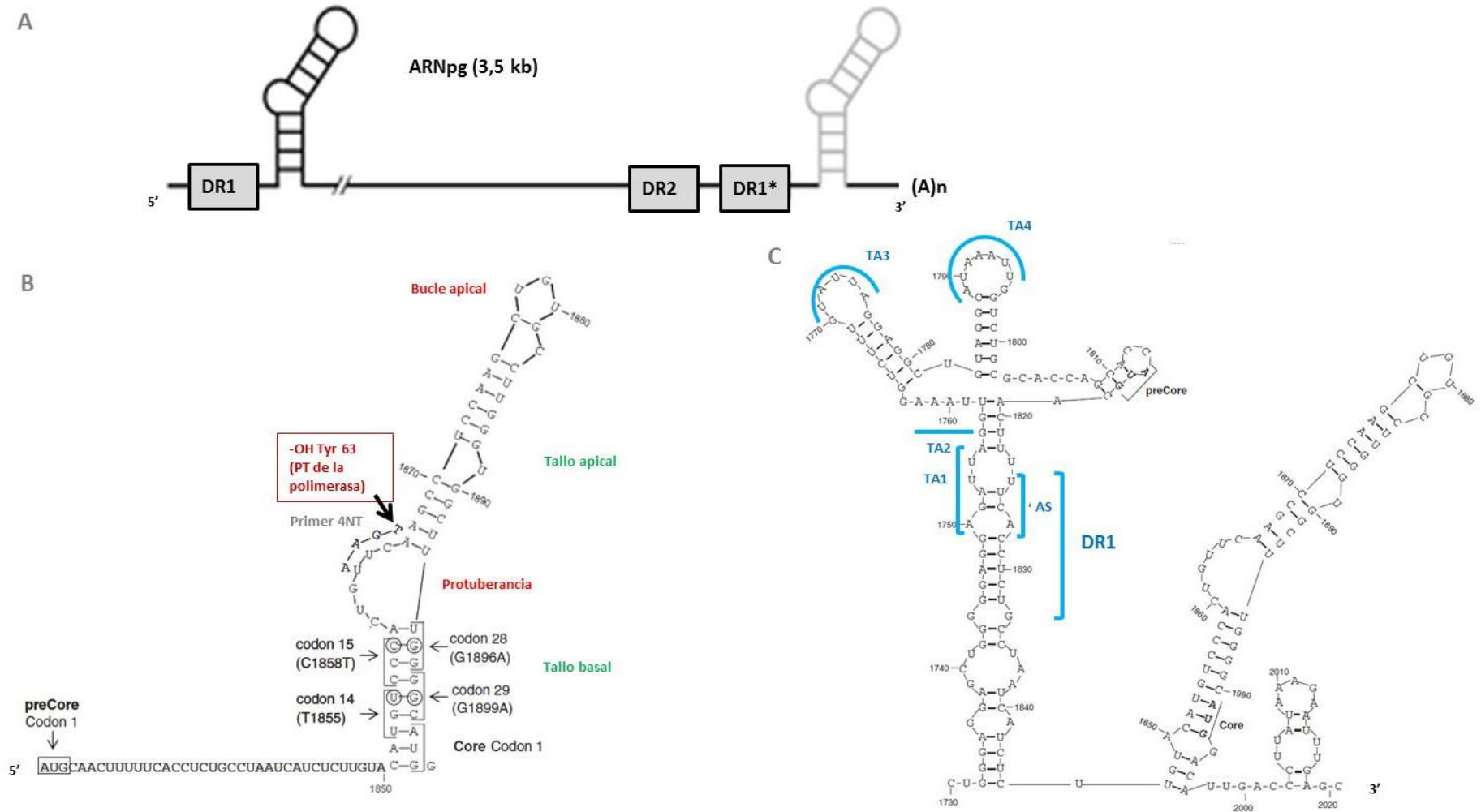
Es el ARNm más pequeño, controlado por el Xp y el ENHI. Xp carece de TATA-box pero tiene diferentes secuencias para la unión de FT ubicuos y específicos del hígado. El ENHI une múltiples activadores de la transcripción de la familia *basic leucine zipper*, incluyendo C/EBP, AP-1 y ATF. Además otros factores implicados en el control del ciclo celular y la apoptosis, incluyendo p53, c-Abl y RFX-1 se unen específicamente al ENHI y regulan su actividad [90]. También responde a muchos factores extracelulares incluyendo: TPA (12-O-Tetradecanoilforbol-13-acetato), ácido okadaico, IL-6 y estímulos genotóxicos que son importantes en la respuesta inflamatoria aguda y la proliferación celular en el hígado, además el ENHI (y Xp) es estimulado por HBx, lo que implica que aparentemente el gen X está bajo autorregulación.

2.1.4.2.2 LA SEÑAL DE ENCAPSIDACIÓN ÉPSILON (nt 1847-1908)

La señal de encapsidación épsilon (ϵ) es el elemento *cis* mejor conocido del ARNpg, corresponde a unas secuencias de 120 pb redundantes en el extremo 5' y 3' que adoptan una estructura de tallo-bucle. Además contiene las DRs, dos regiones cortas muy conservadas de 11 pb implicadas en la replicación viral, DR1 (nt 1824-UUCACCUCUG-1834) en el extremo 5' de la hebra(-) y DR2 (nt 1590-UUCACCUCUGC-1600) en el extremo 3' de la hebra(+) (Figura 14A) [68].

El extremo 5' de la señal de encapsidación ϵ (5'- ϵ), contiene dos tallos, el basal y apical, caracterizados por el emparejamiento de pares de bases, y dos protuberancias, la interna y la apical, que contienen 3 nt (Figura 14B). La alta estabilidad estructural de esta región induce una alta afinidad de la polimerasa para reconocer esta estructura y esta interacción del extremo 5'- ϵ con la polimerasa es el primer paso en el inicio de la transcripción reversa. El extremo 3' de la señal de encapsidación ϵ (3'- ϵ) contiene la DR1 repetida (DR1*), que incluye el sitio aceptor (AS) para el primer 4NT (Figura 14C) (Ver apartado 2.1.5 *Ciclo replicativo*) [69–73].

Figura 14: Señal de encapsidación ϵ (A), extremo 5' (B) y 3' (C).



2.1.4.2.3 OTROS ELEMENTOS REGULADORES

A. ELEMENTOS DE RESPUESTA A GLUCOCORTICOIDES (GRE)

Corresponden a una secuencia corta entre 10-20 nt, 735 *upstream* de la diana EcoRI, que activan la transcripción y la replicación viral, y explican la reactivación del virus en tratamientos basados en esteroides [74].

B. ELEMENTO REGULADOR POST-TRANSCRIPCIONAL (PRE)

Corresponde a una secuencia de entre 365-533 nt en la región 3' común a todos los ARNm, entre el Xp y el Cp. Actúa a nivel post-transcripcional inhibiendo el proceso de splicing y facilitando la exportación y uso de los ARN subgenómicos (preS1, preS2/S y X), ya que no necesitan splicing, aunque pueden experimentarlo (Ver apartado 2.2.4.1.2 *Proteínas virales y HCC*) [75].

C. ELEMENTOS DE RESPUESTA A HORMONAS (ERH)

Se han identificado varios ERH en el genoma del VHB y funcionan a través de la unión a miembros de la superfamilia de receptores nucleares (RN). HFN4 (*Hepatocyte nuclear factor 4*), RXR α (*Retinoid X Receptor α*) y PPAR α (*Peroxisome proliferator activated Receptor α*) se unen al mismo ERH localizado en el ENHI y activan la transcripción [76].

D. NUCLEAR RECEPTOR-RESPONDING ELEMENTS (NRREs)

La superfamilia de los RN tienen efectos en la expresión génica del VHB a través de los NRREs localizados en la región del ENHI y región ENHII/PC. Los NRREs del VHB se unen sobre todo a HFN4 α , RXR α y PPAR α [77].

2.1.5 CICLO REPLICATIVO

El ciclo replicativo del VHB se puede dividir en dos partes, la primera incluye la unión y entrada del virus, el transporte de las cápsidas desde el citoplasma al núcleo, la liberación del ADNrc en el núcleo y la formación del ADNccc. La segunda parte implica la transcripción y traducción de los genes del virus, la encapsidación del ARNpg, la transcripción reversa y la secreción de los viriones y partículas subvirales [78] (Figura 15A).

Figura 15: Ciclo replicativo del VHB. (A) Esquema general en un hepatocito. MVB: *Multivesicular bodies*.

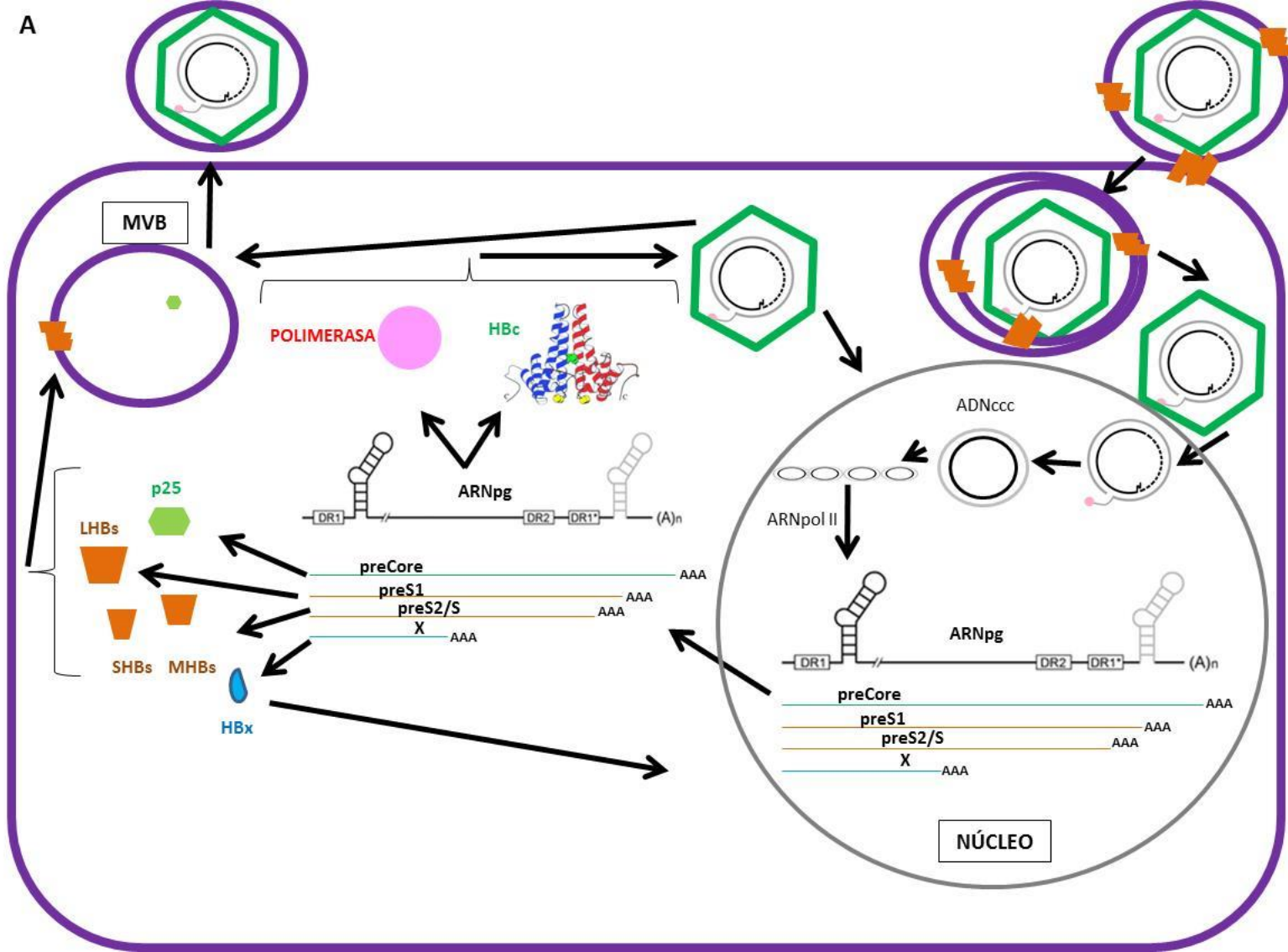
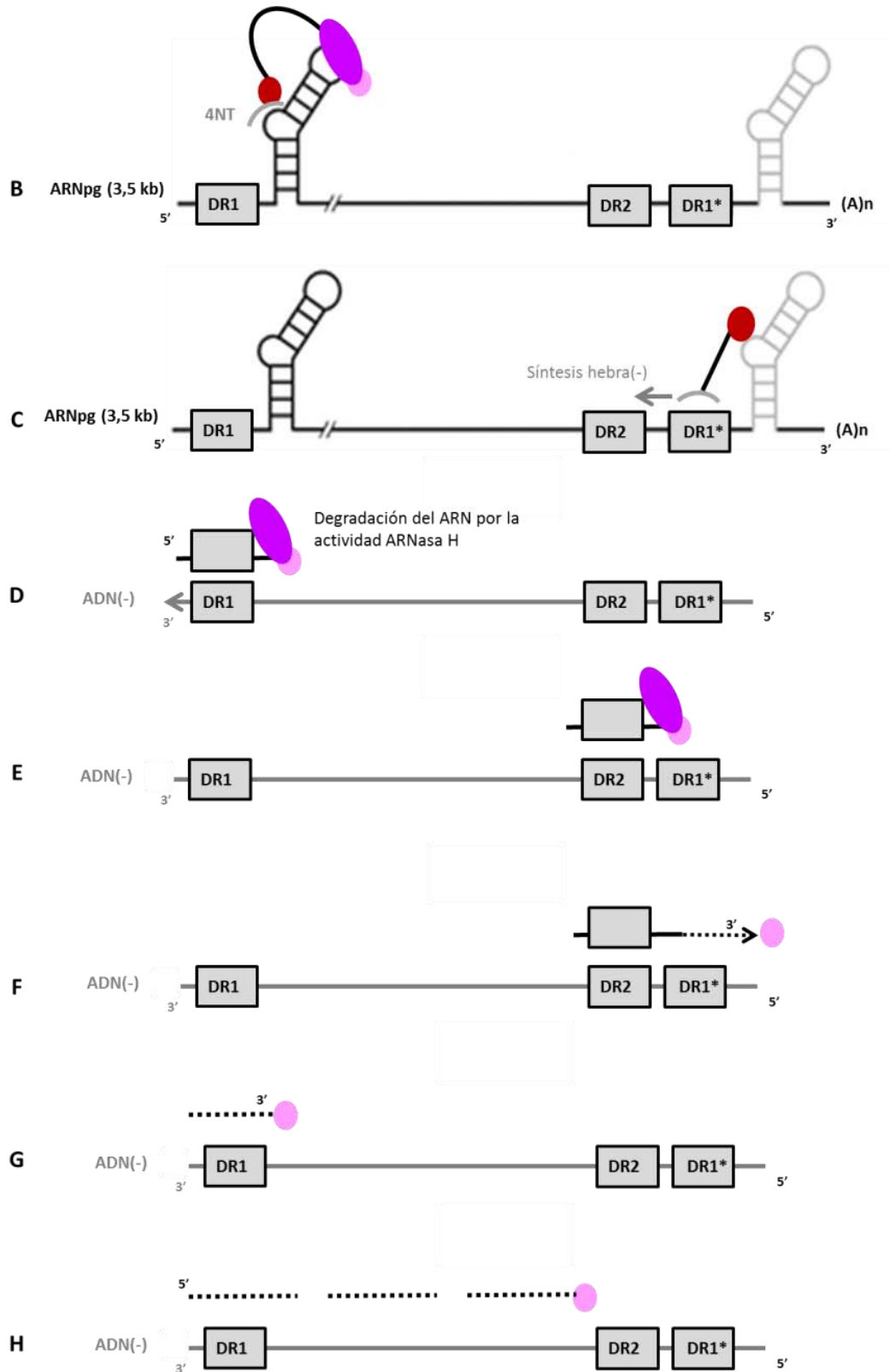


Figura 15: Continuación. (B) Unión polimerasa-e y síntesis 4NT. (C) Translocación de 4NT y polimerasa a DR1. (D) Síntesis de la hebra(-) por RT. (E) Translocación del complejo Pol+ARN a la DR2 en el extremo 5' de la hebra(-). (F) Inicio de la síntesis de la hebra(+). (G) Translocación de la hebra(+) al extremo 3' de la hebra(-). (H) Síntesis de la hebra(+).



2.1.5.1 ENTRADA DEL VIRUS

La unión y entrada del virus comienza con la unión de baja especificidad a heparán sulfato proteoglicanos (HSPG), seguido de la unión altamente específica de LHB al receptor: péptido co-transportador de sodio-taurocolato (NTCP) [13]. El NTCP es miembro de la familia de genes de proteínas transportadoras SLC10, se expresa principalmente en el hígado y se localiza en la membrana basolateral de los hepatocitos [79]. Después de la interacción las partículas virales forman un endosoma en el interior de la célula hepática [80], el subcompartimento subendosómico desde el que las nucleocápsides son liberadas al citoplasma todavía no se ha caracterizado, pero se cree que el dominio TM1 de SHBs (común a la secuencia de las tres proteínas de superficie) podría ser esencial para la hemifusión de las membranas [81]. La nucleocápside es liberada y permanece en el citoplasma con la molécula de ADNrc en el interior unida a la proteína P covalentemente [20].

2.1.5.2 TRANSPORTE DE LAS CÁPSIDES DESDE EL CITOPLASMA AL NÚCLEO Y LIBERACIÓN DEL ADNRC EN EL NÚCLEO

La nucleocápside es transportada activamente por microtúbulos al complejo del poro nuclear. Los dímeros de HBcAg tiene una señal de localización nuclear (SLN) que interacciona con la membrana nuclear, el ADNrc se libera en el núcleo [42], al quedar las nucleocápsides retenidas en el complejo del poro nuclear [82]. La retención en este complejo tiene lugar mediante la interacción con la nucleoporina Nup 153 [83], un mecanismo único entre los virus que atraviesan el complejo del poro nuclear.

2.1.5.3 FORMACIÓN DEL ADNccc

Dentro del núcleo el ADNrc sufre una serie de modificaciones hasta convertirse en ADNccc [84], incluyendo: (1) liberación de la proteína P unida covalentemente al extremo 5' de la hebra(-) y eliminación del oligonucleótido de ARN en el extremo 5' de la hebra(+), (2) eliminación de una de las redundancias terminales de la hebra(-), (3) finalización de la síntesis de la hebra(+) por la polimerasa y (4) unión covalente de los extremos 5' y 3' de ambas cadenas. Los mecanismos concretos por los que tienen lugar estas modificaciones aún no se conocen, sin embargo parece claro que el primer paso para la transformación de ADNrc en ADNccc consiste en la desproteínización del primero [85]. Este proceso podría estar mediado por la topoisomerasa 2 (TOP2), que introduce cortes de doble cadena en el ADN, con lo cual podría liberar la polimerasa viral, ella misma quedaría unida al ADN-VHB simulando un aducto, que podría ser reparado entonces por la maquinaria celular. En concreto el VHB podría utilizar las enzimas tirosil-ADN fosfodiesterasas (TDPs), los aductos causados por TOP2 serían

reparados por la tirosil-ADN fosfodiesterasa 2 (TDP2) [86]. Pero estos resultados son controvertidos, ya que otros estudios han observado que la enzima TDP2 podría no ser imprescindible para la formación del ADNccc [87].

El ADNccc puede permanecer como un minicromosoma no integrado y estable, formando complejos con histonas y proteínas no histonas en el núcleo de los hepatocitos infectados [78], que tienen una vida media de 10-100 días [9]. Este hecho explica la dificultad de eliminar la infección, el riesgo de desarrollar HCC [88], que el control de la estabilidad del ADNccc y la expresión génica viral sea altamente dependiente de modificaciones epigenéticas (acetilación/desacetilación de H3 y H4, metilación de sus islas CpG y ARN de interferencia transcritos desde la propia cadena(+) del cccADN [89]) y que pueda seguir la producción de antígenos virales en ausencia de replicación [84].

2.1.5.4 TRANSCRIPCIÓN VIRAL DEL ADNccc Y TRADUCCIÓN PROTEICA

El ADNccc actúa de molde para la transcripción de los cinco ARNm por la ARNpol II celular, que son transportados al citoplasma y traducidos por los ribosomas celulares para dar lugar a las siete proteínas virales [66]. El ARNpg hace de ARNm para la traducción de la proteína P y las proteínas core [20] y de molde para la transcripción reversa por la cual se replica el VHB [44], su señal de encapsidación ϵ es esencial para la replicación ya que la proteína P se une a 5'- ϵ y desencadena la encapsidación del ARNpg, seguido de la síntesis del ADN dentro de la cápside [20].

2.1.5.5 FORMACIÓN DE LAS PARTÍCULAS CORE Y TRANSCRIPCIÓN REVERSA

2.1.5.5.1 Unión polimerasa- ϵ : encapsidación y cebado proteico (Figura 15B)

La interacción de 5'- ϵ del ARNpg con la polimerasa se requiere para el cebado proteico y la encapsidación.

A. Cebado proteico

La polimerasa cataliza la generación de un primer de cuatro nt usando la protuberancia en 5'- ϵ como molde, la presencia de un grupo -OH en la tirosina 63 de la PT permite añadir un residuo dGMP usando el sitio activo YMDD del dominio RT pol y finalmente lleva a la síntesis de un primer de cuatro nucleótidos TGAA (4NT) unido covalentemente al dominio PT (Figura 14B y 15B) [20]. Para la unión se requiere la asociación de la polimerasa con chaperonas y co-chaperonas como Hsp90, Hsp70, Hsp40, Hop/p60 y p23 para establecer y mantener P en una conformación adecuada para la interacción con ϵ en un proceso dinámico dirigido por la hidrólisis de ATP [20,21].

B. Encapsidación

El complejo ribonucleoproteico polimerasa-ARNpg-chaperonas induce la adición de 120 dímeros de HBc alrededor del complejo lo que lleva a la formación de las nucleocápsides, el primer paso es la formación de los homodímeros HBc por puentes disulfuro entre cys61 [68]. El proceso de encapsidación depende de fuertes interacciones electrostáticas entre las cargas negativas del ARNpg y las cargas positivas de los residuos de Arg en el dominio protamina del extremo Ct de HBcAg (Figura 8) [90].

2.1.5.5.2 Translocación de 4NT y la polimerasa a DR1* (Figura 15C)

El primer 4NT, unido covalentemente a P, transloca a 5'UUCA en la región DR1* en el extremo 3' del ARNpg, al sitio aceptor (AS), donde continua la elongación de la hebra(-) [10]. Como el primer sólo tiene 4 nucleótidos, podría hibridar inespecíficamente en muchos otros sitios del genoma, por eso existen otros elementos que definen a DR1* como el aceptor correcto: ϕ , una secuencia a 30 nt del extremo 5' de DR1* [91] y ω , secuencia que se solapa con el extremo 3' de DR1* [92].

2.1.5.5.3 Síntesis de la hebra(-) por la actividad retrotranscriptasa (Figura 15D)

Dentro de las nucleocápsides, 4NT actúa como primer para la síntesis de la hebra(-) a partir del ARNpg por la actividad retrotranscriptasa de P. Mientras que se sintetiza la hebra(-), la actividad ARNasa H de P degrada el ARNpg, excepto una pequeña secuencia de 11-16 nucleótidos (11-16NT) [10].

2.1.5.5.4 Translocación del complejo P+ARN e inicio síntesis de la hebra(+) (Figura 15E y F)

El pequeño fragmento resistente a la ARNasa H, derivado del extremo 5' del ARNpg y que incluye 5'-DR1 (11-16NT) y el cap, transloca a la DR2 localizada en el extremo 5' de la hebra(-) y actúa como primer para la síntesis de la hebra(+) por la actividad polimerasa de P [21]. Es en este punto donde se origina el ADN lineal de doble cadena (dslADN) por *priming in situ*, el primer de ARN se queda en su posición original en vez de translocar a DR2 y se extiende la hebra(+). Esta forma minoritaria se puede encapsidar y dar lugar a ADNccc, que por retrotranscripción, puede dar lugar a nuevos dslADN [93], y de hecho esta forma del genoma del VHB se integra con más frecuencia en el genoma del huésped que la forma de ADNrc [93].

2.1.5.5.5 Elongación hebra(+) (Figura 15G y H)

El fragmento de ADN(+) en extensión transloca al extremo 3' de la hebra(-) y la síntesis de la hebra(+) prosigue. Este último paso está limitado al momento en el que la cápside se

libera y los nt ya no están disponibles para la replicación, por eso durante la maduración de las cápsides la hebra(+) queda incompleta y el genoma del VHB adopta una conformación de doble hebra parcial (ADNrc) [21].

2.1.5.6 MADURACIÓN Y SECRECIÓN DE VIRIONES Y PARTÍCULAS SUBVIRALES

Las cápsides maduras conteniendo el ADNrc sintetizado *de novo* pueden ser envueltas y secretadas o, alternativamente, liberar el ADNrc al núcleo para amplificar el pool de ADNccc [20]. El transporte nuclear o la formación de viriones con envuelta parecen ser eventos competitivos, en la infección tardía cuando hay disponibilidad de cantidades suficientes de proteínas de la envuelta, la amplificación de ADNccc cesa [94].

2.1.5.6.1 Liberación de partículas virales nuevas

Sólo las formas maduras de las nucleocápsides pueden envolverse ya que son las únicas capaces de interactuar con el dominio específico de las proteínas de superficie LHBs. Parece que el estado de fosforilación de las argininas de la región protamina de HBc forman parte de la señal que indica que una cápside es madura [78]. Las cápsides adquieren la envuelta lipoproteica en cuerpos multivesiculares (MVB) y la gemación depende de la maquinaria ESCRT (*endosomal sorting complexes required for transport*) [95,96].

2.1.5.6.2 Reciclaje de las cápsides

Algunas de las cápsides *de novo* pueden reinfectar el núcleo, aumentando el pool de ADNccc nuclear y la variabilidad de este, este proceso tiene lugar poco después de la infección, cuando los niveles de proteínas de superficie aún son bajos y no se han empezado a producir partículas virales [97]. Los hepatocitos infectados de manera natural por Hepadnavirus contienen entre una y 50 copias de ADNccc [98], sumado a la vida media larga de los hepatocitos asegura que el ADNccc no se pierda durante la división celular [94].

2.2 IMPLICACIONES CLÍNICAS DE LA INFECCIÓN POR EL VHB

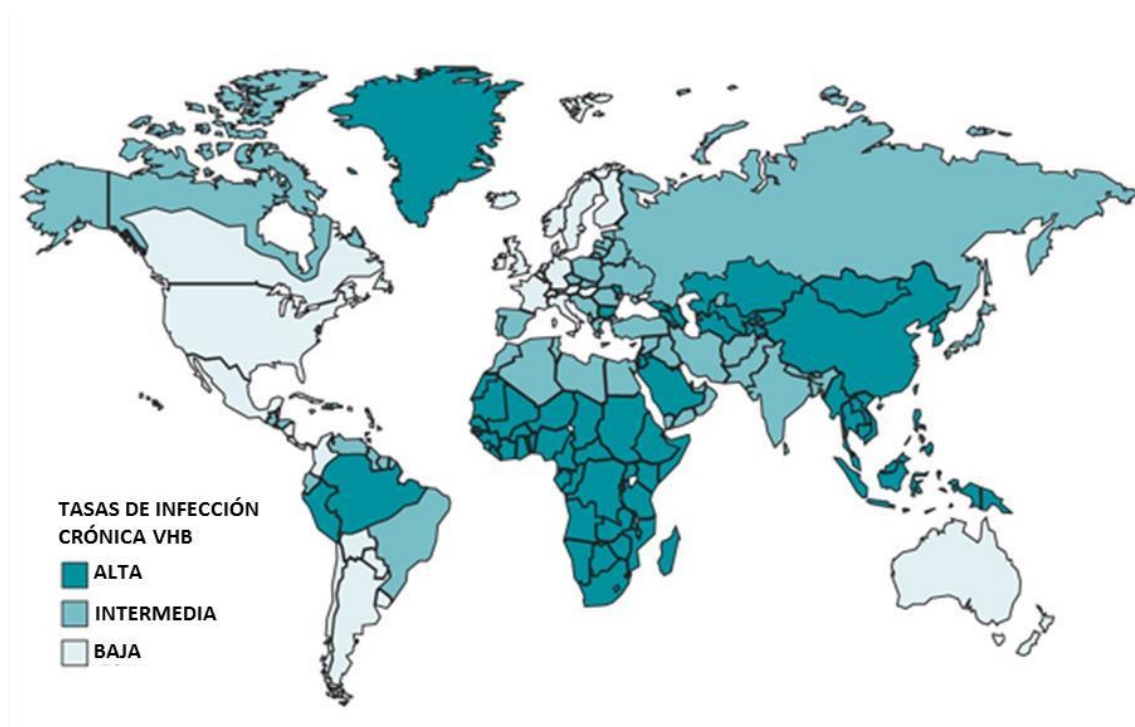
2.2.1 EPIDEMIOLOGÍA

La infección crónica por el VHB es uno de los principales problemas de salud del mundo, se estima que 2.000 millones de personas han estado infectadas en algún momento de su vida y que 240 millones los están en la actualidad. A pesar de la implementación mundial de medidas preventivas, incluyendo la disponibilidad de una vacuna efectiva desde 1982, el VHB sigue siendo una importantísima causa de morbilidad, causa el 30 y 53% de las cirrosis y HCC, respectivamente, y de mortalidad, con 650.000 muertes al año [99,100].

2.2.1.1 PREVALENCIA GLOBAL

La implementación universal de la vacunación de los niños ha causado un descenso drástico de la prevalencia de la infección en muchas partes del mundo, sin embargo, la cobertura de esta vacunación varía entre el 90% en el Pacífico Occidental y el continente Americano hasta el 56% en el Sudeste Asiático, de ahí que las tasas de portadores de HBsAg varíen mucho dentro de las diferentes áreas geográficas y se hayan establecido tres niveles de endemidad: baja ($\leq 2\%$), intermedia (2-8%) y alta ($\geq 8\%$) (Figura 16) [100,101].

Figura 16: Prevalencia global de la infección crónica por el VHB [102].



Aproximadamente, el 45% de los individuos infectados viven en áreas muy endémicas (China, Sudeste Asiático, la mayoría de las islas del Pacífico, África subsahariana, parte de Oriente Medio y la cuenca del Amazonas), el 43% en regiones con endemidad intermedia (países mediterráneos, Europa del este, Asia central, Oriente medio, Rusia y centro y sur de América) y el 12% restante en áreas muy poco endémicas (norte de América, Europa occidental, Australia y Japón) [100].

2.2.1.2 TRANSMISIÓN

El VHB es transmitido por exposición percutánea o de mucosas a sangre u otros fluidos corporales infectados. También se transmite por vía sexual, especialmente en hombres homosexuales no vacunados y personas heterosexuales con múltiples parejas, por la reutilización de agujas y jeringas entre consumidores de drogas inyectables, durante procedimientos médicos, quirúrgicos y dentales y por el uso de agujas para tatuar, cuchillas de afeitar u objetos similares contaminados con sangre infectada. El VHB puede sobrevivir hasta siete días en las superficies por lo que se puede producir el contagio por contacto frecuente con portadores a través de estas superficies contaminadas [8]. El modo de transmisión predominante varía con la prevalencia, en áreas de alta endemidad la transmisión es principalmente vertical (madre-hijo) u horizontal. La transmisión del virus madre-hijo puede ser (1) intra-útero, ocurre en el tercer trimestre del embarazo, posiblemente cuando el VHB de la sangre materna atraviesa la placenta, y supone <5% de las transmisiones de madres HBsAg(+)/HBeAg(+) o (2) peri-natal, ocurre en el momento del parto o cerca (desde la semana 28 de gestación hasta siete días post-parto) por exposición percutánea o a través de mucosas a fluidos infectados de la madre en el canal del parto. Entre el 70-90% de neonatos de madres HBeAg(+) se convertirán en portadores crónicos dentro de los tres meses después del nacimiento y sólo lo harán el 10% de los neonatos de madres HBeAg(-) [35]. En áreas de baja endemidad la infección se adquiere en la adolescencia o edad adulta temprana relacionada con comportamientos de alto riesgo como relaciones sexuales sin protección o el uso de drogas inyectables [99,100].

2.2.2 HISTORIA NATURAL

El VHB puede dar lugar a infecciones agudas o crónicas, que pueden ser desde asintomáticas o leves hasta tener consecuencias severas, raramente fulminantes.

2.2.2.1 HEPATITIS AGUDA

El período de incubación de la hepatitis aguda por el VHB varía entre 1-6 meses y esta puede ser asintomática o sintomática [103]. La infección aguda asintomática está asociada con enfermedad moderada o subclínica, normalmente pasa sin diagnosticarse y es más común en niños (<4 años) que en adultos (>30 años). Los casos con expresión clínica tiene una fase prodrómica (nauseas, vómitos, astenia, anorexia, fiebre y síntomas parecidos a la gripe), que puede ser seguida por una fase icterica (ictericia, heces acólicas, dolor en el cuadrante superior derecho y hepatomegalia) y los síntomas no son distinguibles de otras infecciones hepáticas agudas virales [103]. El 1% de las formas ictericas puede evolucionar a hepatitis fulminantes [104], que tienen una alta mortalidad (80% sin trasplante hepático) [101]. Las manifestaciones extrahepáticas ocurren hasta en el 20% de los pacientes infectados y pueden implicar el sistema gastrointestinal, renal y nervioso. El mecanismo que lleva a estas manifestaciones, incluyendo artralgias, un síndrome parecido a la enfermedad del suero, poliarteritis nodosa y glomerulonefritis, implica inmunocomplejos circulantes conteniendo partículas virales y al sistema del complemento. Varios marcadores serológicos, incluyendo el ADN-VHB, HBsAg, HBeAg e IgM anti-HBc, son detectables durante la fase aguda de la infección, al igual que alteraciones bioquímicas como el aumento de la alanina aminotransferasa (ALT) y la bilirrubina, la seroconversión anti-HBs indica resolución de la infección aguda (Figura 17A).

2.2.2.2 HEPATITIS B CRÓNICA (HBC)

Si el HBsAg persiste más de seis meses en suero se define la progresión a hepatitis crónica (Figura 17B) [101], y esta es inversamente proporcional a la edad de adquisición de la infección: 90% después de una infección perinatal, 30% si se adquiere entre 1-5 años e inferior al 5% en adultos inmunocompetentes [104]. Además de la edad de la infección, factores genéticos del hospedador como polimorfismos en el gen de la IL-18, el promotor del TNF α y variantes en el HLA-DP se han asociado con la persistencia de la infección. Los síntomas en la infección crónica varían desde moderados y no específicos como fatiga y molestias en el cuadrante superior derecho, con daño hepático mínimo, hasta ascitis, edema periférico y encefalopatía en pacientes con enfermedad hepática avanzada [103]. La HBC es un proceso dinámico, que resulta de la interacción entre la replicación del virus y la respuesta inmunológica del hospedador, en la historia natural se pueden diferenciar cinco fases, no necesariamente secuenciales, y por las que no todos los pacientes atraviesan (Figura 18) [117].

2.2.2.2.1 Fase de tolerancia inmune

Es típica de pacientes con infección adquirida durante la infancia. Está asociada con la presencia de HBsAg y replicación viral activa (HBeAg(+)) y niveles altos de ADN-VHB en suero ($\sim 10^7$ UI/ml), por lo que estos pacientes son muy contagiosos, pero niveles de ALT normales o poco elevados e inflamación mínima o ausente en la biopsia hepática, las concentraciones de HBsAg también son altas ($>10^5$ UI/ml). En los pacientes con infección perinatal o adquirida en la infancia esta fase puede durar entre 20-30 años, mientras que en los que adquieren la infección en la adolescencia o en la edad adulta esta fase dura poco. Se caracteriza por un curso moderado, progresión muy lenta, a pesar de la alta carga viral, y tasa de seroconversión espontánea, definida como la pérdida de HBeAg con desarrollo de anti-HBe, muy baja [101,103,104].

2.2.2.2.2 Fase de actividad inmune

Esta fase aparece más rápidamente en personas infectadas en la edad adulta y su duración varía de semanas a años. Está caracterizada por la maduración del sistema inmune del hospedador, aumento de la inflamación hepática con o sin fibrosis, niveles de ADN-VHB y ALT altos, reflejando la replicación activa del VHB, y concentraciones de HBsAg menores (10^3 - 10^4 UI/ml) [101,103,105]. Los pacientes que permanecen durante años en esta fase, sobre todo si presentan brotes repetidos de exacerbaciones agudas, tienen riesgo de desarrollar cirrosis y HCC. En esta fase se produce a menudo la seroconversión espontánea, después la HBC puede evolucionar a portador inactivo o hepatitis crónica HBeAg(-) [117].

2.2.2.2.3 Fase de portador inactivo

Los parámetros bioquímicos y virológicos se normalizan; niveles muy bajos o indetectables de ADN-VHB (generalmente <2.000 UI/ml) y valores de ALT persistentemente normales. Su pronóstico a largo plazo normalmente es bueno y el riesgo de desarrollar cirrosis o HCC es bajo. En ocasiones se producen reactivaciones de la infección con elevaciones de ADN-VHB y ALT. La pérdida de HBsAg con desarrollo de anti-HBs puede ocurrir espontáneamente en el 1-3% de los casos al año, generalmente tras varios años de valores indetectables de ADN-VHB [104,105].

2.2.2.2.4 Hepatitis crónica HBeAg(-)

Está caracterizada por un estatus HBeAg(-), anti-HBe(+) y bajos niveles de ADN-VHB (<2.000 UI/ml). La tasa de seroaclaramiento espontáneo del HBsAg varía entre 0,5% por año hasta una tasa tan alta como 40% a los 25 años del seguimiento [103]. Se caracteriza por

reactivaciones periódicas con patrones de ADN-VHB y ALT fluctuantes y hepatitis activa, durante los cuales hay progresión de la fibrosis hepática. Es la forma más frecuente de HBC en España y su curso clínico es muy variable [104]. A veces es difícil diferenciar esta fase de la de portador inactivo, pero es importante hacerlo, ya que mientras los pacientes en fase de portador inactivo tiene bajo riesgo de experimentar complicaciones, los pacientes con HBC HBeAg(-) tienen una enfermedad hepática activa que puede dar lugar a futuras complicaciones como cirrosis descompensada y HCC. Se deben realizar determinaciones cada tres meses de ALT y ADN-VHB durante al menos un año para catalogar bien a estos pacientes [104].

2.2.2.2.5 Fase de remisión

Se define por HBeAg(-) con o sin anti-HBs y se caracteriza por valores muy bajos o indetectables de ADN-VHB y niveles normales de ALT. Puede persistir una baja replicación viral con ADN-VHB detectable en el hígado, lo que se denomina infección oculta por el VHB (IOB), aunque la relevancia clínica no está bien definida [104], en estos pacientes la inmunosupresión puede dar lugar a la reactivación de la replicación viral [106].

Figura 17: Marcadores del VHB durante el curso de una infección aguda resuelta (A) y durante la transición de una infección aguda a una infección crónica (B). La determinación de HBsAg en suero es el principal test para el diagnóstico de infección aguda por el VHB. IgM anti-HBc es útil durante “la ventana core”, breve período de tiempo entre la pérdida del HBsAg en suero y la aparición de anti-HBs en la resolución de la infección aguda. El ADN-VHB es el marcador detectable más temprano, anterior a la aparición en suero del HBsAg. La aparición de anti-HBe seguida de la aparición de anti-HBs es característica de una infección aguda resuelta. La respuesta anti-HBs permanece detectable en suero varios años seguidos a la recuperación de una infección aguda e indica inmunidad protectora. IgG anti-HBc permanece décadas o incluso de por vida tras la infección aguda [103].

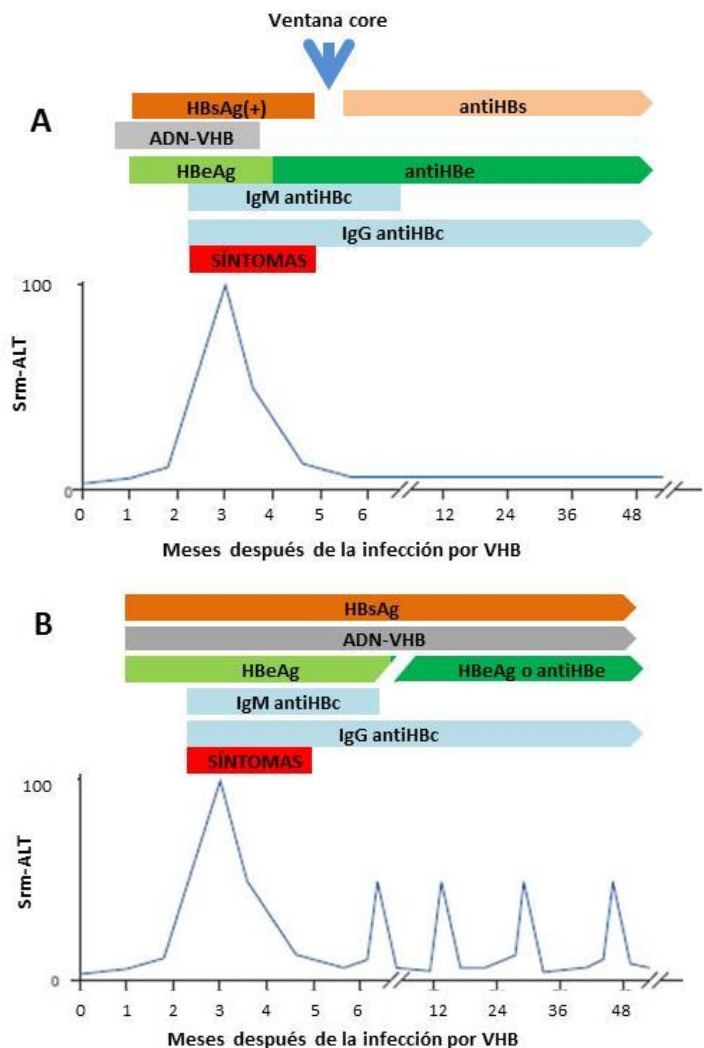
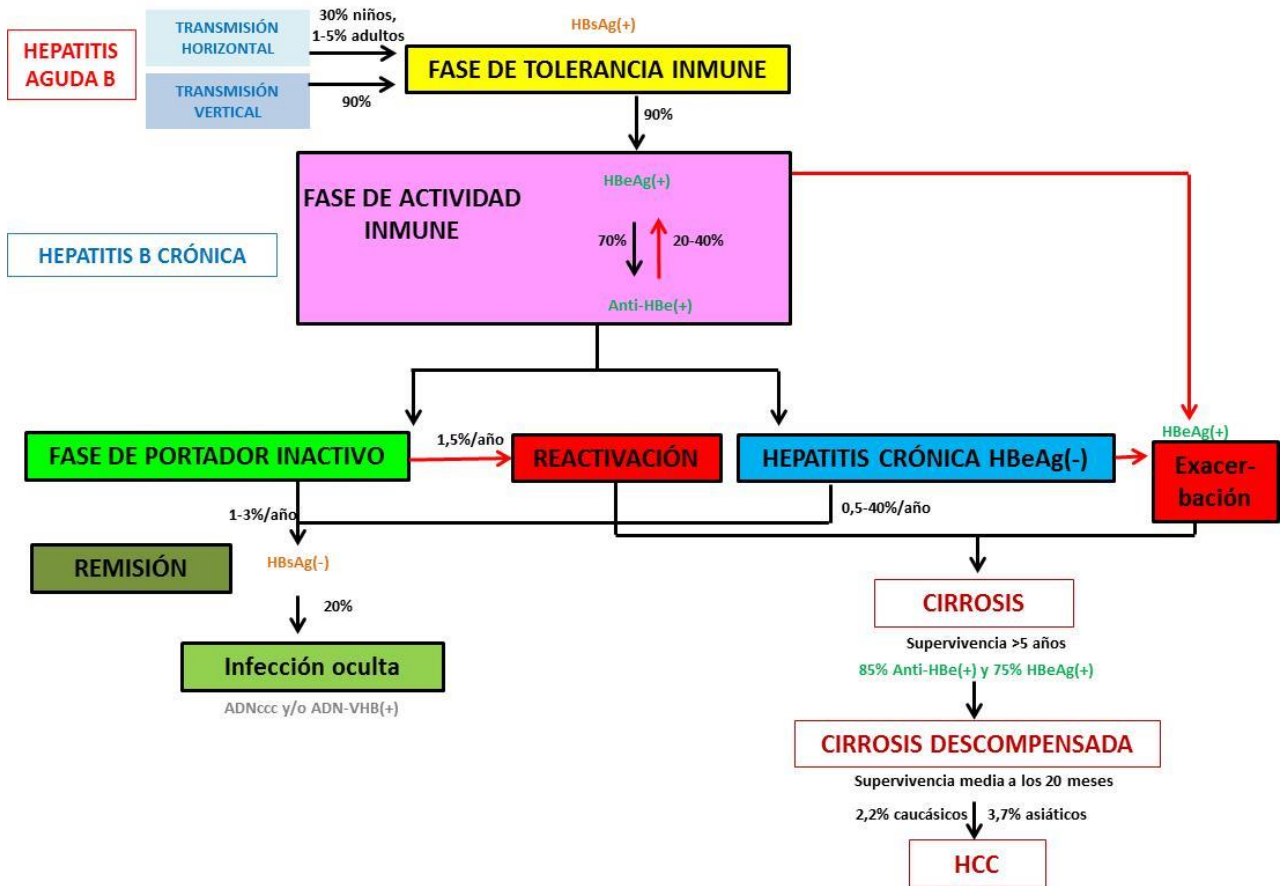


Figura 18: Historia natural de la infección por el VHB.



2.2.3 INMUNOPATOGÉNESIS DEL VHB

El sistema inmune ha desarrollado diferentes mecanismos para eliminar virus extracelulares libres y células infectadas (Figura 19) [107].

2.2.3.1 RESPUESTA INNATA

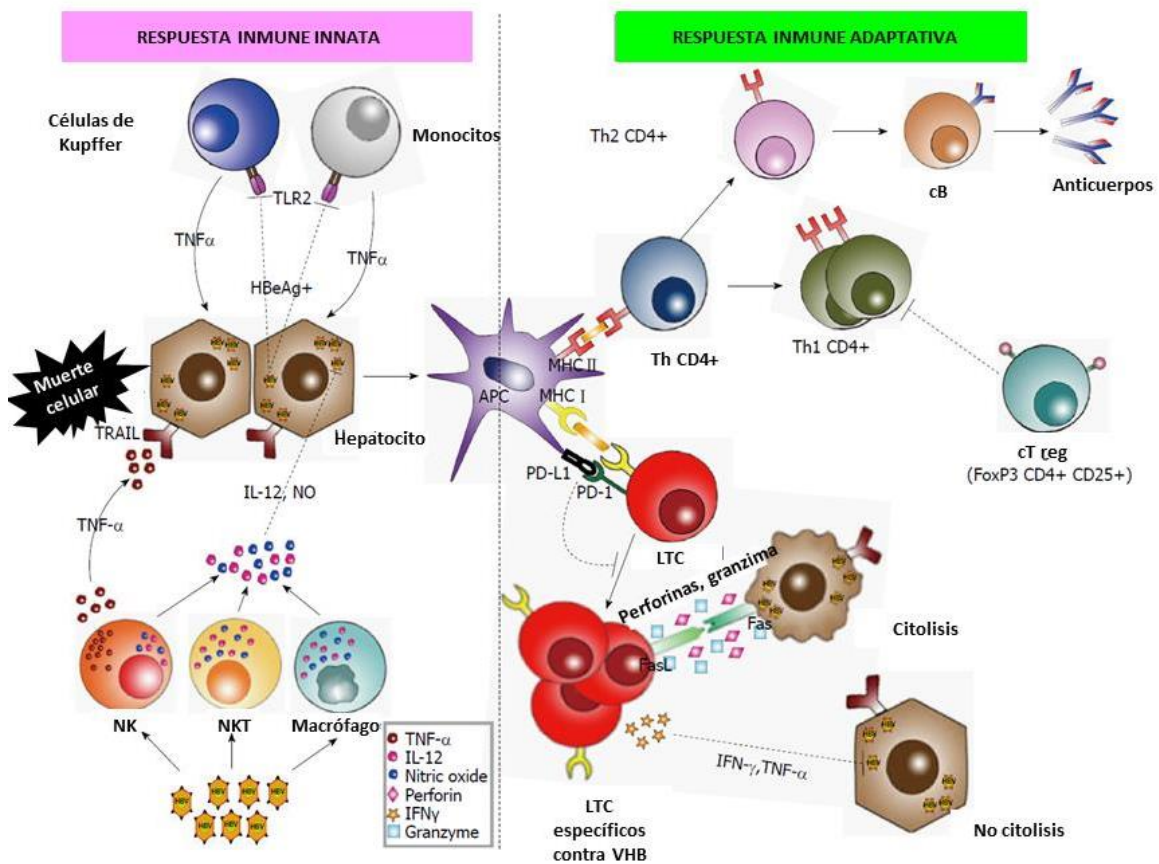
Los hepatocitos infectados y las proteínas virales son reconocidos por células del sistema inmune innato, células dendríticas (CD), monocitos, *Natural killers* (NK), células endoteliales, HSCs y células de Kupffer, la activación de la respuesta innata a través de TLRs (*Toll like receptors*), induce la producción de citoquinas como IFN-1, la destrucción de células infectadas por las NK y el reclutamiento *in situ* de la respuesta adaptativa (Figura 19) [108].

2.2.3.2 RESPUESTA ADAPTATIVA

Las CD maduras migran a áreas de cT en los órganos linfoides y presentan péptidos en HLA clase I a LTC y en HLA clase II a linfocitos T helper (Th) CD4+. Los LTC secretan citoquinas

antivirales que contribuyen a la lisis y muerte de los virus intracelulares, pueden llevar a cabo acciones citolíticas (vía perforinas o sistema Fas); el hepatocito presentador del antígeno experimenta apoptosis, y no citolíticas; secretan linfoquinas, que reclutan células inflamatorias no específicas de antígeno, y citoquinas no citolíticas como IFN- γ o TNF α , que inhiben la expresión de los genes y la replicación viral por desestabilización del ARN [107]. Los Th2 secretan citoquinas que facilitan la activación de macrófagos, citotoxicidad mediada por cT y la producción de anticuerpos por cB (Figura 19) [107].

Figura 19: Interacción entre hepatocitos infectados y células del sistema inmune [109].



2.2.3.3 ESTABLECIMIENTO DE LA INFECCIÓN CRÓNICA

Los factores virales que facilitan el establecimiento de la infección crónica son el HBeAg y el exceso de partículas subvirales, ambos implicados en la inducción de tolerancia inmune. Mientras que HBcAg activa Th1 llevando al ataque inmune, HBeAg desencadena la respuesta Th2 promoviendo la tolerancia inmune, promueve el cambio de respuesta a HBcAg

de Th1 a Th2, atraviesa la placenta para establecer tolerancia inmune en el feto en desarrollo y suprime varias vías de señalización [110].

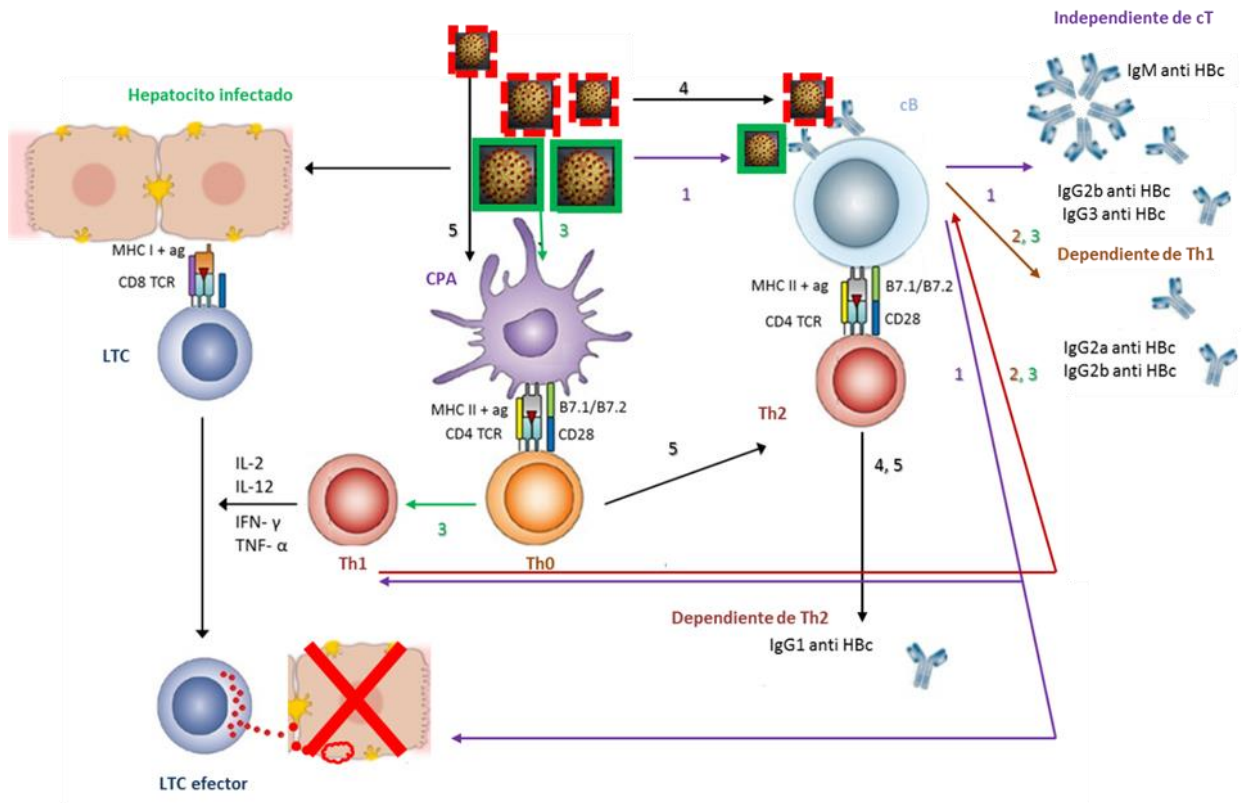
2.2.3.3.1 Rol de HBeAg en la inmunopatogénesis

El HBeAg produce una respuesta principalmente humoral dependiente de cT que da lugar a los anti-HBe ya que induce preferentemente una respuesta tipo Th2. También es capaz de inducir respuesta tipo Th1 (que puede provocar la activación de macrófagos) según las citoquinas del entorno [36]. El HBeAg es capaz de interactuar con TLRs y moléculas de la cascada de señalización, con lo cual suprime la activación de factores inflamatorios, induciendo un estado de inmunosupresión de la respuesta innata que evita la eliminación inmunitaria de la infección [111,112]. Se ha visto que HBeAg reduce la expresión de TLR2 en los hepatocitos, células de Kupffer y monocitos periféricos en pacientes con HBC [108]. En cuanto a la inmunidad adquirida, HBeAg disminuye la presentación antigénica *down*-regulando moléculas co-estimuladoras de las células de Kupffer y macrófagos, con lo cual bloquea la maduración de Th [36]. Puede atravesar la placenta y establecer tolerancia en el útero Th específica para HBeAg y HBcAg, este mecanismo sería el responsable de la alta tasa de cronicidad observada en recién nacidos de madres HBsAg(+) infectados perinatalmente [113].

2.2.3.3.2 Rol de HBcAg en la inmunopatogénesis

Aunque HBcAg y HBeAg comparten la mayor parte de su secuencia y son altamente reactivas a nivel de cT, son reconocidas de manera diferente por cB y la respuesta inmune parece estar regulada de manera independiente. HBcAg es mucho más inmunogénica que HBeAg (y HBsAg), de hecho, de los antígenos del VHB, HBcAg induce la respuesta más fuerte de cB, cT y LTC y actúa como antígeno T dependiente o independiente (Figura 20). Las posibles causas son: (1) plegamiento 3D único, (2) la presencia de una región que interactúa con inmunoglobulinas fuera del sitio de unión antígeno-anticuerpo, (3) la presencia de muchos epítopos para cT CD4+; puede ser presentado por cB como antígeno primario a las cT, desencadenan una fuerte respuesta de LTC durante la infección que se mantiene décadas después de la recuperación clínica y que aparentemente mantiene al virus bajo control y (4) la presencia de ácidos nucleicos encapsados [114,115].

Figura 20: Representación esquemática de las vías del sistema inmune desencadenadas por la nucleocápside del VHB. (1) Vía cT-independiente: la unión de nucleocápsides conteniendo ARN (verde) lleva a la secreción de IgM, IgG2b/3 y la inducción de Th1 y LTC. (2) Seguida de la unión y captura de nucleocápsides por cB *naive*, la interacción entre estas y las cT, resulta en secreción de IgG2a e IgG2b dependiente de cT. (3) La captura de nucleocápsides conteniendo ARN lleva a la activación de células presentadoras de antígeno (CPA), secreción de IL-12 y generación de Th1, esto aumenta la secreción de IgG2a e IgG2b. (4 y 5) la unión de cápsides truncadas, desnaturalizadas o HBeAg (rojo) a cB, junto con la captura por CPA, resulta en la inducción de Th2 y secreción de IgG1[114].



Las partículas HBc son muy inmunogénicas y se desarrolla una respuesta humoral anti-HBc duradera, se producen anticuerpos anti-HBc en todos los pacientes infectados, los del tipo IgM aparecen temprano en la infección y los títulos fluctúan a lo largo de la infección crónica (Figura 17), su presencia correlaciona con persistencia de la replicación y pueden ser usados como indicador de exacerbación aguda en pacientes con HBC (sus niveles se relacionan con daño hepatocelular inmunopatogénico). Los anticuerpos de tipo IgG pueden persistir durante años, con un descenso lento, e incluso de por vida [114]. El principal sitio reconocido por cB en HBc es e1 (aa 74-84), que se localiza en la región MIR de HBc y el siguiente más importante, e2 (aa 126-133), se localiza en la otra región expuesta a la superficie. Los epítomos para cT restringidos por HLA-II son los péptidos: 1-20, 28-47, 50-69, 72-90, 81-105, 90-99, 108-122, 111-125, 117-131, 120-139, 126-146 y 141-165 y los epítomos para LCT restringidos por HLA clase I se localizan en aa 18-27, 141-151 y 88-96 (Tabla 4) [37].

Tabla 4: Posición de los aa en HBc que determinan las regiones epitópicas.

aa HBc	CÉLULA ESTIMULADA	
74-84	cB	
126-133		
1-25	cT CD4+ (Th) *Principal epítipo	
1-20		
28-47		
50-69*		
72-90		
81-105		
90-99		
108-122		
111-125		
117-131		
120-139		
126-146		
141-165		
169-183		
18-27	cT CD8+ (LTC)	Restringido por HLA-A2
141-151		Restringido por HLA-A31 y HLA-Aw68
88-96		Restringido por HLA-A11

2.2.4 HEPATOCARCINOMA RELACIONADO CON EL VHB (HCC-VHB)

El HCC es uno de los tumores sólidos más frecuentes en todo el mundo y la tercera causa de mortalidad por cáncer con más de 250.000 nuevos casos anualmente y 500.000-600.000 muertes cada año. La infección crónica por el VHB es el principal factor de riesgo para este cáncer, siendo responsable de aproximadamente el 50% de los casos en todo el mundo, y este porcentaje aumenta hasta el 80-90% en áreas de alta endemicidad como China y Corea del Sur [56]. Se estima que el riesgo de desarrollar HCC es entre 10-25 veces mayor en portadores de la infección crónica por el VHB [116].

El VHB contribuye a la hepatocarcinogénesis directamente, mediante cuatro mecanismos (1) aumento de la inestabilidad genómica, (2) mutagénesis por inserción (3) modulación de la apoptosis y proliferación y (4) modificaciones epigenéticas por HBx, o indirectamente, promoviendo el daño hepático a largo plazo mediante (1) inflamación crónica, (2) muerte y regeneración celular y (3) daño oxidativo del ADN [116]. Además existen otros factores virales, extravirales y del hospedador asociados con la progresión del HCC-VHB (Tabla 5) [103].

Tabla 5: Factores asociados con la progresión del HCC-VHB.

FACTOR	
DEMOGRÁFICO	Sexo masculino
	Edad (riesgo aumentado con edad avanzada)
	Raza
SOCIAL Y AMBIENTALES	Alcohol
	Tabaco
	Aflatoxina
FACTORES VHB	Niveles de ADN-VHB
	Presencia de HBeAg
	Genotipo (Gen/C más que Gen/B)
	Mutaciones Cp
DAÑO HEPÁTICO	Cirrosis
	Niveles aumentados de transaminasas
OTROS	Co-infección con el Virus de la hepatitis C (VHC)
	Historia familiar de HCC
	Sobre-infección Virus de la hepatitis D (VHD)

2.2.4.1 MECANISMOS MOLECULARES DE LA HEPATOCARCINOGENÉISIS ASOCIADA AL VHB

2.2.4.1.1 Integración del ADN-VHB y hepatocarcinogénesis

La integración del ADN-VHB en los cromosomas del hospedador se ha detectado en el 80% de los HCC, la integración celular del ADN-VHB ocurre cerca del extremo 5' de la hebra(-), alrededor de la región cohesiva final entre los nt 1600-1900, que es el sitio de inicio de la replicación y transcripción viral y donde se localiza el extremo 3' del gen X [50,116], en los cromosomas, los puntos de integración son sobre todo intrones, regiones intergénicas, sitios de fragilidad cromosómica, regiones ricas en secuencias repetitivas o satélite y regiones reguladoras como promotores, y la integración ocurre preferentemente en los cromosomas 3, 10, 11 y 17. Se han encontrado secuencias integradas en el promotor del gen hTERT y MLL4 (*mixed-lineage leukemia protein 4*), en un intrón de Fas, en genes relacionados con vías de señalización de calcio, en genes que codifican para proteínas del ribosoma 69S (*ARHGEF12*, *CYP2C8*, *FN1*, *PHACTR4*, *RBFOX1*, *SMAD5*), en el gen PDGF (*platelet derived growth factor*), *CCNE1* (*Cyclin 1*), *SENP5* (*Sentrin-specific protease 5*), *ROCK1* (*Rho-associated coiled-coil containing protein kinase 1*) y ciclina A y en oncogenes de la familia *myc*, *CCNA*, *RARβ* (*retinoic acid receptor β*), *MK* (Mevalonato kinasas), *CPN2* (carboxipeptidasa N), *SERCA* (sarco/endoplasmic reticulum calcium) y *human minichromosome maintenance protein 8* [116–118].

La integración del VHB en el genoma del hospedador causa mutagénesis insercional; puede alterar la función o regulación de los genes del hospedador, así como inestabilidad genómica, además el ADN-VHB integrado sufre alteraciones, como deleciones, duplicaciones inversas o reordenamientos. A pesar de los reordenamientos, las regiones codificantes de preS2 y X normalmente están conservadas y se pueden transcribir, por lo que LHB y HBx han sido relacionadas con la hepatocarcinogénesis [117–119].

2.2.4.1.2 Proteínas virales y HCC

Las proteínas virales, sobre todo HBx como ya se ha explicado en el apartado 2.1.4.1.4 *ORF X. HBx EN LA INFECCIÓN CRÓNICA Y HCC*, pero también LHB y la proteína de splicing del VHB HBSP (*Hepatitis B spliced protein*), se han asociado a la hepatocarcinogénesis.

Muchos estudios han asociado deleciones en preS2 con HCC, en experimentos de transfección se ha visto que estas deleciones causan una sobreproducción y acumulación de LHB en el RE, desencadenando estrés del RE que induce daño en el ADN, que finalmente lleva a la progresión del ciclo celular y transformación [110], además estas variantes preS2 activan directamente vías promotoras de tumores como VEGF/AKT/mTOR y p27/retinoblastoma/Cdk2/cyclin A, D, entre otras [120].

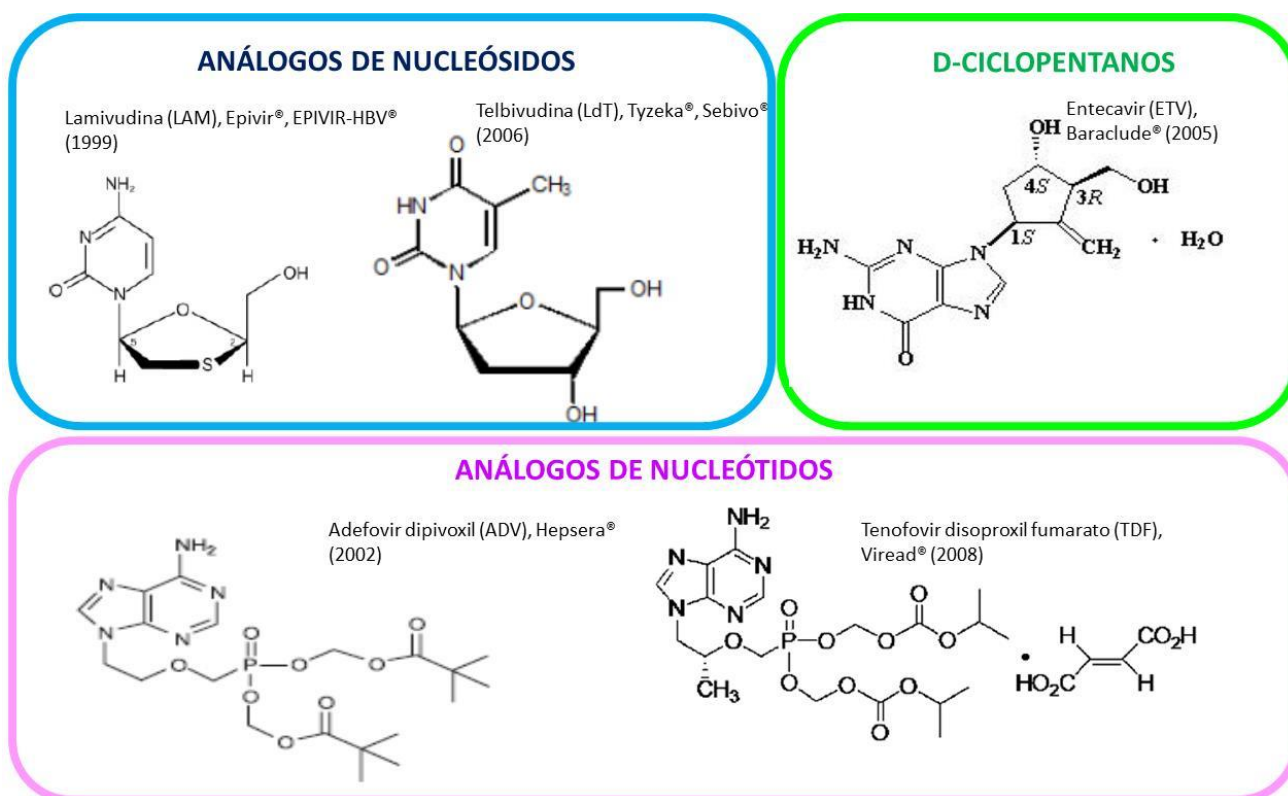
Aunque no es esencial para la replicación, el ARNpg puede sufrir splicing y se han descrito más de una docena de variantes de splicing (spVHB), las más frecuente (sp1) es una variante de 2,2 kb que tiene una deleción de 1,2 kb entre el extremo 3' del ORF C y preS2, que evita la expresión de la proteína P entera, y codifica la proteína de splicing del VHB HBSP (*Hepatitis B spliced protein*), en la que los 46 aa Nt de P se fusionan con 47 aa codificados por otro ORF. Estas variantes, si se les suministra una proteína P intacta, pueden ser empaquetadas y replicarse. Se han detectado en sangre en pacientes crónicos, pero no en pacientes que se recuperan tras una infección aguda, y han sido asociadas con enfermedad hepática avanzada, un ratio elevado spVHB /virus intactos en sangre correlaciona con fibrosis y HCC. La patogenicidad podría atribuirse a que promueven la proliferación y migración de las células, a la expresión aumentada de HBc y HBx y al aumento de la replicación del ADN, además la mayoría de las nuevas variantes de splicing identificadas retienen el gen X, mientras que el 50% retienen los genes preCore/Core y la mayoría carecen de la región S y RT, lo que podría suponer una ventaja en términos de escape de la presión del sistema inmune y del tratamiento [110,121].

2.2.5 TRATAMIENTOS ANTIVIRALES

2.2.5.1 FÁRMACOS APROBADOS Y ESTRATEGIAS DE TRATAMIENTO

Los fármacos aprobados para el tratamiento de la infección por el VHB se dividen en inmunomoduladores; interferón α -2b (IFN α , Intron A[®], aprobado en 1999) y el interferón pegilado 2a (PEG-IFN, Pegasys[®], aprobado en 2005) y cinco antivirales orales directos inhibidores de la polimerasa (DAAs), que se pueden clasificar en análogos de nucleósidos; Lamivudina (LAM) y Telbivudina (LdT), análogos de nucleótidos; Adefovir dipivoxil (ADV) y Tenofovir disoproxil fumarato (TDF) y D-ciclopentanos; Entecavir (ETV) (Figura 21) [122]. *De aquí en adelante se referirá a los cinco DAAs como NUCs (análogos de nucleós(t)idos).

Figura 21: Inhibidores de la polimerasa aprobados para el tratamiento de la infección por el VHB, con sus nombres comerciales y el año en el que fueron aprobados.



2.2.5.1.1 IFN α /PEG-IFN

Son citoquinas con actividad antiviral débil e inmunomoduladora que inhiben la replicación viral y promueven la eliminación de hepatocitos infectados por el sistema inmune. *Up-regulan* una serie de genes antivirales estimulados por interferón ISGs (*interferon stimulated genes*) y la actividad antiviral resulta de la activación de células NK, la inhibición de la transcripción del genoma viral a través de la regulación epigenética del ADNccc, la

desestabilización de la nucleocápside viral y, recientemente se ha sugerido, la destrucción del ADNccc a través de la activación de APOBEC3A (*apolipoprotein B mRNA editing catalytic polypeptide 3A*) [123].

2.2.5.1.2 NUCs

Actúan directamente en la replicación viral, LAM inhibe la elongación de la hebra(-) (transcripción reversa), ADV y TDF evitan el cebado de la transcripción reversa y la elongación de la hebra(-), ETV inhibe las tres actividades sintéticas de la polimerasa evitando la elongación de la hebra(+), además del cebado y la elongación de la hebra(-) y hay pocos datos de LdT, pero también podría inhibir las tres actividades sintéticas de la polimerasa [124].

2.2.5.2 ESTRATEGIAS DE TRATAMIENTO Y OBJETIVO

Existen dos estrategias, el tratamiento de duración finita con IFN o el tratamiento a largo plazo con uno o varios NUCs, en la tabla 6 se resumen los tratamientos en adultos y niños recogidos en la Guía Clínica del 2016 de la AASLD (*American Association for the Study of Liver Diseases*) para el tratamiento de la HBC [125], ambas estrategias tienen ventajas y desventajas que se resumen en la tabla 7 [122].

El objetivo del tratamiento es mejorar la calidad de vida y la supervivencia mediante la prevención de la progresión de la enfermedad a cirrosis, cirrosis descompensada, enfermedad hepática en estadio final ESLD (*end-stage liver disease*), HCC y muerte. Este objetivo se puede alcanzar si se inhibe la replicación del VHB de forma mantenida, lo que se denomina respuesta virológica sostenida (RVS), ya que la reducción de la actividad histológica asociada hace que disminuya el riesgo de cirrosis y HCC. Sin embargo, la infección crónica por el VHB no puede erradicarse por completo debido a que la supresión mantenida de la replicación no lleva a la eliminación del ADNccc, que persiste en el núcleo de los hepatocitos infectados. Si el tratamiento para, la retrotranscripción y, en consecuencia la replicación, puede reactivarse, de ahí que la duración de los tratamientos actuales sea indefinida. No obstante algunos parámetros virológicos y serológicos pueden evaluarse para detener el tratamiento, como se verá en el siguiente apartado [122].

Tabla 6: Tratamientos antivirales aprobados en adultos y niños. A: años, M: meses, S: semana, D: día, *En personas con disfunción renal las dosis tienen que ser ajustadas. ** La dosis de ETV en adultos es 1 mg/día si tratamiento previo con LAM o LdT o cirrosis descompensada. #La dosis de ETV en niños >2 años sin tratamiento previo y ≥10 kg son: 0,15 mg (10-11 kg), 0,2 mg (>11-14 kg), 0,25 mg (>14-17 kg), 0,3 mg (>17-20 kg), 0,35 mg (>20-23 kg), 0,4 mg (>23-26 kg), 0,45 mg (>26-30 kg) y 0,5 mg (>30 kg), con tratamiento previo: 0,30 mg (10-11 kg), 0,4 mg (>11-14 kg), 0,5 mg (>14-17 kg), 0,6 mg (>17-20 kg), 0,7 mg (>20-23 kg), 0,8 mg (>23-26 kg), 0,9 mg (>26-30 kg) y 1,0 mg (>30 kg) [125].

	DOSIS EN ADULTOS*	USO EN NIÑOS*	POSIBLES EFECTOS SECUNDARIOS	MONITORIZACIÓN TRATAMIENTO
PEG-IFN (Adulto) IFN α (Niños)	180 µg/S	≥1A 6x10 ⁶ UI/m ² 3 veces/S	Síntomas tipo gripe, fatiga, alteraciones de ánimo, citopenias, alteraciones autoinmunes en adultos Anorexia y pérdida de peso en niños	Hemograma (cada 1-3M) TSH (cada 3M) Monitorización clínica para complicaciones autoinmunes, isquémicas, neuropsiquiátricas e infecciosas
Lamivudina	100 mg/D	≥2A 3 mg/kg/D Máx 100 mg	Pancreatitis Acidosis láctica	Amilasa si síntomas Niveles de ácido láctico si preocupación clínica
Telbivudina	600 mg/D	-----	↑ CK y miopatía Neuropatía periférica Acidosis láctica	Creatin Kinasa si síntomas Evaluación clínica si síntomas Niveles de ácido láctico si preocupación clínica
Entecavir	0,5-1 mg/D**	≥2A basada en el peso 10-30 kg >30 Kg 0,5 mg/D#	Acidosis láctica	Niveles de ácido láctico si preocupación clínica
Adefovir	10 mg/D	≥12A 10 mg/D	Fallo renal agudo Síndrome Fanconi Diabetes insípida nefrogénica Acidosis láctica	Aclaramiento de creatinina basal Si hay riesgo de fallo renal, aclaramiento de creatinina, fosfato en suero, glucosa en orina y proteínas al menos anualmente Estudio de densidad ósea basal y durante el tratamiento en personas con historial de fracturas o riesgo de osteopenia
Tenofovir	300 mg/D	≥12A 300 mg/D	Nefropatía Síndrome Fanconi Osteomalacia Acidosis láctica	Niveles de ácido láctico si preocupación clínica

Tabla 7: Principales ventajas e inconvenientes de PEG-IFN y NUCs en el tratamiento de la HBC.

	VENTAJAS	INCONVENIENTES
PEG-IFN	Duración finita	Efecto antiviral moderado
	Ausencia de resistencia	Tolerabilidad inferior
	Tasas superiores de seroconversión anti-HBe y anti-HBs con 12 meses de tratamiento	Riesgo de acontecimientos adversos Inyecciones subcutáneas
NUCs	Efecto antiviral potente	Duración indefinida
	Buena tolerancia	Riesgo de resistencia
	Administración oral	Seguridad a largo plazo desconocida

2.2.5.3 VALORACIÓN DEL TRATAMIENTO

La valoración ideal del tratamiento es la pérdida del HBsAg asociada con la seroconversión anti-HBs, lo que llevaría al fin del tratamiento. Sin embargo esto sólo se logra en una minoría de pacientes [123], la EASL (*European association for study of liver disease*) propone un criterio más realista que es la inducción de RVS con normalización de los niveles de ALT (respuesta bioquímica) [122]. Se define RVS en el tratamiento con IFN/PEG-IFN como una concentración de ADN-VHB <2000 UI/ml, que se suele determinar a los 6 meses y al final del tratamiento, así como a los 6 y 12 meses una vez finalizado el tratamiento. La RVS en el tratamiento con NUCs se define como ADN-VHB indetectable mediante un ensayo analítico de PCR sensible, generalmente se evalúa cada 3-6 meses durante el tratamiento, según la gravedad de la hepatopatía y el tipo de NUCs. En la tabla 8 se muestran las tasas de RVS con estos fármacos resumidos en diferentes guías clínicas internacionales [122,123].

Tabla 8: Resultados de los principales estudios del tratamiento de pacientes con HBC HBeAg(+) o HBeAg(-) a los 6 meses y tras 12 meses (48 ó 52 semanas) de tratamiento con PEG-IFN y a los 12 meses (48 ó 52 semanas) de tratamiento con NUCs. pt: post-tratamiento, M: meses, A: años [122].

	PEG-IFN 2a		LAM		LdT		ETV		ADV		TDF	
	HBe(+)	HBe(-)	HBe(+)	HBe(-)	HBe(+)	HBe(-)	HBe(+)	HBe(-)	HBe(+)	HBe(-)	HBe(+)	HBe(-)
Supresión ADN-VHB	10-40%	20-40%	36-44%	72-73%	60%	88%	60%	90%	13-21%	51-63	75%	90%
Seroconversión anti-HBe	30-35%		16-18%		22%		20%		12-18%		20%	
Normalización ALT	35-50%	60%	41-72%	71-79%	77%	74%	70-80%	80-90%	48-54%	72-77	70%	80%
Pérdida HBsAg	5% (6M pt) 10% (3A pt)	4% (6M pt) 6% (3A pt)	0-1%	0%	0,5%	0%	3% (1A pt) 5% (2A pt)	<1%	0%	0%	3% (1A pt) 8% (3A pt)	<1%

La administración intravenosa y los efectos secundarios del PEG-IFN, además de su dependencia de la eficacia con el genotipo (Ver apartado 2.3.1.6 *Implicación del genotipo en el tratamiento*) limitan el tratamiento a un pequeño porcentaje de casos altamente seleccionados (pacientes jóvenes con genotipo A) [126]. ETV y TDF son los fármacos más potentes disponibles actualmente, en monoterapia durante un período ≥ 3 años alcanzan una RVS en la inmensa mayoría de pacientes y por eso se recomiendan como tratamiento de primera línea. Los otros tres (LAM, ADV y LdT) pueden usarse solamente si no se dispone de fármacos más potentes o los que existen no son los apropiados [122].

2.2.5.4 RESISTENCIAS A NUCs

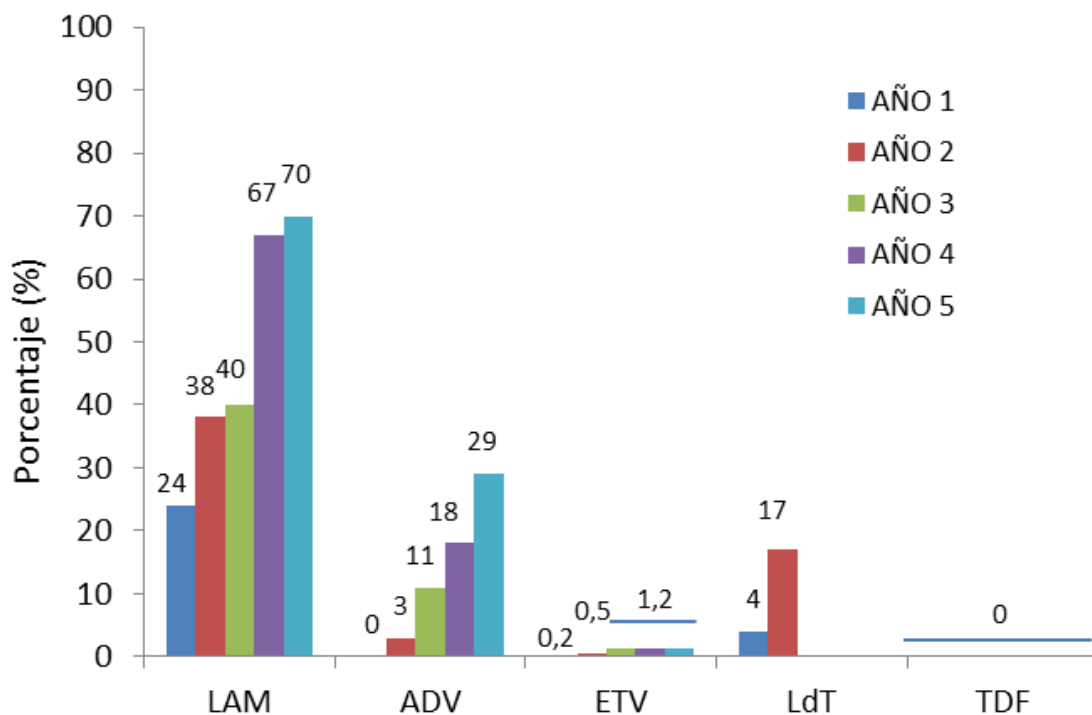
La principal desventaja de los tratamientos con NUCs es la selección de mutaciones en el gen de la polimerasa que confieren resistencias. Las mutaciones de resistencia se clasifican en mutaciones primarias, son responsables directas de la resistencia, y mutaciones secundarias

o compensatorias, asociadas con mejoras en el *fitness* viral de las variantes con mutaciones primarias. La selección de mutaciones de resistencia primaria pueden reducir el *fitness* viral, es decir, la capacidad de un virus para replicarse en un determinado ambiente [127], entonces la aparición de mutaciones compensatorias mejoran la eficacia de la polimerasa viral para restaurar los niveles de replicación [124]. Otro concepto importante es la barrera genética de resistencia a un fármaco, que se define como el número de mutaciones primarias necesarias para que aparezca la resistencia, así un fármaco con una alta barrera tiene baja probabilidad de desarrollar resistencias o porque necesita múltiples mutaciones primarias o porque estas mutaciones son más difíciles de seleccionar [124]. Las tasas de resistencias (Figura 22) y el patrón de mutaciones (Tabla 9) son diferentes para cada NUCs.

2.2.5.4.1 TASAS DE RESISTENCIA

Las tasas de resistencia a los cinco años de cada uno de los NUCs utilizados en el tratamiento de la infección por el VHB se muestran en la figura 22 [122].

Figura 22: Incidencia acumulada de resistencia del VHB a lamivudina (LAM), adefovir (ADV), entecavir (ETV), telbivudina (LdT) y tenofovir (TDF) en ensayos clínicos en pacientes con HBC no tratados anteriormente con NUCs (*naive*) [122].



2.2.5.4.2 PATRONES DE MUTACIÓN (Tabla 9)

Las dos mutaciones asociadas a la resistencia primaria a LAM son rtM204V/I en el motivo YMDD en el dominio funcional C del dominio RT de la polimerasa. La mutación rtL180M en el dominio funcional B, que aparece sobre todo en combinación con la rtM204V, restaura la capacidad replicativa a niveles semejantes de la polimerasa intacta y aumenta más la resistencia a LAM. A esta combinación rtL180M/rtM204V se pueden añadir mutaciones compensatorias como rtL80V/I, rtI169T, rtV173L, rtL180M, rtT184S/G, rtS202I y rtQ215S, no aumentan la resistencia a LAM pero aumentan la eficacia replicativa de la polimerasa, otras mutaciones compensatorias propuestas son rtV84M, rt214, rtL217P, rtL229M, rtI233V y rtN238H [128]. La mutación rt181T/V en el dominio B también da lugar a resistencia primaria a LAM aunque con una frecuencia muy baja [129].

Las mutaciones rtA181T/V, localizada en el dominio B de la RT, y rtM204V/I también se han asociado a resistencias primarias a LdT [120,130].

El ETV tiene una alta barrera genética de resistencia, para la resistencia es necesario la preselección de la combinación de resistencias a LAM rtL180M+rtM204V, junto con mutaciones adicionales en algunas de estas posiciones: rtI169T, rt184G/S, rtS202I/G y rtM250V [131].

La resistencia a ADV está caracterizada por rtN236T y/o rt181V/T en los dominios D y B de la polimerasa, respectivamente [120], la mutación rtI233V también puede conferir resistencia aunque los resultados son contradictorios [128].

No se han descrito resistencias a TDF después de tres y cuatro años de tratamiento (Figura 22) pero rtA181T/V y rtN236T, las principales mutaciones de resistencia asociadas a ADV, reducen su sensibilidad [120]. Por otro lado, estudios *in vitro* han apuntado que la mutación rtA194T también podría estar relacionada con la resistencia a TDF [132], sin embargo no se ha descrito la selección de esta mutación en la práctica clínica.

Tabla 9: Mutaciones de resistencia primaria en el gen de la polimerasa. *Número de mutaciones necesaria para la resistencia. ETV es el que tiene una barrera más alta, se requieren tres mutaciones para la resistencia (rtL180M, rtM204V/I y una mutación simple en rtT184, rtS202 o rtM250), por el contrario LAM, LdT o ADV sólo necesitan una mutación [124].

		DOMINIO DE LA POLIMERASA					
		A	B	C	D	E	Nº mutaciones*
LAM y LdT			rtV173L rtL180M rt181T/V	rtM204V/I			1
ETV			rtI169 rtL180M rtT184	rtM204V/I rtS202		rtM250I/V	3
ADV			rtA181T/V		rtN236T		1
TDF							

Además existen resistencias cruzadas, es decir, resistencia a un fármaco al que el virus no ha estado expuesto. Este fenómeno se caracteriza por un aumento rápido y repentino de la actividad viral conocido como VBK (*Virological Breakthrough*) y se debe iniciar un tratamiento de rescate con el antiviral disponible más efectivo que no tenga resistencias cruzadas (Tabla 10) [124]. En el caso de resistencia a LAM o LdT, se recomienda el cambio a TDF o añadirlo (o ADV si no se dispone de TDF). Sólo un 1% de pacientes sin tratamiento previo presentan resistencias a ETV, comparado con el 40% si ha estado tratado previamente con LAM, cuando aparecen se debe cambiar a TDF o añadirlo (o ADV si no se dispone de TDF). En el caso de resistencia a ADV, si el paciente es *naive*, no ha recibido NUCs antes, se debe cambiar a ETV o TDF, si el paciente ha presentado una resistencia previa a LAM, cambiar a TDF o añadirlo.

Tabla 10: Datos de resistencia cruzada para las variantes del VHB más frecuentes. Se muestran los perfiles de sustitución de aminoácidos en la columna izquierda y el nivel de sensibilidad para cada fármaco: S (sensible), I (sensibilidad intermedia/reducida) y R (resistente).

	LAM	LdT	ETV	ADV	TDF
Sin resistencias	S	S	S	S	S
M204V	R	S	I	I	S
M204I	R	R	I	I	S
L180M + M204V	R	R	I	I	S
A181T/V	I	S	S	R	S
N236T	S	S	S	R	I
L180M+ M204V/I ± I169T ± V173L ± M250V	R	R	R	S	S
L180M+ M204V/I ± T184G ± S202I/G	R	R	R	S	S

2.2.5.5 NUEVAS ESTRATEGIAS PARA CURAR LA INFECCIÓN POR EL VHB

Como se ha dicho anteriormente, actualmente no existe cura para el VHB ya que no se puede eliminar el ADNccc de los hepatocitos infectados y el tratamiento debe ser mantenido de por vida en la mayoría de los pacientes. No obstante, los avances recientes en entender la biología molecular del ciclo replicativo, incluyendo la identificación del receptor de entrada del virus (NCTP), y el desarrollo de modelos *in vitro* de infección, han dado nuevas perspectivas para el desarrollo de terapias emergentes [126]. Las estrategias se resumen en la tabla 11 y se basan en atacar directamente al virus con DAAs y/o aumentar la respuesta inmune del hospedador con HTAs (*host targeting agents*).

2.2.5.5.1 AGENTES ANTIVIRALES DIRECTOS (DAAs)

A. INHIBIDORES DEL ADNccc

Para reducir la aparición de resistencias, han surgido inhibidores potentes de la polimerasa como Besifovir (LB80380), análogo de guanósina, con una rápida fosforilación intracelular que inhibe la replicación del VHB independientemente de las resistencias que tenga a otros NUCs, Tenofovir alafenamida (GS-7340 o TAF), profármaco de segunda

generación de TDF, de metabolización intracelular a tenofovir difosfato que, comparado con TDF, muestra mayores niveles intracelulares del metabolito y menores concentraciones plasmáticas de tenofovir, además no tiene transportador en el riñón con lo que puede reducir el riesgo de afectación renal y CMX157, conjugado hexadeciloxipropil de Tenofovir [109,133]. Myrcludex B es un Péptido preS1 miristoilado que inhibe la entrada del virus [133]. Se están modificando nucleasas de unión a ADN que tienen como diana el ADNccc para lograr su eliminación física, como nucleasas con dedos de zinc, estrategias TALENs (*transcription activator-like effector nucleases*) y CRISPR (*RNA-guided clustered regulatory interspaced short palindromic repeats*). CCC-0975 y CCC-0346 son sulfonamidas disustituidas que bloquean directamente la conversión de ADNrc a ADNccc. Se están desarrollando moléculas que tienen como diana las enzimas modificadoras de histonas para conseguir la inactivación epigenética del ADNccc [126]. Derivados fenilpropanamida (AT-61, AT130) y sulfamoilbenzamida evitan la encapsidación de ARNpg y heteroarildihidropirimidinas (Bay 41-4109 y GLS4) también inhiben la replicación por el incorrecto ensamblaje de la nucleocápside[133]. Existen fármacos en desarrollo para inhibir factores celulares o virales (HBx y HBc) que contribuyen a la estabilidad/formación del ADNccc [133].

B. INHIBICIÓN ARN-VHB

Otra estrategia es inhibir la expresión de los genes virales, ya que los altos niveles de antígenos pueden llevar al agotamiento y disfunción de cT específicas, por tanto, la disminución de la expresión génica puede restaurar la respuesta inmune [133]. ARC-520 son ARNs pequeños de interferencia (siARNs) cuya diana es la transcripción [133]. REP9-AC (REP 2055) es un polímero de ácidos nucleicos (NAPs) de 40 nt poli-citidina anfipático, que inhibe tanto la entrada como la liberación de HBsAg de los hepatocitos infectados, se ha visto que evitando la inmunosupresión mediada por HBsAg, se puede restaurar la inmunidad antiviral [133].

2.2.5.5.2 AGENTES CONTRA EL HOSPEDADOR (HTAs)

El otro gran mecanismo consiste en actuar sobre el sistema inmune del hospedador.

A. ESTIMULACIÓN INMUNIDAD INNATA

Los agonistas del receptor de Linfotoxina- β activan las citidina desaminasas APOBEC3A (*apolipoprotein B mRNA editing catalytic polypeptide 3A*) y APOBEC3B en células infectadas induciendo la degradación no citolítica de ADNccc [133]. GS-9620 es un agonista de TLR7 que en modelos pre-clínicos *in vivo* estimula la producción de IFN α y otras citoquinas, activa los genes estimulados por interferón, las células NK y linfocitos [109].

B. ESTIMULACIÓN INMUNIDAD ADAPTATIVA

Las vacunas terapéuticas intentan romper la tolerancia de cT a las proteínas del virus (HBsAg y HbCag) y estimular las cT [126], GS-4774/DV601 pueden estimular cT CD8+, NASVAC contiene HBsAg y HbCag y TG1050 codifica tres dominios de antígenos del VHB, core, polimerasa y superficie y podría inducir células específicas de memoria T CD8+ [109,133]. GI-13020 es una quimera de los antígenos X, S y C inmunogénica e induce respuesta específica de cT [109]. En pacientes con HBC, la alta carga viral lleva a la disfunción de las cT por inducción de inhibidores de *checkpoints* como PD-1 (*programmed death-1*) y CTLA-4 (*cytotoxic T lymphocyte antigen-4*), Nivolumab (BMS-936558) y pembrolizumab (MK-3475) son anticuerpos monoclonales ant-PD1, al bloquear PD-1 aumentan la proliferación de cT CD8+ y favorecen la producción de IFN en linfocitos intrahepáticos [109,133].

Tabla 11: Terapias emergentes para el tratamiento de la infección crónica por el VHB y su estado actual [109,126,133].

ESTRATEGIA		AGENTE TERAPÉUTICO	ESTADO	
NUCs	INHIBIDORES ADNccc	NUEVOS NUCs	Besifovir (LB770380)	Fase III
			Tenofovir alafenamide (TAF, GS-7340)	Fase III
			CMX157	Fase II
		INHIBICIÓN DE LA ENTRADA	Myrcludex B	Fase II
		SITIOS DE CORTE ESPECÍFICOS DEL ADN	ZFNs (<i>zinc finger nucleases</i>)	Preclínico
			TALENs (<i>transcription activator-like effector nucleases</i>)	Preclínico
			CRISPR	Preclínico
		INHIBICIÓN DE LA CONVERSIÓN A ADNccc	CCC-0975/CCC-0346	Preclínico
		CONTROL EPIGENÉTICO ADNccc		
	INHIBIDORES ENSAMBLAJE	AT-61/AT-130	Fase I	
		Bay 41-4109/GLS4	Fase I	
		NVR-1221	Fase I	
	UNIÓN ADNccc A PROTEÍNAS			
INHIBICIÓN ARN-VHB	<i>KNOCK-DOWN</i>	ARC-520	Fase II/III	
	INHIBICIÓN LIBERACIÓN HBsAg	REP9-AC (REP 2055)	Fase II	
HTAs	INMUNIDAD INNATA	AGONISTAS RECEPTOR LINFOTOXINA- β	Preclínico	
		AGONISTAS TLRs	GS-9620	Fase II
	INMUNIDAD ADAPTATIVA	VACUNAS TERAPÉUTICAS	GS-4774/DV-601	Fase II
			NASVAC	Fase III
			TG1050	Fase I
		ACTIVACIÓN cT ESPECÍFICAS	GI-13020	Preclínico
INHIBIDORES DE LOS <i>CHECKPOINTS</i>	Nivolumab (BMS-936558)/pembrolizumab (MK-3475)	Preclínico		

2.3 VARIABILIDAD GENÉTICA

2.3.1 GENOTIPOS Y SUBGENOTIPOS

2.3.1.1 CLASIFICACIÓN

El VHB se divide en ocho genotipos confirmados (A-H) y dos propuestos (I y J), basados en una diferencia >8% en toda la secuencia del genoma [134]. A su vez la diversidad dentro de los genotipos ha llevado a su división en numerosos subgenotipos (A1-A7, B1-B9, C1-C16, D1-D9, F1-F4 e I1-I2) cuando la divergencia está entre 4-8%, el Genotipo C (Gen/C) tiene 16 subgenotipos y probablemente es el más antiguo, por el contrario los Gen/E, Gen/G y Gen/H no tienen subgenotipos, lo que sugiere un origen reciente [110]. La letra/número viene dado por el orden en el que fueron descubiertos y no corresponde con la evolución, por ejemplo A4 y D5 son los más antiguos de los Gen/A y D, respectivamente [135]. La clasificación de los supuestos Gen/I y J ha sido controvertida, en el 2000 se aislaron tres genomas completos en Vietnam con una región preS y S correspondiente al Gen/A y G, respectivamente, y el resto de la secuencia C4, asignar a este recombinante el nuevo Gen/I tuvo oposición ya que la divergencia con el Gen/C era <8%, pero al aislarse virus similares en Laos, sur de China y una tribu primitiva del este de India, el punto de corte para la divergencia entre genotipos se redujo a 7,5% y el Gen/I fue reconocido [110]. El supuesto Gen/J fue aislado en un japonés de 88 años que probablemente contrajo la infección en Borneo, una isla del Sudeste Asiático, durante la II Guerra Mundial, el genoma completo difiere de los genotipos A-I al menos en 9,9%, pero presenta menor divergencia con el genoma del VHB de gibón y probablemente es un recombinante entre este virus en los genes P y S y el Gen/C del VHB en el core [110]. Antes del análisis molecular la clasificación se basaba en la heterogeneidad inmunológica del HBsAg, lo que llevó a la categorización en 10 serotipos (ayw1, ayw2, ayw3, ayw4, ayr, adw2, adw3, adw4q-, adr y adrq-), después se ha visto que los cambios de aa en la región MHR son los responsables de esta clasificación (Tabla 12) [135] y están parcialmente correlacionados con los genotipos [136].

Tabla 12: Posiciones más importantes de la SHBs para definir los serotipos y subtipos. El determinante “a” y al menos dos parejas de epítomos variables y mutuamente excluyentes: “d” o “y” y “w” o “r”, permiten la identificación de cuatro subtipos antigénicos: adw, ayw, adr y ayr, la designación d/y hace referencia al aa 122 de HBsAg (Lys o Arg) y las letras w/r al aa 160 (Lys o Arg). A su vez se dividen en 10 serotipos menores, ayw1, ayw2, ayw3, ayw4, ayr, adw2, adw3, adw4q-, adr y adrq-, la designación w1-4 obedece a la identidad del aa 127 (Pro, Thr o Leu) y el subtipo q- corresponde al aa 177 (Ala) y aa 178 (Gly) [219]. *Reactividad w1 también requiere Arg122, Phe134 y/o Ala159.

Posición HBsAg	aa	Especificidad
122	Lys	d
	Arg	y
127	Pro	w1*/w2
	Thr	w3
	Leu	w4
134	Tyr	ayw2/ayw3
	Phe	ayw1/ayw4/adw2/adw4
160	Lys	w
	Arg	r

2.3.1.2 DISTRIBUCIÓN GEOGRÁFICA

La distribución geográfica y étnica de genotipos y subgenotipos está bien caracterizada (Figura 23, Tabla 13) [110,135–139]. El Gen/A presenta una distribución global y predomina en el noroeste de Europa, norte de América, sur de África y Brasil. El Gen/B aparece sobre todo en Asia pero también en Oceanía y alrededor del Polo Norte. El Gen/C también aparece en Asia, sobre todo en las islas del Pacífico, con diferentes tasas de prevalencia en los distintos países, y Oceanía. El Gen/D es ubicuo, es el más frecuente en el sur de Europa, países mediterráneos, India y Rusia. El Gen/E está restringido al África subsahariana, siendo el más prevalente en esta región. El Gen/F co-circula en población indígena de América del Sur y Centroamérica, probablemente originado en las poblaciones amerindias en Alaska, América Central y del Sur. El Gen/G es poco frecuente, pero a pesar de su baja prevalencia está muy extendido y aparece en Europa (Francia y Alemania), América (EE.UU), Asia y África. El Gen/H se da en Centroamérica, probablemente originado en las poblaciones amerindias. El Gen/I está restringido a Laos, Vietnam, India y China y el Gen/J a Japón. En algunos casos la distribución de genotipos/subgenotipos está relacionada con el origen étnico, por ejemplo B6 está confinado a poblaciones indígenas del Ártico canadiense, C4 a los aborígenes australianos, C3 y C5-C10 a población indígena de Indonesia y se ha propuesto que los (sub)genotipos han evolucionado en diferentes corrientes de migraciones humanas [135]. No obstante, la distribución geográfica de los genotipos se está alterando debido a los viajes y a la migración, como se demuestra por la aparición de genotipos europeos, asiáticos y africanos en países de baja endemicidad como Australia [110].

Figura 23: Distribución geográfica de los genotipos del VHB [137].

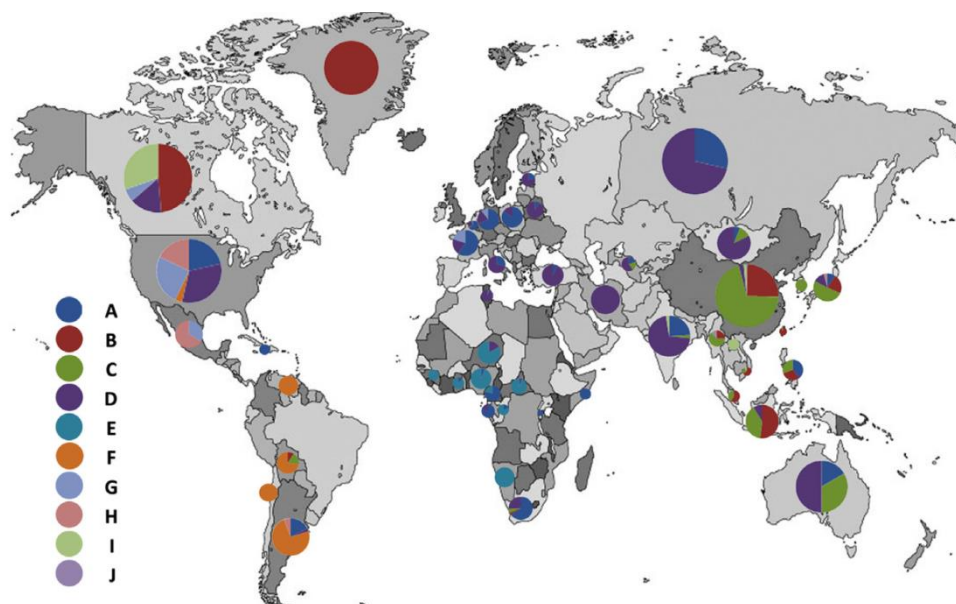


Tabla 13: Distribución geográfica de los subgenotipos del VHB.

SUBGENOTIPO	DISTRIBUCIÓN GEOGRÁFICA
A1	Restringido a África e India. Prevalente en África subsahariana
A2	Europa y norte de América
A3	Camerún y norte de África, Gambia y oeste de África
A4	Mali
A5	Nigeria
A6	Pacientes africano-belgas
A7	Camerún
B1	Restringido a Japón, sin evidencias de recombinación
B2	En el resto de Asia, es un recombinante B/C (gen preCore/core derivado del Gen/C). Prevalente en Taiwán
B3	Indonesia
B4	Vietnam y Francia
B5	Filipinas
B6	En población indígena del Ártico, incluyendo Alaska, norte de Canadá y Groenlandia, sin evidencias de recombinación, y Sur de China
B7	Este de Indonesia (islas Nusa Tenggara)
B8, B9	Indonesia
C1	Este de Asia, Taiwán sobre todo
C2	China y Sudeste asiático
C3	Oceanía
C4	Aborígenes australianos
C5	Filipinas
C6	Indonesia
C7	Filipinas, Indonesia
C8-C16	Indonesia
D1	Irán y países fronterizos
D2	Europa del este y Rusia
D3	Serbia, Sur de África y Alaska
D4	Oceanía y Somalia
D5	Este de India
D6	Indonesia
D7	Túnez
D8	Níger
D9	Este de India
F1	Centroamérica, Argentina, Chile y Alaska
F2	Centro y sur de América, Venezuela y Brasil
F3	Norte de Sudamérica, Panamá, Venezuela y Colombia
F4	Sur de Sudamérica, Brasil, Argentina y Bolivia.
I1	Laos, Vietnam, noroeste de China
I2	Laos, Vietnam

2.3.1.3 RECOMBINACIÓN INTERGENOTÍPICA

En todo el mundo se han documentado nuevas variantes generadas por recombinación entre diferentes genotipos, por ejemplo, el gen preCore/Core de B2-B5 tiene su origen en el Gen/C [140], el VHB dominante en el Tíbet es un recombinante C/D [141], el Gen/I es un recombinante A/C/G y se han identificado recombinantes A/D, G/C y D/E en África [142], también hay evidencias de recombinación entre genotipos de VHB humanos y VHB de simios,

como en el Gen/J. Además el análisis a gran escala de secuencias completas en bases de datos ha evidenciado incongruencias en árboles filogenéticos construidos desde diferentes regiones del genoma del VHB, que han llevado a la detección de nuevos genomas recombinantes [138].

2.3.1.3.1 MECANISMO DE RECOMBINACIÓN

El mecanismo exacto no se conoce pero existen una serie de factores biológicos y epidemiológicos que afectan la probabilidad de recombinación como (1) la co-circulación en la misma área geográfica de varios genotipos, (2) la homología genética entre genotipos, (3) el tamaño de la población viral en el hospedador, (4) la prevalencia de la infección por el VHB, (5) la tasa de co-infección, (6) un alto grado de cronicidad del VHB y (7) un bajo nivel de salud pública, de ahí que la mayoría de recombinantes se hayan descrito en el este de Asia o África [135,143]. Gracias al desarrollo de las técnicas de secuenciación masiva NGS (*Next Generation Sequencing*) se ha visto que la tasa de infecciones mixtas en pacientes con HBC es más alta de lo esperado [144] y esta co-existencia de diferentes genotipos infectando la misma célula podría llevar al intercambio de material genético, en Asia se ha descrito una alta prevalencia de infecciones mixtas [145], por ejemplo mezclas Gen/B+C en Taiwán [12], y en Europa Central y del Este se han descrito infecciones mixtas (Gen/A+D, Gen/A+D+F y Gen/D+F) en el 8% de los individuos infectados [146]. Parece poco probable que la recombinación ocurra durante la transcripción reversa, ya que el proceso tiene lugar en la nucleocápside implicando sólo una molécula de ARN_{pg}, con lo que probablemente ocurre en el núcleo donde la recombinación homóloga (HR) ocurriría entre moléculas de ADN_{ccc} de diferentes genotipos co-infectantes [138]. En el mismo estudio donde se sugirió el papel de la TDP2 como la enzima que inicia el proceso de transformación de ADN_{rc} a ADN_{ccc} [86], ya se propuso el complejo MRN (formado por MRE11, Rad50 y NBS) como un posible candidato para sustituir TDP2. Este complejo tiene un papel destacado en vías de reparación del ADN por HR y vías clásicas y alternativas de unión de extremos no homólogos (NHEJ) [147]. La transformación del ADN_{rc} en ADN_{ccc} por estos mecanismos implicaría que una molécula de ADN_{ccc} que ya existe en el núcleo podría proporcionar el fragmento de ADN homólogo para que tuviera lugar la HR, aunque hoy en día no hay evidencias experimentales de esto.

2.3.1.3.2 PUNTOS DE RECOMBINACIÓN (*BREAKPOINTS*)

Aunque los puntos de recombinación están distribuidos a lo largo de todo el genoma, se han identificado ciertos sitios preferentes entre las posiciones nt 1600-2000 (cerca de la región DR1), nt 2100-2500 (extremo 3' del gen core), nt 3000-300 (región preS/S) y nt 700-

1000 (extremo 3' del gen S) [138]. Además se ha visto que el Gen/C tiende a contribuir con su región preCore/Core en los genotipos recombinantes asociados [148].

2.3.1.3.3 DESCRIPCIÓN DETALLADA DE LOS GENOMAS RECOMBINANTES

Los recombinantes intergenotípicos identificados hasta ahora (Tabla 14) incluyen todos los genotipos excepto el Gen/H, los recombinante B/C y C/D representan la mayoría de recombinantes (60%) y los otros intergenotipos incluyen A/B/C, A/C, A/C/G o U/C, A/D, A/E, A/G, B/C/U, C/F, C/G, C/J, D/E, D/F y F/G (U=*unknown*) [136,138]. En general, los intergenotipos muestran el mismo patrón de circulación que los genotipos parentales [138].

Tabla 14: Descripción detallada de genotipos recombinantes, con la región donde se han aislado [138]. U=*unknown*.

GENOTIPO	PAÍS	BREAKPOINTS
A/B/C	Filipinas	B2 con Gen/A en X
	Sur de África	B2 con Gen/A en preS1
A/C	China, Taiwán	Gen/C con Gen/A integrado en preS1/S2, preCore/Core o región S-X
	Sur de África	Gen/A con pequeño fragmento de Gen/C en regiones X y PC
A/D	Argentina	A2/D3 (Gen/A en región S)
	India	Gen/A con Gen/D en preS/S, Gen/A con Gen/D en proCore/Core y P-X, Gen/D con Gen/A en preS/S y S
	Italia	Gen/D con Gen/A en S-X y recombinantes más complejos con 4 <i>breakpoints</i>
	Polonia	Gen/A con pequeño fragmento Gen/D en X
	Rusia	Gen/A con Gen/D en preCore/Core
	Sur de África	Gen/A con Gen/D en preS/S
	Túnez	Gen/D con Gen/A en preCore/Core
	Uzbekistán	Gen/D con Gen/A en S
A/E	Camerún	A3 con Gen/E en P (nt 874-1062) y X
	Guinea	Gen/E con A3 en P (nt 908-1026) y X-core. A3 con Gen/E en preCore/Core. Gen/E con A3 en X
	Francia	A3 con Gen/E en preCore/Core
	Gana	A3 con Gen/E en gen core
A/G	Brasil	Gen/G con Gen/A desde preS2 hasta S (nt 366)
	Canadá	4 y 6 <i>breakpoints</i>
	EEUU	Gen/G con pequeño fragmento Gen/A en preS1/S2
B/C	China, Camboya, Indonesia, Japón, Malasia, Filipinas, Suiza, Taiwán, Tailandia, Vietnam	Gen/B con Gen/C en preCore/Core (son los recombinantes B/C dominantes)
	China	Gen/C entre preS2 o S-core con Gen/B en el resto. Gen/B entre preS2 o S-X con Gen/C en el resto. Gen/C con Gen/B en X, S-X o S. Gen/B con Gen/C en X-core. Gen/B con Gen/C en gen core-preS2
	Indonesia, Taiwán y China	Gen/C con pequeño fragmento Gen/B en P (entre gen core-preS1). Complejos con 4 <i>breakpoints</i>
	Vietnam	Gen/C con Gen/B en X
	Filipinas	Gen/C con Gen/B en S
	Japón	Gen/C con Gen/B en preCore/Core
	Taiwán	Gen/C con Gen/B en preS1. Gen/C con Gen/B en S

Tabla 14: Continuación.

GENOTIPO	PAÍS	BREAKPOINTS
B/C/U o B2/U	Vietnam	B2 (B/C) con Gen/U entre final S-comienzo gen core
C/D	Mongolia, China	Gen/C con Gen/D en preS2/S
	China, Bélgica	Gen/C en X-preS1 y Gen/D en el resto
	China	Gen/D con Gen/C en S-X y S-PC. Gen/D con pequeño fragmento Gen/C en preS2-S
	India	Clasificados como D9 (Gen/D con Gen/C entre nt 1675-2149)
C/F	Bolivia	Gen/F con un pequeño fragmento de Gen/C en nt 1630-1779
C/G	Indonesia	Gen/C con Gen/G en preS1
	Tailandia	Gen/C con Gen/G en S. Gen/G con Gen/C en mosaico en preCore/Core
C/J	Australia	C4 con Gen/J en nt 217/330-830
D/E	Níger, Gabón, Gana, Sudán	Gen/D con Gen/E en preS1
	Irlanda	Gen/D con Gen/E en preCore/Core
	Níger	Gen/E con Gen/D en nt 85-505 y nt 796-1306
D/F	Argentina	F4 con D2 en X
F/G	Argentina	Gen/G con Gen/F en gen core-P
	Brasil	Gen/F con Gen/G en nt 482-1816. Gen/F con Gen/G en nt 1842-2100. Gen/G con Gen/F en preCore/Core
I (U/C)	Vietnam	Gen/C con fragmentos preS1-X (nt 2866-1800) de Gen/U
	China	Gen/C en nt 1670-3100 y resto Gen/U, nt 1-1250 similar a Gen/G
I (A/C/G)	India, Laos	3 <i>breakpoints</i> (1) nt 1400-3000 similar a Gen/C (2) nt 3000-400 similar a Gen/A (3) nt 400-1400 similar a Gen/G
	China	4 <i>breakpoints</i> (1) nt 1548-3051 Gen/C (2) nt 3052-295 Gen/A (3) nt 296-847 Gen/G (4) nt 848-1547 Gen/A
	Vietnam	5 <i>breakpoints</i> (1) nt 1386-3215 Gen/C (2) nt 1-396 Gen/A (3) nt 396-666 Gen/G (4) nt 666-872 Gen/C (5) nt 872-1104 Gen/G

2.3.1.4 IMPLICACIÓN DEL GENOTIPO EN LA VIROLOGÍA

Los genotipos presentan características virológicas particulares, que podrían explicar la diferente patogénesis y respuesta al tratamiento como se verá en el siguiente apartado.

2.3.1.4.1 TAMAÑO GENOMA

La mayoría de genotipos tienen un genoma de 3215 pb, excepto los genotipos A, D, E, G y J debido a inserciones (ins) o deleciones (del) en pauta (Tabla 15) [110].

Tabla 15: Longitud del genoma de los genotipos del VHB.

GENOTIPO	GENOMA (pb)
A	3221 (ins de 6 nt cerca del extremo 3' gen core)
B	3215
C	3215
D	3182 (del de 33 nt en extremo 5' preS1)
E	3212 (del de 3 nt en extremo 5' preS1)
F	3215
G	3248 (ins de 36 nt en extremo 5' gen core y del de 3 nt en 5' preS1)
H	3215
I	3215
J	3182 (del de 33 nt en 5' preS1)

2.3.1.4.2 MODO DE TRANSMISIÓN

El genotipo influye en el modo de transmisión, los Gen/B y C son transmitidos fundamentalmente por vía vertical perinatal llevando a una alta prevalencia de infección crónica en el este de Asia, mientras que los Gen/A, D y E se adquieren normalmente por transmisión horizontal. En el caso del Gen/G suele ocurrir por contacto sexual en hombres homosexuales [110]. No obstante, este aspecto no está claro ya que estas son las vías de transmisión habituales en las regiones donde dichos genotipos son mayoritarios, por lo que no se puede confirmar que el genotipo sea el responsable de la vía de transmisión *per se* o simplemente que esté utilizando la vía más común en el área correspondiente.

2.3.1.4.3 CARACTERÍSTICAS GENÉTICAS PARTICULARES DE LOS GENOTIPOS

La mutación G1896A, que elimina la expresión de HBeAg por una stop en el PC (Ver apartado 2.3.2.4.1 *Variantes en el PC*), es rara en los Gen/A, F y H y tiene su máxima prevalencia en el Gen/D [135], es común en países mediterráneos y orientales con alta prevalencia de Gen/D y B, respectivamente y es menos prevalente en Europa Occidental y norte de América donde el Gen/A predomina. Esta asociación con el genotipo se debe a las secuencias solapadas del PC y la señal de encapsidación ϵ (Figura 14B), en la región ϵ las posiciones 1896 y 1858 están emparejadas, la sustitución G1896A resulta en la estabilización de esta estructura en los Gen/B, D, E y G en los que la que la posición 1858 es una T, en cambio en los Gen/A, C, F y H hay una C en posición en 1858 y la sustitución G1896A resulta en una pérdida de estabilidad termodinámica, lo que produce una disminución de la encapsidación y replicación [149]. El Gen/A tiene alterado el sitio de corte de los 29 aa residuales Ct en p25 para dar lugar al HBeAg, lo que produce formas adicionales de la proteína más largas, aunque el significado funcional es incierto [110]. En el Gen/G existen dos codones stop en el gen core y no secreta HBeAg, esto podría explicar porque la infección no ocurre sola sino es asociación

con otros genotipos fuente de HBeAg, normalmente Gen/A [120], por otro lado, la inserción de 36 nt en el gen core causa una superproducción de proteína HBc [135].

2.3.1.5 IMPLICACIÓN DEL GENOTIPO EN LA PATOGÉNESIS

Cada vez existen más evidencias del impacto del genotipo en los resultados a corto y largo plazo de la infección crónica por el VHB [139]. En general, los Gen/A y B muestran mejores resultados clínicos que los Gen/D y C, respectivamente, incluyendo seroconversión más temprana, mayor probabilidad de remisión sostenida después de la seroconversión, menor actividad de la enfermedad en la histología y riesgo reducido de HCC [135,150].

2.3.1.5.1 GENOTIPOS B y C

En general, individuos infectados por Gen/C experimentan seroconversión HBeAg más tardía y tiene un riesgo aumentado de inflamación hepática, brotes de hepatitis, fibrosis hepática y cirrosis y se ha sido identificado como factor de riesgo para el desarrollo de HCC [120]. Estudios llevados a cabo en China han mostrado que el Gen/C es más frecuente en la IOB, en embarazadas es un factor de riesgo para la transmisión madre-hijo y C2 es un factor independiente de riesgo de cronificación [136]. En cuanto al Gen/B, en Japón B1 se ha relacionado con hepatitis fulminantes [137], en el este de Asia B2 con HCC o recurrencia de HCC en pacientes jóvenes [137], B6 con una clínica moderada y B1 con infección aguda y fulminante y desarrollo de enfermedad hepática a edad más avanzada que B2-B5 [135].

2.3.1.5.2 GENOTIPOS A, D, F Y G

Comparativamente, en Europa el Gen/A está asociado con infección crónica y el Gen/D con infección aguda [110], la infección crónica por el Gen/A presenta tasas significativamente mayores de remisión bioquímica sostenida, aclaramiento del ADN-VHB y HBsAg, comparado con la infección crónica con Gen/D [21]. En India el Gen/D está asociado con enfermedad hepática más severa y HCC en pacientes más jóvenes comparado con Gen/A [120]. En cuanto al Gen/A, en Europa pacientes infectados con A2 tienen una clínica moderada y una alta probabilidad de aclaramiento HBsAg con pérdida de ADN-VHB y en el sur y oeste de África pacientes con A1 presentan peor pronóstico que pacientes infectados con A2 [135]. Respecto al Gen/D, en EEUU es un factor de riesgo independiente de hepatitis fulminante [120,135] y en India se ha asociado con complicaciones hepáticas más severas que cualquier otro genotipo [135]. En EEUU, el Gen/G y F están asociados con un riesgo aumentado de fibrosis [110] y mayores tasas de mortalidad, respectivamente, que los Gen/A o D [135]. En un estudio se ha visto que pacientes infectados con F2 o C2 desarrollan más HCC que aquellos con A2, B6 o

subgenotipos D [135]. También se ha visto que la muerte por enfermedad hepática es más frecuente en pacientes con Gen/D y F que con Gen/A [139]. Un estudio en Alaska, donde se encuentran 5 de los 10 genotipos, ha mostrado que las tasas de complicaciones, incluyendo HCC, de los pacientes infectados con Gen/A son menores que las encontradas en pacientes infectados con Gen/D, C o F1 [120].

2.3.1.6 IMPLICACIÓN DEL GENOTIPO EN EL TRATAMIENTO

Los genotipos están significativamente asociados con la RVS a los tratamientos basados en IFN y cada vez existen más evidencias de la asociación entre el genotipo y la respuesta a NUCs, aunque existen algunos estudios contradictorios.

2.3.1.6.1 IFN

En general los Gen/A y B muestran una mejor respuesta a IFN que los Gen/D y C [139]. En pacientes HBeAg(+), la incidencia de ALT normalizada mantenida, seroconversión HBeAg entre 6-12 meses después de parar el tratamiento y la incidencia de aclaramiento HBeAg duradero después de 3 años, es significativamente mayor en los Gen/A y B que en los Gen/D o C [139]. En pacientes con Gen/A HBeAg(-) tratados con PEG-IFN la incidencia de aclaramiento HBsAg (20%) es mayor que en los Gen/B, C y D (6, 9 y 6%, respectivamente) [139] y en pacientes chinos HBeAg(-), el Gen/B parecen responder mejor a IFN que el Gen/C [136]. Por tanto, pacientes con Gen/A con ALT alta o ADN-VHB bajo y pacientes con Gen/B o C con ALT alta y ADN-VHB bajo tienen una alta probabilidad de RVS y son candidatos a tratamiento con IFN [139].

2.3.1.6.2 NUCs

En general el genotipo no tiene un impacto fuerte y claro en la respuesta a NUCs, pero algunos estudios indican cierta asociación [136]. Pacientes con Gen/B o C tratados con TDF presentan una menor probabilidad de alcanzar la RVS que aquellos con Gen/A y D [136], en un estudio se demostró que el mejor predictor de pérdida de HBsAg en el tratamiento con TDF era el Gen/A o D en pacientes de raza caucásica [150]. En un estudio en Tokio en pacientes con HBC que recibieron tratamiento de rescate con NUCs tras el fallo de LAM, el Gen/A fue independientemente asociado con la seroconversión HBsAg tanto en pacientes HBeAg(+) como HBeAg(-) comparado con otros genotipos [136]. Sin embargo, en el estudio BE-LOW (51% pacientes caucásicos y 49% asiáticos) se encontró que en aquellos tratados con ETV, con o sin TDF, durante 100 semanas, el Gen/A estaba asociado con un descenso de HBsAg más prolongado durante el tratamiento que los genotipos no-A [136]. En otro estudio con

pacientes en tratamiento con ETV o LAM, la pérdida de HBsAg era más probable en pacientes caucásico con Gen/A o D [150]. El Gen/B se ha asociado con detección más temprana de resistencias a LAM [139] y en un estudio japonés, en el que los pacientes recibieron ADV tras el fallo de LAM, la respuesta al tratamiento fue más temprana en pacientes con Gen/B que C [136].

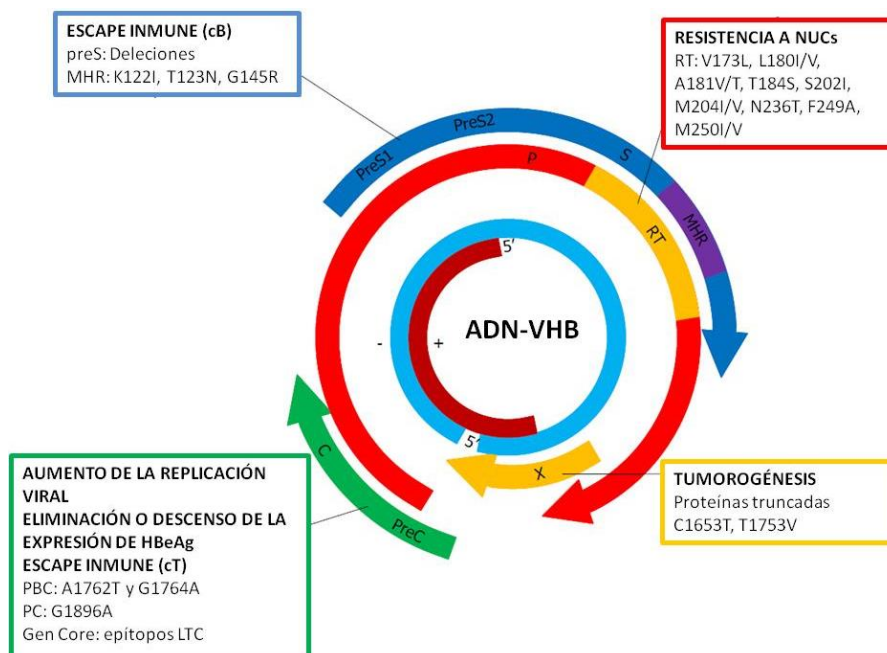
2.3.1.7 IMPACTO DE LOS GENOTIPOS EN LAS MEDIDAS PROFILÁCTICAS

La vacuna actual es un péptido recombinante súper antígeno del HBsAg que cubre el determinante “a”. Se considera que un título de anti-HBs > 10 UI/l es protector y se desarrolla en el 95-99% de los niños sanos y adultos jóvenes [151]. A pesar del éxito en la mayoría de los casos, la vacunación a veces falla debido al genotipo. Las mutaciones de escape inmune en el gen S (Ver apartado 2.3.2.1 *Variabilidad en el ORF preS/S*) varían entre genotipos, de hecho algunas de estas mutaciones de escape son la secuencia original en otros genotipos [135]. Además en Europa dos pacientes desarrollaron infección por Gen/F a pesar de tener anti-HBs y no tenían mutaciones de escape inmune [135].

2.3.2 PRINCIPALES VARIANTES GENÉTICAS DE INTERÉS CLÍNICO EN EL VHB

A continuación se describen las principales variantes del VHB en los diferentes ORFs, así como su implicación clínica (Figura 24).

Figura 24: Principales tipos de variantes en los diferentes ORFs del VHB y su significado clínico [120].



2.3.2.1 VARIABILIDAD EN EL ORF preS/S

Las mutaciones puntuales y/o deleciones en la principal región hidrofílica MHR del gen preS/S, pueden llevar al escape inmune y a la IOB [120], ya que es la región donde se localiza el determinante “a” (aa 124-147), principal diana para la respuesta neutralizante de cB y los anticuerpos usados en los test serológicos [128]. G145R es la principal mutación de escape de respuesta inmune inducida por vacunación [128] y se ha encontrado en niños que adquieren la infección por el VHB a pesar de la inmunoprofilaxis activa y pasiva [120]. Otras mutaciones de escape inmune son T116N, P120S/E, I/T126A/N/I/S, Q129H/R, M133L, K141EP142S y D144A/E [128]. Se han encontrado mutaciones similares en pacientes inmunocomprometidos (M103I, L109I, T118K, P120A, Y134H, S143L, D144E y S171F) y se consideran responsables de la reactivación de la infección en pacientes previamente anti-HBs(+). K122I se ha asociado con la IOB, ya que elimina la reactividad del HBsAg en todos los inmunoensayos comercializados [120]. En pacientes que reactivan la infección durante la inmunosupresión, además de mutaciones en el epítipo de cB, se han encontrado variantes en los epítipos de cT restringidas por HLA clase I (C48G, V96A, L175S, G185E y V190A) [128]. Variantes en preS/S también se han asociado con la capacidad de ensamblaje y replicación, las sustituciones T123N y A159G reducen el ensamblaje de los viriones y la sustitución K160N parece que compensa el efecto negativo de A159G en la producción de viriones, W172L y L173F puede aumentar la replicación viral [120]. Mutaciones puntuales y/o deleciones en esta región también podrían contribuir a la hepatocarcinogénesis [128]. La mutación W4P/R en la región preS1 podría estar asociada con enfermedad hepática severa (HCC y cirrosis) en hombres con HBC por Gen/C [128]. Se han detectado grandes deleciones en pauta en la región preS, normalmente agrupadas en la región 3' de preS1 y 5' de preS2 [110], y muchos estudios han asociado estas deleciones, sobre todo en preS2, con HCC. Al menos tres mecanismos podrían estar implicados en la patogénesis del HCC asociado a estas deleciones en preS: (1) una respuesta inmune alterada, ya que las regiones preS1 y preS2 contienen varios epítipos para las cB y cT [120], (2) la acumulación de LHB en el RE, que desencadena estrés del RE [110] y (3) los mutantes preS2 activan directamente vías promotoras de tumores como las vías VEGF/AKT/mTOR, p27/retinoblastoma/Cdk2/cyclin A, D, entre otras [120].

2.3.2.2 VARIABILIDAD EN EL ORF P

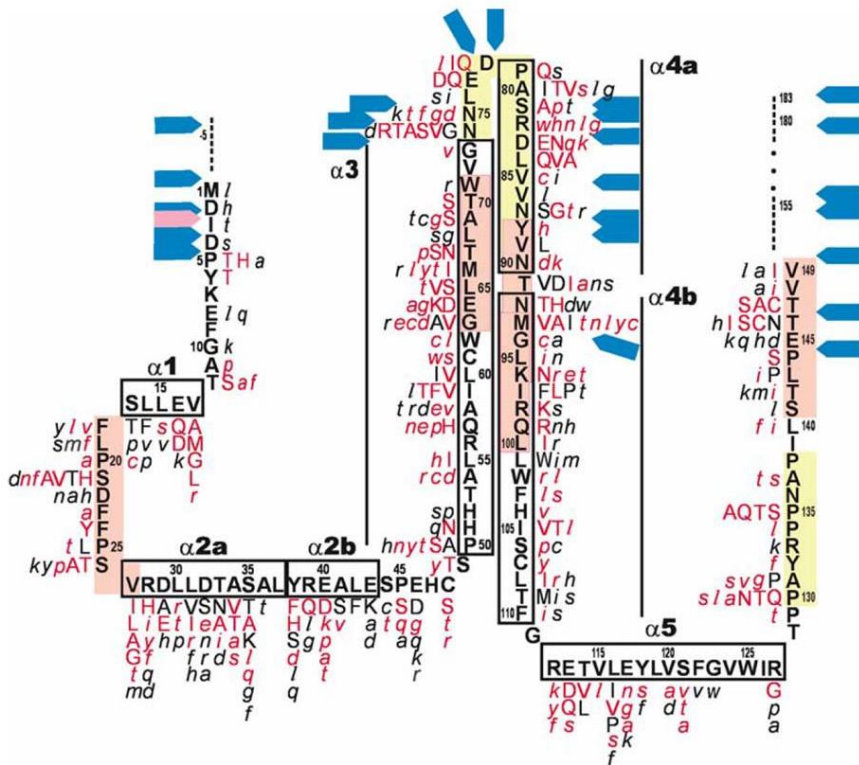
Como se ha visto en el apartado 2.2.5.4 *Resistencias a NUCs*, las mutaciones en el dominio RT de la polimerasa pueden llevar a la aparición de resistencias a los tratamientos antivirales después del uso de NUCs durante un tiempo prolongado [120]. Como el ORF P y S se solapan a nivel de los aa 8-236 de RT, algunas mutaciones de resistencias afectan al HBsAg

solapante. Por ejemplo, el mutante rtA181T seleccionado por ADF, LAM o LdT resulta en un codón stop en el gen S (sW172stop) causando un defecto en la secreción de HBsAg y esta mutación se encuentra con frecuencia en HCC [128], las mutaciones de resistencia rtV173L + rtL180M + rtM204V resultan en las mutaciones sE164D + sI195M en HBsAg que reducen la unión a anticuerpos [120].

2.3.2.3 VARIABILIDAD EN EL CORE

Aunque el gen core es el más conservado de los genes del VHB, se han encontrado mutaciones puntuales o deleciones en la mayor parte de su secuencia, la mayoría localizadas en las regiones epitópicas (Figura 25) [37]. En un estudio se han encontrado que las tres mutaciones C2288A/T, C2304A y A/G2525C/T correlacionan con enfermedad hepática severa, entre estas, tanto C2288A/T (P130T/S) como C2304A (P135Q) se localizan en la región hipervariable que cubre los aa 130-135 de la proteína core y podrían alterar los epítomos reconocidos por cB y cT, llevando al escape inmune [120], la otra mutación A/G2525C/T, en el promotor preS1, introduce sitios de unión adicionales a NF-1 y C/EBP α , sugiriendo una mayor producción de HBsAg, que podría contribuir a la enfermedad hepática avanzada [152].

Figura 25: Estructura secundaria del monómero core con la distribución de las mutaciones, hélices α en cajas, en minúscula sustituciones de aa encontradas en sólo una secuencia HBc (compilación de secuencias en GenBank y SwissProt), en rojo sustituciones sugeridas en relación con un rasgo particular en el curso de la infección. Epítomos de cB en amarillo y cT en rosa. Flechas azules: supuestos sitios de inserción y deleción. La flecha rosa indica inserción natural de 12 aa (RRTLPGPLGLD) dentro del gen core en el Gen/G [37].



2.3.2.4 VARIABILIDAD EN LA REGIÓN X

La región X del VHB hace referencia a la parte del genoma donde se solapan el extremo 3' del ORF X y el extremo 5' del ORF preCore/Core y que incluye el Cp y los elementos reguladores enhancers, DRs y TATA boxes-like (Figura 10), a continuación se detallan las variantes encontradas en esta región.

2.3.2.4.1 VARIANTES EN EL PC

Son variantes clínicamente relevantes ya que eliminan la expresión de HBeAg, las principales se resumen en la tabla 16.

Tabla 16: Variantes en el PC y PBC que llevan a la no expresión o niveles reducidos de HBeAg, respectivamente [128,149]. *Normalmente acompañando a G1896A, se ha asociado a cirrosis [149].

Región	Variante	Efecto
Codón 28 preCore	G1896A	Codón stop
Codón 28 preCore	G1897A	Codón stop
Codón 29 preCore	G1899A*	pcG29D
Primer codón del preCore	A1814C	Pérdida primer codón de inicio ATG
	T1815	Pérdida primer codón de inicio ATG
	G1816A	Pérdida primer codón de inicio ATG
Segundo codón del preCore	C1817A	Codón stop
PBC	A1762T/G1764A	↓ expresión ARN preCore

2.3.2.4.2 VARIANTES EN EL PBC

Las mutaciones en el PBC reducen pero no eliminan completamente la expresión de HBeAg y están asociadas con enfermedad hepática severa (cirrosis y HCC) independientemente del genotipo y la carga viral. Un estudio reciente con técnicas de NGS ha demostrado que pacientes con una alta proporción de mutaciones en el PBC ($\geq 45\%$) un año antes de la seroconversión HBeAg tienen una mayor probabilidad de progresar a cirrosis comparado con los de menor porcentaje ($< 10\%$) [110]. Al solaparse con el gen X, las mutaciones en el PBC producen cambios en HBx, que también explicaría que las mutaciones en el PBC promuevan el HCC [110]. La más común es la doble mutación A1762T y G1764A que reduce la expresión de HBeAg hasta en un 70%, aumenta la replicación viral [120], provee de un mecanismo a los genotipos que no pueden desarrollar la mutación G1896A para suprimir la expresión de HBeAg y lleva a dos cambios en HBx, K130M y V131I, mutaciones más frecuentes en el HCC-VHB [47]. La mutación T1753C aumenta la replicación, su prevalencia es más alta en pacientes con HCC o cirrosis y provoca el cambio I127T en HBx, lo cual provoca un cambio de la afinidad de HBx para BCL2 y afecta a la apoptosis inducida por el VHB [47]. Las mutaciones

C1766T y T1768A también reducen la síntesis de HBeAg y aumentan la replicación, además T1768A induce el cambio F132Y en HBx [110].

2.3.2.4.3 VARIANTES EN LA REGIÓN X/Cp/PC

Estudios en Corea con pacientes con HBC y subgenotipo C2 han reportado que el riesgo de HCC aumenta en presencia de ≥ 6 de 8 de estas mutaciones: G1613A, C1653T, T1753C, A1762T, G1764A, A1846T, G1896A y G1899A localizadas a lo largo del Cp y la porción proximal del PC [120], la doble mutación en el PBC podría ser la base para la acumulación de estas mutaciones e influenciarían la expresión de HBeAg y la replicación del ADN-VHB [153]. La mutación C1653T, que provoca el cambio H94Y en HBx, se ha descrito recientemente como un factor predictivo de HCC en Japón [47], localizada en la box α del Cp, puede aumentar la afinidad de unión y la actividad del ENHII/Cp [47]. La mutación G1613A, a pesar de ser sinónima en el aa 80 de HBx, se ha asociado con HCC en pacientes chinos infectados con Gen/B y C [47]. Varios tipos de mutaciones en el motivo tipo BH3 del gen X; T/G1719G/T (cambio en HBx L/V116V/L), T1724C (sinónima en el aa 117 de HBx), T1727C (sinónima en el aa 118 de HBx) y T1741C (cambio en HBx L123S), también se han relacionado con HCC, estas mutaciones podrían interferir en la interacción de HBx con miembros de la familia Bcl-2 (Bcl-2 y Bcl-xL) y afectar a la replicación viral y la muerte celular [47].

2.3.2.4.4 VARIANTES EN EL ORF X

Debido a las múltiples funciones de la proteína HBx, las variantes en el gen podrían tener importantes consecuencias [149]. De hecho, a través de estudios epidemiológicos varias mutaciones se han asociado con la progresión de la enfermedad hepática a cirrosis y HCC [47,154]. Destaca el aumento continuo en la prevalencia de estas mutaciones relacionadas con el HCC conforme se acerca al establecimiento del mismo, mostrando el mayor número de acumulación de mutaciones a los 4,5 años antes del diagnóstico de HCC [155]. Existen algunos *hotspots* de mutación como el codón 127, 131 ó 133 [156].

A. Mutaciones puntuales en el dominio regulador negativo de HBx (aa 1-50)

La mutación A1383T (sinónima en el aa 4) se ha asociado con HCC en una cohorte coreana, en una cohorte de pacientes chinos infectados con Gen/B y C con peor pronóstico después del trasplante hepático y como factor de riesgo independiente de HCC en el Gen/C. Esta mutación puede llevar a una capacidad de unión reducida de miR-15a/16 a HBx, que puede inhibir las vías de señalización celular de Bcl-2, y por lo tanto contribuir a la hepatocarcinogénesis [47]. La mutación G1386A/C (V5M/L), también descrita por primera vez

en una cohorte coreana infectada con el subgenotipo C2, es más frecuente en pacientes con HCC y posiblemente juega un papel clave en la progresión desde cirrosis hepática a HCC. Recientemente, la combinación de la doble mutación en el PBC y V5M o V5L en HBx, ha sido reportada por aumentar el riesgo de HCC [47]. La mutación G1461A/T/C (sinónima en el aa 30), se ha asociado a un riesgo aumentado de HCC en una cohorte china infectada con Gen/B y C [47]. La mutación C1485T (P38S) ha sido descrita como factor de riesgo independiente de HCC en una cohorte japonesa con Gen/C y también se ha descrito en coreanos con infección por subgenotipo C2, como esta mutación está localizada en la región del epítipo para cB, podría llevar a la infección persistente por evasión de la respuesta inmune humoral en el hospedador [47].

B. Mutaciones puntuales en el ORF X fuera del dominio regulador negativo

La mutación T1544A/C (sinónima en el aa 57) se ha asociado a la supervivencia de pacientes con HCC en una cohorte china infectada con Gen/B y C [47]. En 2016 se ha encontrado la doble mutación A10R/S144R en tejidos tumorales y se ha visto que para el ciclo celular, disminuye la apoptosis y se une menos a p53 que HBx intacta [154].

C. Variantes con inserciones /deleciones (indel)

En 1996, Günther *et al.* encontraron en trasplantados renales con HBC inserciones y/o deleciones (indel) en el Cp/ENHII entre los nt 1760-1780 y la persistencia de estas indel estaba asociada a enfermedad hepática severa [157]. En 1997 se detectaron deleciones de 8 pb (nt 1763-1770) entre la TA-2 y TA-3 y de 20 pb (nt 1753-1772) entre la TA-1 y TA-2 en el Cp en pacientes con HBC [158]. Desde entonces se han encontrado diferentes variantes por indel que se resumen en la tabla 17.

Tabla 17: Variantes por inserción (ins)/delección(del) descritas en la región X. DUP: duplicación.

Variante indel	Cambio aa HBx	Referencia
Del: 1396-1403 (8 nt)	135 aa (Hbx truncado)	[159]
Del: 1575-1627 Dup: 1660-1666 Ins: 1775 15nt		[160]
Del: 1633-1664 + 1725-1749 Dup: 1720-1766 Ins: 1765 TA Doble mutación PBC		[157]
Del: 1725-1749 Dup: 1720-1766 Ins: 1765 TA Doble mutación PBC		[157]
Del: 1746 + 1756-1763		[161]
Del: 1746 + 1763-1770		[157]
Del: 1746 Ins: 1765 TA		[161]
Del: 1747-1757		[161]
Del: 1747-1767		[161]
Del: 1748-1768 Ins: 1674-1675 (DUP 1651-1674)		[161]
Del: 1748-1768 Ins: 1678-1679 (DUP 1649-1678)		[161]

Tabla 17: Continuación.

Variante indel	Cambio aa HBx	Referencia
Del: 1750-1770		[161]
Del: 1750-1770 Ins: 1832 CT		[161]
Del: 1750-1780		[161]
Del: 1751-1770		[161]
Del: 1752-1770		[161]
Del: 1752-1770 Ins: 1668-1669 (DUP 1642-1668)		[161]
Del: 1752-1770 Ins: 1668-1669 (DUP 1642-1668) Ins: 1827-1828 (DUP 1822-1827)		[161]
Del: 1753+1756 Ins: 1765 TA Doble mutación PBC		[157]
Del: 1754-1755 + 1763-1770		[161]
Del: 1754-1755 + 1763-1770 Ins: 1677-1678 (DUP 1647-1677)		[161]
Del: 1754-1755 + 1763-1770 Ins: 1678-1679 (DUP 1649-1678) Ins: 1820 CT		[161]
Del: 1754-1763		[161]
Del: 1755-1773	127 aa (HBx truncado)	[159]
Del: 1755-1774	127 aa (HBx truncado)	[159]
Del: 1756-1763		[161]
Del: 1756-1772		[161]
Del: 1756-1773		[161]
Del: 1756-1773 Ins: 1669-1670 (DUP 1647-1669)		[161]
Del: 1756-1773 Ins: 1673-1674 (DUP 1647-1673)		[161]
Del: 1756-1773 Ins: 1678-1679 (DUP 1650-1678)		[161]
Del: 1756-1773 Ins: 1676-1677 (DUP 1649-1676)		[161]
Del: 1757-1766		[161]
Del: 1758-1766		[161]
Del: 1758-1777		[161]
Del: 1758-1777 Ins: 1659A		[161]
Del: 1758-1777 Ins: 1832-1833 (DUP 1828-1833)		[161]
Del: 1760-1772		[161]
Del: 1762-1768		[161]
Del: 1763-1770		[157]
Del: 1764-1770		[161]
Del: 1765-1772		[157]
Del: 1768-1775		[161]
Del: 1768-1775 Ins: 1825 TA		[161]
Del: 1770-1776		[157]
Del: 1770-1778	137 aa (HBx truncado)	[159]
Del: 1825		[161]
Del: 1825 Ins: 1653-1654 (DUP 1642-1653)		[161]
Del: 1828-1841	158 aa (HBx más largo)	[159]
Dup: 1641-1668 Del: 1763-1770		[47]
Dup: 1641-1668 Del: 1770-1776		[47]
Dup: 1644-1674 Del: 1763-1770		[157]
Dup: 1648-1663 Del: 1753 + 1763-1770		[157]
Dup: 1648-1663 Del: 1763-1770		[157]
Dup: 1660-1666 Ins 1775 15nt		[157]
Dup: 1723-1789 Del: 1763-1770		[157]
Ins: 1577 AGGCC Mutación G1633A + G/C/T1637A	154 aa	[159]
Ins: 1608-1613	Ins: aa 79-80 GM	[159]
Ins: 1650-1661	Ins: aa 93-94 PKLL	[159]
Ins: 1653-1654 (DUP 1642-1653)		[161]
Ins: 1653-1654 (DUP 1642-1653) Ins: 1768 GTT		[161]
Ins: 1668-1669 (DUP 1642-1668)		[161]

Tabla 17: Continuación

Variante indel	Cambio aa HBx	Referencia
Ins: 1674-1675 (DUP 1651-1674)		[161]
Ins: 1765 TA		[161]
Ins: 1765 TA Doble mutación PBC		[157]
Ins: 1768 GTT		[157]
Ins: 1768 GTT Ins: 1827-1828 (DUP 1819-1827)		[161]
Ins: 1768-1769 (DUP 1642-1668)		[161]
Ins: 1788 TA		[161]
Ins: 1824-1825 (2nt)	176 aa (HBx más largo)	[159]
Ins: 1825 TA		[161]
Ins: 1827-1828 (DUP 1819-1827)		[161]
Ins: 1837 A		[161]
Ins: 1924-1832	Ins: aa 151-152 FFN	[159]
Ins:1827-1828 (DUP 1821-1827)		[161]

2.3.3 ESTRUCTURA EN QUASIESPECIES DE LAS POBLACIONES DEL VHB

2.3.3.1 DEFINICIÓN DE CUASIESPECIE (QS)

Se define cuasiespecie (QS) como un conjunto de genomas virales no idénticos pero estrechamente relacionados entre sí que co-infectan un mismo paciente y que están sometidos a un continuo proceso de variación genética y selección por procesos de competencia y colaboración entre ellas [162]. El término surgió para referirse a virus ARN que genéticamente evolucionan como una distribución compleja de variantes [127].

2.3.3.2 ORIGEN DE LA CUASIESPECIE VIRAL

Los factores que hacen que el VHB circule como una QS son [149]:

1. La replicación incluye un paso de transcripción reversa, similar a la de retrovirus, llevado a cabo por la ADN polimerasa viral que carece de actividad correctora de errores exonucleasa 3'-5'. Los genomas son muy proclives a acumular errores y tienen una alta tasa de mutación ($3,2 \times 10^{-5} - 7,9 \times 10^{-5}$ nt sustituidos/ciclo replicativo), cien veces superior a otros virus ADN.
2. La alta tasa de replicación viral (10^{12} partículas virales/día), sumado al pequeño tamaño del genoma (3,2 kb), hace que se produzcan unas 10^{10-11} mutaciones puntuales al día en individuos con replicación activa.
3. Factores del hospedador que están asociados con una hipermutación del genoma. La actividad de la familia de citidina desaminasas APOBEC3, que desaminan citidinas a uridinas en la hebra(-), resulta en una hipermutación G por A en la hebra(+).

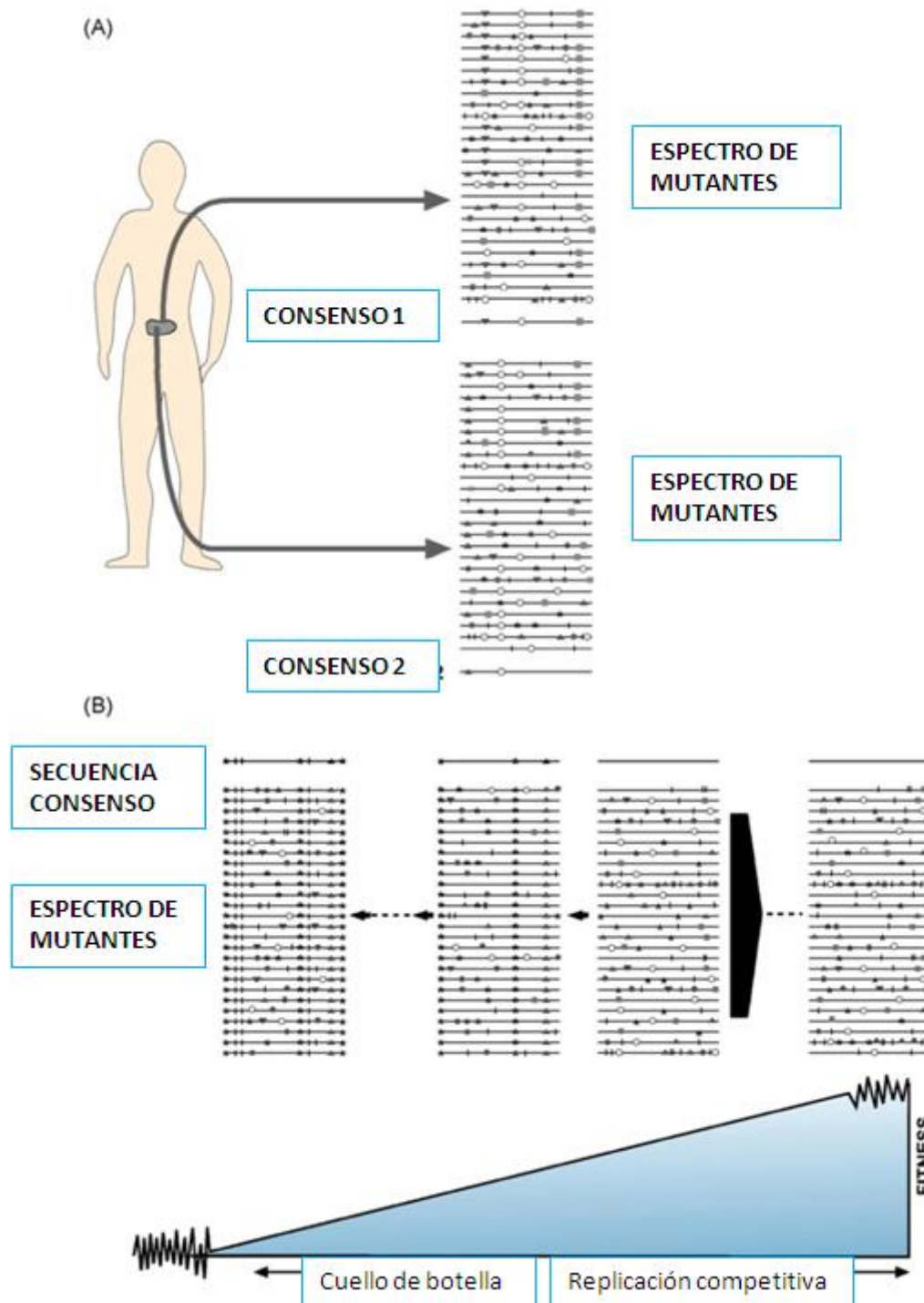
4. Las nuevas cápsides contienen genomas diferentes al que infectó el núcleo y el proceso de reinfección genera una especie de QS intracelular, además de la circulante, aumentando la complejidad del VHB.
5. Una misma célula hepática puede ser infectada por múltiples viriones simultáneamente o secuencialmente durante su vida.
6. La presencia de regiones homopoliméricas podría estar asociado con un aumento del riesgo de error de la polimerasa por *sliding*, que resultarían en inserciones y/o deleciones [163].
7. El alto grado de solapamiento, por el que la sustitución en un ORF puede inducir cambios en otro. No obstante, a diferencia del VIH o el VHC, la variabilidad de la QS del VHB está limitada precisamente por el alto grado de solapamiento de los ORFs.

2.3.3.3 IMPORTANCIA DEL ESTUDIO DE LA QS

La naturaleza de QS del VHB le permite adaptarse rápidamente a los cambios del ambiente replicativo, la QS sería una unidad de selección, un espectro de mutantes que poseen diferentes niveles de *fitness** en determinados ambientes, aquellas variantes que suponen una ventaja ante una presión selectiva, como el tratamiento antiviral o el sistema inmune del hospedador, son seleccionadas durante la replicación (Figura 26) [149]. Por lo que muchas de las propiedades importantes del virus no se pueden explicar por una secuencia consenso y requieren el conocimiento de las variantes presentes en la población viral [164]. El estudio de la dinámica de QS es importante para entender la adaptabilidad, el poder patogénico y la persistencia de los virus y así diseñar estrategias para prevenir y tratar la enfermedad [149].

**Fitness* viral: capacidad de un virus para replicarse en un determinado ambiente [127].

Figura 26: Representación esquemática de la QS y variaciones en el *fitness*. La evolución de la QS está guiada por el gradiente de *fitness* del ambiente, es decir, los mutantes con mayor *fitness* pueden dominar por replicación competitiva [165,166]. A: De manera simplificada se describe una QS con dos espectros de mutantes con su secuencia consenso en un individuo infectado. Cada línea horizontal representa un genoma y cada símbolo un tipo de mutación. Una QS real puede contener miles de genomas continuamente mutando para formar un conjunto enorme. B: La QS actúa como una unidad de selección, las flechas pequeñas indican pasos repetidos de cuello de botella que resultan en un descenso del *fitness*. La flecha grande representa pasos a poblaciones más grandes que generalmente resultan en una ganancia de *fitness*. Las fluctuaciones en el *fitness* mostrados a valores altos y bajos en los ángulos del triángulo se han visto experimentalmente y pueden ser interpretados como efectos estocásticos de las mutaciones en el *fitness* cuando la ganancia de *fitness* está limitada por el tamaño de la población o cuando el *fitness* es muy bajo.



2.3.4 COMPLEJIDAD DE QS

2.3.4.1 DEFINICIÓN Y SIGNIFICADO CLÍNICO

Se define complejidad de una QS como una propiedad que cuantifica la diversidad y la frecuencia de haplotipos (conjunto de genomas con la misma secuencia nucleotídica), independientemente del tamaño de la población viral. La complejidad y diversidad del conjunto de genomas que constituye la QS (espectro de mutantes) puede influir en el potencial patogénico, la evolución clínica, la seroconversión a anti-HBe o la respuesta al tratamiento antiviral [165,167–170], con lo cual es importante cuantificarla mediante los índices de diversidad (Ver apartado 2.3.5 *Índices para el estudio de la diversidad de cuasiespecie*). De manera ideal, la complejidad de la QS debería basarse en todo el repertorio de mutantes de la población pero esto no es posible con la metodología actual y se extrapolan los resultados del análisis de una muestra limitada [171].

2.3.4.2 EVOLUCIÓN DE LA COMPLEJIDAD DE QS EN LA HISTORIA NATURAL DE LA INFECCIÓN

La complejidad y diversidad de la QS del VHB varían a lo largo de la historia natural de la enfermedad hepática crónica o aguda por diferentes niveles de presión selectiva del sistema inmune. Se ha visto que la diversidad y complejidad de la QS en pacientes crónicos es mayor que en aquellos con infección aguda y mayor en las regiones LHB, MHB, HBcAg, P y RT que en X y Cp [172]. En pacientes que seroconvierten tienen significativamente mayor diversidad de la QS en la región preCore/Core y la diversidad viral es más extrema justo antes de la seroconversión HBcAg [170]. Se ha observado que el aumento gradual de la diversidad de la QS está asociado con un aumento en las mutaciones de stop del PC, seguido de una reducción de los niveles de ADN y HBcAg, así como la selección positiva de variantes en los aa 13 y 135 de la proteína core [169].

2.3.4.3 EVOLUCIÓN DE LA COMPLEJIDAD DE QS DURANTE EL TRATAMIENTO

En general, la complejidad de la QS y diversidad dentro del dominio RT en respondedores al tratamiento antiviral es más baja que en no respondedores durante las fases tempranas del tratamiento. Además los cambios dinámicos dentro de RT en la QS en los estadios tempranos del tratamiento, muestran diferentes patrones de evolución de la QS entre respondedores y no respondedores y se podrían usar para predecir la eficacia del tratamiento y la aparición de resistencias a largo plazo [172]. Por ejemplo, en pacientes en tratamiento con LAM, la complejidad y diversidad de QS es significativamente menor en respondedores que en no respondedores en la semana 4, independiente del genotipo y estatus HBcAg [168], en pacientes tratados con ETV los respondedores muestran un descenso de la complejidad de QS

durante las cuatro primeras semanas del tratamiento estadísticamente significativo comparado con los respondedores parciales y los patrones de evolución de la QS del VHB son un buen parámetro para la predicción de la respuesta a largo plazo del tratamiento [165].

2.3.5 ÍNDICES PARA EL ESTUDIO DE LA DIVERSIDAD DE CUASIESPECIE

La complejidad del conjunto de genomas que constituye la QS se puede calcular con los índices de diversidad usados en Ecología [167,171]. Estos índices de diversidad se clasifican en tres grupos: Incidencia, abundancia y función.

2.3.5.1 INCIDENCIA (total):

Los índices basados en incidencia son aquellos que corresponden al total de entidades en el alineamiento múltiple de los haplotipos, son índices de riqueza, número de especies en una comunidad [171]:

Número de haplotipos: número de secuencias diferentes obtenidas, cada una de las cuales representa una variante genética [171].

Número de sitios polimórficos o de mutaciones: número de posiciones en un conjunto de secuencias alineadas donde se puede observar al menos una mutación [171].

2.3.5.2. ABUNDANCIA (frecuencias):

Los índices basados en abundancia consideran tanto las entidades observadas como su frecuencia en la población. Miden diversidad o uniformidad de la distribución de haplotipos (nº y frecuencia de diferentes haplotipos) y varían entre 0 y un máximo de 1, cuando todos los haplotipos observados tienen idéntica abundancia [171].

Entropía de Shannon (H_s): originalmente desarrollada en el campo de la teoría de la información, en virología se usa como medida global de la complejidad de QS, es una función del número de haplotipos en la población viral y sus frecuencias relativas [171].

$$H_s(p) = - \sum_{i=1}^H p_i \log(p_i)$$

p_i; frecuencia de cada haplotipo en la QS

Entropía de Shannon normalizada (H_{SH} o S_n): la entropía de Shannon mide la diversidad o uniformidad de una QS cuando es normalizada a la diversidad máxima. La H_{SH} permite medir la diversidad de la QS teniendo en cuenta la cantidad de haplotipos o variantes que se han encontrado y la frecuencia relativa entre ellos. Varía entre 0 y 1, valor mínimo cuando todas las variantes son idénticas y máximo cuando todas son diferentes, pero no tiene en cuenta la heterogeneidad de las diferentes variantes de la QS porque es insensible al número de mutaciones [171].

$$H_{SH}(p) = -\sum_{i=1}^H p_i \log(p_i) / \log(H)$$

H; número de haplotipos, p_i ; frecuencia de cada haplotipo en la QS

2.3.5.3. FUNCIÓN (diferencias entre haplotipos):

Los índices funcionales son aquellos basados en las diferencias entre haplotipos dentro de la QS y pueden incluir o no la frecuencia de cada uno de ellos en la población. Miden la heterogeneidad intra-población, como de diferente son los miembros de una población. Se pueden subdividir entre índices funcionales basados en incidencia o abundancia [171].

INCIDENCIA (total de diferencias): frecuencia de mutación (Mf)

ABUNDANCIA (frecuencia de diferencias): diversidad nucleotídica (P_i o π)

Frecuencia de mutación (Mf): heterogeneidad genética respecto a la variante dominante (máster) o una secuencia consenso de la QS. Proporción de sitios mutados en una población de genomas virales con respecto al haplotipo dominante (máster) o la secuencia consenso, el valor es normalizado por el total de nt de la secuencia. Cuanto mayor es el valor más diferentes son los individuos de la población con respecto a la de referencia [171].

$$Mf = \frac{1}{lN} \sum_{i=1}^h n_i m_{1i} = \frac{1}{l} \sum_{i=1}^h \hat{p}_i m_{1i} = \frac{1}{l} \langle \hat{P} | M_1 \rangle$$

l; longitud amplicón, N; tamaño muestra (nº de secuencias), n_i ; número de haplotipos, m_{1i} ; nº de sustituciones entre los haplotipos y la secuencia máster, \hat{p}_i ; Maximum likelihood estimator (MLE) de p_i , h; nº de haplotipos, M_1 ; nº de mutaciones

Diversidad nucleotídica (P_i o π): heterogeneidad genética global de la población teniendo en cuenta la media del número de mutaciones entre cada pareja de variantes de la QS. Mide el promedio de nucleótidos diferentes entre dos genomas de la QS, da más información que la Mf porque tiene en cuenta las diferencias entre dos genomas cualesquiera en la población [171].

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

d_{ij} ; distancia genética entre haplotipos i y j , p_i ; frecuencia de cada haplotipo i en la QS, p_j ; frecuencia de cada haplotipo j en la QS

2.3.6 MÉTODOS PARA EL ESTUDIO DE LAS VARIANTES GENÉTICAS DEL VHB

2.3.6.1 MÉTODOS PARA EL GENOTIPAJE

2.3.6.1.1 MÉTODOS MOLECULARES PARA EL GENOTIPAJE

A. Secuenciación y análisis filogenético del genoma completo

El análisis filogenético del genoma completo es el método más concluyente para el genotipaje y se considera el *gold standard*, aunque no puede detectar infecciones mixtas y es una técnica relativamente cara y laboriosa.

B. Secuenciación y análisis filogenético del gen preS/S

El análisis filogenético de la región preS1/preS2 (nt 3025-80 desde el sitio EcoRI) muestra una capacidad para discriminar entre distintos genotipos similar a la del genoma completo [135].

C. INNO-LiPA [Fujirebio (Innogenetics), Ghent, Bélgica]

Se basa en la hibridación reversa, fragmentos amplificados biotinilados hibridan en unas sondas específicas de oligonucleótidos inmovilizadas en líneas paralelas en tiras de nitrocelulosa (Figura 27) [174]. En términos de coste-beneficio es la más ventajosa por su bajo coste, alta especificidad y sencillez y, además, es capaz de detectar mezclas de genotipos, la principal desventaja es la baja sensibilidad si existen mutaciones en la secuencia analizada [150].

D. PCR-RFLP (*Polymerase chain*

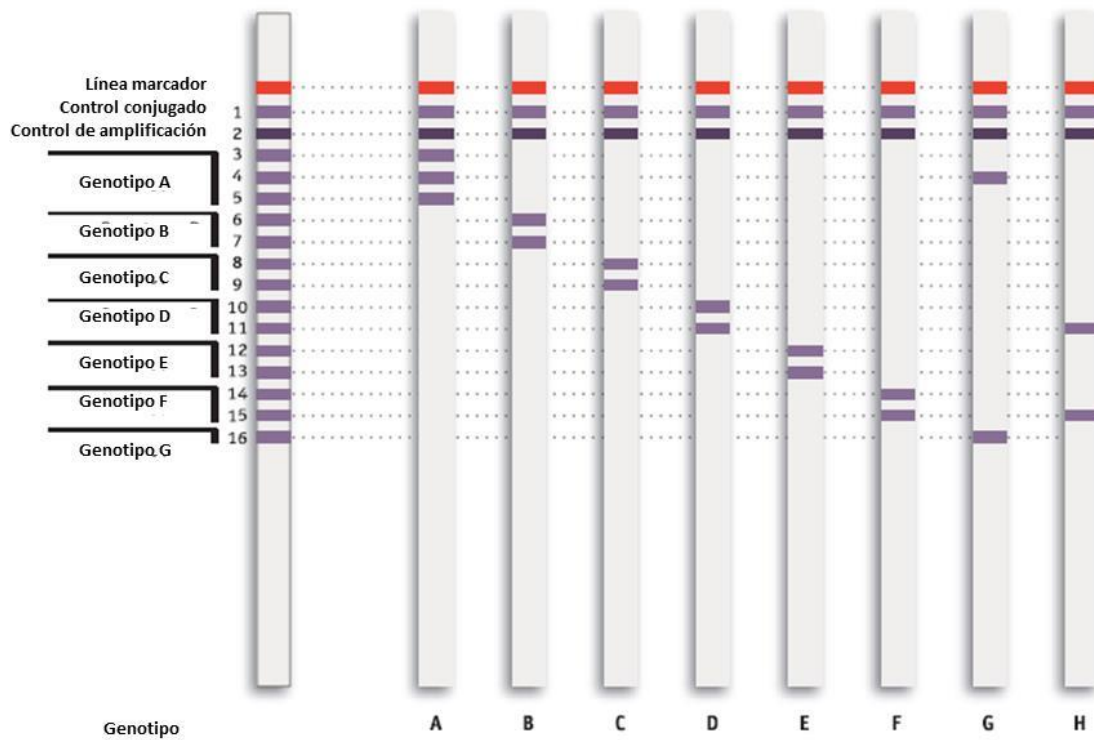
reaction-restriction fragment length polymorphism)

En el análisis de fragmentos de restricción de longitud polimórfica, la muestra se digiere con un conjunto de endonucleasa de restricción y se analiza el patrón de migración de los fragmentos resultantes en un gel de electroforesis, específico para cada genotipo, ya que debido a las diferencias entre ellos cada uno tiene diferentes dianas, pero se pueden generar patrones atípicos que requiere re-análisis por secuenciación y filogénesis [175].

E. OTRAS TÉCNICAS

Las técnicas *Oligonucleotide Microarray chips*, primers específicos de genotipo en Multiplex-PCR, *PCR-invader assay*, PCR a tiempo real, Dot blot reverso, FT-RDB (*Flow reverse dot blot*) y FRMP (*Restriction fragment mass polymorphism*), son precisas, baratas y simples, pero no muy sensibles ya que cualquier mutación puede afectar al resultado [35].

Figura 27: Representación gráfica de la tira del ensayo de genotipaje INNO-LiPA, las bandas representan sondas de oligonucleótidos específicas para controles y cada uno de los genotipos. Las sondas están diseñadas en base a la secuencia específica de los motivos de cada genotipo en la región amplificada, que abarca la principal región hidrofílica del HBsAg entre las posiciones nt 456-798, tres motivos específicos para el Gen/A, dos para Gen/B-F y uno para el Gen/G. Las líneas se comparan con una tarjeta y se interpretan según las indicaciones de la casa comercial, no hay sonda específica para el Gen/H, pero una combinación de líneas 11 y 15 se considera específica de este genotipo [174].



2.3.6.1.2 MÉTODOS ON LINE PARA GENOTIPAR

Existen varias herramientas de genotipaje del VHB basadas en la web NCBI, además de otras desarrolladas por expertos como: *Hepatitis B virus database* (HBVdb), *HBV STAR*, *BioAfrica-Oxford HBV Automated Subtyping Tool*, *HepSEQ Genotyper* y *Jumping profile Hidden Markov Model* (jpHMM) [35].

2.3.6.2 HERRAMIENTAS PARA EL ESTUDIO DE LA QS

El estudio de las variantes que forman parte de la QS tiene un gran interés clínico y virológico. A lo largo de los años se han desarrollado diferentes técnicas.

2.3.6.2.1 TÉCNICAS MOLECULARES INDIRECTAS

Al igual que para el genotipaje, existen ensayos de hibridación reversa INNO-LiPA para detectar mutaciones en el dominio RT relacionadas con resistencias al tratamiento presentes en pequeñas proporciones (5-10%) [176], también para mutaciones en el PBC y en el PC relacionadas con la transición de HBeAg(+) a HBeAg(-) [177]. La técnica PCR-RFLP sirve para detectar mutaciones previamente caracterizadas [175] y otras técnicas menos frecuentes son *5'-nuclease assays* [178], *melting point analysis* [179], *mass spectrometry* [180], *DNA chip technology* [181] y PCR a tiempo real con primers específicos de mutación [182]. Aunque estos métodos pueden detectar mezclas, no permiten la detección simultánea de sustituciones en el mismo genoma, con lo que no se pueden estudiar variantes complejas con dos o más cambios y por lo tanto no sirven para describir correctamente la composición de la QS [149].

2.3.6.2.2 TÉCNICAS DE SECUENCIACIÓN

A. Secuenciación Sanger

Los primeros métodos para secuenciar el ADN se describieron en los años 70, los más importantes fueron la secuenciación química de Maxam y Gilbert y los métodos de terminación de cadena de Sanger. Ambos basados en el marcaje isotópico de oligonucleótidos y electroforesis en gel de cuatro pools de reacciones para reconstruir la secuencia de ADN. El primer punto de inflexión en la secuenciación del ADN se produjo en 1980 cuando Leroy substituyó el marcaje radioactivo por fluorescencia y automatizó el método de Sanger, esta mejora supuso la base para el desarrollo y comercialización de secuenciadores de electroforesis capilar [183]. A pesar de su abrumador éxito, las técnicas de primera generación o secuenciación directa no tienen un alto rendimiento, tienen limitada adaptabilidad y no son coste-efectivas para la secuenciación de múltiples muestras y/o largos genomas [184]. Presentan dos limitaciones principales para el análisis de mezclas complejas: la baja sensibilidad para detectar variantes minoritarias (>20%) y la imposibilidad del análisis de haplotipos [149].

B. Clonaje y secuenciación

Es el método tradicional para estudiar una población genética heterogénea, en ella se clonan los amplicones obtenidos por PCR en plásmidos con los que se transforman bacterias

competentes y se analizan cada una de las colonias por secuenciación directa [185]. Es una técnica tediosa, que requiere muchos y complejos pasos experimentales y muchos clones, más de 100, para el análisis de la QS [149].

C. Secuenciación masiva o *next generation sequencing* (NGS)

La secuenciación directa sin clonaje previo ha experimentado una gran revolución gracias a las técnicas de NGS, en las que se generan un número muy elevado de *reads* (lecturas de la secuencia de una molécula individual de ADN), desde cientos/miles hasta millones [184]. No obstante presentan una serie de limitaciones; (1) la principal es la longitud de los fragmentos, menor que en la secuenciación Sanger, (2) las tecnologías de NGS son proclives al error aleatorio, inserciones, deleciones y/o mutaciones puntuales ocurren durante la amplificación por PCR o la secuenciación en sí misma, aunque esta limitación se puede compensar secuenciando moléculas diferentes de la misma región varias veces, es decir, aumentando la cobertura [186] y (3) la PCR previa puede sesgar la frecuencia relativa de las variantes. Por lo tanto se debe llevar a cabo un buen cribaje de los datos, eliminar las secuencias de baja calidad y corregir errores para separar las variantes verdaderas de aquellas debidas a artefactos experimentales [164].

SECUENCIACIÓN DE SEGUNDA GENERACIÓN

En 2004 se introdujeron las técnicas de segunda generación en la que el material de partida es fragmentado, amplificado clonalmente y cargado en microchips o pocillos microscópicos para la secuenciación masiva en paralelo [183], usan una técnica *wash-and-scan*, los datos de la secuencia se van adquiriendo como un conjunto de imágenes fluorescentes o luminiscentes capturadas en la superficie de una celda de flujo en cada ciclo repetitivo de secuenciación [187]. En 2004, 454 Life Sciences (Branford, Connecticut) lanzó el primer instrumento de pirosecuenciación, que después fue adquirido por Roche (Basel, Suiza); 454™ (454 Life Sciences/Roche, Branford, Connecticut) [184]. En 2005 apareció la secuenciación por síntesis; Genome Analyzer (GA, Solexa, Chesterford, Reino Unido) ahora llamado Illumina® (Illumina, Inc. San Diego, California) [184]. En 2007 el sistema de secuenciación SOLiD fue lanzado por Applied Biosystems (Foster City, California) [184]. En 2013 Complete Genomics fue adquirida por el BGI (*Beijing Genomics Institute*) para ofrecer un servicio de secuenciación por la tecnología *Combinatorial Probe-Anchor Ligation*™ (cPAL™). En 2015 se lanzó GeneReader, que en una única plataforma incorpora un sistema de preparación de la muestra (QUIcube) y la plataforma Qiagen Clinical Insight para el análisis de variantes usando la misma aproximación que Illumina [188].

SECUENCIACIÓN DE TERCERA GENERACIÓN o NNGS (*Next-next Generation Sequencing*)

En la tercera generación de técnicas la señal se registra en tiempo real directamente durante la reacción enzimática y no es necesaria la amplificación clonal previa. En 2011 el secuenciador SMRT (*single molecule real-time*) fue introducido por Pacific Biosciences (Menlo Park, California) [184]. En 2015 Oxford Nanopore Technologies (Oxford, Reino Unido) lanzó un secuenciador de molécula única larga [184].

SECUENCIACIÓN ENTRE SEGUNDA Y TERCERA GENERACIÓN (Sólo cumplen un criterio de los de tercera generación)

En 2011 Life Technologies (Carlsbad, California) comercializó la plataforma Ion Torrent™ (Ion Torrent/Life Technologies, South San Francisco, California), no utiliza una técnica *wash-and-scan* pero requiere amplificación previa [184].

SECUENCIACIÓN DE CUARTA GENERACIÓN

Ya se ha descrito la secuenciación de cuarta generación, con métodos que utilizan el fundamento de los de segunda generación para secuenciar ácidos nucleicos directamente en células fijadas o tejidos *in situ*; ISS (*In situ sequencing*) [183].

En líneas generales todas las plataformas constan de tres pasos en su procedimiento:

1. Preparación de la librería: Se pueden usar tres tipos de librerías [186].

- Librería de amplicones: se lleva a cabo una PCR con ADN purificado amplificando un fragmento concreto.

- Librería *shotgun*: se lleva a cabo una fragmentación aleatoria de ADN, de ahí el nombre, después se añade un dATP al extremo 3' de cada fragmento, seguido de la unión de adaptadores específicos de cada plataforma necesarios para la secuenciación.

- Librería *mate-pair*: se pueden usar secuencias que están físicamente unidas entre ellas pero a cierta distancia en el mismo fragmento de ADN, para ello, tras una fragmentación inicial, se añaden adaptadores que permiten la circularización de los fragmentos de ADN, estas moléculas circulares son re-fragmentadas como en el *shotgun*.

2. Captura del ADN y enriquecimiento: a las moléculas cortas se les añaden adaptadores para la captura y separación física de cada fragmento único en un soporte sólido. Cada molécula única actúa como molde para la amplificación clonal. En la técnicas de tercera generación no se requiere este paso de amplificación previa.

3. Secuenciación/Captura de imagen: La secuenciación se basa en la polimerización del ADN combinada con la detección de los nucleótidos incorporados, que ocurre a la vez en una multitud de fragmentos.

4. Procesamiento de los datos: Existen softwares desarrollados para las diferentes aplicaciones de NGS.

2.3.6.3 DESCRIPCIÓN DE LAS PLATAFORMAS DE NGS

A continuación se describen las aproximaciones para la secuenciación de fragmentos cortos y después para largos, haciendo especial énfasis en la tecnología 454™ usada en esta tesis.

A. NGS DE FRAGMENTOS CORTOS

Las aproximaciones para la secuenciación de fragmentos cortos se pueden dividir en: SECUENCIACIÓN POR LIGACIÓN (*sequencing by ligation*, SBL)

Una sonda marcada con un fluoróforo hibrida con el fragmento de ADN a secuenciar (template), a su vez el template es ligado a un oligonucleótido adyacente para la captura de la imagen. El espectro de emisión del fluoróforo indica la identidad de base o bases complementarias a la posición específica en la sonda [188].

SECUENCIACIÓN POR SÍNTESIS (*sequencing by synthesis*, SBS)

Se usa una polimerasa y la señal, fluorescencia o cambio en la concentración iónica, identifica la base incorporada. A su vez se diferencian dos aproximaciones [188]:

- Terminadores cíclicos reversible (*cyclic reversible termination*, CRT)
- Adición de nucleótidos individuales (*single-nucleotide addition*, SNA)

A.1 AMPLIFICACIÓN CLONAL

En ambas aproximaciones el ADN es amplificado clonalmente en una superficie sólida, el anclaje se realiza mediante un adaptador en el extremo del template (Figura 28) [188].

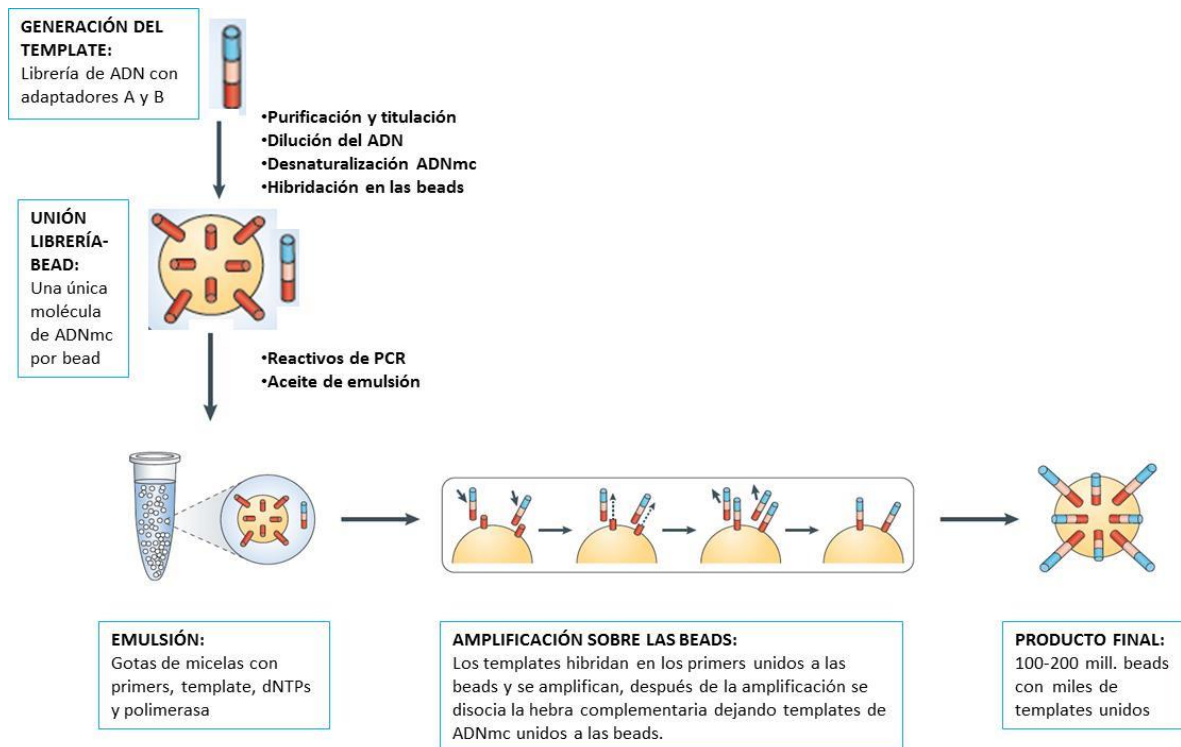
Figura 28: Los adaptadores usados en la tecnología 454 Roche son primers de 44 nt añadidos a cada extremo de los template, las primeras 20 pb en el extremo 5' son los primers para la PCR (amarillo), las siguientes 20 pb son el primer de secuenciación (azul) y las 4 pb en el extremo 3' **AGTC**, son una secuencia no palindrómica "llave" que contiene los cuatro tipos de nt, muy importante para el procesado final de los datos, para calibrar la señal. Después está la secuencia MID (*Multiplex Identifier*), un identificador único de 10 nt establecida por el fabricante y que se utiliza durante el análisis bioinformático para identificar la muestra a la que pertenece cada *read*. Opcionalmente, permite mezclar en el mismo experimento los amplicones de diferentes muestras. En el extremo 3' está la secuencia específica del template. Hay dos clases de adaptadores A y B, que se diferencian en la secuencia de nt y la presencia de una biotina 5' en el adaptador B [189].

A: 5'-	CCATCTCATCCCTGCGTGTCCATCTGTTCCCTCCCTGTC	TCAG-MID-	Secuencia específica template-3'
B: 5'-	/5BioTEG/CCTATCCCTGTGTGCCTTG	CCTATCCCTGTTGCGTGTCTCAG-MID-	Secuencia específica template-3'

A.1.1 emPCR (454™) (Figura 29)

El sistema de amplificación clonal de la tecnología de *Ultra-deep pyrosequencing* (UDPS), utilizada en los estudios de esta tesis, se basa en la PCR en emulsión (emPCR). Mediante esta técnica las moléculas de ADN de la librería se desnaturalizan para obtener el template monocatenario (ADNmc), esta molécula se hibrida con esferas (beads) cubiertas con la secuencia complementaria al adaptador de manera que una sola molécula de ADN se une a cada bead [189]. Para ello previamente la librería de ADN es purificada, cuantificada y comprobada su calidad, se diluye para garantizar el ratio: una molécula de ADN por cada bead [189]. Una vez que la librería está inmovilizada en las beads, es emulsionada con los reactivos necesarios para la amplificación (dNTPs, primers y la ADNpol) en gotas acuosas embebidas en aceite (micelas), de manera que cada gota contiene una única bead con una única molécula de ADN y actúa como un microreactor para la PCR [189]. Con la emulsificación se consiguen aproximadamente 1000 microreactores de PCR por microlitro y gotas de diámetro 100-150 µm [189], dentro de las cuales se produce la amplificación por PCR para crear beads recubiertas con miles de copias de la misma secuencia [190]. Después de la amplificación se elimina la hebra complementaria para dejar el template en forma monocatenaria y se lleva a cabo el enriquecimiento de las beads [190]. Millones de estas beads se inmovilizan en pocillos de una placa PTP (PicoTiterPlate) (Roche/454) donde tiene lugar la secuenciación [188,190].

Figura 29: emPCR.



A.2 SECUENCIACIÓN POR LIGACIÓN (SBL) (SOLiD, COMPLETE GENOMICS)

La secuenciación por ligación se basa en la hibridación y ligación de dos tipos de sondas al template: (1) sonda marcada y (2) sonda de anclaje. La sonda marcada tienen una o dos bases conocidas (sonda de una base/sonda de dos bases) y bases degeneradas o universales y dirigen la unión complementaria entre sonda y template. La sonda de anclaje tiene una secuencia complementaria al adaptador y aportan el sitio para iniciar la ligación, después de la cual se captura la imagen del template y la(s) base(s) conocidas de la sonda son identificadas. Un nuevo ciclo comienza después de generar un sitio de ligación por eliminación completa del complejo sonda de anclaje-sonda marcada o por eliminación del fluoróforo de la sonda marcada.

A.3 SECUENCIACIÓN POR SÍNTESIS (SBS)

A.3.1 TERMINADORES REVERSIBLES CÍCLICOS (CRT, *cyclic reversible termination*) (Illumina, Gene Reader)

Usan moléculas terminadoras similares a las usadas en la secuenciación de Sanger, en las que el extremo 3'-OH está bloqueado evitando la extensión. Al template se une un primer complementario al adaptador, que inicia la unión de la polimerasa. En cada ciclo se añade una mezcla de cada dNTP marcado y con el extremo 3' bloqueado, después de que se añade el

complementario en la hebra en extensión, los dNTPs que no están unidos se lavan y se lleva a cabo la captura de la imagen. El fluoróforo y el grupo bloqueante pueden ser eliminados para que comience un nuevo ciclo.

A.3.2 Adición de nucleótidos únicos (SNA, *single-nucleotide addition*) (454™, Ion Torrent)

A diferencia de la anterior se basa en una única señal para marcar la incorporación de dNTPs, por tanto cada uno de los cuatro nt debe ser añadido alternativamente en cada reacción de secuenciación. No requiere que los dNTPs estén bloqueados, ya que la ausencia del siguiente nt complementario bloquea la reacción de secuenciación siguiente y evita la elongación, con excepción de las regiones homopoliméricas donde se añaden dNTPs idénticos, en este caso la identificación de la secuencia se basa en el aumento de señal proporcional a los dNTPs incorporados.

A.3.2.1 Pirosecuenciación 454™ o UDPS (*Ultra-deep pyrosequencing*)

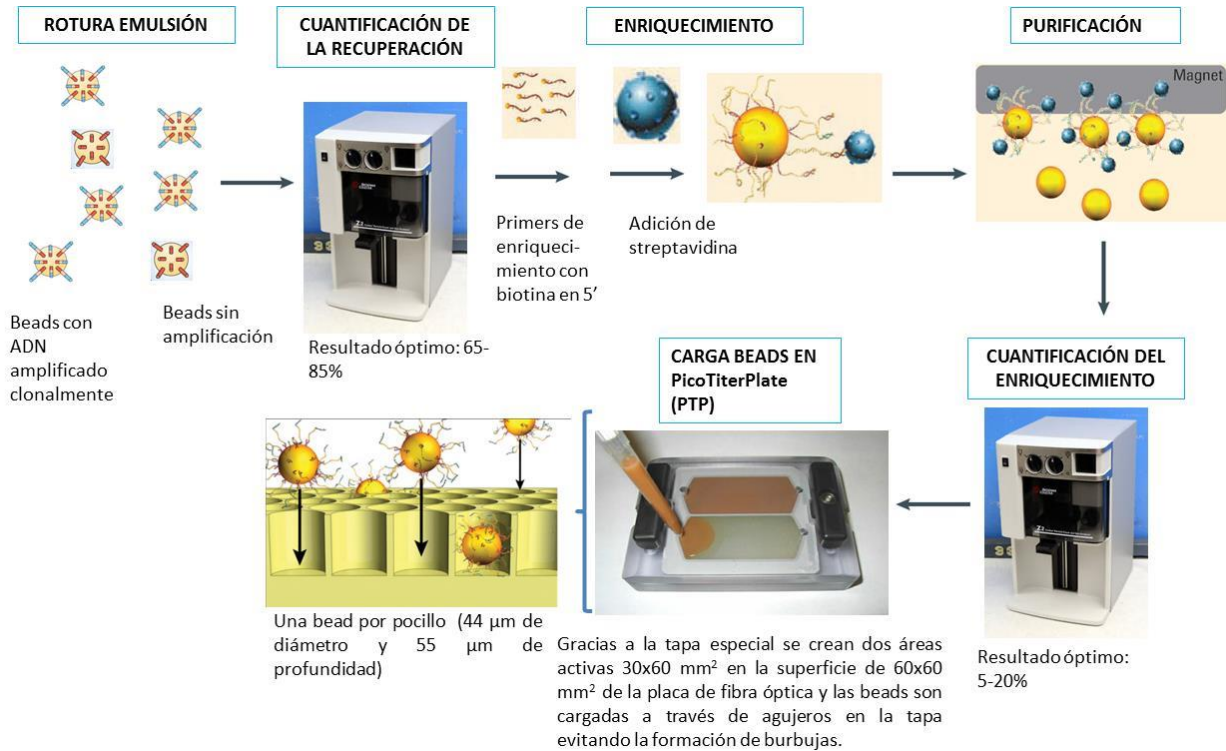
La pirosecuenciación es un método no electroforético bioluminiscente que mide la liberación de pirofosfato inorgánico (PPi) que es convertido proporcionalmente en luz visible a través de una serie de reacciones enzimáticas [190].

A.3.2.1.1 PASOS PREVIOS A LA REACCIÓN DE SECUENCIACIÓN (Figura 30)

Después de la emPCR se rompe la emulsión y se cuentan las beads, la recuperación óptima está entre el 65 y 85%. Se obtienen dos tipos de beads, beads con ADN amplificado y beads nulas sin producto amplificado, el proceso de enriquecimiento consiste en capturar las beads con ADN y eliminar las nulas, para ello se añade un buffer de hibridación que contiene un primer de enriquecimiento: 5' *Biotin-18-atom hexa-etilenglicol espaciador* - CGTTTCCCCTGTGTGCCTTGCCATCTGTTCCCTCCCTGTC-3' complementario al adaptador en el extremo 3' del template inmovilizado en las beads. Esta mezcla de beads se incuba con beads paramagnéticas recubiertas de estreptavidina que se unirán al primer de enriquecimiento biotinilado. El tubo se introduce en un imán para que se unan las beads paramagnéticas y el sobrenadante con beads nulas se eliminan. A continuación se libera la hebra no biotinilada y las beads que permanecen en el imán son usadas como librería de templates de cadena única. A las beads recuperadas se les añade el buffer de hibridación y el primer de secuenciación (5'-CCATCTGTTCCCTCCCTGTC-3'), se cuantifican y pueden almacenarse a 4°C. El enriquecimiento óptimo está entre el 5-20% de las 10 millones de beads originales añadidas a la emulsión (500.000-2.000.000 beads). Las beads se depositan en una placa de fibra óptica PTP (PicoTiterPlate), una bead por pocillo. En la placa PTP las fibras son empaquetadas

hexagonalmente en haces, cada core de fibra óptica tiene un diámetro de 29 μm rodeado de un revestimiento de 2-3 μm , el grabado de cada core crea pocillos de reacción con una distancia de centro a centro de 34 μm , que resulta en una densidad de pocillo de 3,2 millones en una superficie de 60x60 mm^2 [189].

Figura 30: Pasos previos a la reacción de secuenciación en el instrumento 454TM [191].



A.3.2.1.2 PIROSECUENCIACIÓN o UDPS (*Ultra-deep pyrosequencing*) (Figura 31)

Seguido a la carga de las beads con ADN amplificado en los pocillos de la PTP, se añaden beads recubiertas de sulfurilasa y luciferasa y dNTPs fluyen a través de los pocillos, entonces puede ocurrir la reacción de pirosecuenciación que se realiza en cinco pasos [189]:

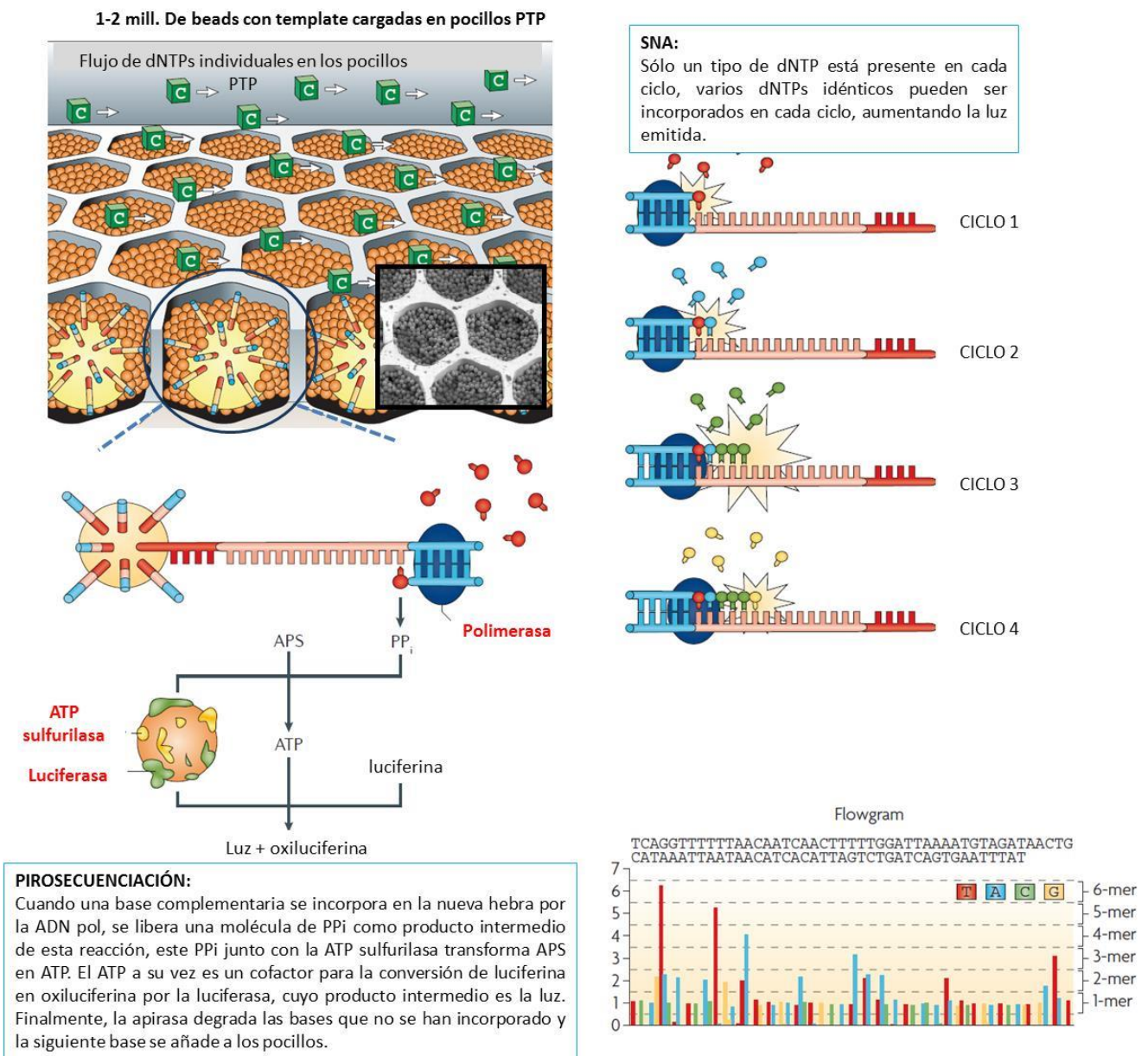
1. El ADN de cadena simple, amplificado por PCR, hibrida con el primer de secuenciación y se incuba con las enzimas ADN polimerasa, ATP sulfurilasa, luciferasa y apirasa, más los sustratos adenosina-5'-fosfosulfato (APS) y luciferina.
2. La adición de uno de los cuatro dNTPs inicia el segundo paso, en el que la ADN polimerasa cataliza la incorporación del dNTP al molde si es complementario. Si hay incorporación se libera PPI equivalente a la cantidad de dNTPs incorporados.
3. La ATP-sulfurilasa convierte cuantitativamente el PPI en ATP en presencia de APS. El ATP generado permite la conversión de la luciferina en oxiluciferina por acción de la luciferasa,

generando luz visible en cantidades proporcionales a la cantidad de ATP presente. La luz emitida es detectada por una cámara de alta resolución CCD (*charge-coupled device*). Cada señal luminosa es proporcional a la cantidad de nucleótidos incorporados.

4. Para continuar con la secuenciación, es esencial la degradación de aquellos dNTPs que no han sido incorporados. La apirasa es la enzima encargada de ello.

5. Nuevos dNTPs pueden ser añadidos para iniciar un nuevo ciclo.

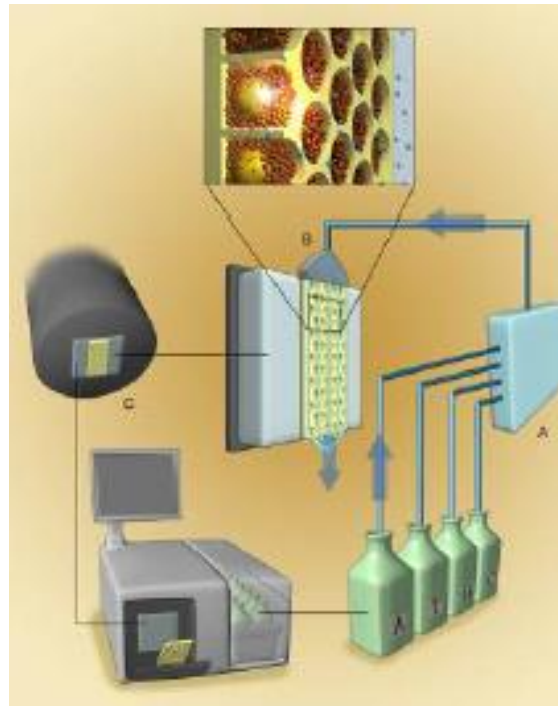
Figura 31: Pirosecuenciación [188].



A.3.2.1.3 CAPTURA DE LA IMAGEN

El instrumento 454TM consiste en tres montajes principales (Figura 32) [189]. La placa se monta en la cámara de flujo diseñada para crear un canal 300 μm encima de la apertura de los pocillos a través de la cual fluyen los reactivos de secuenciación. La parte de abajo de la placa de fibra óptica está directamente en contacto con la cámara CCD que permite la detección de la luz generada cuando tiene lugar la reacción de pirosecuenciación. El orden e intensidad de los picos de luz se graban en un *flowgrama* (Figura 31) que revela la secuencia de ADN [190].

Figura 32: El instrumento de secuenciación 454TM consiste en tres subsistemas principales [191]. (A) Subsistema de fluidica: dispensa los reactivos, conectado a un suministro de cada uno de los cuatro dNTPs. (B) Cámara de flujo: donde se coloca la placa PTP para ser sometida a flujos iterativos de dNTPs del dispensador de fluidos. (C) Subsistema óptico: formado por una cámara CDC (*charge-coupled device*) que capta una imagen de la placa después de cada flujo de un dNTP.



A.3.2.1.4 ANÁLISIS DE LA IMAGEN

Cada flujo en cada pocillo resulta en no incorporación, incorporación de uno o dos o tres o más nt, la intensidad de señal, calibrada a partir de la secuencia llave (Figura 28), indica directamente el número de nt incorporados, esta linealidad en la señal existe hasta al menos homopolímeros de ocho nt. La plataforma 454TM dispone de un software (Amplicon Variant Analyzer, AVA, Roche) que le permite convertir estos datos en *reads* en un proceso conocido como asignación de base (*base calling*), los parámetros de cada imagen (intensidad de señal,

ruido de fondo), se analizan por el software que aplica una serie de pasos de normalización, corrección y filtrado de la señal obtenida y la transforma en una gráfica de flujos (*flowgrama*) (Figura 31), el cual da lugar a un *read* [192]. Pero a veces se requiere de softwares más avanzados para la eliminación de bases de baja calidad y filtrado de los errores. Un proceso muy efectivo en reducir la tasa de error es aumentar la cobertura, es decir, cada *read* se alinea a una secuencia de referencia con un bajo nivel de exigencia para detectar mutaciones y variaciones genómicas, una vez que este alineamiento se ha llevado a cabo, se hace una media de todas las señales de varios *reads* que corresponden a la misma localización, y se identifica cada base individual [189].

B. NGS DE FRAGMENTOS LARGOS

Existen dos tipos de tecnologías para secuenciar fragmentos largos, por un lado las basadas en secuenciación a tiempo real de molécula única (SMRT, *Single-Molecule Real-Time Sequencing*), en las que se fundamentan las plataformas Pacific Biosciences (PacBio) y Oxford Nanopore Technologies (ONT) [188] y por otro lado, las tecnologías basadas en *reads* largos sintéticos (SLR, *Synthetic Long-Reads*), que se fundamentan en tecnologías de fragmentos cortos para construir *reads* largos *in silico*. En realidad no generan fragmentos largos, hacen uso de librerías de fragmentos cortos preparadas con códigos de barras que permiten el ensamblaje computacional en fragmentos largos, esta tecnología es la usada por la plataforma Illumina synthetic long-read sequencing y el sistema 10X Genomics (GemCode and Chromium).

2.3.6.4 COMPARACIÓN DE LAS PLATAFORMAS

Cada plataforma presenta unas características en cuanto a número y longitud de *reads*, precisión, tiempo por run y coste (Tabla 18), ventajas y desventajas, que hace que tengan diferentes aplicaciones (Tabla 19).

Tabla 18: Especificaciones de las plataformas de NGS. LONG; longitud, B; billón, d; días, h; horas, indel; inserciones y deleciones, M; millón, ND; no disponible. PE; paired-end sequencing y SE; single-end sequencing, los reads en la secuenciación se obtienen secuenciando desde ambos extremos o sólo desde uno, respectivamente [188].

NGS	PLATAFORMA	LONG READ (pb)	THROUGHPUT (bases/run)	Nº READS	TIEMPO RUN	TASA DE ERROR	COSTE INSTRUMENTO (US \$)	COSTE/Gb (US \$)	
SOLiD (Thermo Fisher)	SOLiD 5500 Wildfire	50 (SE)	80 Gb	~700 M	6 d	≤0,1%	ND	\$130	
		75(SE)	120 Gb						
		50 (SE)	160 Gb						
	SOLiD 5500 xl	50 (SE)	160 Gb	~1,4 B	10 d	≤0,1%	\$251000	\$70	
		75(SE)	240 Gb						
		50 (SE)	320 Gb						
Complete Genomics (BGI)	BGISEQ-500 FCS155	50-100 (SE/PE)	8-40 Gb	ND	24 h	≤0,1%	\$250	ND	
	BGISEQ-500 FCL155	50-100 (SE/PE)	40-200 Gb	ND	24 h	≤0,1%	\$250000	ND	
Illumina	Illumina MiniSeq Mid output	150 (SE)	2,1-2,4 Gb	14-16 M	17 h	<1%	\$50000	\$200-300	
		Illumina MiniSeq High output	75(SE)	1,6-1,8 Gb	22-25 M	7 h	<1%	\$50000	\$200-300
			75(PE)	3,3-3,7 Gb	44-50 M	13 h			
	150 (PE)		6,6-7,5 Gb	44-50 M	24 h				
	Illumina MiSeq v2	36 (SE)	540-610 Mb	12-15 M	4 h	0,10%	\$99000	\$1000	
		25(PE)	750-850 Mb	24-30 M	5,5 h			\$996	
		150(PE)	4,5-5,1 Gb	24-30 M	24 h			\$212	
		250 (PE)	7,5-8,5 Gb	24-30 M	39 h			\$142	
	Illumina MiSeq v3	75(PE)	3,3-3,8 Gb	44-50 M	21-56 h	0,10%	\$99000	\$250	
		300(PE)	13,2-15 Gb					\$110	
	Illumina NextSeq 500/550 Mid output	75 (PE)	16-20 Gb	Hasta 260 M	15 h	<1%	\$250	\$42	
		150 (PE)	32-40 Gb		26 h			\$40	
	Illumina NextSeq 500/550 High output	75 (SE)	25-30 Gb	400 M	11 h	<1%	\$250	\$43	
		75 (PE)	50-60 Gb	800 M	18 h			\$41	
		150 (PE)	100-120 Gb	800 M	29 h			\$33	
	Illumina HiSeq 2500 v2 Rapid run	36 (SE)	9-11 Gb	300 M	7 h	0,10%	\$690	\$230	
		50 (PE)	25-30 Gb	600 M	16 h			\$90	
		100 (PE)	50-60 Gb	600 M	27 h			\$52	
		150 (PE)	75-90 Gb	600 M	40 h			\$45	
		250 (PE)	125-150 Gb	600 M	60 h			\$40	
	Illumina HiSeq 2500 v3	36 (SE)	47-52 Gb	1,5 B (SE)	2 d	0,10%	\$690	\$180	
		50 (PE)	135-150 Gb	3 B (PE)	5,5 d			\$78	
		100 (PE)	270-300 Gb		11 d			\$45	
	Illumina HiSeq 2500 v4	36 (SE)	64-72 Gb	2 B (SE)	29 h	0,10%	\$690	\$150	
		50 (PE)	180-200 Gb	4 B (PE)	2,5 d			\$58	
		100 (PE)	360-400 Gb		5 d			\$45	
		125 (PE)	450-500 Gb		6 d			\$30	
	Illumina HiSeq 3000/4000	50 (SE)	105-125 Gb	2,5 B (SE)	1-3,5 d	0,10%	\$740/\$900	\$50	
		75 (PE)	325-375 Gb					\$31	
		150 (PE)	650-750 Gb					\$22	
Illumina HiSeq X	150 (PE)	800-900 Gb	2,6-3 B (PE)	<3 d	0,10%	\$1000	\$7		

Tabla 18: Continuación.

NGS	PLATAFORMA	LONG READ (pb)	THROUGHPUT (bases/run)	Nº READS	TIEMPO RUN	TASA DE ERROR	COSTE INSTRUMENTO (US \$)	COSTE/Gb (US \$)	
Gene Reader (Quiagen)	Quiagen GeneReader	ND	12 genes: 1250 mutaciones	ND	Varios días	Similar a otros sistemas SBS	ND	\$400-600/panel	
454 (Roche)	454 GS Junior	Hasta 600, 400 de media (SE, PE)	35 Mb	0,1 M	10 h	1%	ND	\$40000	
	454 GS Junior+	Hasta 1000, 700 de media (SE, PE)	70 Mb	0,1 M	18 h	1%	\$108000	\$19500	
	454 GS FLX Titanium XLR70	Hasta 600, 450 de media (SE, PE)	450 Mb	0,1 M	10 h	1%	ND	\$15500	
	454 GS FLX Titanium XL+	Hasta 1000, 700 de media (SE, PE)	700 Mb	1 M	23 h	1%	\$450000	\$9500	
Ion torrent	Ion PGM 314	200 (SE) 400 (SE)	30-50 Mb 60-100 Mb	400000- 550000	23 h 3,7 h	1%	\$49	\$25-3500	
	Ion PGM 316	200 (SE) 400 (SE)	300-500 Mb 600 Mb	2-3 M	3 h 4,9 h	1%	\$49	\$700-1000	
	Ion PGM 318	200 (SE) 400 (SE)	600 Mb 1-2 Gb	4-5,5 M	4 h 7,3 h	1%	\$49	\$450-800	
	Ion Proton	Hasta 200 (SE)	Hasta 10 Gb	60-80 M	2-4 h	1%	\$224	\$80	
	Ion S5 520	200 (SE) 400 (SE)	600 Mb 1,2-2 Gb	3-5 M	2,5 h 4 h	1%	\$65	\$2400 \$1200	
	Ion S5 530	200 (SE) 400 (SE)	3-4 Gb 6-8 Gb	15-20 M	2,5 h 4 h	1%	\$65	\$950 \$475	
	Ion S5 540	200 (SE)	10-15 Gb	60-80 M	2,5 h	1%	\$65	\$300	
	PacBio	Pacific BioSciences RS II	20 kb	500 Mb- 1 Gb	55000	4 h	13% paso único, pero ≤1% read circular consenso	\$695	\$1000
Pacific BioSciences Sequel		8-12 Kb	3,5-7 Gb	350000	0,5-6 h	ND	\$350	ND	
ONT	Oxford Nanopore MK 1 MinION	Hasta 200 Kb	Hasta 1,5 Gb	>100000	Hasta 48 h	12%,	\$1000	\$750	
	Oxford Nanopore PromethION	ND	Hasta 4 Tb	ND	ND	ND	\$75	ND	
Illumina synthetic long-read	Illumina Synthetic long-read	100 kb	Ver 2500	HiSeq 2500	Ver HiSeq 2500	HiSeq 2500	Ver 2500	HiSeq 2500	No se requiere instrumento adicional \$1000
10X Genomics	10X Genomics	Hasta 100 kb	Ver 2500	HiSeq 2500	Ver HiSeq 2500	HiSeq 2500	Ver 2500	HiSeq 2500	Ver HiSeq 2500 +\$500/muestra

Tabla 19: Ventajas, desventajas y aplicaciones de las plataformas de NGS [164,187,188,193]. FC; fragmentos cortos, FL; fragmentos largos, SBL; *sequencing by ligation*, SBS; *sequencing by synthesis*, CRT; *cyclic reversible termination*, SNA; *single-nucleotide addition*, SMRT; *Single-Molecule Real-Time Sequencing*, SLR; *Synthetic Long-Reads*. Indel: inserciones y/o deleciones.

NGS		VENTAJAS	DESVENTAJAS	APLICACIONES		
FC	SBL	SOLiD (Thermo Fisher)	Muy precisas	Longitud corta <i>reads</i> y tiempo de run largo. Sustituciones, imprecisión regiones ricas AT y GC	Resequenciación, transcryptómica cuantitativa, genotipaje	
		Complete Genomics (BGI)	Muy precisas	Longitud corta <i>reads</i> . Imprecisión regiones ricas AT		
	SBS	CRT	ILLUMINA	HiSeq X es la plataforma de mayor rendimiento actualmente, menos susceptible a errores de homopolímeros	Sustituciones, imprecisión regiones ricas AT y GC	Secuenciación genoma y exoma, aplicaciones epigenómicas: ChIP-seq (<i>chromatin immunoprecipitation followed by sequencing</i>), ATAC-seq (<i>assay for transposase-accessible chromatin using sequencing</i>) y metilseq (<i>DNA methylation sequencing</i>) y aplicaciones de transcriptómica: RNA-seq (<i>RNA sequencing</i>)
			Gene Reader (Quiagen)	Similar a Illumina Miniseq	Similar a Illumina Miniseq	Paneles de genes relacionados con cáncer
	SNA	454 (Roche)	<i>Reads</i> más largos	Indel, homopolímeros	Resequenciación dianas (amplicones), genotipaje	
		Ion torrent	<i>Reads</i> más largos y run rápidos	Alta tasa de error (Indel predominante), rendimiento limitado, alto coste	Paneles de genes y aplicaciones clínicas	
FL	SMRT	PacBio	Sin PCR y fragmentos largos		Secuenciación <i>de novo</i> , resecuenciación dianas, muestras no amplificables	
		ONT	Portabilidad, no manipulación humana	Indel, homopolímeros	Respuesta clínica rápida	
	SLR	ILLUMINA synthetic long-read	Se pueden aprovechar las infraestructuras de Illumina	Posible error de código de barras	Como Illumina, pero además fragmentos largos	
		10X Genomics	Se pueden aprovechar infraestructuras, poco ADN de partida	Posible error de código de barras	Biopsias	

2.4 HIPÓTESIS Y OBJETIVOS

2.4.1 HIPÓTESIS

El VHB exhibe una tasa de mutación de $3,2 \times 10^{-5} - 7,9 \times 10^{-5}$ nt sustituidos/ciclo replicativo, cien veces superior a otros virus ADN. Resultando en una diversidad genética en forma de genotipos, subgenotipos, cuasiespecies y variantes en diferentes regiones del genoma. Existen evidencias de que la diversidad genética influye en las tasas de seroconversión HBeAg, los niveles de ADN-VHB, el escape del sistema inmune, la severidad de la lesión hepática, la respuesta y la aparición de resistencias al tratamiento antiviral y los resultados de la vacunación [120], lo que evidencia la gran relevancia del estudio de esta diversidad. Por razones metodológicas, hasta hace unos años el estudio de la variabilidad del VHB se ha limitado al estudio observacional de ciertas variantes cuando están presentes en forma mayoritaria, como las mutaciones de resistencia a NUCs, gracias al desarrollo de las técnicas de NGS, en la actualidad se puede estimar cuantitativamente la composición de una QS y analizar proporciones minoritarias de variantes. La elección de una tecnología u otra de NGS depende del objetivo final, en este caso al requerirse fragmentos relativamente largos y un elevado número de *reads*, es decir, una alta cobertura, se ha utilizado UDPS (*Ultra-deep pyrosequencing*). La principal desventaja de esta tecnología es la alta tasa de error en zonas homopoliméricas, no obstante en el laboratorio se han desarrollado y validado algoritmos bioinformáticos que permiten un correcto tratamiento de los datos [185], confirmando el carácter cuantitativo de esta tecnología y su utilidad en el estudio evolutivo de la QS viral.

La dinámica de la QS permite al VHB evolucionar fácilmente y adaptarse a situaciones ambientales cambiantes como la presión del sistema inmune o la aplicación de un determinado tratamiento, y su complejidad podría estar asociada con diferente curso de la infección, la probabilidad de respuesta al tratamiento o la seroconversión anti-HBe. Por lo tanto, el estudio de las variantes de una QS y su dinámica, así como la complejidad, es fundamental para entender la adaptabilidad, el poder patogénico y la persistencia del VHB y así desarrollar estrategias para prevenir y tratar la enfermedad que causa.

La variabilidad del VHB se expresa tanto a nivel poblacional (QS) como a nivel genotípico, en este último caso, ya se ha descrito previamente por nuestro propio grupo la presencia de mezclas de diferentes genotipos en un mismo paciente [194]. No obstante, estos datos han sido reportados tras el análisis de unas regiones concretas del genoma viral y no se han realizado estudios comparativos de genotipaje viral y detección de mezclas en diferentes regiones. Por otro lado, múltiples evidencias recientes indican una alta tasa de recombinación

en los genomas del VHB que incluso cuestiona la clasificación actual genotípica de los mismos [148], por lo que se puede esperar que el análisis de dos regiones diferentes del genoma resulte en genotipos diferentes como consecuencia de la posible recombinación.

Dentro del genoma, la región X y preCore que incluye parte del ORF X, el preCore/Core y los principales elementos reguladores PBC y ENHII (región X/PC/Core/PBC), parece la más indicada para los estudios de complejidad y variabilidad genética.

La primera hipótesis de este proyecto de tesis doctoral, que se desarrolla en el primer estudio, es que la complejidad de la QS se asocia con características clínicas tales como el proceso de seroconversión HBeAg/anti-HBe, clave en la historia natural de la infección, así como la respuesta virológica al tratamiento antiviral. El estudio de complejidad a nivel de QS requiere de una región que no presente demasiadas restricciones por solapamiento a la evolución, por este motivo se analiza la complejidad en un fragmento de la región PC/Core/PBC que presenta muy bajo grado de solapamiento y se asume, por tanto, que tiene una mayor tolerancia a la variabilidad, considerándose óptima para el estudio de la evolución de la QS que se desarrolla en el primer estudio de esta tesis. En la figura 33 se detalla la región seleccionada.

La segunda hipótesis de este proyecto de tesis doctoral es que la complejidad de la población viral también existe a nivel de los genotipos, es decir, la presencia de mezclas de diferentes genotipos del VHB en un mismo paciente no es un fenómeno infrecuente y además existen evidencias recientes de una alta tasa de recombinación [148], cuestionándose incluso la actual clasificación genotípica del VHB. En este sentido se plantea la hipótesis de que el análisis de dos regiones diferentes del genoma viral en una misma muestra puede dar lugar a resultados genotípicos diferentes. Este hecho iría en la línea de la necesidad de replantear la clasificación del genoma del VHB, paso esencial para asociar el genotipo viral con posibles implicaciones clínicas, resulta poco fundamentado pretender asociar un genotipo u otro con una consecuencia clínica o la eficacia de un proceso preventivo o terapéutico si la clasificación genotípica no es la adecuada. Para el desarrollo de este segundo objetivo de distribución de genotipos virales y la posible presencia de mezclas de los mismos, su composición y proporción, así como la comparación en dos regiones diferentes del genoma, se han seleccionado dos fragmentos con una restricción a la evolución por solapamiento similar, para así poder comparar sus complejidades. El primero de ellos corresponde a la región P/S, donde el gen S y el dominio RT de la polimerasa se solapan, y el segundo a la región X/PC, también con un alto grado de solapamiento, que incluye los principales elementos reguladores del ciclo viral y en la que se han descrito numerosas variantes con relevancia en la patogénesis y la respuesta al tratamiento, ambos se detallan en la figura 33.

2.4.2 OBJETIVOS

Para el desarrollo de las hipótesis indicadas en el apartado anterior se han planteado los siguientes objetivos:

1. Estudiar mediante técnicas de NGS (*Ultra-Deep Pyrosequencing*) la variabilidad genética del VHB a nivel de la QS viral (secuencias de la población que infecta a un individuo concreto) mediante diferentes parámetros de complejidad. Este objetivo se subdivide en dos:

1.1 La caracterización en profundidad de la complejidad de la QS del VHB en la región PC/Core/PBC, por ser la región del genoma viral con un menor grado de solapamiento y por lo tanto con una menor restricción a la variabilidad, y su evolución tanto en la dinámica natural como bajo la presión del tratamiento antiviral con NUCs.

1.2 Relacionar esta complejidad con el estatus HBeAg y la respuesta al tratamiento antiviral con NUCs.

2. Estudiar mediante técnicas de NGS (*Ultra-Deep Pyrosequencing*) la variabilidad genética y complejidad del VHB a nivel de genotipos, para detectar la presencia de infecciones mixtas y estudiar la evolución de esta complejidad en ausencia y presencia de tratamiento antiviral con NUCs, comparando los resultados en dos regiones del genoma (P/S y X/PC) con un solapamiento similar y por tanto, una restricción a la variabilidad similar.

Figura 33: Genoma del VHB indicando las regiones de los estudios (A) y de manera detallada para PC/Core/PBC y X/PC (B).

Fragmento PC/Core/PBC del primer estudio (nt 1757-2152): abarca la región PC/Core/PBC que incluye el preCore completo (nt 1814-1900), asociado con la expresión de la proteína inmunomoduladora HBeAg y con la señal de encapsidación ϵ (nt 1847-1908) esencial para el inicio de la replicación, los 84 primeros codones del core (nt 1901-2142), que incluyen los dos dominios epitópicos principales Th 50-69 (nt 2048-2107) y B74-84 (nt 2120-2152) y la única región sin solapamiento del genoma (nt 1838-2357), y la parte Ct del ORF X (nt 1374-1838). Además incluye diferentes regiones reguladoras, la principal el PBC (nt 1742-1849), esencial para la transcripción del ARNpg, el extremo 5' del ENHII (nt 1627-1774), implicado en la replicación del VHB, 3 de las 4 cajas TATA-like (TA2, nt 1758-1762; TA3, nt 1771-1775; TA4, nt 1788-1795), la región de escaneo ribosomal kozak (nt 1809-1812) y la DR1 (nt 1824-1834), necesaria para la unión del primer e inicio de la síntesis de la hebra(-).

Fragmento P/S del segundo estudio (nt 615-969): Cubre un fragmento donde el gen S y el dominio RT de la polimerasa se solapan.

Fragmento X/PC del segundo estudio (nt 1590-1912): abarca la región X/PC, que cubre la mayor parte del ORF X (nt 1374-1838), incluyendo toda la parte Ct transactivadora, se solapa con el codón stop del gen P (nt 1623) y con el inicio del preCore (nt 1814-1900), también incluye un pequeño fragmento del gen core (nt 1901-2142). Esta región agrupa los principales elementos reguladores del VHB, parte de la DR2 (nt 1590-1600), donde comienza la síntesis de la hebra(+), el ENHII (nt 1627-1774) completo, las cuatro cajas TATA-like (TA1, nt 1750-1755; TA2, nt 1758-1762; TA3, nt 1771-1775; TA4, nt 1788-1795), la región de escaneo ribosomal kozak (nt 1809-1812), la DR1 (nt 1824-1834) y el PBC (nt 1742-1849) completo. Todo ello hace que tenga una gran relevancia clínica.

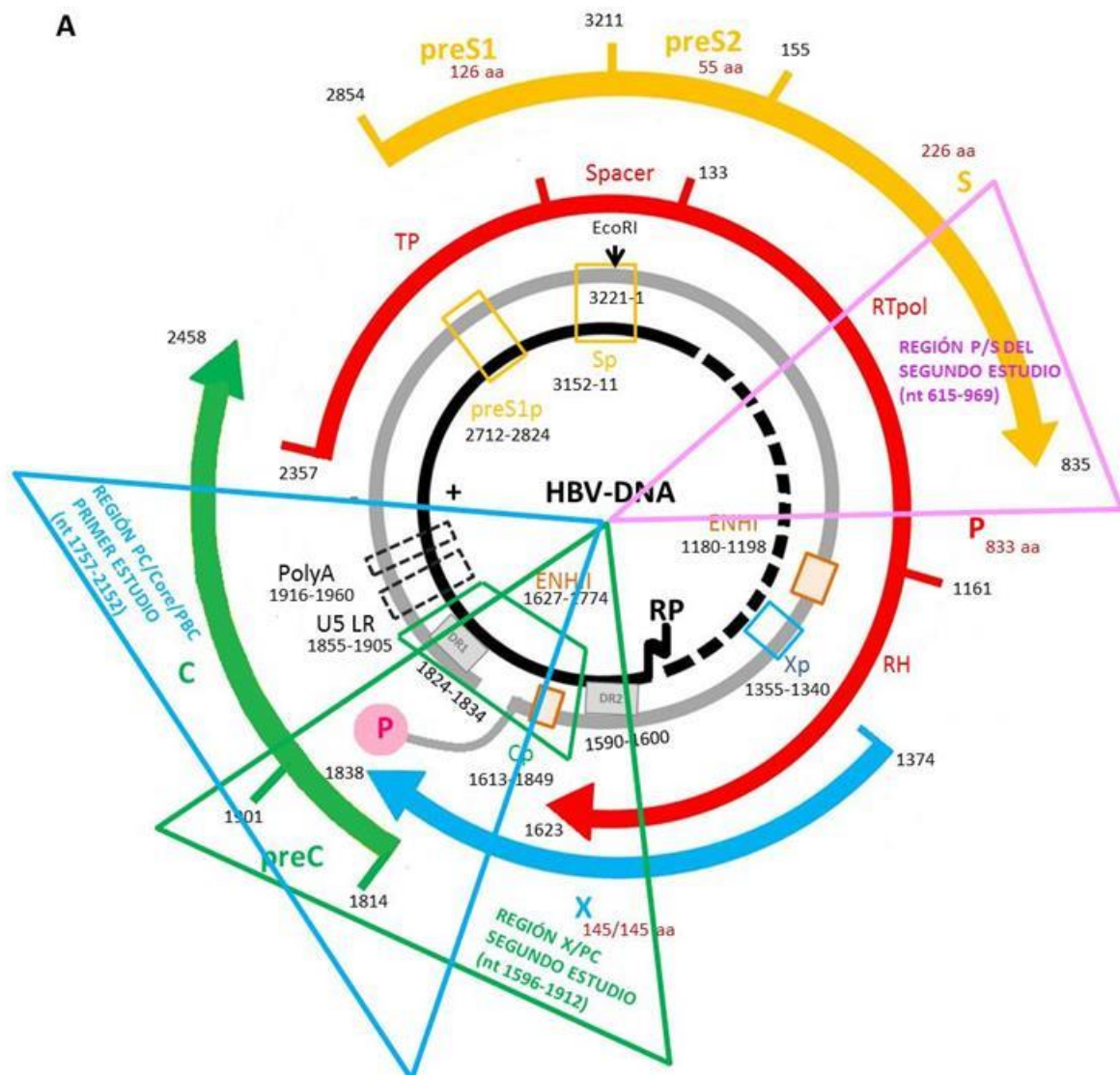
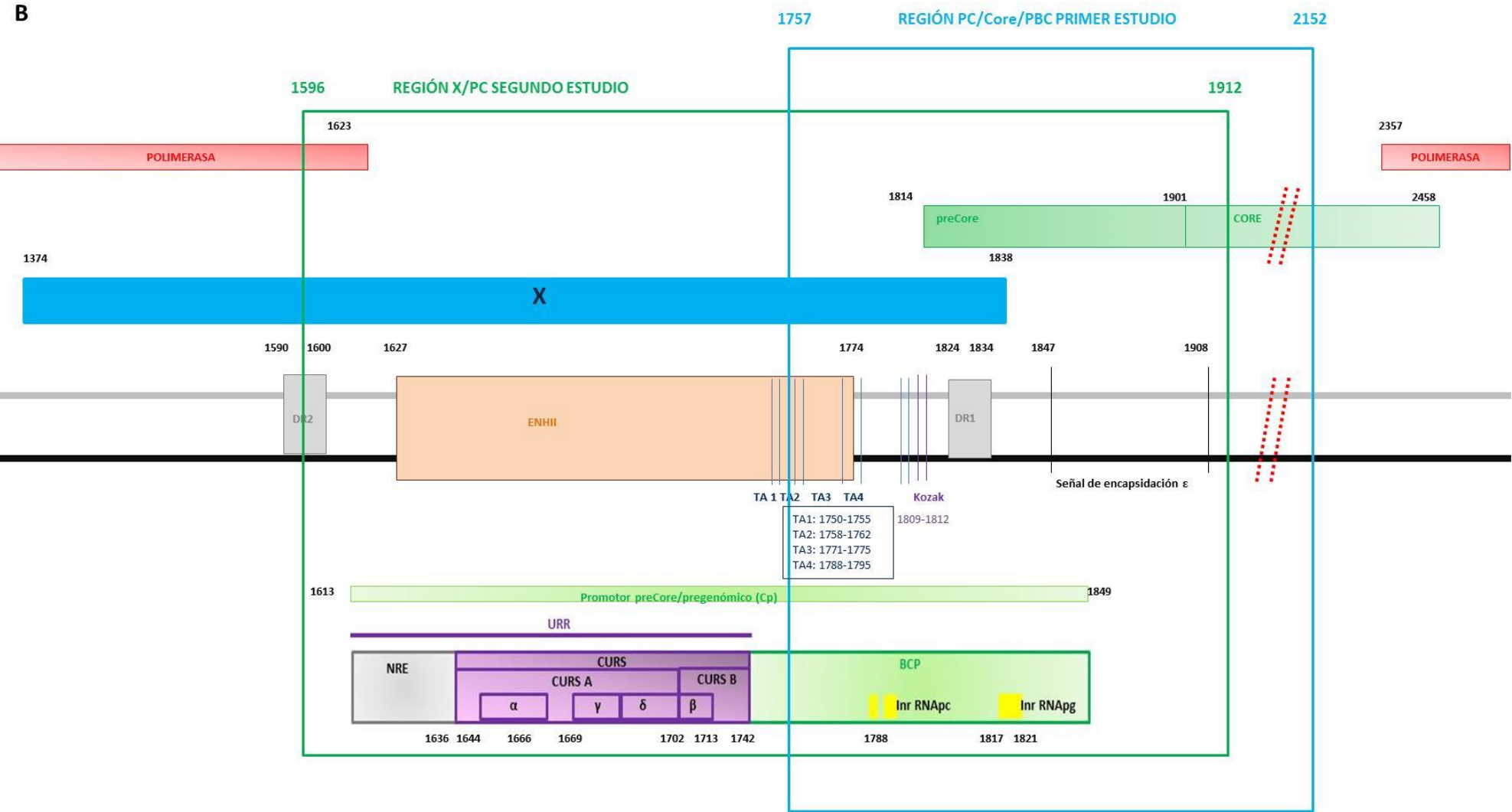


Figura 33: Continuación..



Primer estudio

3. PRIMER ESTUDIO: Clinical Application of Estimating Hepatitis B Virus Quasispecies Complexity by Massive Sequencing: Correlation between Natural Evolution and On-Treatment Evolution

3.1 Introducción

La QS del VHB es capaz de modificar su composición para adaptarse a las condiciones cambiantes del ambiente como pueden ser la presión del sistema inmune o un determinado tratamiento antiviral, con lo que su estudio es importante para entender la adaptabilidad, el poder patogénico y persistencia del VHB y así optimizar las estrategias en el manejo y prevención de la infección, para ello existen varios índices de diversidad (S_n , M_f y P_i) que sirven para cuantificar la complejidad de la QS.

Estudios recientes apoyan el concepto de que la complejidad de la QS es un factor clínicamente relevante, varios grupos han descrito como la complejidad y diversidad pueden influir en el potencial patogénico, la evolución clínica, la seroconversión anti-HBe o la respuesta al tratamiento antiviral. Por ejemplo, pacientes que seroconvierten tienen significativamente mayor diversidad de la QS en la región preCore/Core y esta diversidad es más extrema justo antes de la seroconversión HBeAg [170], la complejidad de la QS y diversidad dentro de la región RT en respondedores al tratamiento antiviral con LAM o ETV es más baja que en no respondedores o respondedores parciales, respectivamente, durante las fases tempranas del tratamiento [165,168,172] y se ha observado mayor diversidad en seroconvertidores anti-HBe naturales o inducidos por IFN o LAM que en no seroconvertidores [170,195].

Durante el curso natural de la infección o bajo el tratamiento con NUCs se monitorizan varios marcadores virológicos y bioquímicos, entre ellos el HBeAg, que está asociado con diferente curso de la infección y respuesta al tratamiento y su expresión depende de variantes localizadas en el PC y PBC. La región PC/core, además, regula la replicación, incluye la única zona no solapante del genoma y en el gen core se encuentran dominios epitópicos que juegan un papel clave en la respuesta del sistema inmune, todo ello hace que esta región sea óptima para investigar la evolución natural de la QS bajo la presión del sistema inmune.

3.2 Hipótesis y objetivos

3.2.1 Hipótesis

En base a los estudios previos que muestran una alta complejidad de la QS en la región PC/core y que asocian esta complejidad, estimada con parámetros objetivos como la S_n , con la severidad de la lesión hepática, la probabilidad de la seroconversión anti-HBe y el grado de respuesta al tratamiento antiviral, realizados con métodos clásicos de clonaje, se decidió analizar esta región mediante UDPS. Además existían pocos estudios de esta región bajo tratamiento con NUCs.

3.2.2 Objetivos

Los objetivos de este estudio fueron:

1. Caracterizar la complejidad de la QS del VHB en las región PC/Core/PBC durante la dinámica natural y bajo el tratamiento antiviral con NUCs, con una profundidad superior a la obtenida con los métodos de clonaje clásicos.
2. Evaluar la asociación entre el estatus HBeAg y la complejidad de la QS del VHB en la región PC/Core/PBC.

3.3 Resumen del estudio

Se amplificó por UDPS en el secuenciador FLX (454Life Sciences) la región PC/Core/PBC (nt 1757-2152) (Figura 33) en muestras secuenciales de 10 pacientes con HBC en los que siendo *naive* previamente falló el tratamiento con LAM, fueron seleccionados por presentar diferente estatus de HBeAg a lo largo del tiempo; grupo HBeAg(+): 4 pacientes con HBeAg(+) permanentemente, grupo HBeAg(-): 2 pacientes HBeAg(-) permanentemente y grupo HBeAg (+/-): 4 pacientes con HBeAg fluctuante en los que se observan períodos HBeAg(+) y HBeAg(-) (Figura 34). De cada paciente se analizaron tres muestras; una basal obtenida en el momento del diagnóstico: B, una posterior a la basal y antes del inicio del tratamiento con NUCs: TF (*treatment-free*) y otra después del fallo del tratamiento: TNR (*treatment nonresponse*), que definen dos períodos: (1) un período sin tratamiento entre B y TF y (2) un período bajo la presión del tratamiento con NUCs entre TF y TNR. Se obtuvieron 986.014 secuencias con una alta cobertura por paciente (mediana 98.494) y una tasa de error del 0,25%. Como parámetros de complejidad se calcularon la entropía de Shannon normalizada (S_n), la frecuencia de mutación (M_f) y la diversidad nucleotídica (P_i), y se observó una correlación significativa entre las tres, con lo que cualquiera de ellas puede representar la complejidad de la QS. Además se evaluó la complejidad de QS a nivel de aa con la frecuencia de mutación (M_{fAA}), de manera

independiente para la región preCore (48 aa) y core (84 aa). Las diferencias en los parámetros de complejidad entre la muestra B y TF se consideran medidas de la dinámica natural de la QS, es decir, cambios en el *fitness* de haplotipos por factores intrínsecos (como la respuesta inmune) y sin presiones externas, y entre TF y TNR medidas de la evolución bajo la presión externa del tratamiento.

3.2.1. Complejidad de la QS y su relación con el período de toma de la muestra, genotipo, ADN-VHB, niveles ALT y estatus HBeAg

La comparación de la complejidad de QS entre las muestras B, TF y TNR mostró que la Sn, Mf y Pi eran significativamente más altas en B comparada con TNR, también más alta en TF comparado con TNR pero no significativamente y con valores similares entre B y TF (Figura 34), lo que sugiere un descenso en la evolución de la QS asociado con el fallo del tratamiento con LAM, la QS presenta menor heterogeneidad. Se ha descrito que la respuesta inmune aumenta cuando el tratamiento controla la replicación del virus [196], en este caso al ser pacientes no respondedores, estaríamos en la situación contraria, la mayoría de los pacientes muestra un descenso en la complejidad de QS en TNR que podría reflejar una atenuación de la actividad del sistema inmune, al haber una menor presión a la evolución, la complejidad se reduce.

No se observaron diferencias significativas en los parámetros de complejidad entre Gen/A y D, a pesar de la restricción a la selección de las principales mutaciones en el preCore por el genotipo, ni correlación entre estos parámetros y los niveles de ADN-VHB o ALT (Figura 34). Aunque se asume que los niveles de ALT son un estimador de la intensidad de la respuesta del sistema inmune contra la infección, actividad necro-inflamatoria, sus niveles pueden estar influenciados por muchos otros factores y una medida puntual puede no ser indicativa del estatus del sistema inmune a largo plazo. En cuanto a la carga viral, el análisis a nivel de aa sí que mostró que la MfAA en el PC tenía una correlación negativa significativa con los niveles de ADN-VHB y la MfAA en el core aumentaba conforme los niveles de ADN-VHB disminuían, pero no llegaba a alcanzar la significación estadística, quizás porque la alta carga viral de los pacientes puede actuar como factor de confusión. Estos resultados están de acuerdo con hallazgos recientes [169,170] que demuestran que cuanto menor es la complejidad en el PC/core mayor es la carga viral, lo que sugiere que el aumento de complejidad en el PC está asociado con un descenso de la replicación viral. Las variantes PC con mutaciones podrían disminuir el *fitness* del preCore para regular la replicación viral, el PC contiene la señal de encapsidación ϵ , esencial para la replicación, y las sustituciones podrían interferir en su funcionalidad, esto llevaría a una menor replicación viral y por tanto a una población más pequeña (bajos niveles de ADN-VHB) pero muy heterogénea (alta complejidad).

De manera global la complejidad de la QS fue significativamente menor en muestras HBeAg(+) que HBeAg(-) tanto a nivel nucleotídico (Sn, Mf y Pi), como a nivel de aa en el preCore (MfAA PC) y core (MfAA core). Para evitar el efecto del tratamiento también se realizaron los cálculos sin las muestras TNR y los resultados fueron los mismos para la Sn y Pi pero Mf, MfAA PC y core no mostraron diferencias significativas entre HBeAg(+) y HBeAg(-). Teniendo en cuenta los diferentes grupos de pacientes según el estatus HBeAg, la complejidad de QS fue significativamente menor en el grupo HBeAg(+) que en el grupo HBeAg(-) o HBeAg(+/-) y no se observaron diferencias significativas entre HBeAg(-) y HBeAg(+/-) (Figura 34). Para intentar definir si la variabilidad estaba localizada en el PC o en el core, se calculó de MfAA por separado en ambas regiones en cada grupo, la MfAA tanto en el PC como en el core fueron significativamente menores en el grupo HBeAg(+) que HBeAg(-) o HBeAg(+/-) y no se observaron diferencias entre HBeAg(-) y HBeAg(+/-). Esta mayor complejidad encontrada en el grupo HBeAg(-) comparado con HBeAg(+) está de acuerdo con estudios previos y en conjunto se puede decir que el aumento de la diversidad viral está asociada con la seroconversión, el aumento de la presión de evolución debido a una respuesta inmune más intensa, probablemente debido a la falta de HBeAg y su efecto inmunomodulador, llevaría a este aumento de la complejidad.

3.2.2 Complejidad de QS y estatus HBeAg: estudio longitudinal

Para explorar diferencias en la complejidad de QS entre los diferentes tiempos de muestreo, los parámetros fueron calculados en cada grupo según el estatus HBeAg en las tres muestras analizadas (B, TF y TNR) (Figura 34). El estudio longitudinal permitió deducir algunos patrones de evolución en ausencia y presencia de tratamiento. La complejidad de QS fue sistemáticamente más alta en el grupo HBeAg(-) y HBeAg (+/-) que en el HBeAg(+) en los tres puntos, pero estadísticamente significativa sólo en la muestra basal entre el grupo HBeAg(+) y HBeAg(+/-), con una complejidad significativamente menor en el grupo HBeAg(+) que HBeAg(+/-). La mayor complejidad en pacientes HBeAg(-) encontrada a nivel global, no es significativa cuando analizamos la complejidad teniendo en cuenta la evolución, pero esto quizás sea debido al pequeño número de pacientes HBeAg(-), lo que sí se observa es que el grupo HBeAg(+/-) tienen mayor complejidad que el HBeAg(+) en la muestra B a nivel nucleotídico y en B y TF cuando la complejidad se analiza a nivel de aa en el core (MfAA Core), pero no hay diferencias significativas entre ninguno de los grupos en MfAA en el preCore (Figura 35). Esto podría indicar que el sistema inmune actúa principalmente contra los epítomos del core en pacientes con HBeAg fluctuante.

En conjunto estos hallazgos (1) mayor complejidad en el grupo HBeAg(-) comparado con HBeAg(+) a nivel global, (2) mayor complejidad sistemáticamente en el grupo HBeAg(-) y HBeAg (+/-) que en el HBeAg(+) en los tres puntos, (3) mayor complejidad en el grupo HBeAg(+/-) que en HBeAg(+) en las muestra B a nivel nucleotídico y en B y TF a nivel de aa en el core, aportan evidencias de que el aumento de la diversidad está asociado con la seroconversión HBeAg. El aumento de la complejidad en pacientes HBeAg(-) sugiere un aumento de la presión de evolución debido a una respuesta inmune más intensa, o incluso al efecto del tratamiento en el gen core, que es más evidente en el estatus HBeAg fluctuante, y que probablemente se deba a la ausencia de HBeAg y su efecto inmunomodulador.

Figura 34: Distribución de la complejidad de la QS, representada con la Sn, carga viral y estatus HBeAg de cada uno de los 10 pacientes de los 10 pacientes. Eje izquierdo, y líneas negras punteadas: HBV-DNA (log UI/ml). *NUC: indica los niveles de ADN-VHB durante el tratamiento con LAM. Eje derecho, y líneas grises: Sn en las muestras B, TF y TNR. Barras encima del gráfico: estatus HBeAg (gris oscuro HBeAg(+), gris claro HBeAg(-)). Pacientes agrupados según estatus HBeAg, siempre HBeAg(+) en rojo, siempre HBeAg(-) en verde y con HBeAg fluctuante en azul.

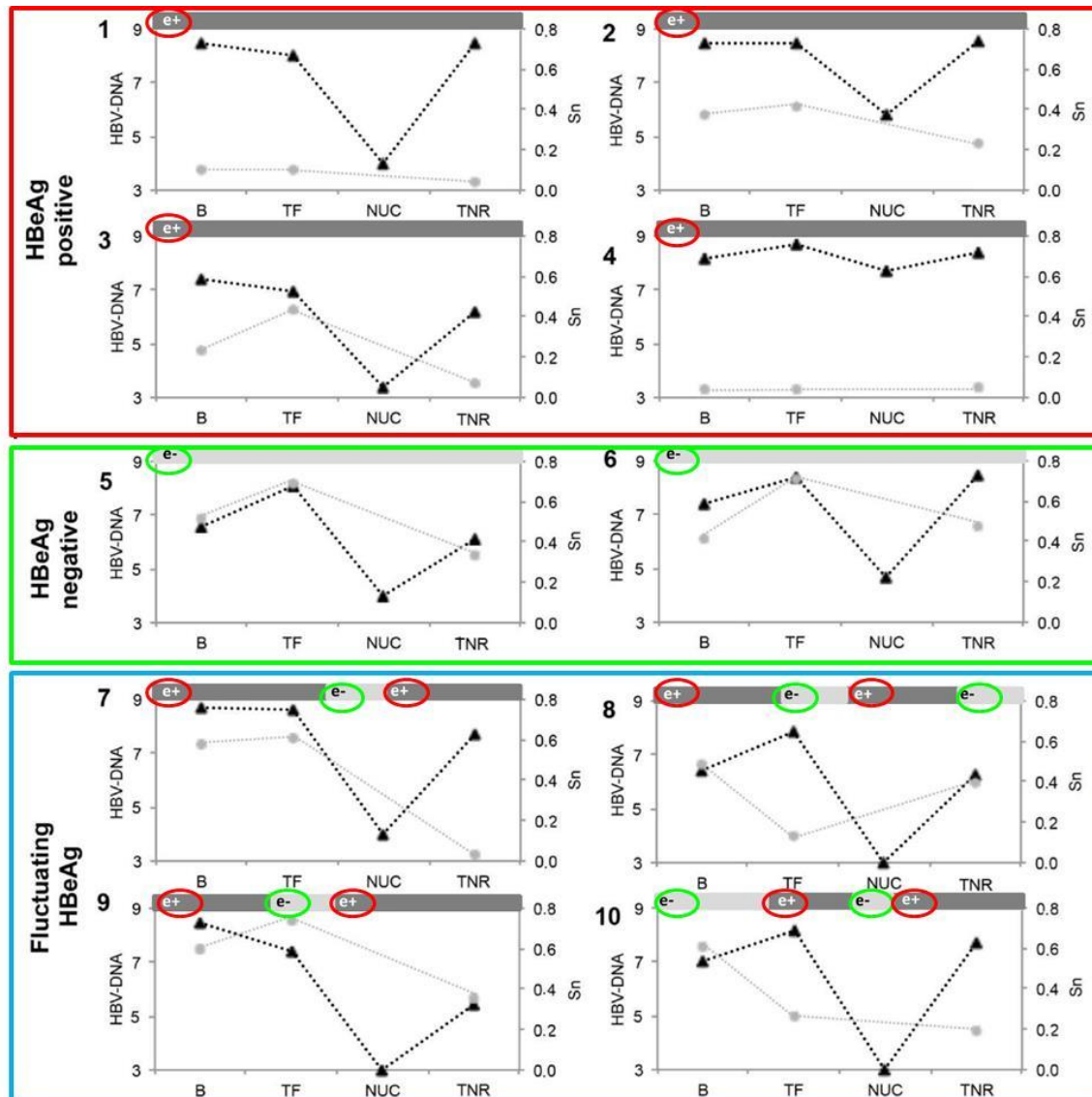
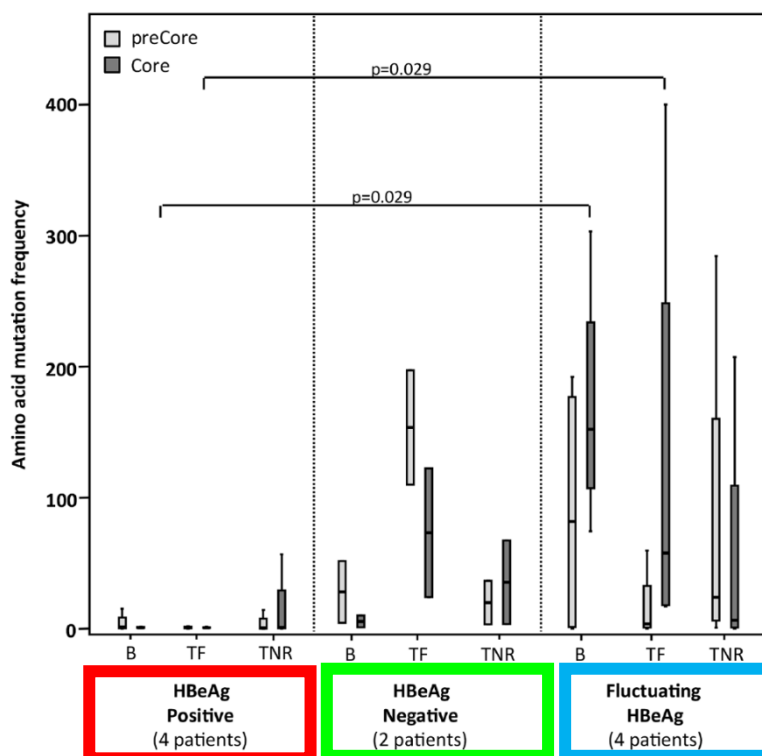


Figura 35: Distribución de MfAA en la región PC (gris claro) y core (gris oscuro) en cada una de las tres muestras (B, TF y TNR) en los tres grupos según el estatus HBeAg.



3.2.3. Distribución de los cambios en preCore y core

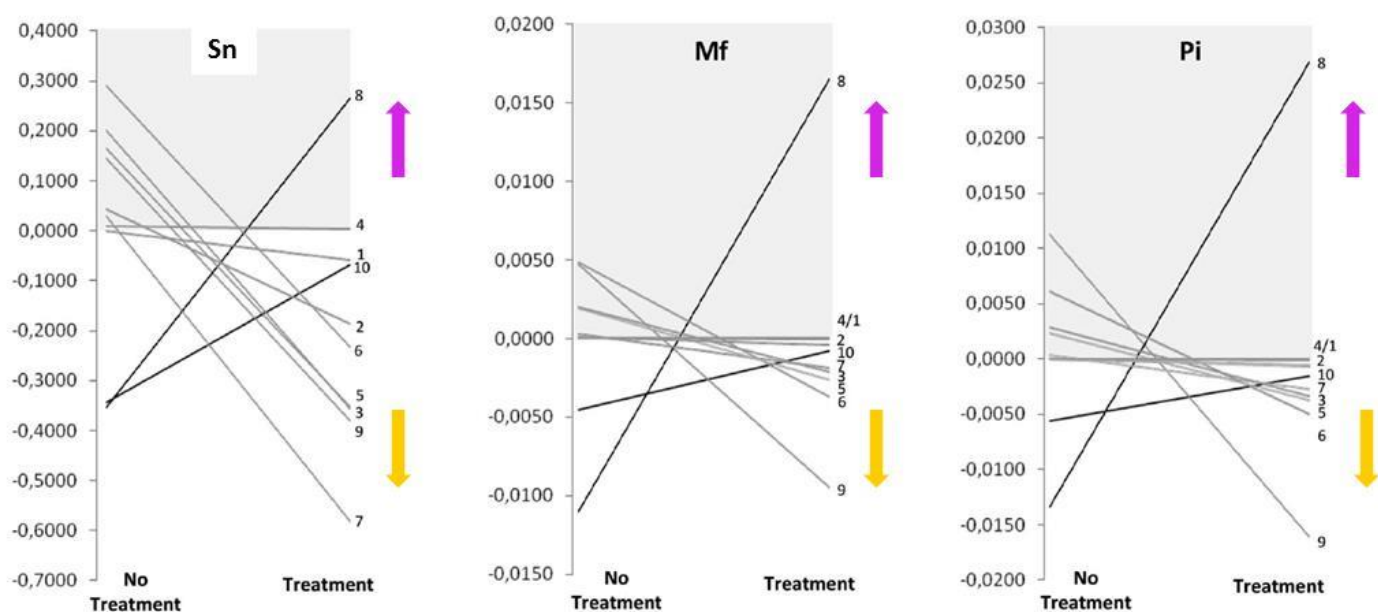
Se analizó la distribución de los cambios de nt encontrados en ambas regiones y se observó que en 5 pacientes los cambios de nt se agrupaban sobre todo en la región core, en 4 pacientes, en ambas regiones y en uno en la región PC principalmente, sin patrón específico según la muestra (B, TF y TNR). Las principales mutaciones responsables de la disminución o no expresión de HBeAg en el PBC y PC, respectivamente, fueron más frecuente en el grupo HBeAg(+/-) que en pacientes HBeAg(+) y eran las secuencias normales del grupo HBeAg(-). La mitad de los pacientes mostraron una pequeña tendencia a acumular variabilidad en las regiones epitópicas del core (Th50-69 y B74-84), lo que está de acuerdo con su función inmunoestimuladora. Estos resultados, sumados a que la variabilidad a nivel de aa en el core (MfAA core) fue mayor en el grupo HBeAg(+/-) en las muestras B y TF que en el grupo HBeAg(+), apoyan la idea de que el sistema inmune actúa principalmente contra los epítomos del core en pacientes con HBeAg fluctuante. Sin embargo, la selección positiva de variantes core en la QS, definida como la selección de una nueva secuencia máster diferente de la secuencia máster en la muestra previa, era más frecuente en el grupo HBeAg(-) y HBeAg(+/-) que en HBeAg(+), lo que sugiere el aumento de la presión del sistema inmune por la ausencia de HBeAg.

3.2.4 Evolución de la complejidad de la QS sin tratamiento y bajo la presión de NUCs

La evolución de la complejidad de la QS se analizó en dos períodos, en la dinámica natural de la QS y bajo la presión de NUCs, la evolución de cada uno de los 10 pacientes en el período TF y TNR se muestra en la figura 36. En la mayoría de los casos (7/10), se observó un aumento en la complejidad durante la evolución natural y un descenso bajo el tratamiento. Sólo en un caso no se observaron cambios en la complejidad entre TF y TNR y dos pacientes HBeAg(+/-) mostraron un patrón inverso, probablemente debido al estatus HBeAg fluctuante.

Pero a pesar de los diferentes patrones, se encontró una correlación negativa entre la complejidad de la QS en ambos períodos, lo que sugiere que los cambios ocurridos durante la evolución natural podrían estar afectados por la respuesta inmune y determinar la evolución de la misma región bajo el tratamiento con NUCs, es decir, cuanto mayor es la complejidad durante la evolución natural, más homogénea es la población después del tratamiento. En la evolución natural, bajo la presión del sistema inmune, se desarrollaría una población compleja con muchas variantes, pero durante el tratamiento un número limitado de variantes con mutaciones de resistencia aparecería y predominaría, llevando a una QS menos compleja en un fenómeno tipo “cuello de botella”.

Figura 36: Patrón de la complejidad de la QS (Sn, Mf y Pi) en el período sin y con tratamiento en cada paciente. En el eje vertical aparecen las diferencias entre los parámetros de complejidad (Sn, Mf y Pi) durante el período de tiempo indicado en el eje horizontal (sin y con tratamiento). Las áreas grises indican los valores positivos e indican un aumento en la complejidad (flecha rosa) de la QS y la parte sin color indica un descenso de la complejidad (flecha amarilla).



3.2.5 Conclusiones

1. Este estudio muestra la utilidad de la tecnología UDPS para estudiar la evolución de la QS del VHB.
2. Se ha encontrado una mayor variabilidad de QS en pacientes HBeAg(-) que HBeAg(+), evidenciando que el aumento de la diversidad está asociado a la seroconversión.
3. La alta complejidad en la región preCore está asociada con una baja replicación viral, esto es lógico debido al papel clave de esta región en la misma, y sugiere una respuesta inmune aumentada en pacientes HBeAg(-), probablemente debido a la ausencia de actividad inmunomoduladora de HBeAg.
4. La selección positiva de variantes core en pacientes HBeAg(-) y HBeAg(+/-) sugiere un mecanismo potencial de escape del sistema inmune por cambios en la secuencia de la nucleocápside.
5. La correlación negativa de la evolución de la QS en TF y TNR sugiere la importancia de estudiar la complejidad de la QS antes de iniciar el tratamiento y usarla como un posible factor predictivo de la evolución del virus en el caso de no respondedores a NUCs.
6. Con la consolidación de los métodos de NGS se podría llevar a cabo el estudio de haplotipos de manera reproducible y así usar los parámetros de complejidad de QS para el manejo clínico de la infección por el VHB.

3.4 Manuscrito completo

OPEN ACCESS Freely available online

PLOS ONE



Clinical Application of Estimating Hepatitis B Virus Quasispecies Complexity by Massive Sequencing: Correlation between Natural Evolution and On-Treatment Evolution

Maria Homs^{1,2,3}, Andrea Caballero², Josep Gregori^{1,4}, David Tabernero^{1,2,3}, Josep Quer^{1,4}, Leonardo Nieto^{2,3}, Rafael Esteban^{1,5}, Maria Buti^{1,5}, Francisco Rodriguez-Frias^{1,2,3*}

1 Centro de investigación biomédica en red: enfermedades hepáticas y digestivas (CIBERehd), Instituto de Salud Carlos III, Barcelona, Spain, **2** Liver Pathology Unit, Departments of Biochemistry and Microbiology, Hospital Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain, **3** Virology Unit, Department of Microbiology, Hospital Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain, **4** Liver Diseases, Research Institute Hospital Vall d'Hebron, Barcelona, Spain, **5** Liver Unit, Department of Internal Medicine, Hospital Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain

Abstract

Aim: To evaluate HBV quasispecies (QA) complexity in the preCore/Core regions in relation to HBeAg status, and explore QA changes under natural evolution and nucleoside analogue (NUC) treatment.

Methods: Ultra-deep pyrosequencing of HBV preCore/Core regions in 30 sequential samples (baseline [diagnosis], treatment-free, and treatment-nonresponse) from 10 retrospectively selected patients grouped according to HBeAg status over time: HBeAg+ (N = 4), HBeAg- (N = 2), and fluctuating HBeAg (transient seroreversion/seroconversion pattern) (N = 4). QA complexity was defined by Shannon entropy, mutation frequency, nucleotide diversity, and mutation frequency of amino acids (MfAA) in preCore and Core.

Results: The QA was less complex in HBeAg+ than in HBeAg- or fluctuating HBeAg. High complexity in preCore was associated with decreased viral replication (preCore MfAA negatively correlated with HBV-DNA, $p = 0.005$). QA complexity in the treatment-free period negatively correlated with values seen during treatment. Specific variants were mainly selected in the Core region in HBeAg- and fluctuating HBeAg patients, suggesting higher immune pressure than in HBeAg+.

Conclusions: The negative correlation between QA natural evolution and on-treatment evolution indicates the importance of pre-treatment QA study to predict QA changes in NUC nonresponders. Study of QA complexity could be useful for managing HBV infection.

Citation: Homs M, Caballero A, Gregori J, Tabernero D, Quer J, et al. (2014) Clinical Application of Estimating Hepatitis B Virus Quasispecies Complexity by Massive Sequencing: Correlation between Natural Evolution and On-Treatment Evolution. PLoS ONE 9(11): e112306. doi:10.1371/journal.pone.0112306

Editor: Anna Kramvis, University of the Witwatersrand, South Africa

Received: April 2, 2014; **Accepted:** October 14, 2014; **Published:** November 13, 2014

Copyright: © 2014 Homs et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files. In addition, UDPS sequencing data are available from the GenBank SRA database (BioProject and BioSample accession numbers are provided within the manuscript).

Funding: This study was funded by Instituto de Salud Carlos III (grant PI12/01893), cofinanced by the European Regional Development Fund (ERDF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: frarodri@vhebron.net

Introduction

Hepatitis B virus (HBV, *Hepadnaviridae* family) causes acute and chronic infection in humans and chimpanzees. Despite the development of successful vaccination programs and effective antiviral therapies, there are more than 240 million carriers of HBV surface antigen (HBsAg) worldwide [1]. Around 150 million of these individuals have active infection and are at a high risk of progressing to cirrhosis or hepatocellular carcinoma [2].

The HBV genome (3.2-kb length) is a partially double-stranded DNA molecule, with four highly overlapping open reading frames: the polymerase, surface, Core, and X [3]. Although the HBV genome has a considerable degree of overlapping (67%) [4],

several factors contribute to its significant variability (1.36×10^{-3} – 6.62×10^{-4} substitutions/site/year) [5]. Some of these are the lack of proofreading of the viral polymerase, the high viral replication rates, and various host enzyme factors, such as guanine to adenine hypermutation activity of the APOBEC3 enzyme [6]. HBV variability yields a swarm of variants that are genetically closely related, but not identical. These evolve, show complex distributions in hosts, and are known as quasispecies (QA) [7]. The HBV QA rebalances its composition to fit to environmental conditions, including host immune stimulation and antiviral treatment [8].

QA study with accurate models and techniques is important to understand the adaptability, pathogenic power, and persistence of

HBV, and to optimize strategies to manage and prevent HBV infection [9]. In this sense, Shannon entropy is a useful parameter to study QA complexity in relation to patients' clinical evolution [10,11]. Various groups have used this approach to study the HBV QA by classic clonal methods, in order to determine the implications of QA complexity on the outcome of infection and the treatment response [12–14]. However, the small number of clones analyzed in these studies may significantly bias the Shannon entropy results obtained [15]. Fortunately, a large number of clonal sequences can now be obtained by next-generation technologies, particularly ultra-deep pyrosequencing (UDPS), thereby enabling accurate, quantitative QA description [15–17].

Current CHB treatment is mainly based on inhibition of viral polymerase activity by nucleos(t)ide analogues (NUCs), mainly entecavir and tenofovir, both of which have a high genetic barrier, an extremely low probability of resistant variant selection, and considerable suppression of viral replication [18]. However, this antiviral strategy does not affect the HBV intrahepatic reservoir, known as cccDNA. Hence, it is assumed that HBV infection cannot be cured, and the precise duration of treatment remains undefined [18]. Moreover, a significant percentage of patients who previously experienced viral breakthrough after lamivudine, telbivudine, or adefovir treatment are more prone to developing resistance [18].

Several biochemical and virological parameters should be monitored during natural disease evolution and particularly, during antiviral treatment [18]. One of these, HBeAg expression, is associated with a differing course of infection and with the probability of response to antiviral therapy. The presence of HBeAg in serum depends on variants located in the preCore region (main preCore mutation, position 1896) or in the basic core promoter (BCP, positions 1762–1764) [3]. In addition, the Core gene contains epitopic domains that play a central role in the immune response against the virus [19,20]. Therefore, the preCore/Core is an optimal region to investigate QA evolution in relation to host immune system stimulation. In addition, the preCore/Core regulates HBV replication and includes the only non-overlapping sequence in the HBV genome [3].

The evolution of the preCore/Core regions under NUC treatment has been little investigated. Studies involving molecular cloning of this region have reported that HBeAg seroconversion is associated with increased viral diversity [12,14]. There is also a recent study in which the preCore/Core was analyzed by UDPS [21] using new bioinformatic tools.

The aim of this study was to evaluate associations between HBeAg status and HBV QA complexity in the preCore/Core region, and to explore QA complexity under natural evolution and under NUC antiviral treatment. To this end, we UDPS-analyzed HBV variability and QA complexity in the preCore/Core region at baseline (at diagnosis of the infection), during a period before starting NUC treatment, and during a period of treatment nonresponse.

Materials and Methods

Patients and samples

This is a retrospective study approved by the Ethics Committee of Vall d'Hebron Research Institute, including sequential samples from 10 chronic hepatitis B patients, who gave written consent for participation. The 10 patients were older than 18 years, had detectable HBsAg for more than 6 months, and tested negative for HCV, HIV, and HDV. They were treatment-naïve before starting LVD and failed treatment within two years: 9 patients presented viral breakthrough and one was a primary nonresponder.

Furthermore, these patients were selected to include different HBeAg status over time: HBeAg-positive (N = 4), HBeAg-negative (N = 2), and fluctuating HBeAg (ie, a transient seroreversion/seroconversion pattern) (N = 4). Patients with fluctuating HBeAg included 2 patients with seroreversion (HBeAg-negative to positive) and 2 with different periods of seroreversion/seroconversion.

A total of 30 serum samples were UDPS-analyzed. Three samples per patient were analyzed at three time points when HBV DNA was higher than 5 log IU/mL: one at baseline, another after a treatment-free period, and another at treatment nonresponse (Table 1). The baseline (B) sample was the first sample, obtained at the time of the diagnosis. The sample corresponding to the treatment-free (TF) period was one taken after the baseline sample and before starting treatment (median 17 months). The sample at treatment nonresponse (TNR) was subsequent to the treatment-free sample and after treatment nonresponse, with viral breakthrough defined as a confirmed increase in HBV DNA of more than 1 decimal logarithm of IU/mL (log IU/mL) compared to the nadir HBV DNA level on therapy. Primary non-response was defined as less than 1 log IU/ml decrease in HBV DNA level from baseline at 3 months of therapy.

The patients' characteristics, the biochemical and virological results for each sample, and the rescue treatment prescribed in each case are presented in Table 1. HBV DNA was quantified with the COBAS Ampliprep/COBAS Taqman HBV Test (Roche Diagnostics). HBV serological markers were determined by commercial immunoassays.

Amplification and UDPS

The region analyzed by UDPS, clustered positions 1757 to 2152 [16], covered the complete preCore (1814–1900) and the first 84 codons of the Core gene (1901–2152), which include the two main immunodominant epitopes flanked by amino acids 50 to 69 (Ih 50–69) and by amino acids 74 to 84 (B74–84) [22].

Briefly, HBV DNA was extracted from 200 μ L of serum with the QIAamp DNA MiniKit (QIAGEN, Hilden, Germany). The amplicon library corresponding to the HBV preCore/Core region was obtained after two PCR runs (nested). The first PCR primers were as follows: sense (position 1721–43) 5' GTTTAAA/GGACTGGGAGGAGC/TTGG 3' and antisense (position 2804–23) 5' TGTTCCTCAA/GGAATAA/TGGTGA 3'. The nested PCR primers included the recognition site for UDPS, shown in italics. The sequence of the sense primer (position 1737–56) was 5' CGTATCGCCTCCCTCGCGCCATCAGGAGC/TTGGGGGAGGAGAC/TTAG 3' and the antisense primer (position 2153–80) was 5' CTATGCGCCTTGCCAGCCCGCT-CAGCCATA/GTTAGTA/GTTA/GACATAAC/TTA/C/GA-CTAC 3'. To minimize the error rate of the PCR process, high fidelity polymerase (Pfu Ultra-II, Stratagene, La Jolla, USA) was used. The nested PCR products had a length of 494 bp and were isolated from 0.9% agarose gel with the QIAquick extraction Kit (QIAquick Spin Handbook, QIAGEN, Hilden, Germany) and quantified using Quan-IT Picogreen dsDNA reagent (Invitrogen).

Before the UDPS sequencing reaction, each amplicon was pooled to obtain a concentration of 4×10^6 molecules of the HBV region. This working solution was enriched with the capture beads needed for sequencing. After optimal enrichment, clonal amplification in beads was done in forward and reverse directions (emPCR kits II and III, 454 Life Sciences). UDPS was performed using the Genome Sequencer FLX system (454 Life Sciences).

Table 1. Description of the patients and samples included in the study.

Pt	Sex	Year birth	Baseline sample				Treatment-free sample				Treatment-nonresponse sample				Rescue Treatment
			HBV gen	HBV DNA (logIU/mL)	ALT (IU/L)	HBeAg	TF months	HBV DNA (logIU/mL)	ALT (IU/L)	HBeAg	Monthson LVD	HBV DNA (logIU/mL)	ALT (IU/L)	HBeAg	
1	M	1958	A	6.1	92	+	13	6.4	75	+	15	>8	84	+	TDF
2	M	1990	D	>8	64	+	18	>8	147	+	13	>8	280	+	TDF+ETV
3	F	1981	D	>8	67	+	11	5.8	60	+	15	5.5	43	+	TDF
4	M	1947	D	7.6	21	+	17	7.2	178	+	48	>8	74	+	TDF
5	M	1943	D	5.5	65	-	15	7.3	117	-	13	5.1	106	-	LVD+TDF
6	M	1980	D	6.2	124	-	32	5.2	200	-	12	>8	210	-	TDF
7*	M	1947	A	7.5	452	+	16	7.5	206	+	17	5.6	125	+	ADV
8	M	1950	D	6	40	+	30	6.6	140	-	13	5.4	43	-	ETV+TDF
9	M	1978	A	>8	83	+	24	5	33	-	33	5.2	18	+	TDF
10	M	1962	A	>8	40	-	18	>8	95	+	15	6.5	65	+	LT

ADV, adefovir; HBV gen, HBV genotype; ETV, entecavir; log IU/mL, decimal logarithm of IU/mL; LVD, lamivudine; TDF, tenofovir; TF, treatment-free; LT, liver transplantation.

*Patient 7 experienced transient HBeAg seroconversion while on treatment, but HBeAg seroreversion occurred before the treatment-nonresponse sample was taken. Seroreversion coincided with virological breakthrough, as is depicted in Figure 1.

doi:10.1371/journal.pone.0112306.t001

UDPS data analysis

FLX 454 data processing was carried out on the open source R environment, using the Biostrings library for pattern matching and sequence alignment, and in-house R scripts [15,16]. Sequences considered of low quality and those seen in either sense or antisense strands alone were filtered out. Haplotypes found in greater than 0.1% abundance and common to both sense and antisense strands were considered forward and reverse consensus haplotypes (FRCH). Because coverage was not below 10,000 reads, FRCH with a population abundance above 0.25% were considered error-free haplotypes [16]. The HBV genotype of each haplotype was obtained by phylogenetic analysis, using GenBank reference sequences (Table 2).

Parameters for QA complexity

The HBV QA was evaluated including all haplotypes accepted after filter analysis. QA complexity was quantified by three parameters [16]: normalized Shannon entropy (Sn), mutation frequency (Mf), and nucleotide diversity (Pi). Sn measures haplotype diversity attending to the number of haplotypes and their frequency, Mf measures the genetic diversity with respect to the most prevalent haplotype, and Pi measures the population genetic diversity as the average number of mutations per site between each pair of haplotypes in the viral population [15].

In addition, QA complexity was evaluated in terms of the mutation frequency of amino acids in the regions studied. This parameter was calculated separately for the preCore region (48 amino acids) and Core region (84 amino acids).

Study of QA evolution

To examine QA evolution, two periods were defined: a treatment-free period (median 17 months), and a period under treatment pressure (median 15 months). The differences found for each parameter (Sn, Mf, and Pi) between the B and TF samples were considered measures of natural evolution. Differences between the TF and TNR samples were considered measures of evolution occurring under treatment pressure.

Statistical analysis

All data are presented as the median and range. Statistical analyses were carried out using IBM SPSS 20 (SPSS Inc., Chicago, USA). Correlation analyses were performed with Spearman's rho test. The Mann-Whitney *U* test and Kruskal-Wallis test

were used for comparisons of two independent variables. Significance was set at $p \leq 0.05$.

In a previous study we demonstrated the applicability of the *t* or *Z* test in statistical inference with indices of viral quasispecies diversity [15]. In the present study, we used the more stringent Spearman rho correlation test for this purpose, in keeping with the analyses used in recent studies on QA evolution under antiviral treatment and the treatment response [12,13].

Results

Thirty samples from 10 patients were analyzed (3 samples/patient at 3 time points: B, TF, and TNR). After applying the quality filters (FRCH above 0.25%) 986,014 sequences were analyzed, and high coverage per patient was obtained (median 98,494 sequences per patient, range 59,039–136,660). UDPS sequencing data have been submitted to the GenBank SRA database (BioProject accession number PRJNA260562, BioSample accession numbers in Table 3). HBV DNA, alanine aminotransferase (ALT) levels, and HBeAg status of each sample are presented in Table 1. All changes detected in the 30 samples are presented in Table S1 and the dominant haplotypes of all samples are aligned in File S1. HBV genotype of the 10 patients, which is also indicated in Table 1, showed no changes over the sequential study.

Overall, QA complexity in all samples yielded the following median values: Sn 0.366 (range, 0.026–0.738), Mf 7.7×10^{-4} (range, 0.7×10^{-5} – 1.7×10^{-2}), and Pi 1.43×10^{-3} (range, 1.3×10^{-5} – 2.8×10^{-2}). Significant correlations were found between the three variables used to study HBV QA complexity ($R_{s_{Sn-Mf}}$ 0.914, $p < 0.001$; $R_{s_{Sn-Pi}}$ 0.910, $p < 0.001$; and $R_{s_{Mf-Pi}}$ 0.998, $p < 0.001$), indicating that the three approaches provided very similar estimations. In the overall population, median preCore mutation frequency of amino acids (MfAA) was 2.827 (range, 0–284.35) and median Core MfAA was 6.682 (0–399.92).

QA complexity and its relationship with sample time points, ALT levels, HBV genotype, HBV DNA quantification, and HBeAg status

The QA distribution, HBV DNA level, and HBeAg status of each patient are depicted in Figure 1. To simplify presentation of the results, the QA is represented by Sn. In each patient, the distribution of Mf and Pi was the same as that of Sn, but at a smaller scale (data not shown), as all parameters significantly correlated. Analysis of differences in QA complexity between B,

Table 2. NCBI GenBank accession numbers of the reference sequences used in HBV genotyping.

HBV Genotype A	HBV Genotype D
AJ309369	U95551
AJ309370	M32138
AJ344115	X65257
AM282986	X59795
S50225	X02496
V00866	X97848
X02763	X72702
X70185	V01460
Z72479	HE815465
JX507080	

doi:10.1371/journal.pone.0112306.t002

Table 3. Biosample accession numbers for each sample analyzed.

Patient	Sample	Biosample accession number
1	Baseline	SAMN03023656
1	Treatment-free	SAMN03023657
1	Treatment nonresponse	SAMN03023658
2	Baseline	SAMN03023659
2	Treatment-free	SAMN03023660
2	Treatment nonresponse	SAMN03023661
3	Baseline	SAMN03023662
3	Treatment-free	SAMN03023663
3	Treatment nonresponse	SAMN03023664
4	Baseline	SAMN03023665
4	Treatment-free	SAMN03023666
4	Treatment nonresponse	SAMN03023667
5	Baseline	SAMN03023668
5	Treatment-free	SAMN03023669
5	Treatment nonresponse	SAMN03023670
6	Baseline	SAMN03023671
6	Treatment-free	SAMN03023672
6	Treatment nonresponse	SAMN03023673
7	Baseline	SAMN03023674
7	Treatment-free	SAMN03023675
7	Treatment nonresponse	SAMN03023676
8	Baseline	SAMN03023677
8	Treatment-free	SAMN03023678
8	Treatment nonresponse	SAMN03023679
9	Baseline	SAMN03023680
9	Treatment-free	SAMN03023681
9	Treatment nonresponse	SAMN03023682
10	Baseline	SAMN03023683
10	Treatment-free	SAMN03023684
10	Treatment nonresponse	SAMN03023685

doi:10.1371/journal.pone.0112306.t003

TF, and TNR samples showed significantly higher Sn (and greater diversity) in B (0.45, range 0.0367–0.607) compared to TNR (0.21, range 0.026–0.478) ($p = 0.035$), nonsignificantly higher values in TF (0.42, range, 0.045–0.738) relative to TNR (0.21, range 0.026–0.478) ($p = 0.075$), and similar values between B and TF samples.

There were no significant differences in any complexity parameter (Sn, Pi or Mi) between genotypes A and D, nor were there significant correlations between these parameters and HBV DNA or ALT levels. On separate analysis of the preCore and Core regions, preCore MfAA showed a significant negative correlation with HBV DNA level (R_s preCoreMfAA-DNA -0.5 , $p = 0.005$), whereas MfAA values in Core increased as HBV DNA decreased, but a significant correlation was not found. Keeping in mind the small sample size, these findings suggest that increased preCore complexity is associated with decreased HBV viral replication. To avoid possible effects of treatment, only B and TF samples were analyzed. No significant differences in QA complexity parameters were observed between HBV genotypes, and complexity did not correlate with HBV DNA or ALT levels.

On assessment of QA complexity and HBeAg status in the 30 samples, we found significantly lower QA complexity in HBeAg-

positive patients than in HBeAg-negative ones, as determined by Sn, Mi, and Pi, and by MfAA in preCore and Core (Table 4). To explore differences in QA complexity according to HBeAg positive or negative status, and to avoid possible effects of treatment on HBeAg, only B and TF samples were analyzed. As was seen in the total of samples, Sn and Pi results showed greater complexity in HBeAg-negative than positive samples (Sn, $p = 0.03$; Pi, $p = 0.037$, Table 4). preCore- and Core-MfAA did not significantly differ between HBeAg-positive and negative samples.

QA complexity and HBeAg status: longitudinal study

To explore possible associations between HBeAg status and QA variability over time, patients were divided into three groups: the HBeAg+ group included patients with persistently positive HBeAg (cases 1–4); the HBeAg- group, patients with persistently negative HBeAg (cases 5 and 6), and HBeAg+/-, patients with fluctuating HBeAg (cases 7–10) (Figure 1). Sn, Mi, and Pi results for each sample are presented in Table S2. QA complexity was significantly lower in HBeAg+ than HBeAg- patients (Sn $p = 0.001$, Mi $p = 0.002$, and Pi $p = 0.002$) or the HBeAg+/- group (Sn $p = 0.014$, Mi $p = 0.004$, and Pi $p = 0.003$). There were no

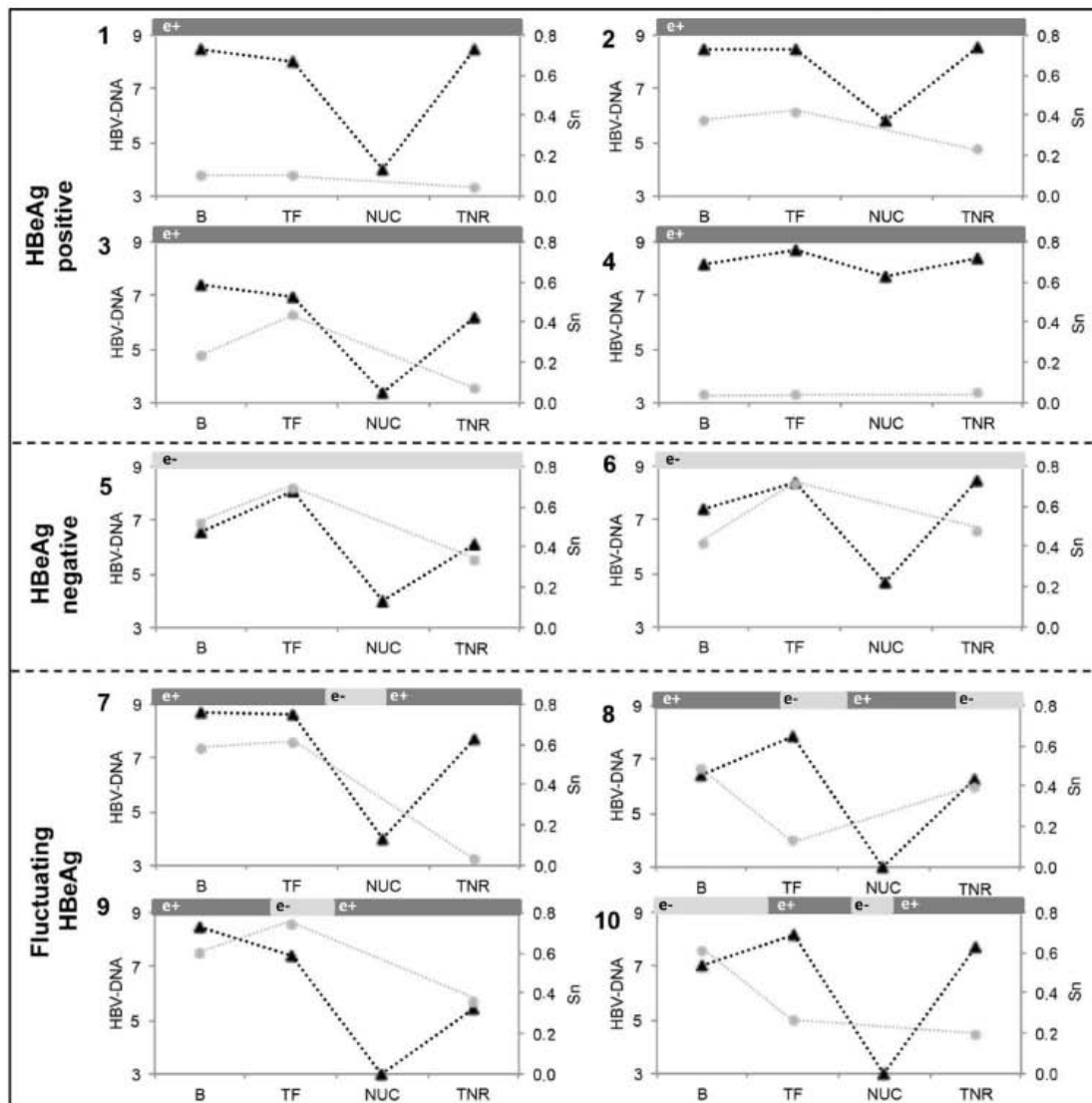


Figure 1. Viral replication and QA evolution of the 10 patients included in the study. The left axis and black-dotted lines represent HBV-DNA (log IU/mL), and the right axis and grey lines represent S_n in baseline (B), treatment-free (TF), and treatment-nonresponse (TNR) samples. HBV DNA level during LVD treatment is indicated as NUC. HBeAg status is shown in bars over the graph: dark grey bar, HBeAg-positive and light grey bar, HBeAg-negative.
doi:10.1371/journal.pone.0112306.g001

significant differences between HBeAg- and HBeAg+/- patients. In an attempt to define whether the variability was located in the preCore or Core region, MfAA was calculated for these regions separately in each HBeAg status group (Figure 2a). preCore and Core MfAA were both significantly lower in HBeAg+ than in HBeAg- (preCore MfAA, $p=0.001$ and Core MfAA, $p=0.005$) or HBeAg+/- (preCore MfAA, $p=0.045$ and Core MfAA, $p=0.001$) patients, and there were no differences between the HBeAg- and HBeAg+/- groups.

To explore possible differences in QA complexity between the sampling time points, parameters were calculated according to HBeAg status over time and the sample analyzed (B, TF, and TNR). QA complexity was systematically higher in HBeAg- and HBeAg+/- than in HBeAg+ at all three time points (Table 5). However, statistically significant differences in S_n , Mf, and P were only found at baseline between HBeAg+ and HBeAg+/- patients ($p=0.029$). preCore and Core MfAA results at the three time points are depicted in Figure 2b. Interestingly, both parameters indicated significantly greater complexity in HBeAg+/- cases

Table 4. HBV QA complexity attending to HBeAg status in the 30 samples analyzed and in the 20 without treatment pressure.

QA parameter	All samples N = 30			Samples without treatment N = 20		
	HBeAg (+)	HBeAg (-)	P	HBeAg (+)	HBeAg (-)	P
Sn	0.23 (0.03–0.61)	0.49 (0.13–0.74)	0.003	0.35 (0–1.7 × 10 ⁻²)	0.42 (0.04–0.74)	0.03
Mf	2.8 × 10 ⁻⁴ (0–1.1 × 10 ⁻²)	2.8 × 10 ⁻³ (5.4 × 10 ⁻⁴ –1.7 × 10 ⁻²)	0.003	7.4 × 10 ⁻⁴ (0–1.2 × 10 ⁻³)	1.35 × 10 ⁻³ (0–1.7 × 10 ⁻²)	NS
Pi	5.3 × 10 ⁻⁴ (0–1.4 × 10 ⁻²)	4.0 × 10 ⁻³ (10 ⁻² –2.8 × 10 ²)	0.002	1.3 × 10 ⁻⁴ (0–1.7 × 10 ⁻⁴)	2.37 × 10 ⁻³ (0–2.8 × 10 ⁻²)	0.037
pcMfAA	1.41 (0–192.06)	44 (0–284)	0.024	1.67 (0–192.1)	51.75 (0–197)	NS
CoreMfAA	1.35 (0–399.94)	45.74 (0.95–207.2)	0.011	1.32 (0.34–399.9)	24.15 (0.95–164.42)	NS

doi:10.1371/journal.pone.0112306.t004

than HBeAg+ ones in B and TF samples (p = 0.029). With regard to the lack of significance in HBeAg- cases, it should be remembered that only two patients always tested HBeAg-negative.

Distribution of preCore and Core changes

In 5 patients (cases 2, 3, 5, 7, and 10), the nucleotide changes encountered were mainly located in the Core region, 4 patients (cases 1, 4, 8, and 9) showed changes in both the preCore and Core, and in the last patient (case 6), changes were mainly located in the preCore region. No specific patterns of variability were observed at the different time points (B, TF, and TNR).

The detection of preCore variants was specific for each patient and each sample (Table S1), but we have summarized the main mutations responsible for HBeAg expression (BCP and preCore mutations) (Table 6). These particular mutations were more frequent in the HBeAg+/- group (patients 7, 8, 9 and 10) than in HBeAg+ or HBeAg- patients.

Five cases (patients 1, 2, 4, 5, and 7; Table S1) showed a slight tendency to accumulate variability in Core epitopic regions (Th50–69, nucleotide positions 2048–2107, and B74–84, positions 2120–2152), in agreement with the role of Core as an immunostimulating region. However, positive selection of Core variants in the HBV QA, defined as selection of a new master sequence that differed from the master sequence in the previous sample, was more frequent in HBeAg- (2 of 2) and HBeAg+/- (4 of 4) patients than in HBeAg+ patients (1 of 4) (Table 7), a finding that may suggest enhancement of the host immune response due to the lack of HBeAg.

Evolution of QA complexity without treatment and under NUC pressure

QA evolution in each of the 10 patients in the TF and TNR periods is shown in Figure 3. In most cases (7/10), there was an increase in QA complexity during natural evolution (grey areas in Figure 3) and a decrease in QA complexity under treatment. One case (patient 4) showed no changes in QA complexity between TF and TNR using any of the three parameters, and two patients, both HBeAg+/- (cases 8 and 10), showed an inverse behavior: a decrease in QA complexity without treatment and an increase during treatment (black lines, Figure 3).

Despite the different patterns observed among the 10 patients, QA complexity at TF showed a significant negative correlation with the complexity at TNR regarding Sn (Rs -0.661, p = 0.038), Mf (Rs -0.891, p = 0.01) and Pi (Rs -0.903, p < 0.001), suggesting that the evolution of preCore/Core QA complexity during the time without treatment determines the evolution of these regions during treatment.

Discussion

The HBV QA composition and the changes that occur over QA evolution are important factors related to controlling and treating chronic HBV infection. Hence, acquiring accurate knowledge of HBV QA complexity is currently a major challenge for managing chronic hepatitis B patients. Recent reports support the concept that QA complexity is a clinically relevant factor in the course and prognosis of this disease and in the response to treatment [12–14,23]. In this line, HBV QA complexity has been associated with the antiviral response in ETV-treated patients, in whom lower complexity was seen in responders than in partial responders [13]. Cheng et al. [12,14] reported higher viral diversity in HBeAg natural or treatment-induced (IFN or LVD) seroconverters than in non-seroconverters, thereby providing evidence that increased viral diversity is associated with HBeAg seroconversion, in

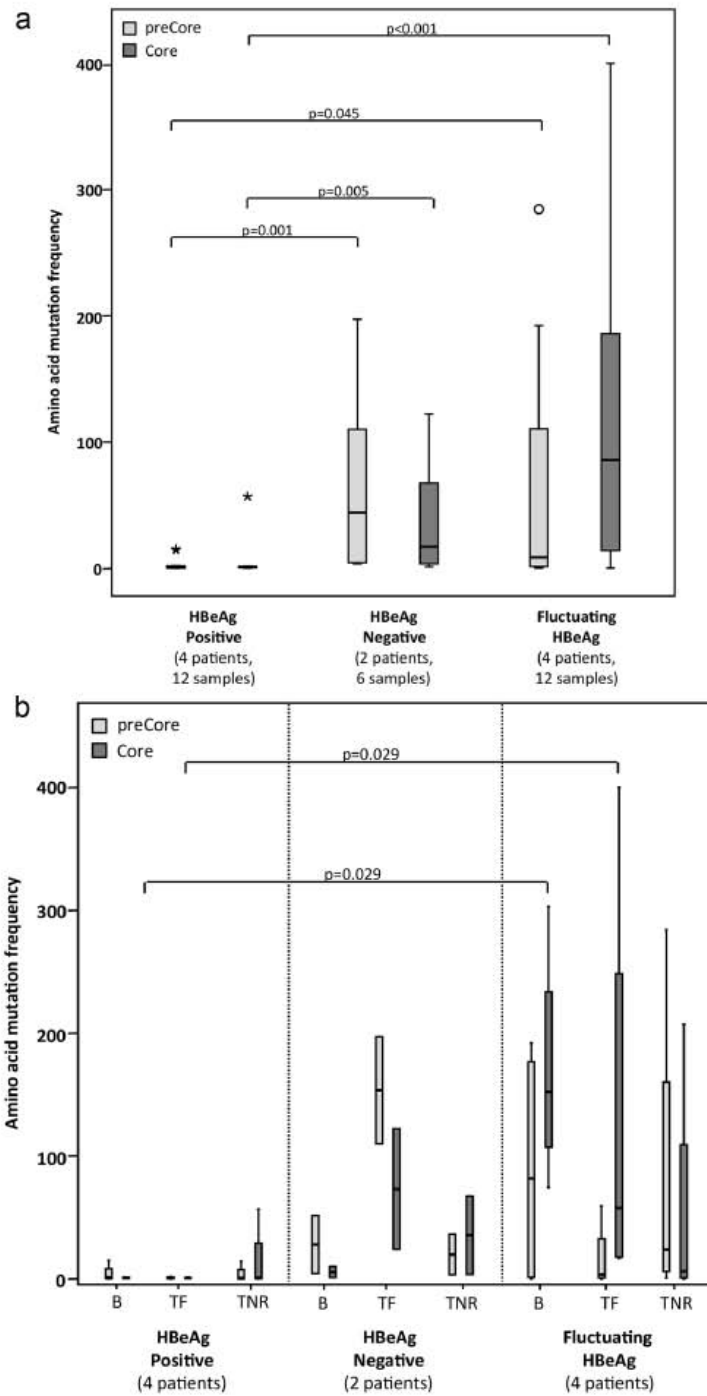


Figure 2. Distribution of mutated amino acids in the preCore and Core regions (a) in patients grouped by HBeAg evolution and (b) splitting into HBeAg evolution and the three time points: baseline (B), treatment-free (TF) and treatment-nonresponse (TNR).
 doi:10.1371/journal.pone.0112306.g002

Table 5. Shannon entropy (Sn), mutation frequency (Mf), and nucleotide diversity (Pi) at three time points in patients grouped by HBeAg evolution.

	HBeAg-Positive (N = 4)			HBeAg-Negative (N = 2)			Fluctuating HBeAg (N = 4)		
	Baseline*	TF	TNR	Baseline	TF	TNR	Baseline*	TF	TNR
Sn	Median 0.1664	0.2597	0.06165	0.4695	0.69735	0.40545	0.58545	0.43615	0.2778
	Range 0.037–0.374	0.045–0.431	0.044–0.230	0.420–0.519	0.685–0.710	0.333–0.478	0.48–0.61	0.128–0.74	0.026–0.39
Mf	Median 1.95×10^{-4}	4.11×10^{-4}	3.70×10^{-5}	1.3×10^{-3}	4.63×10^{-3}	1.44×10^{-3}	5.61×10^{-3}	1.49×10^{-3}	5.16×10^{-4}
	Range 0.15×10^{-4} – 1.0×10^{-4}	0.19 – 22.3×10^{-4}	1.9 – 26×10^{-5}	1.27 – 1.35×10^{-3}	3.22 – 6.05×10^{-3}	0.55 – 2.33×10^{-3}	1.65 – 11.7×10^{-3}	0.68 – 10×10^{-3}	0.06 – 17.0×10^{-4}
Pi	Median 3.81×10^{-4}	6.24E-04	7.35×10^{-5}	1.31×10^{-3}	6.55×10^{-3}	2.23×10^{-3}	7×10^{-3}	2.37×10^{-3}	9.96×10^{-4}
	Range 0.3 – 10×10^{-4}	0.38 – 35×10^{-4}	3.7 – 4.8×10^{-4}	1.27 – 2.43×10^{-3}	4.74 – 8.36×10^{-3}	1.05 – 3.42×10^{-3}	2.37 – 14.7×10^{-3}	1.33 – 17×10^{-3}	0.1 – 282×10^{-4}

*Significant differences (p=0.029) between HBeAg-positive and fluctuating HBeAg patients in baseline samples.
TF, treatment-free; TNR, treatment-nonresponse.
doi:10.1371/journal.pone.0112306.t005

agreement with our observations in the present study. Although these studies [12,14] have provided valuable findings, the techniques used analyzed only small numbers of clones, and the results may not be representative of the overall viral population, which contains billions of particles. In contrast, next-generation sequencing methods, particularly UDPS used in the present study, enable clonal analysis of thousands of sequences in a single sample [8,24–29], provide a number of clonal sequences to lend reliability to calculation of QA complexity parameters [15], and have robust quantitative and reproducibility values, making them viable alternatives to molecular cloning for QA study [16].

In this study, we applied UDPS to determine QA complexity in the HBV preCore and Core regions. More than 19,500 sequences per sample were analyzed, a number considerably higher than the numbers in recent reports on the HBV QA (mean 20–26 clones/sample) [12–14,23] and clearly sufficient to guarantee QA calculations [15]. To determine HBV QA complexity, three parameters were used: Shannon entropy (Sn), mutation frequency (Mf), and nucleotide diversity (Pi). The high correlation between the results obtained with these parameters indicates that Sn, Mf and Pi equally represented HBV QA complexity. In addition, MFAA was used to independently explore variability at the amino acid level in both the preCore and Core regions.

We found significant differences in the QA complexity parameters according to HBeAg status, with greater complexity in HBeAg- than HBeAg+ samples, in agreement with previous studies [12,14]. However, the higher complexity (Sn, Mf and Pi) seen in HBeAg- cases lacked significance when analyzing QA complexity attending to HBeAg evolution, likely because of the small number of HBeAg-negative patients. Interestingly, both the HBeAg- and HBeAg+/- groups showed significantly higher QA complexity than HBeAg+ patients when complexity was analyzed at the amino acid level (MFAA) in the preCore and Core regions separately. Therefore, our data provide evidence that increased viral diversity is associated with HBeAg seroconversion and strongly suggest significant evolutionary enhancement that was even more evident in fluctuating HBeAg status. These findings may indicate an increase in evolutionary pressure due to a more intense immune response in HBeAg-negative status, likely associated with the lack of HBeAg and its immunomodulatory effect [30].

Overall, there were no significant differences in QA complexity between A and D genotypes in the 30 samples analyzed, despite the constraints on main preCore mutation selection and HBV genotype [31]. Moreover, no correlation was observed between ALT levels and QA variability. Although it is assumed that ALT status provides an estimate of the strength of the immunological response against viral infection [32], ALT can be influenced by many factors and a single point measurement may not be indicative of the long-term immune status of a host [32]. Furthermore, the aim of this study was to sequentially analyze HBV QA complexity with deep clonal sequence coverage to guarantee the complexity calculations, and because of the huge amount of data involved, only ten patients were included.

The negative correlation between HBV DNA levels and QA complexity in the present study agrees with recent findings [12,14]. Although the correlations did not achieve statistical significance, the three main QA complexity parameters (Sn, Mf and Pi) showed that the higher the HBV DNA level, the lower was QA complexity in preCore/Core. However, it should be remembered that all samples included had a high viral load (HBV DNA >5 log IU/mL). This potentially confounding factor may have contributed to the absence of significance in the correlation studies. Nonetheless, in the separate analysis of the

Table 6. Percentage of main preCore variants in the samples analyzed.

Pt	Sample	HBeAg status	preCore variants	
			BCP	main preCore variants
1	B	+	wt	wt
	TF	+	wt	wt
	TNR	+	wt	wt
2	B	+	wt	wt
	TF	+	wt	wt
	TNR	+	wt	wt
3	B	+	A1762T/G1764A (0.27%)	wt
	TF	+	wt	wt
	TNR	+	wt	wt
4	B	+	wt	wt
	TF	+	wt	wt
	TNR	+	wt	wt
5	B	-	A1762T/G1764A (100%)	G1896A/G1899A (100%)
	TF	-	A1762T/G1764A (100%)	G1896A/G1899A (100%)
	TNR	-	A1762T/G1764A (100%)	G1896A/G1899A (100%)
6	B	-	A1762T/G1764A (100%)	G1896A (89.1%)/G1899A (10.92%)
	TF	-	A1762T/G1764A (100%)	G1896A (73.1%)/G1899A (26.9%)
	TNR	-	A1762T/G1764A (100%)	G1896A (100%)
7	B	+	A1762T/G1764A (99.44%)	wt
	TF	+	A1762T/G1764A (100%)	wt
	TNR	+	A1762T/G1764A (100%)	wt
8	B	+	A1762T/G1764A (100%)	G1896A (38.23%)/G1899A (58.83%)
	TF	-	A1762T/G1764A (100%)	G1896A (98%)/G1899A (0.77%)
	TNR	-	A1762T/G1764A (80.41%)	G1896A (80.41%)
9	B	+	A1762T/G1764A (100%)	G1899A (0.34%)
	TF	-	A1762T/G1764A (100%)	G1896A (11.64%)/G1899A (0.49%)
	TNR	+	A1762T/G1764A (100%)	wt
10	B	-	A1762T/G1764A (100%)	wt
	TF	+	A1762T/G1764A (100%)	wt
	TNR	+	A1762T/G1764A (100%)	wt

BCP, Basal core promoter; wt, wild-type: samples that showed 100% of wild-type main BCP or preCore variants.
doi:10.1371/journal.pone.0112306.t006

Table 7. Patients with Core positive selection. Cases in which the new variant was detected in previous samples as minor variants are also indicated.

Pt	Sample	Core positive selection	Minor variant	Sample
3	TNR	V63, 100%	-	-
5	TNR	E83, 100%	D83E	TF (13.97%)
6	TF*	S12, 73.1% - Q79, 60.52%	-	-
7	TF	S79, 54.15%	P79S	B (30.63%)
	TNR	G63-M66-P79, 100%	V63M-T66M-S79P	B (0.32%) and TF (0.47%)
8	TF*	Q14-A35-H57-V58-T59-D64-N67, 98%	-	-
9	TF	H5-I60, 88.36%	P5H-L60I	B (47.39%)
	TNR	P5-L60, 98.09%	H5P-I60L	TF (11.64%)
10	TF*	E40, 98.03% -S41, 92.29%	D40E-A41S	B (0.32)

*Detected in the treatment-free (TF) and maintained in the treatment-nonresponse (TNR) sample.
doi:10.1371/journal.pone.0112306.t007

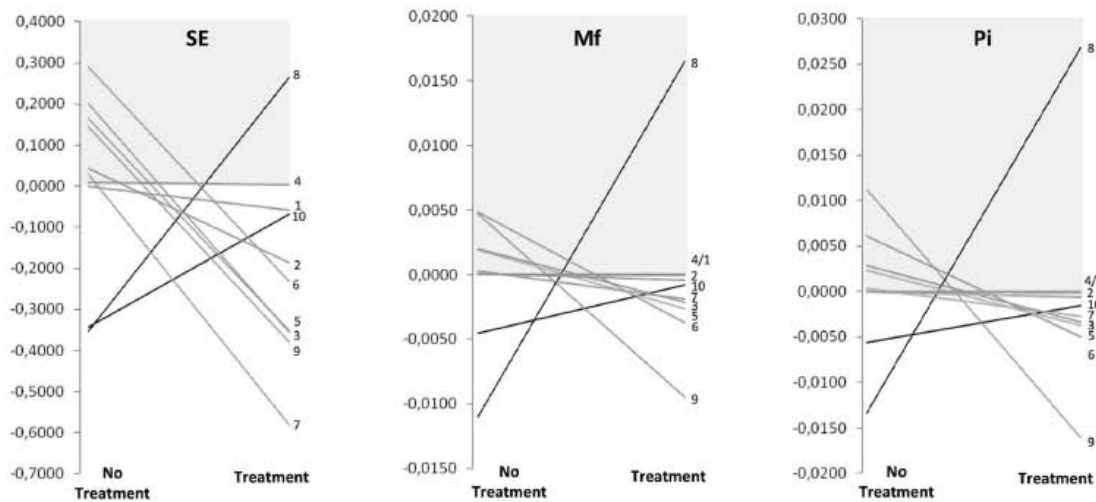


Figure 3. Patterns of QA complexity in the period without and with treatment, according to Sn, Mf, and Pi. The vertical axis represents differences between the complexity parameter (Sn, Mf or Pi) during the time period indicated in horizontal axis, No Treatment (TF-B) and Treatment (TNR-TF). The grey areas highlight positive values and represent an increase in HBV QA complexity. doi:10.1371/journal.pone.0112306.g003

preCore and Core regions at the amino acid level (MfAA calculation), the significant negative correlation with HBV DNA observed in preCore MfAA suggests that the preCore mutated variants could confer a decrease in preCore fitness to regulate HBV replication. The preCore contains the essential encapsidation signal for viral replication, and substitutions in this region may interfere with its functionality [3].

The Sn in preCore/Core was significantly higher at B than at TNR, and was non-significantly higher in TF than at TNR. These findings suggest a decrease in QA evolution, associated with LVD treatment failure. Boni et al [33] reported that the immune response is enhanced when antiviral treatment controls HBV replication, but our patients were treatment nonresponders and therefore in a situation contrary to that of Boni's population [33]. In fact, most of our cases showed a decrease in QA complexity at TNR, which likely reflects attenuation of immune system activity at nonresponse.

The possible relationship between QA complexity and antiviral therapy response was not tested in this study because responders, who have undetectable HBV-DNA levels after treatment, were not included. However, longitudinal UDPS study of our patients enabled examination of evolutionary patterns in the absence and presence of antiviral treatment, and this yielded conclusive results according to HBeAg status. Some patterns were identified: at baseline, QA complexity in HBeAg-positive patients was significantly lower than in those with fluctuating HBeAg (Table 5). In addition, HBeAg-positive patients showed lower Core gene variability (Core MfAA) than those with fluctuating HBeAg at baseline and TF. In contrast, preCore MfAA was not significantly different between any of the groups, which could indicate that the host immune response mainly acts against Core epitopes in patients with fluctuating HBeAg. In fact, preCore variants and positive selection of Core variants were common in HBeAg-negative and fluctuating patients, in keeping with results from our previous studies [8,22]. These findings may result from HBV adaptation under host immune pressure, or even be due to an effect of antiviral treatment on the Core gene.

HBV QA evolution was analyzed in two periods, natural evolution and under NUC pressure. The significant negative correlation in QA complexity between the two periods suggests that changes occurring in natural evolution might be affected by the host immune response and determine the evolution of the same region under NUC treatment. We found that in most patients, the greater the complexity during natural evolution, the more homogeneous was the population after treatment, indicating that NUCs might also have some indirect effects on the preCore/Core region. However, this pattern was not observed in 2 of the 10 patients studied, probably due to the transient seroconversion/seroreversion status of these cases (fluctuating HBeAg). Apart from these differences, the main inverse pattern of QA evolution in the two periods may indicate the following: under immune pressure, a complex population evolves, but during antiviral therapy pressure, a limited number of HBV variants carrying resistant mutations emerge and predominate, rendering the QA less complex in a type of bottleneck phenomenon.

The main limitation of this study is the small sample of 10 patients sequentially analyzed, which was related to our aim to characterize the HBV QA as accurately as possible by very high coverage (median, 98,494 sequences per patient). All cases were longitudinally studied, and 3 samples per patient were included to enable examination of evolutionary patterns in the absence and presence of antiviral treatment. Thus, sample size was mainly determined by the huge amount of data to process and the cost of UDPS.

In conclusion, the results of this study provide further evidence of the utility of UDPS for investigating the evolution of the HBV QA. In addition, they provide confirmatory data for previous findings in studies with lower analytical coverage indicating greater QA variability in HBeAg-negative than HBeAg-positive patients. Our results show that high complexity in the preCore region is associated with low viral replication, in keeping with the key role of this region in HBV replication, and suggest an enhanced immune response in HBeAg-negative patients, probably related to the lack of HBeAg immunomodulatory activity. In the

same direction, the positive selection of Core variants in HBeAg-negative and fluctuating status can be understood as a potential mechanism to escape the host immune system by nucleocapsid sequence changes. Finally, the strong negative correlation of QA evolution in the treatment-free period and under treatment shows the importance of studying the QA before treating patients, as a potential predictive factor of HBV evolution in cases of NUC nonresponse. With the consolidation of next-generation sequencing methods that enable the reproduction of viral haplotype study, QA complexity parameters could be useful for clinical management of HBV infection.

Supporting Information

Table S1 Percentages of changes observed in baseline (B), treatment-free (TF) and treatment non-response (TNR) sample of each patient. Nucleotide changes are registered in relation to the dominant haplotype of each sample. (PDF)

References

1. WHO Hepatitis B (2014) Available at <http://www.who.int/mediacentre/factsheets/fs204/en/>. Accessed 29 August 2014.
2. Ilcoje UH, Yang H-L, Su J, Jen C-L, You S-L, et al. (2006) Predicting cirrhosis risk based on the level of circulating hepatitis B viral load. *Gastroenterology* 130: 678–686. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16530509>. Accessed 29 August 2014.
3. Rodriguez-Frias F, Buti M, Tabernero D, Homs M (2013) Quasispecies structure, cornerstone of hepatitis B virus infection: mass sequencing approach. *World J Gastroenterol* 19: 6995–7023. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3819535&tool=pmcentrez&rendertype=abstract>. Accessed 3 April 2014.
4. Kidd-Ljunggren K, Miyakawa Y, Kidd AH (2002) Genetic variability in hepatitis B viruses. *J Gen Virol* 83: 1267–1280. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12029141>.
5. Tedder RS, Bissett SL, Myers R, Ijaz S (2013) The “Red Queen” dilemma—running to stay in the same place: reflections on the evolutionary vector of HBV in humans. *Antivir Ther* 18: 489–496. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23792884>. Accessed 29 August 2014.
6. Noguchi C, Ishino H, Tsuge M, Fujimoto Y, Imamura M, et al. (2005) G to A hypermutation of hepatitis B virus. *Hepatology* 41: 626–633. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15726649>. Accessed 18 December 2012.
7. Domingo E, Holland JJ (1997) RNA virus mutations and fitness for survival. *Annu Rev Microbiol* 51: 151–178. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9343347>.
8. Homs M, Buti M, Tabernero D, Quer J, Sanchez A, et al. (2012) Quasispecies dynamics in main core epitopes of hepatitis B virus by ultra-deep-pyrosequencing. *World J Gastroenterol* 18: 6096–6105. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3496886&tool=pmcentrez&rendertype=abstract>. Accessed 29 August 2014.
9. Lam SG, Cheng Y, Guindon S, Sect BL, Lee LY, et al. (2007) Viral quasi-species evolution during hepatitis B antigen seroconversion. *Gastroenterology* 133: 951–958. Available: <http://www.ncbi.nlm.nih.gov/pubmed/17854598>. Accessed 29 August 2014.
10. Cabot B, Martell M, Esteban JI, Sauleda S, Otero T, et al. (2000) Nucleotide and amino acid complexity of hepatitis C virus quasispecies in serum and liver. *J Virol* 74: 805–811. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=111600&tool=pmcentrez&rendertype=abstract>. Accessed 29 August 2014.
11. Nasu A, Marusawa H, Ueda Y, Nishijima N, Takahashi K, et al. (2011) Genetic heterogeneity of hepatitis C virus in association with antiviral therapy determined by ultra-deep sequencing. *PLoS One* 6: e24907. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3178558&tool=pmcentrez&rendertype=abstract>. Accessed 29 August 2014.
12. Cheng Y, Guindon S, Rodrigo A, Wee LY, Inoue M, et al. (2013) Cumulative viral evolutionary changes in chronic hepatitis B virus infection precedes hepatitis B e antigen seroconversion. *Gut* 62: 1347–1355. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23242209>. Accessed 25 August 2014.
13. Liu F, Chen L, Yu D-M, Deng L, Chen R, et al. (2011) Evolutionary patterns of hepatitis B virus quasispecies under different selective pressures: correlation with antiviral efficacy. *Gut* 60: 1269–1277. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21292683>. Accessed 25 August 2014.
14. Cheng Y, Guindon S, Rodrigo A, Lim SG (2013) Increased viral quasispecies evolution in HBeAg seroconverter patients treated with oral nucleoside therapy. *J Hepatol* 58: 217–224. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23023011>. Accessed 29 August 2014.

Table S2 Normalized Shannon Entropy (Sn), mutation frequency (Mf) and nucleotide diversity (Pi), and mutation frequency of amino acids in the preCore (pre Core MfAA) and Core (Core MfAA) regions for each sample. (PDF)

File S1 Alignment of the dominant haplotypes of the 30 samples analyzed. (FAS)

Acknowledgments

We thank Celine Cavallo for English language support and helpful editing suggestions.

Author Contributions

Conceived and designed the experiments: MH MB FRF. Performed the experiments: MH AC. Analyzed the data: MH JG FRF. Contributed reagents/materials/analysis tools: JG DT JQ LN. Wrote the paper: MH RE MB FRF.

15. Gregori J, Salicrú M, Domingo E, Sanchez A, Esteban JI, et al. (2014) Inference with viral quasispecies diversity indices: clonal and NGS approaches. *Bioinformatics* 30: 1104–1111. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24389655>. Accessed 17 July 2014.
16. Ramirez C, Gregori J, Buti M, Tabernero D, Camós S, et al. (2013) A comparative study of ultra-deep pyrosequencing and cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. *Antivir Res* 98: 273–283. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23523552>. Accessed 11 August 2013.
17. Gregori J, Esteban JI, Cubero M, Garcia-Cehic D, Perales C, et al. (2013) Ultra-deep pyrosequencing (UDPS) data treatment to study amplicon HCV minor variants. *PLoS One* 8: e83361. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3877031&tool=pmcentrez&rendertype=abstract>. Accessed 3 April 2014.
18. European Association for the Study of the Liver (2012) EASL clinical practice guidelines: Management of chronic hepatitis B virus infection. *J Hepatol* 57: 167–185. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22436845>. Accessed 7 August 2013.
19. Torre F, Cramp M, Owsianka A, Dornan E, Marsden H, et al. (2004) Direct evidence that naturally occurring mutations within hepatitis B core epitope alter CD4+ T-cell reactivity. *J Med Virol* 72: 370–376. Available: <http://www.ncbi.nlm.nih.gov/pubmed/14748060>. Accessed 29 August 2014.
20. Belnap DM, Watts NR, Conway JF, Cheng N, Stahl SJ, et al. (2003) Diversity of core antigen epitopes of hepatitis B virus. *Proc Natl Acad Sci U S A* 100: 10884–10889. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=196897&tool=pmcentrez&rendertype=abstract>. Accessed 29 August 2014.
21. Yousef M, Bell TG, Mudawi H, Glebe D, Kramvis A (2014) Analysis of ultra-deep pyrosequencing and cloning based sequencing of the basic core promoter/precore/core region of hepatitis B virus using newly developed bioinformatics tools. *PLoS One* 9: e95377. Available at <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3989311&tool=pmcentrez&rendertype=abstract>. Accessed 29 August 2014.
22. Homs M, Jardi R, Buti M, Schaper M, Tabernero D, et al. (2011) HBV core region variability: effect of antiviral treatments on main epitope regions. *Antivir Ther* 16: 37–49. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21311107>. Accessed 13 March 2011.
23. Lavocat F, Dény P, Pichoud C, Al Hawajri N, Kitirinos K, et al. (2013) Similar evolution of hepatitis B virus quasispecies in patients with incomplete adefovir response receiving tenofovir/emtricitabine combination or tenofovir monotherapy. *J Hepatol* 59: 684–695. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23742912>. Accessed 29 August 2014.
24. Rodriguez-Frias F, Tabernero D, Quer J, Esteban JI, Ortega I, et al. (2012) Ultra-deep pyrosequencing detects conserved genomic sites and quantifies linkage of drug-resistant amino acid changes in the hepatitis B virus genome. *PLoS One* 7: e37874. Available: <http://dx.plos.org/10.1371/journal.pone.0037874>. Accessed 23 May 2014.
25. Homs M, Buti M, Quer J, Jardi R, Schaper M, et al. (2011) Ultra-deep pyrosequencing analysis of the hepatitis B virus preCore region and main catalytic motif of the viral polymerase in the same viral genome. *Nucleic Acids Res* 39: 8457–8471. Available: <http://www.nar.oxfordjournals.org/cgi/doi/10.1093/nar/gkr451>. Accessed 3 April 2014.
26. Nishijima N, Marusawa H, Ueda Y, Takahashi K, Nasu A, et al. (2012) Dynamics of hepatitis B virus quasispecies in association with nucleos(t)ide analogue treatment determined by ultra-deep sequencing. *PLoS One* 7: e35052. Available: <http://www.pubmedcentral.nih.gov/articlerender>.

- fcgi?artid=3327662&tool=pmcentrez&rendertype=abstract. Accessed 27 August 2014.
27. Solimone M, Vincenti D, Prosperi MCF, Bruselles A, Ippolito G, et al. (2009) Use of massively parallel ultradeep pyrosequencing to characterize the genetic diversity of hepatitis B virus in drug-resistant and drug-naïve patients and to detect minor variants in reverse transcriptase and hepatitis B S antigen. *J Virol* 83: 1718–1726. Available: <http://jv.asm.org/cgi/content/abstract/83/4/1718>. Accessed 28 December 2012.
 28. Margeridon-Thermet S, Shulman NS, Ahmed A, Shahriar R, Liu TF, et al. (2009) Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naïve patients. *J Infect Dis* 199: 1275–1285. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19301976>. Accessed 30 June 2010.
 29. Margeridon-Thermet S, Svarovskaia ES, Babrzadeh F, Martin R, Liu TF, et al. (2013) Low-level persistence of drug resistance mutations in hepatitis B virus-infected subjects with a past history of Lamivudine treatment. *Antimicrob Agents Chemother* 57: 343–349. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23114756>. Accessed 23 May 2014.
 30. Chen MT, Billaud J-N, Sällberg M, Guidotti LG, Chisari FV, et al. (2004) A function of the hepatitis B virus precore protein is to regulate the immune response to the core antigen. *Proc Natl Acad Sci U S A* 101: 14913–14918. Available: <http://www.pubmedcentral.nih.gov/artidrender.fcgi?artid=522042&tool=pmcentrez&rendertype=abstract>. Accessed 29 August 2014.
 31. Rodríguez-Frias F, Buti M, Jardí R, Cotrina M, Viladomiu L, et al. (1995) Hepatitis B virus infection: precore mutants and its relation to viral genotypes and core mutations. *Hepatology* 22: 1641–1647. Available: <http://www.ncbi.nlm.nih.gov/pubmed/7489968>. Accessed 29 August 2014.
 32. Wang H-Y, Chien M-H, Huang H-P, Chang H-C, Wu C-C, et al. (2010) Distinct hepatitis B virus dynamics in the immunotolerant and early immunoclearance phases. *J Virol* 84: 3454–3463. Available: <http://www.pubmedcentral.nih.gov/artidrender.fcgi?artid=2838120&tool=pmcentrez&rendertype=abstract>. Accessed 29 August 2014.
 33. Boni C, Laccabue D, Lampertico P, Giuberti T, Viganò M, et al. (2012) Restored function of HBV-specific T cells after long-term effective therapy with nucleos(t)ide analogues. *Gastroenterology* 143: 963–73.e9. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22796241>. Accessed 17 July 2014.

Segundo estudio

4. SEGUNDO ESTUDIO: Complex Genotype Mixtures Analyzed by Deep Sequencing in Two Different Regions of Hepatitis B Virus

4.1 Introducción

En el segundo estudio se quiso analizar la diversidad genética y complejidad del VHB en forma de genotipos, como en el estudio anterior, en dos períodos de tiempo, uno libre de tratamiento y otro bajo la presión de NUCs.

El VHB se ha clasificado en ocho genotipos bien caracterizados (A-H) y dos más propuestos (I-J) y su distribución geográfica y étnica está bien caracterizada, en España, el 90% de los casos corresponden a Gen/A y D, en proporciones similares, con un 6% de Gen/F [194]. Además se han descrito fenómenos de recombinación intergenotípica, deducida por las controversias en el análisis filogenético de diferentes zonas del genoma (el 30% de las 3300 secuencias del GenBank presentan evidencias de recombinación [148]), por ejemplo, el Gen/I es un triple recombinante A/C/G [197], el gen preCore/Core de los subgenotipos B2-B5 tiene su origen en el Gen/C [140] y el VHB dominante en el Tíbet es un recombinante C/D [141]. El mecanismo de recombinación no se conoce pero se considera que el primer paso es la co-infección con varios genotipos, de hecho los recombinantes tienden a tener la misma distribución geográfica que los genotipos parentales [138]. En Asia se ha descrito una alta prevalencia de infecciones mixtas [145], por ejemplo mezclas Gen/B+C en Taiwán [12], en Europa Central y del Este se han descrito infecciones mixtas (Gen/A+D, Gen/A+D+F y Gen/D+F) en el 8% de los individuos infectados [146] y en España nuestro grupo ya encontró una tasa considerable (22%) de infecciones mixtas (Gen/A+D y Gen/A+D+F) en pacientes tratados con NUCs [194,198].

4.2 Hipótesis y objetivos

4.1.1 Hipótesis

Cada vez existen más estudios que asocian el genotipo con la seroconversión HBeAg, la evolución clínica, el pronóstico y la respuesta al tratamiento con IFN y NUCs [135,136,139]. Por ejemplo, los Gen/C y D están asociados con seroconversión HBeAg tardía o ausente, enfermedad hepática más severa, peor respuesta a IFN y resistencia más temprana a LAM que los Gen/B y A, respectivamente. También las infecciones mixtas podrían correlacionar con la

clínica, en Asia la mezclas de genotipos se han asociada con mayores niveles de ADN-VHB y un curso más severo de la enfermedad que en pacientes monoinfectados por Gen/C [145]. El primer paso para poder establecer estas asociaciones genotipo-evolución clínica/respuesta tratamiento es el correcto genotipaje, el mejor método para ello es el análisis filogenético de toda la secuencia nucleotídica, como esto no es posible en los laboratorios de análisis clínicos, se ha optado por utilizar una parte representativa del genoma, pero tampoco existe consenso acerca de qué región usar para este propósito y, además, la filogénesis local de diferentes fragmentos del genoma podría llevar a una clasificación incorrecta de los genotipos por los fenómenos de recombinación. No obstante, actualmente la región S del genoma del VHB es la que más se usa, ya que se ha visto que la secuenciación y análisis filogenético de la región pre-S1/pre-S2 (nt 3025-80 desde el sitio EcoRI) muestra un genotipaje similar al del genoma completo [135], sin embargo no detecta infecciones mixtas. Los ensayos comerciales de hibridación reversa INNO-LiPA, utilizan la región entre las posiciones nt 456-798 donde se solapan el gen P y S y detectan mezclas de genotipos cuando están presentes al menos en un 5%, pero cualquier variante en la región puede alterar el resultado [176,199,200]. La UDPS puede superar estas limitaciones, ya que permite el análisis de miles de secuencias clonales, aumentado la sensibilidad para la detección de variantes en una determinada región del genoma, sin embargo la relativa corta longitud de los fragmentos (en torno a 400 pb) limita el estudio de la recombinación intergenotípica.

4.2.2 Objetivos

Con el objetivo principal de estudiar la complejidad del VHB a nivel genotípico en diferentes regiones del genoma y bajo diferentes condiciones del ambiente, en un período sin tratamiento y otro bajo la presión de NUCs, se analizaron dos regiones del genoma del VHB (P/S y X/PC) mediante UDPS y genotipaje por filogénesis para:

1. Cuantificar las mezclas de genotipos
2. Evaluar los cambios en el patrón de genotipos de las QS del VHB en presencia/ausencia de tratamiento con NUCs.

También se quiso caracterizar la complejidad a nivel de QS en las región X/PC en la evolución natural y bajo tratamiento antiviral con NUCs y compararla con la región P/S.

4.3 Resumen del estudio

Como en el primer estudio, muestras secuenciales de diez pacientes con HBC y en los que falló el tratamiento con LAM fueron analizadas, tres muestras por paciente, una basal

obtenida en el momento del diagnóstico (1ª), otra muestra tomada antes del inicio de tratamiento (2ª) y otra en el momento de respuesta virológica parcial o VBK (3ª).

El genotipo se determinó mediante UDPS en el secuenciador FLX (454Life Sciences) y análisis filogenético posterior, en dos fragmentos del genoma (Figura 33); (1) región P/S (nt 615-969), que cubre un fragmento donde el gen S y el dominio RT de la polimerasa se solapan e incluye un fragmento de 183 pb de la región amplificada en INNO-LiPA y (2) región X/PC (nt 1596-1912), que incluye numerosos elementos reguladores que hace que tenga una gran relevancia clínica, y sobre todo, y a diferencia del primer estudio, el ENHII completo.

La complejidad de QS en ambas regiones en la dinámica natural y bajo tratamiento antiviral con NUCs, se evaluó mediante los índices de diversidad Mf y Pi, como en el estudio anterior.

4.3.1. Validación de las regiones P/S y X/PC para el genotipaje

El primer paso fue verificar la capacidad del método utilizado para genotipar aplicándolo a 113 secuencias de referencia del genoma del VHB del GenBank (16 de Gen/A, 19 de Gen/B, 25 de Gen/C, 17 de Gen/D, 8 de Gen/E, 17 de Gen/F, 4 de Gen/G, 4 de Gen/H, 2 de Gen/I y 1 de Gen/J), todas las secuencias se clasificaron perfectamente excepto la secuencia de Gen/J. Además utilizando ventanas de 400 pb desplazadas 40 pb hasta cubrir todo el genoma (*Sliding Windows*), se observó que la precisión de la clasificación genotípica es altamente dependiente de la región seleccionada y que la región preS es la que muestra mayor poder de discriminación. La capacidad de las dos regiones de este estudio para diferenciar entre genotipos se analizó utilizando el mismo conjunto de secuencias de referencia y se observó que la región P/S presentaba una alta capacidad discriminante de los ocho genotipos principales (Gen/A-H), sólo dos secuencias de referencia de Gen/C fueron clasificadas como Gen/A, por el contrario la región X/PC mostraba una clara mezcla entre los Gen/D y E y Gen/C e I. Esto no supuso un problema para este estudio dado que el Gen/E está restringido a África y ninguno de los 10 pacientes tenía origen africano, no obstante para ser precisos los resultados de Gen/D en esta región se representan como Gen/DE, y por otro lado ninguna de las secuencias fue Gen/C.

4.3.2 Resultados del genotipaje por UDPS

Se observaron los tres genotipos circulantes en España (Gen/A, D y F), el porcentaje de cada uno de ellos en cada región (P/S y X/PC) y en cada muestra (1ª, 2ª y 3ª) de cada uno de los pacientes se muestra en la tabla 20.

4.3.2.1 Genotipaje en la región P/S

La mayoría de las muestras (70%) mostraron un único genotipo: Gen/A en la mayoría (81%) o Gen/D (19%), y el 30% restante mostraba una infección mixta: Gen/A+D o Gen/A+D+F (con prevalencias similares, 56 y 44%, respectivamente). De las 9 muestras con mezclas, INNO-LiPA sólo detectó mezclas en dos y en cuatro el porcentaje era teóricamente detectable, estas diferencias entre INNO-LiPA y UDPS en la detección de mezclas pueden ser debidas a que el fragmento analizado es diferente y a que ambas técnicas tienen una metodología diferente, UDPS es un método de secuenciación y el genotipo se obtiene por filogénesis mientras que INNO-LiPA es un método indirecto de hibridación, y de hecho incongruencias entre ambos métodos ya han sido descritas previamente [201,202].

4.3.2.2 Genotipaje en la región X/PC

A diferencia de la región P/S sólo una muestra presentaba infección por un único genotipo (Gen/A en la 1ª muestra de un paciente), el 96,7% restante mostraban mezclas de genotipos: triple mezcla Gen/A+DE+F en la mayoría (90%) y doble mezcla Gen/A+DE (6,7%) y Gen/A+F (3,3%) en el resto.

4.3.2.3 Comparación de la región P/S vs. X/PC para el genotipaje

La comparación de los resultados del genotipaje entre ambas regiones mostró varias discrepancias: (1) un número significativamente más elevado de muestras presentaban mezclas de genotipos en la región X/PC (96.7%) comparado con P/S (30.0%), (2) la distribución del genotipo prevalente también fue significativamente diferente, en X/PC el Gen/A predominaba en todas las muestras, mientras que en P/S el Gen/A predominaba en el 80% y Gen/D y F predominaban en el 16.7% y 3,3% de las muestras, respectivamente y (3) también fue diferente la prevalencia de los diferentes genotipos en ambas regiones: el porcentaje de muestras con mezclas de Gen/D+F en la región X/PC (90%) fue significativamente mayor que en P/S (13,3%). En cualquier caso los Gen/A, D y F son los más prevalentes en nuestra población y son las mezclas que se han encontrado en ambas regiones estudiadas, la presencia de variantes minoritarias y las mezclas de genotipos podría haber estado subestimadas por la baja sensibilidad de los métodos utilizados hasta ahora. El por qué son más frecuentes en X/PC requiere futuros estudios pero podría deberse a la alta complejidad de esta región, incluyendo la DR1 y el gap de la hebra(-), que actuarían como posibles puntos de recombinación y llevarían a la interpretación incorrecta de una infección mixta. Siguiendo en esta línea, además de las infecciones mixtas, en 6 muestras de 4 pacientes el genotipo principal en P/S fue

diferente del obtenido en X/PC (para más detalle ver tabla 20), sugiriendo una posible recombinación intergenotípica. No se puede descartar que las discrepancias sean consecuencia de la variabilidad debida a UDPS o a la baja cobertura de secuencias, pero hay muestras donde la diferencia de porcentaje es notable, con un complejo patrón de recombinación, incluyendo un triple recombinante A/D/F, lo que está de acuerdo con otros estudios en que los recombinantes observados incluyen los genotipos más frecuentes que coexisten en la población [203]. La prevalencia del Gen/A y D similar en nuestro país sumado a la vida media larga de los hepatocitos infectados (de 2,4 a más de 70 días durante el tratamiento con LAM [204]) podría facilitar la infección de un único hepatocito con diferentes genotipos y favorecer los fenómenos de recombinación. No obstante se debería confirmar mediante UDPS en un fragmento más largo que abarcase ambas regiones en la misma secuencia.

Tabla 20: % de cada genotipo obtenido en P/S y X/PC en las tres muestras de cada uno de los 10 pacientes. En la región P/S se muestran en amarillo las mezclas detectadas por UDPS, en rojo mezclas detectadas por UDPS y LiPA y en azul las teóricamente detectables por LiPA, sin color las infecciones por un único genotipo. El porcentaje del genotipo principal aparece en ***negrita, cursiva y subrayado*** cuando es discordante en ambas regiones:

- En la 1ª muestra del paciente 2, el genotipaje en P/S lo clasifica como Gen/F mayoritariamente y en X/PC como Gen/A mayoritariamente.

- En la 2ª y 3ª muestra del paciente 4, en P/S es clasificado como Gen/D mayoritariamente y en X/PC como Gen/A mayoritariamente.

- En la 2ª muestra del paciente 9, en P/S es clasificado como Gen/D y en X/PC como Gen/A mayoritariamente.

- En el paciente 10, el genotipaje en X/PC lo clasifica como Gen/A mayoritariamente en las tres muestras, mientras que en P/S, se detecta únicamente Gen/D en la 1ª y 2ª muestra y únicamente Gen/A en la 3ª.

*Baja cobertura de secuencias, no se pueden descartar mezclas.

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

4.3.3 Cambios de genotipo durante los períodos de tiempo estudiados

El porcentaje en el que se enriquecía un genotipo en ambas regiones durante la dinámica natural de la QS (diferencia de % entre la 1ª y 2ª muestra) y durante el tratamiento (diferencia de % entre la 2ª y la 3ª) se describe en la tabla 21.

La dinámica de genotipos, definida como cambios en el genotipo principal, es decir, cuando se selecciona un genotipo en un $\% \geq 50$, sólo se observó en P/S. Durante la dinámica natural de la QS el genotipo cambió en 3 pacientes, en un paciente el cambio fue de Gen/A+D a Gen/A y en dos pacientes de Gen/A pasaron a Gen/D y durante el tratamiento el genotipo cambió en 2 pacientes, que pasaron de Gen/D a A (este cambio también se detectó por LiPA pero no en X/PC).

En cuanto a los cambios en la proporción de genotipos, es decir, el cambio en el porcentaje no implica la selección de un nuevo genotipo (0-49%), en la región P/S sólo cambió en tres pacientes durante la dinámica natural de la QS o durante el tratamiento, sin embargo en la región X/PC se observaron cambios en las proporciones en todos los pacientes en ambos períodos.

Analizado por regiones, en P/S durante la dinámica natural de la QS el Gen/A aumentó en tres pacientes, el Gen/D en otro y en otro paciente tanto el Gen/A como D, durante el tratamiento, en un paciente aumentó el Gen/D, en otro aumentaron los Gen/A y F, en otro el Gen/A y dos pacientes pasaron de ser Gen/D en la 2ª muestra a ser Gen/A en la 3ª. En X/PC durante la dinámica natural de la QS, el porcentaje de Gen/A aumentó en cuatro pacientes, en uno el Gen/F, en dos pacientes aumentó tanto el Gen/D como F, en otro el Gen/D, en otro los Gen/A y D y en un paciente aumentaron los Gen/A y F. Durante el tratamiento en X/PC aumentaron los Gen/D y F en seis pacientes, en un paciente aumentó el Gen/D y en otro el Gen/A.

La selección de un genotipo sobre otro en la dinámica natural de la QS ocurriría debido a las condiciones inmunológicas, en el segundo período debido al efecto del tratamiento antiviral, en este caso la selección o mantenimiento del Gen/A en P/S después del tratamiento en 8/10 pacientes, concuerda con nuestras observaciones anteriores [198] y sugiere que esta región estaría bajo presión evolutiva durante el tratamiento y que las fluctuaciones de genotipo podrían ser un fenómeno común en P/S.

Tabla 21: Se describe la dinámica de genotipos (cambio de genotipo principal, %≥50) en azul y cambio de proporciones (variación del % de genotipos <50%) sin color durante la dinámica natural de la QS (Natural QD) (diferencia de % de cada genotipo entre la 1ª y 2ª muestra, marcado en negrita) y durante el tratamiento (diferencia entre la 2ª y 3ª muestra).

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

4.3.4 Complejidad de la QS

Se estudió la complejidad en las 30 muestras de los 10 pacientes en ambas regiones, con la frecuencia de mutación (Mf) y la diversidad nucleotídica (Pi), no se observó ninguna tendencia general, en cada paciente la QS parece evolucionar según un patrón individual gobernado por su propio sistema inmune. Lo que sí se vio es que los dos índices eran significativamente mayores en la región X/PC que en P/S, de acuerdo con el mayor solapamiento de P/S y que la región X/PC podría estar bajo mayor presión, lo que está en concordancia con la alta complejidad observada en el primer estudio en la región PC/core/PBC.

4.3.5 Conclusiones

1. Mediante UDPS se evidencia la existencia de mezclas de genotipos que cambian a lo largo del tiempo, lo que pone de manifiesto la compleja dinámica de la QS del VHB como un mecanismo adicional para adaptarse a nuevas situaciones.
2. Las discrepancias entre genotipos en ambas regiones sugieren un fenómeno de recombinación intergenotípica.
3. Estos nuevos paradigmas evidencian la necesidad de consensuar una región y una tecnología para el genotipaje del VHB y así entender mejor el significado clínico de la clasificación genotípica del VHB.

4.4 Manuscrito completo

RESEARCH ARTICLE

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

Andrea Caballero¹✉, Josep Gregori^{2,3,4}✉, Maria Homs^{1,2}, David Tabernero^{1,2}, Carolina Gonzalez¹, Josep Quer^{2,3}, Maria Blasi^{1,2}, Rosario Casillas³, Leonardo Nieto¹, Mar Riveiro-Barciela^{2,5}, Rafael Esteban^{2,5}, Maria Buti^{2,5}, Francisco Rodriguez-Frias^{1,2}*

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

✉ These authors contributed equally to this work.

* frarodri@vhebron.net



CrossMark
click for updates

OPEN ACCESS

Citation: Caballero A, Gregori J, Homs M, Tabernero D, Gonzalez C, Quer J, et al. (2015) Complex Genotype Mixtures Analyzed by Deep Sequencing in Two Different Regions of Hepatitis B Virus. PLoS ONE 10(12): e0144816. doi:10.1371/journal.pone.0144816

Editor: Golo Ahlenstiel, University of Sydney, AUSTRALIA

Received: June 29, 2015

Accepted: November 24, 2015

Published: December 29, 2015

Copyright: © 2015 Caballero et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Dr F Rodríguez-Frias received a fund from the Instituto de Salud Carlos III (grant PI12/01893) for supplies and reagents. The study was cofinanced by the European Regional Development Fund (ERDF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Roche Diagnostics provided support in the form of a salary for JG, but did not have any role in the study design or execution, data collection or analysis, decision to

Abstract

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

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

publish, or preparation of the manuscript. The specific roles of JG are described in the Author Contributions section.

Competing Interests: J. Gregori is employee of Roche Diagnostics Spain. Roche Diagnostics Spain did not participate in the study design, data collection, data analysis, data interpretation, or writing of the report. This commercial affiliation does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials. All other authors have no conflict of interest to declare.

Introduction

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

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

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

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

Patients and Methods

Patients

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

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

Serological and virological determinations

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

HBV genome regions analyzed by UDPS

HBV-DNA was extracted from serum by QIAampMiniSpin columns (QIAampDNA Mini Kit, QIAGEN, Hilden, Germany), according to the manufacturer's instructions. Two HBV genomic fragments were selected for UDPS (GS FLX, 454 Life Sciences, Roche) ([Fig 1](#)). One was a 355 bp (nt 615–969) fragment including the reverse transcriptase domain and overlapping the surface gene (P/S region); P/S includes 183 bp of the region amplified by LiPA. The other was a 356 bp (nt 1580–1936) fragment including the 3'-end of the X gene and the 5'-end of the pre-Core/Core gene (X/preCore region). The X/preCore region was chosen because it is clinically relevant, as it includes the ENHII region, the main motifs involved in HBV replication and regulation, and the sequence where synthesis of the HBV DNA minus strand is initiated (DR1). In addition, it is likely to be a region in which the viral quasispecies would show high complexity. Two PCRs with high-fidelity polymerase (Pfu Ultra-II, Stratagene, La Jolla, United States) were performed for P/S and X/preCore. PCR products were isolated from 1.5% agarose gel and their quality and quantity were checked. A detailed description of the reactions, concentrations, and primer sequences used is provided in supplementary materials and methods in [S1 Protocol](#). Amplicons were pooled to a concentration of 2×10^6 molecules. This working solution was enriched with capture beads for forward and reverse clonal amplification (emPCR kits II and III, 454 Life Sciences). UDPS was carried out on the FLX Genome Sequencer system (454Life Sciences).

UDPS data treatment

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

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

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

Pt	Sex	First sample at diagnosis (baseline)					Second sample (before starting treatment)					Third sample (under treatment)				
		Age, years	ALT, IU/mL	HBV DNA, IU/mL	HBeAg	Gen	Months after 1st sample	ALT, IU/mL	HBV DNA, IU/mL	HBeAg	Gen	Months after 2nd sample	ALT, IU/mL	HBV DNA, IU/mL	HBeAg	Gen
1	M	51	105	3.13E+06	N	A	7	364	2.32E+07	N	A	25	38	1.64E+04	N	A
2	M	56	391	2.9E+07	N	DF	25	656.9	1.16E+08	N	A	13	19.5	6.67E+04	N	A
3	M	30	46.4	1.68E+08	N	A	20	101.8	9.67E+07	N	A	14	34.7	1.18E+04	N	A
4	M	29	116	1.37E+03	N	A	28	70.7	8.64E+06	N	D	5	49.5	2.93E+04	N	D
5	M	56	75	2.65E+05	N	A	7	60	7.87E+08	P	A	19	55.9	1.03E+09	P	A
6	M	45	92	7.80E+08	N	A	36	129	2.78E+09	N	A	32	38	2.38E+03	N	A
7	F	20	59	1.66E+10	P	A	4	60	2.23E+09	P	A	14	75	2.36E+09	P	A
8	M	32	78	9.64E+04	N	AF	16	242	4.42E+05	N	A	22	57	3.17E+04	N	A
9	M	48	65	6.12E+05	N	A	43	109	7.43E+06	N	D	11	50	9.73E+03	N	A
10	M	35	262	9.66E+08	P	D	17	137	8.99E+06	P	D	18	93.8	1.31E+04	N	A

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

doi:10.1371/journal.pone.0144816.t001

Genotyping and quasispecies complexity indexes

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

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

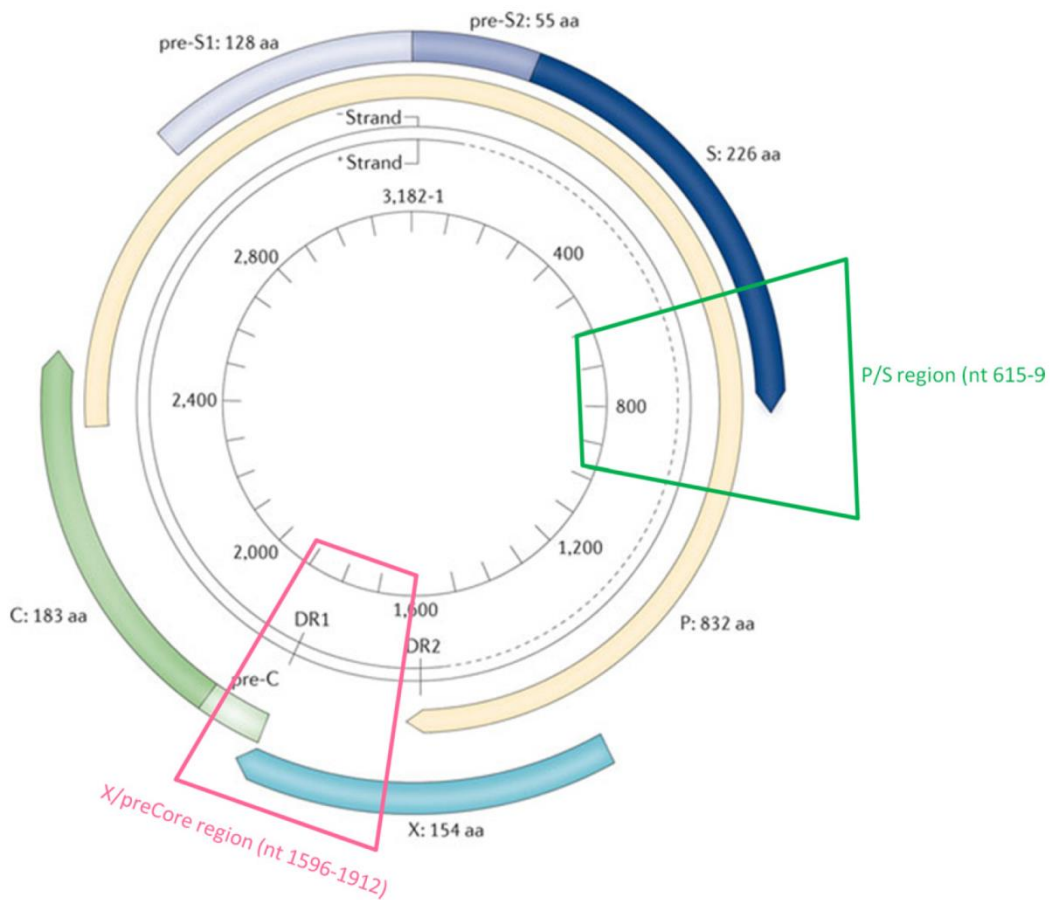


Fig 1. Regions of the HBV genome amplified for the present study: P/S is indicated in green and X/preCore in pink.

doi:10.1371/journal.pone.0144816.g001

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

Statistical analysis

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

Results

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

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

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

Genotyping

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

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

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

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

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

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

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

doi:10.1371/journal.pone.0144816.t002

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

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

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

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

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

doi:10.1371/journal.pone.0144816.t003

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

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

HBV genotype changes over the study period

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

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

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

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

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

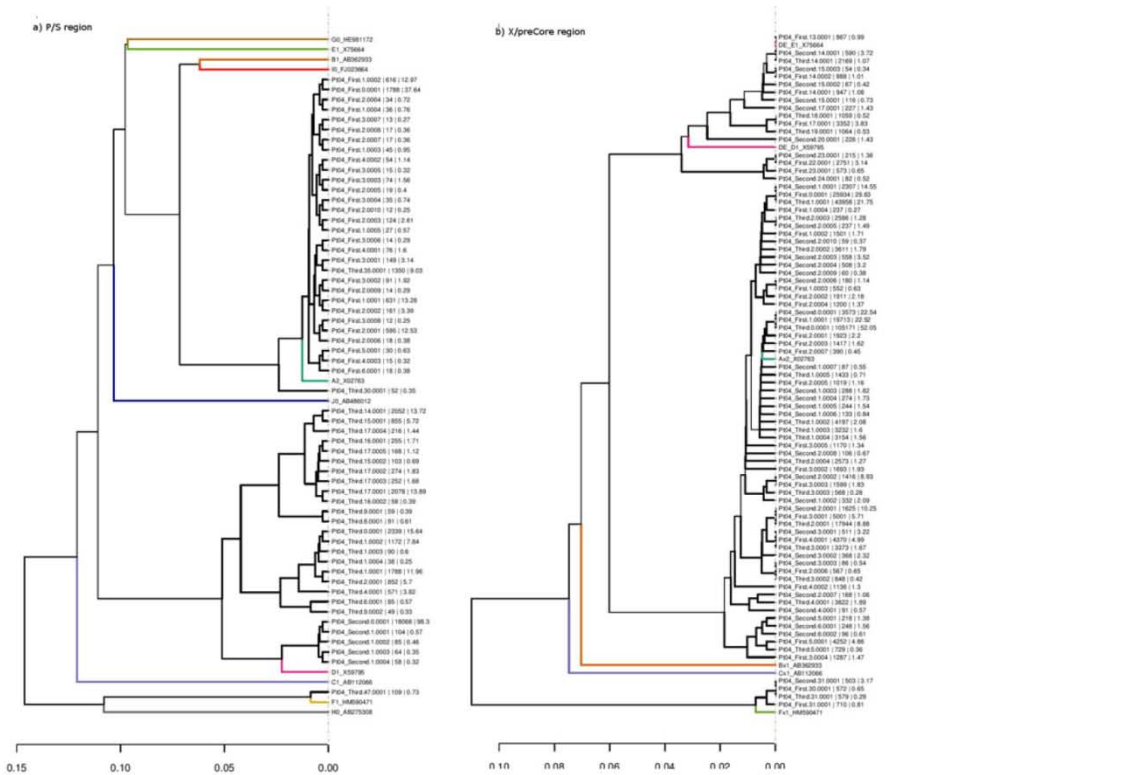


Fig 2. UPGMA tree obtained from the polymerase/surface (a) and the X/preCore (b) regions of the three samples from Pt 4. The main haplotype of the first sample (First.0.0.0001) was genotyped as A, whereas the main haplotype of the second and third sample (Second.0.0.0001 and Third.0.0.0001) was genotype as D. At the X/preCore region, the main genotype was A, but Gen/DE and F haplotypes were also observed. Genotype reference sequences for the two regions: A2_X02763; B1_AB362933; C1_AB112066; D1_X59795 (for P/S region); DE_D1_X59795 (for X/preCore region); E1_X75664 (for P/S region); DE_E1_X75664 (for X/preCore region); F1_HM590471; G0_HE981172; H0_AB275308; IO_FJ023664; JO_AB486012.

doi:10.1371/journal.pone.0144816.g002

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

Quasispecies complexity

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

Discussion

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

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

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

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

[#]X/preCore region showed an intermixing between Gen/D and E (S3 File)

doi:10.1371/journal.pone.0144816.t004

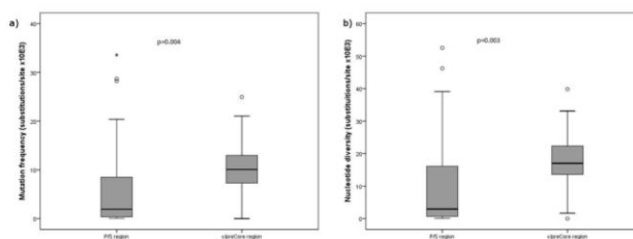


Fig 3. Distribution of quasispecies complexity in all the samples analyzed according to mutation frequency (a) and nucleotide diversity (b) in the polymerase/surface (P/S) and X/preCore regions.

doi:10.1371/journal.pone.0144816.g003

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

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

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

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

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

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

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

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

Supporting Information

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

(PDF)

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

(PDF)

S3 File. Genotyping capacity of the reference sequences trimmed to positions 1596 to 1 912 (covering the P/S amplicon).

(PDF)

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

(PDF)

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

(PDF)

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

(PDF)

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

(PDF)

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

(DOCX)

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

(DOCX)

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

(DOCX)

Acknowledgments

The authors thank Celine Cavallo for English language support and helpful editing suggestions.

Author Contributions

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

References

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

3. Kramvis A. Genotypes and genetic variability of hepatitis B virus. *Intervirology*. 2014; 57: 141–50. doi: [10.1159/000360947](https://doi.org/10.1159/000360947) PMID: [25034481](https://pubmed.ncbi.nlm.nih.gov/25034481/)
4. Rodriguez-Frias F, Jardi R, Buti M, Schaper M, Hermosilla E, Valdes A, et al. Hepatitis B virus genotypes and G1896A precore mutation in 486 Spanish patients with acute and chronic HBV infection. *J Viral Hepat*. 2006; 13: 343–350. doi: [10.1111/j.1365-2893.2005.00691.x](https://doi.org/10.1111/j.1365-2893.2005.00691.x) PMID: [16637866](https://pubmed.ncbi.nlm.nih.gov/16637866/)
5. Hsieh T-H, Tseng T-C, Liu C-J, Lai M-Y, Chen P-J, Hsieh H-L, et al. Hepatitis B virus genotype B has an earlier emergence of lamivudine resistance than genotype C. *Antivir Ther*. 2009; 14: 1157–63. doi: [10.3851/IMP1454](https://doi.org/10.3851/IMP1454) PMID: [20032545](https://pubmed.ncbi.nlm.nih.gov/20032545/)
6. Lin C-L, Kao J-H. The clinical implications of hepatitis B virus genotype: Recent advances. *J Gastroenterol Hepatol*. 2011; 26 Suppl 1: 123–30. doi: [10.1111/j.1440-1746.2010.06541.x](https://doi.org/10.1111/j.1440-1746.2010.06541.x) PMID: [21199523](https://pubmed.ncbi.nlm.nih.gov/21199523/)
7. Buti M, Elefsiniotis I, Jardi R, Vargas V, Rodriguez-Frias F, Schapper M, et al. Viral genotype and baseline load predict the response to adefovir treatment in lamivudine-resistant chronic hepatitis B patients. *J Hepatol*. 2007; 47: 366–72. doi: [10.1016/j.jhep.2007.04.011](https://doi.org/10.1016/j.jhep.2007.04.011) PMID: [17561304](https://pubmed.ncbi.nlm.nih.gov/17561304/)
8. Kramvis A, Kew MC. Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy. *J Viral Hepat*. 2005; 12: 456–64. doi: [10.1111/j.1365-2893.2005.00624.x](https://doi.org/10.1111/j.1365-2893.2005.00624.x) PMID: [16108759](https://pubmed.ncbi.nlm.nih.gov/16108759/)
9. Osiowy C, Giles E. Evaluation of the INNO-LiPA HBV genotyping assay for determination of hepatitis B virus genotype. *J Clin Microbiol*. 2003; 41: 5473–7. PMID: [14662927](https://pubmed.ncbi.nlm.nih.gov/14662927/)
10. Araujo NM, Araujo OC, Silva EM, Villela-Nogueira CA, Nabuco LC, Parana R, et al. Identification of novel recombinants of hepatitis B virus genotypes F and G in human immunodeficiency virus-positive patients from Argentina and Brazil. *J Gen Virol*. 2012; 94: 150–158. doi: [10.1099/vir.0.047324-0](https://doi.org/10.1099/vir.0.047324-0) PMID: [23079380](https://pubmed.ncbi.nlm.nih.gov/23079380/)
11. Georgi-Geisberger P, Berns H, Loncarevic IF, Yu ZY, Tang ZY, Zentgraf H, et al. Mutations on free and integrated hepatitis B virus DNA in a hepatocellular carcinoma: footprints of homologous recombination. *Oncology*. 1992; 49: 386–95. PMID: [1326735](https://pubmed.ncbi.nlm.nih.gov/1326735/)
12. Shi W, Carr MJ, Dunford L, Zhu C, Hall WW, Higgins DG. Identification of novel inter-genotypic recombinants of human hepatitis B viruses by large-scale phylogenetic analysis. *Virology*. 2012; 427: 51–59. doi: [10.1016/j.virol.2012.01.030](https://doi.org/10.1016/j.virol.2012.01.030) PMID: [22374235](https://pubmed.ncbi.nlm.nih.gov/22374235/)
13. Yang J, Xing K, Deng R, Wang J, Wang X. Identification of Hepatitis B virus putative intergenotype recombinants by using fragment typing. *J Gen Virol*. 2006; 87: 2203–15. doi: [10.1099/vir.0.081752-0](https://doi.org/10.1099/vir.0.081752-0) PMID: [16847116](https://pubmed.ncbi.nlm.nih.gov/16847116/)
14. Shi W, Zhu C, Zheng W, Carr MJ, Higgins DG, Zhang Z. Subgenotype reclassification of genotype B hepatitis B virus. *BMC Gastroenterol*. 2012; 12: 116. doi: [10.1186/1471-230X-12-116](https://doi.org/10.1186/1471-230X-12-116) PMID: [22925657](https://pubmed.ncbi.nlm.nih.gov/22925657/)
15. Deterding K, Constantinescu I, Nedelcu FD, Gervain J, Nemecek V, Situncky O, et al. Prevalence of HBV genotypes in Central and Eastern Europe. *J Med Virol*. 2008; 80: 1707–11. doi: [10.1002/jmv.21294](https://doi.org/10.1002/jmv.21294) PMID: [18712830](https://pubmed.ncbi.nlm.nih.gov/18712830/)
16. Jardi R, Rodriguez-Frias F, Schaper M, Giggi E, Tabernero D, Homs M, et al. Analysis of hepatitis B genotype changes in chronic hepatitis B infection: Influence of antiviral therapy. *J Hepatol*. 2008; 49: 695–701. doi: [10.1016/j.jhep.2008.03.030](https://doi.org/10.1016/j.jhep.2008.03.030) PMID: [18490074](https://pubmed.ncbi.nlm.nih.gov/18490074/)
17. Kao JH, Chen PJ, Lai MY, Chen DS. Acute exacerbations of chronic hepatitis B are rarely associated with superinfection of hepatitis B virus. *Hepatology*. 2001; 34: 817–23. doi: [10.1053/jhep.2001.28188](https://doi.org/10.1053/jhep.2001.28188) PMID: [11584381](https://pubmed.ncbi.nlm.nih.gov/11584381/)
18. Ghedin E, Fitch A, Boyne A, Griesemer S, DePasse J, Bera J, et al. Mixed infection and the genesis of influenza virus diversity. *J Virol*. 2009; 83: 8832–41. doi: [10.1128/JVI.00773-09](https://doi.org/10.1128/JVI.00773-09) PMID: [19553313](https://pubmed.ncbi.nlm.nih.gov/19553313/)
19. Bull RA, Luciani F, McElroy K, Gaudieri S, Pham ST, Chopra A, et al. Sequential Bottlenecks Drive Viral Evolution in Early Acute Hepatitis C Virus Infection. *Ou JJ*, editor. *PLoS Pathog*. 2011; 7: e1002243. doi: [10.1371/journal.ppat.1002243](https://doi.org/10.1371/journal.ppat.1002243) PMID: [21912520](https://pubmed.ncbi.nlm.nih.gov/21912520/)
20. Bunnik EM, Swenson LC, Edo-Matas D, Huang W, Dong W, Frantzell A, et al. Detection of inferred CCR5- and CXCR4-using HIV-1 variants and evolutionary intermediates using ultra-deep pyrosequencing. *PLoS Pathog*. 2011; 7: e1002106. doi: [10.1371/journal.ppat.1002106](https://doi.org/10.1371/journal.ppat.1002106) PMID: [21731496](https://pubmed.ncbi.nlm.nih.gov/21731496/)
21. Hoffmann C, Minkah N, Leipzig J, Wang G, Arens MQ, Tebas P, et al. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic Acids Res*. 2007; 35: e91. doi: [10.1093/nar/gkm435](https://doi.org/10.1093/nar/gkm435) PMID: [17576693](https://pubmed.ncbi.nlm.nih.gov/17576693/)
22. Homs M, Buti M, Quer J, Jardi R, Schaper M, Tabernero D, et al. Ultra-deep pyrosequencing analysis of the hepatitis B virus preCore region and main catalytic motif of the viral polymerase in the same viral genome. *Nucleic Acids Res*. 2011; 39: 8457–71. doi: [10.1093/nar/gkr451](https://doi.org/10.1093/nar/gkr451) PMID: [21742757](https://pubmed.ncbi.nlm.nih.gov/21742757/)
23. Homs M, Caballero A, Gregori J, Tabernero D, Quer J, Nieto L, et al. Clinical application of estimating hepatitis B virus quasispecies complexity by massive sequencing: correlation between natural

- evolution and on-treatment evolution. *PLoS One*. 2014; 9: e112306. doi: [10.1371/journal.pone.0112306](https://doi.org/10.1371/journal.pone.0112306) PMID: [25393280](https://pubmed.ncbi.nlm.nih.gov/25393280/)
24. Solmone M, Vincenti D, Prosperi MCF, Bruselles A, Ippolito G, Capobianchi MR. Use of massively parallel ultradeep pyrosequencing to characterize the genetic diversity of hepatitis B virus in drug-resistant and drug-naïve patients and to detect minor variants in reverse transcriptase and hepatitis B S antigen. *J Virol*. 2009; 83: 1718–26. doi: [10.1128/JVI.02011-08](https://doi.org/10.1128/JVI.02011-08) PMID: [19073746](https://pubmed.ncbi.nlm.nih.gov/19073746/)
 25. European Association For The Study Of The Liver. EASL Clinical Practice Guidelines: management of chronic hepatitis B. *J Hepatol*. European Association for the Study of the Liver; 2009; 50: 227–42. doi: [10.1016/j.jhep.2008.10.001](https://doi.org/10.1016/j.jhep.2008.10.001)
 26. Ramírez C, Gregori J, Buti M, Taberner D, Camós S, Casillas R, et al. A comparative study of ultra-deep pyrosequencing and cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. *Antiviral Res*. 2013; 98: 273–83. doi: [10.1016/j.antiviral.2013.03.007](https://doi.org/10.1016/j.antiviral.2013.03.007) PMID: [23523552](https://pubmed.ncbi.nlm.nih.gov/23523552/)
 27. Gregori J, Esteban JI, Cubero M, Garcia-Cehic D, Perales C, Casillas R, et al. Ultra-Deep Pyrosequencing (UDPS) Data Treatment to Study Amplicon HCV Minor Variants. *PLoS One*. 2013; 8: e83361. doi: [10.1371/journal.pone.0083361](https://doi.org/10.1371/journal.pone.0083361) PMID: [24391758](https://pubmed.ncbi.nlm.nih.gov/24391758/)
 28. R: The R Project for Statistical Computing [Internet]. [Accessed 14 Apr 2015]. Available: <http://www.r-project.org/>
 29. Pages H, Aboyoun P, Gentleman R, DebRoy S. Bioconductor—Biostrings [Internet]. [Accessed 14 Apr 2015]. Available: <http://www.bioconductor.org/packages/release/bioc/html/Biostrings.html>
 30. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. 2004; 20: 289–90. PMID: [14734327](https://pubmed.ncbi.nlm.nih.gov/14734327/)
 31. Cuadras C. Distance analysis in discrimination and classification using both continuous and categorical variables. In: Dodge Y, editor. *Statistical AData analysis and Interference*. Amsterdam: Elsevier; 1989. pp. 459–473.
 32. Cuadras C. A distance approach to discriminant analysis and its properties. *Mathematics preprint series*. Barcelona; 1991.
 33. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980; 16: 111–20. PMID: [7463489](https://pubmed.ncbi.nlm.nih.gov/7463489/)
 34. Kumar S, Gadagkar SR. Efficiency of the neighbor-joining method in reconstructing deep and shallow evolutionary relationships in large phylogenies. *J Mol Evol*. 2000; 51: 544–53. PMID: [11116328](https://pubmed.ncbi.nlm.nih.gov/11116328/)
 35. Borg I, PJF G. *Modern multidimensional scaling: theory and applications*. 2nd ed. New York: Springer; 2005.
 36. Bekondi C, Olinger CM, Boua N, Talarmin A, Muller CP, Le Faou A, et al. Central African Republic is part of the West-African hepatitis B virus genotype E crescent. *J Clin Virol*. 2007; 40: 31–7. doi: [10.1016/j.jcv.2007.05.009](https://doi.org/10.1016/j.jcv.2007.05.009) PMID: [17689139](https://pubmed.ncbi.nlm.nih.gov/17689139/)
 37. Zhang Q, Cao G. Genotypes, mutations, and viral load of hepatitis B virus and the risk of hepatocellular carcinoma: HBV properties and hepatocarcinogenesis. *Hepat Mon*. 2011; 11: 86–91. PMID: [22087123](https://pubmed.ncbi.nlm.nih.gov/22087123/)
 38. Simmonds P, Midgley S. Recombination in the genesis and evolution of hepatitis B virus genotypes. *J Virol*. 2005; 79: 15467–76. doi: [10.1128/JVI.79.24.15467-15476.2005](https://doi.org/10.1128/JVI.79.24.15467-15476.2005) PMID: [16306618](https://pubmed.ncbi.nlm.nih.gov/16306618/)
 39. Mese S, Arkan M, Cakiris A, Abaci N, Gumus E, Kursun O, et al. Role of the line probe assay INNO-LiPA HBV DR and ultradeep pyrosequencing in detecting resistance mutations to nucleoside/nucleotide analogues in viral samples isolated from chronic hepatitis B patients. *J Gen Virol*. 2013; 94: 2729–38. doi: [10.1099/vir.0.053041-0](https://doi.org/10.1099/vir.0.053041-0) PMID: [24045109](https://pubmed.ncbi.nlm.nih.gov/24045109/)
 40. Neumann AU. Hepatitis B viral kinetics: A dynamic puzzle still to be resolved. *Hepatology*. 2005; 42: 249–254. doi: [10.1002/hep.20831](https://doi.org/10.1002/hep.20831) PMID: [16025494](https://pubmed.ncbi.nlm.nih.gov/16025494/)

S1File. Genotyping capacity of the reference sequences using UPGMA trees and plotting (complete HBV genome).

HBV reference sequences by genotype, with accessions, used in full genome exploration

A	B	C	D	E	F	G	H	I	J
A1_AB241115	B1_AB073858	C0_D23683	D0_X65259	E1_X75664	F1_AY090459	G0_AF160501	H0_AB179747	I0_FJ023660	J0_AB486012
A1_AY233278	B1_AB362933	C0_L08805	D0_X68292	E2_X75657	F1_DQ823095	G0_EF464098	H0_AB275308	I0_FJ023664	
A2_AJ309371	B1_D00329	C0_M38636	D1_X59795	E3_AM494694	F1_HE981184	G0_HE981172	H0_AB516395		
A2_AM282986	B2_AP011084	C0_X14193	D1_X80926	E3_FJ349237	F1_HM590471	G0_HE981176	H0_AP007261		
A2_AY738141	B2_AY596111	C1_AB031265	D2_X97848	E4_FJ349226	F1_HQ378247				
A2_X02763	B2_GQ924653	C1_AB112066	D2_Z35716	E4_HM363569	F2_AY090455				
A2_X51970	B2_GU815751	C10_AB540583	D3_AY233291	E5_DQ060828	F2_AY311369				
A2_Z72479	B3_AP011085	C2_AB033553	D3_V01460	E5_JQ000008	F2_X69798				
A3_AB194951	B3_M54923	C2_AF533983	D3_X65258		F3_AB036910				
A3_AB194952	B4_AB073835	C2_AY123041	D4_AB033559		F3_AB036911				
A4_AM180623	B4_AB115551	C2_D16665	D4_AB048702		F3_AB036915				
A4_AY934764	B5_AB219427	C2_D23681	D5_AB033558		F3_FJ589066				
A5_FJ692609	B5_AP011086	C2_X52939	D5_DQ315779		F3_X75663				
A5_FJ692613	B6_AB287316	C3_X75656	D6_AB493846		F4_AB166850				
A6_GQ331047	B6_DQ463787	C3_X75665	D6_AB554023		F4_DQ823090				
A6_GQ331048	B7_AP011091	C4_AB048704	D7_AM494716		F4_EU366116				
	B7_EF473977	C4_AB048705	D7_FJ904430		F4_HE974368				
	B8_AP011093	C5_AB241109							
	B8_AP011094	C5_AP011099							
		C6_AP011102							
		C6_AP011103							
		C7_EU670263							
		C8_AP011104							
		C8_AP011107							
		C9_AP011108							
16	19	25	17	8	17	4	4	2	1

Outgroup
X.chimpanzee_AF222323
X.gibbon_AB037927
X.orangutan_AF193863
X.wmonkey_AF046996

HBV Reference Sequences Discriminating Power

HBV - Whole Genome

Genotype	RefSeq No.	Geometric Variability	Mean Squared Distances Within Genotype				Nearest Dif. Gen.	Nearest Genotype
			Min	Median	Max	Max Corrected		
A	16	0,000919	0,001320	0,001800	0,004760	0,003841	0,00660	I
B	19	0,001070	0,001540	0,001990	0,003460	0,002390	0,00637	C
C	25	0,001260	0,001470	0,002430	0,004450	0,003190	0,00505	I
D	17	0,000963	0,001250	0,002120	0,002960	0,001997	0,00637	E
E	8	0,000099	0,000188	0,000219	0,000293	0,000194	0,00586	D
F	17	0,000994	0,001630	0,002230	0,002440	0,001446	0,00675	H
G	4	0,000037	0,000055	0,000088	0,000161	0,000124	0,01540	A
H	4	0,000168	0,000256	0,000348	0,000836	0,000668	0,00852	F
I	2	0,000218	0,000872	0,000872	0,000872	0,000654	0,00758	C

Mean squared distances within genotype $\frac{1}{n_I - 1} \sum_{i \in I} d_{i,j}^2; \quad j: 1..n_I$

Genotype geometric variability $\hat{V}_I = \frac{1}{2n_I^2} \sum_{i,j \in I} d_{i,j}^2$

DB rule $\min_I \left(\hat{\phi}_I^2(k) = \frac{1}{n_I} \sum_{i \in I} d_{i,k}^2 - \hat{V}_I \right)$

HBV Reference Sequences Classification by DB rule

Whole genome sequences

Smpl	$\Phi_A^2(k)$	$\Phi_B^2(k)$	$\Phi_C^2(k)$	$\Phi_D^2(k)$	$\Phi_E^2(k)$	$\Phi_F^2(k)$	$\Phi_G^2(k)$	$\Phi_H^2(k)$	DB Rule
									Type
A1_AB241115	0,001100	0,008020	0,006870	0,008210	0,007650	0,018200	0,011800	0,020300	A
A1_AY233278	0,001180	0,008760	0,007100	0,008990	0,008430	0,019500	0,013300	0,021700	A
A2_AJ309371	0,000632	0,007940	0,007040	0,008580	0,009390	0,019900	0,012600	0,021600	A
A2_AM282986	0,000360	0,007540	0,006240	0,007860	0,008780	0,018400	0,011700	0,020400	A
A2_AY738141	0,000727	0,008030	0,007250	0,008710	0,009600	0,019900	0,012900	0,021600	A
A2_X02763	0,000599	0,008020	0,007010	0,008200	0,009270	0,019700	0,012400	0,021600	A
A2_X51970	0,000702	0,007940	0,006700	0,008560	0,009000	0,019100	0,012700	0,021200	A
A2_Z72479	0,004030	0,015800	0,014200	0,015700	0,016800	0,030400	0,021000	0,031700	A
A3_AB194951	0,001110	0,008290	0,006970	0,009440	0,009260	0,020500	0,013200	0,022300	A
A3_AB194952	0,001010	0,008390	0,006510	0,009380	0,008930	0,019900	0,013000	0,021600	A
A4_AM180623	0,000946	0,008060	0,005830	0,008430	0,008010	0,018700	0,011600	0,022000	A
A4_AY934764	0,000893	0,007740	0,006160	0,008580	0,008290	0,019200	0,011800	0,022200	A
A5_FJ692609	0,000756	0,008180	0,006750	0,008600	0,008720	0,020000	0,012900	0,022500	A
A5_FJ692613	0,000802	0,008390	0,006910	0,008600	0,008640	0,019400	0,013100	0,021300	A
A6_GQ331047	0,000841	0,007580	0,005770	0,008310	0,008090	0,019300	0,011500	0,021300	A
A6_GQ331048	0,001030	0,007970	0,006140	0,008710	0,008540	0,019600	0,011900	0,022300	A
B1_AB073858	0,008050	0,002150	0,009020	0,010800	0,011800	0,020300	0,016200	0,022800	B
B1_AB362933	0,007920	0,001620	0,008120	0,010100	0,011100	0,018900	0,015900	0,021400	B
B1_D00329	0,006710	0,001490	0,007580	0,008870	0,009970	0,019100	0,015300	0,021300	B
B2_AP011084	0,008360	0,000697	0,006440	0,010500	0,011100	0,019500	0,015800	0,021600	B
B2_AY596111	0,009940	0,001330	0,008080	0,012000	0,012700	0,021700	0,018300	0,022800	B
B2_GQ924653	0,008030	0,000624	0,006400	0,010000	0,010700	0,019100	0,015100	0,021200	B
B2_GU815751	0,007790	0,000515	0,006150	0,009700	0,010600	0,018900	0,015200	0,020800	B
B3_AP011085	0,008040	0,000439	0,005610	0,009110	0,010400	0,019100	0,014600	0,021200	B
B3_M54923	0,009020	0,000721	0,006190	0,009900	0,011400	0,019900	0,015600	0,021400	B
B4_AB073835	0,007980	0,000882	0,006000	0,009580	0,010700	0,018400	0,014000	0,020900	B
B4_AB115551	0,009040	0,001320	0,006820	0,011100	0,011700	0,019900	0,015100	0,022900	B
B5_AB219427	0,009870	0,001230	0,007060	0,011400	0,012200	0,021300	0,015800	0,024000	B
B5_AP011086	0,009380	0,000866	0,006440	0,010600	0,011600	0,019200	0,015600	0,021700	B
B6_AB287316	0,009000	0,002470	0,009000	0,010300	0,011600	0,020000	0,016100	0,020900	B
B6_DQ463787	0,008980	0,002450	0,008460	0,010200	0,011200	0,020700	0,015100	0,021000	B
B7_AP011091	0,008840	0,000759	0,005930	0,009000	0,010500	0,018500	0,015300	0,020900	B
B7_EF473977	0,009450	0,000911	0,006460	0,010200	0,011000	0,019300	0,015200	0,021400	B
B8_AP011093	0,008640	0,000900	0,006270	0,009730	0,010800	0,019500	0,016500	0,021400	B
B8_AP011094	0,010100	0,001200	0,007080	0,010600	0,011400	0,020800	0,016800	0,021900	B
C0_D23683	0,007850	0,007300	0,001070	0,010300	0,011000	0,018700	0,015900	0,020200	C
C0_L08805	0,007260	0,006650	0,000861	0,010100	0,010200	0,017700	0,014500	0,019200	C
C0_M38636	0,007950	0,007580	0,000860	0,010700	0,011200	0,019100	0,016600	0,021900	C
C0_X14193	0,007880	0,007510	0,000828	0,010700	0,011200	0,019000	0,016500	0,021800	C
C1_AB031265	0,006910	0,004970	0,001920	0,009170	0,010600	0,017900	0,014300	0,019700	C
C1_AB112066	0,008000	0,007320	0,001810	0,010200	0,011500	0,018300	0,016700	0,020200	C
C10_AB540583	0,009030	0,008700	0,002460	0,011300	0,011000	0,020600	0,015500	0,021700	C
C2_AB033553	0,006350	0,006020	0,000160	0,008830	0,008910	0,016700	0,013800	0,018800	C
C2_AF533983	0,007320	0,006610	0,000600	0,009740	0,010100	0,018200	0,015500	0,020600	C
C2_AY123041	0,007140	0,006760	0,000512	0,009770	0,009660	0,018400	0,015300	0,020500	C
C2_D16665	0,005990	0,007530	0,001350	0,009060	0,008780	0,018200	0,014900	0,019500	C
C2_D23681	0,007950	0,007270	0,000692	0,010300	0,011000	0,018700	0,015700	0,020700	C
C2_X52939	0,007710	0,007600	0,000870	0,009560	0,010200	0,018300	0,015500	0,019800	C
C3_X75656	0,006840	0,006400	0,001110	0,008900	0,008490	0,016500	0,014500	0,019500	C
C3_X75665	0,006500	0,006930	0,001370	0,008640	0,008770	0,017000	0,013900	0,019000	C
C4_AB048704	0,008800	0,009480	0,003270	0,010800	0,011300	0,018800	0,018600	0,022000	C

HBV Reference Sequences Classification by DB rule

Whole genome sequences

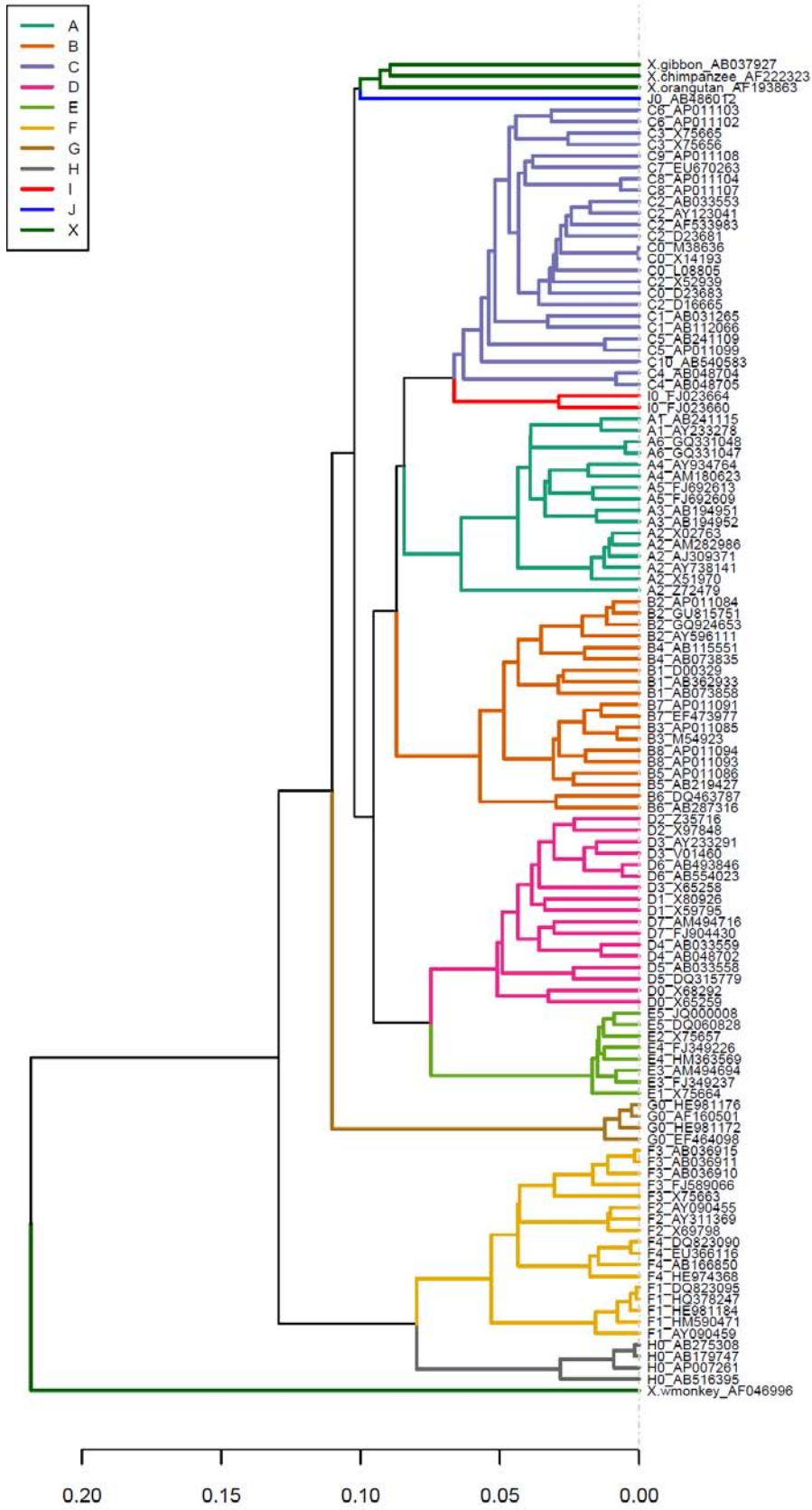
Smpl	$\Phi_A^2(k)$	$\Phi_B^2(k)$	$\Phi_C^2(k)$	$\Phi_D^2(k)$	$\Phi_E^2(k)$	$\Phi_F^2(k)$	$\Phi_G^2(k)$	$\Phi_H^2(k)$	DB Rule
									Type
C4_AB048705	0,008310	0,008740	0,002800	0,010200	0,010600	0,017900	0,017900	0,020700	C
C5_AB241109	0,007600	0,007240	0,002200	0,009970	0,009740	0,017700	0,015000	0,022800	C
C5_AP011099	0,007160	0,007840	0,002010	0,009340	0,009810	0,017900	0,015300	0,023400	C
C6_AP011102	0,007460	0,007220	0,001560	0,010100	0,010300	0,017700	0,016300	0,019600	C
C6_AP011103	0,007320	0,007540	0,001520	0,010500	0,009500	0,017000	0,014700	0,019400	C
C7_EU670263	0,007220	0,006880	0,001100	0,009280	0,009700	0,017400	0,016000	0,019100	C
C8_AP011104	0,007230	0,006860	0,001160	0,009710	0,010100	0,016900	0,016000	0,018600	C
C8_AP011107	0,007470	0,007080	0,001170	0,009760	0,010100	0,017400	0,016500	0,019100	C
C9_AP011108	0,006570	0,006000	0,000934	0,008800	0,009320	0,017700	0,015100	0,018900	C
D0_X65259	0,006540	0,009560	0,009410	0,001610	0,007000	0,019300	0,013100	0,022100	D
D0_X68292	0,005670	0,009640	0,009670	0,002050	0,007770	0,019600	0,013300	0,020800	D
D1_X59795	0,010600	0,011200	0,011000	0,001330	0,007360	0,022400	0,014400	0,024800	D
D1_X80926	0,009970	0,010100	0,010100	0,001160	0,006410	0,020600	0,015300	0,023000	D
D2_X97848	0,009220	0,009100	0,009260	0,000552	0,005030	0,019000	0,012400	0,020900	D
D2_Z35716	0,009480	0,009940	0,009300	0,000623	0,005490	0,020300	0,013900	0,022500	D
D3_AY233291	0,008730	0,009610	0,008980	0,000437	0,005320	0,019800	0,012600	0,020000	D
D3_V01460	0,008920	0,009470	0,008760	0,000246	0,005180	0,019500	0,012800	0,020600	D
D3_X65258	0,009040	0,011000	0,010700	0,001210	0,007700	0,022000	0,014700	0,023700	D
D4_AB033559	0,008960	0,009730	0,008270	0,000844	0,005350	0,018600	0,013300	0,020200	D
D4_AB048702	0,008830	0,009730	0,008820	0,000972	0,005690	0,019300	0,013600	0,021000	D
D5_AB033558	0,010000	0,011300	0,009620	0,001770	0,007090	0,020000	0,014600	0,021600	D
D5_DQ315779	0,010500	0,011300	0,010400	0,001640	0,006780	0,019400	0,014000	0,021200	D
D6_AB493846	0,009170	0,009410	0,009090	0,000588	0,005750	0,019000	0,013300	0,020500	D
D6_AB554023	0,009330	0,009340	0,009240	0,000534	0,005660	0,019400	0,013300	0,020900	D
D7_AM494716	0,009420	0,010000	0,009150	0,001200	0,005400	0,019800	0,013500	0,022200	D
D7_FJ904430	0,010300	0,011000	0,010200	0,001690	0,005920	0,021100	0,014300	0,022400	D
E1_X75664	0,008790	0,010500	0,008850	0,005440	0,000206	0,018100	0,012200	0,020400	E
E2_X75657	0,007890	0,009840	0,008660	0,005120	0,000094	0,017300	0,011900	0,020100	E
E3_AM494694	0,008770	0,010700	0,009280	0,005390	0,000133	0,018200	0,012200	0,020800	E
E3_FJ349237	0,008200	0,009900	0,008740	0,005130	0,000092	0,017700	0,011900	0,020100	E
E4_FJ349226	0,008380	0,010100	0,009380	0,005330	0,000162	0,018000	0,011800	0,020600	E
E4_HM363569	0,008230	0,010200	0,008920	0,005550	0,000110	0,018200	0,012400	0,021000	E
E5_DQ060828	0,008620	0,010500	0,009240	0,005450	0,000146	0,018200	0,011900	0,021600	E
E5_JQ000008	0,008260	0,009990	0,008620	0,005040	0,000086	0,018000	0,012100	0,020800	E
F1_AY090459	0,019000	0,020400	0,017800	0,019900	0,017800	0,001470	0,021600	0,008060	F
F1_DQ823095	0,018400	0,019200	0,017300	0,019900	0,017100	0,001320	0,021100	0,007060	F
F1_HE981184	0,018500	0,019700	0,017600	0,020000	0,017300	0,001440	0,021400	0,007310	F
F1_HM590471	0,018600	0,019400	0,017400	0,019600	0,017000	0,001200	0,021500	0,006880	F
F1_HQ378247	0,018200	0,019100	0,017200	0,020000	0,016900	0,001330	0,020900	0,007040	F
F2_AY090455	0,020900	0,019600	0,018100	0,020000	0,019100	0,001250	0,021000	0,007330	F
F2_AY311369	0,021800	0,020300	0,018400	0,020500	0,019700	0,001240	0,021200	0,007490	F
F2_X69798	0,020700	0,019100	0,016800	0,019400	0,018300	0,000985	0,021100	0,006970	F
F3_AB036910	0,020800	0,019000	0,017300	0,019500	0,019400	0,000606	0,021600	0,006230	F
F3_AB036911	0,021400	0,019400	0,017700	0,020000	0,019700	0,000717	0,021400	0,006220	F
F3_AB036915	0,021200	0,019400	0,017800	0,020000	0,020000	0,000800	0,021300	0,006490	F
F3_FJ589066	0,021900	0,019800	0,018700	0,020500	0,020500	0,001030	0,022400	0,006760	F
F3_X75663	0,018400	0,019100	0,017100	0,019200	0,018000	0,001430	0,020900	0,007760	F
F4_AB166850	0,020900	0,019800	0,018200	0,020300	0,020200	0,001360	0,021600	0,007690	F
F4_DQ823090	0,020300	0,019300	0,017300	0,019900	0,019300	0,000664	0,021200	0,006590	F
F4_EU366116	0,020800	0,019700	0,017700	0,020100	0,019900	0,000826	0,021500	0,006810	F
F4_HE974368	0,021200	0,021100	0,018800	0,020400	0,020400	0,001400	0,023000	0,008020	F

HBV Reference Sequences Classification by DB rule
Whole genome sequences

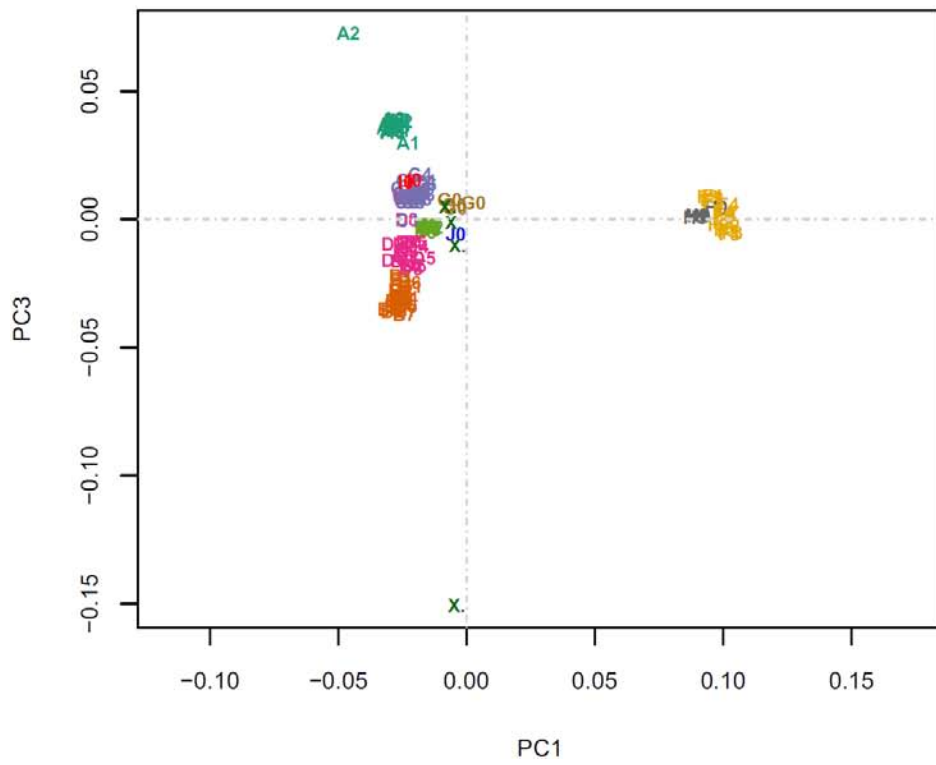
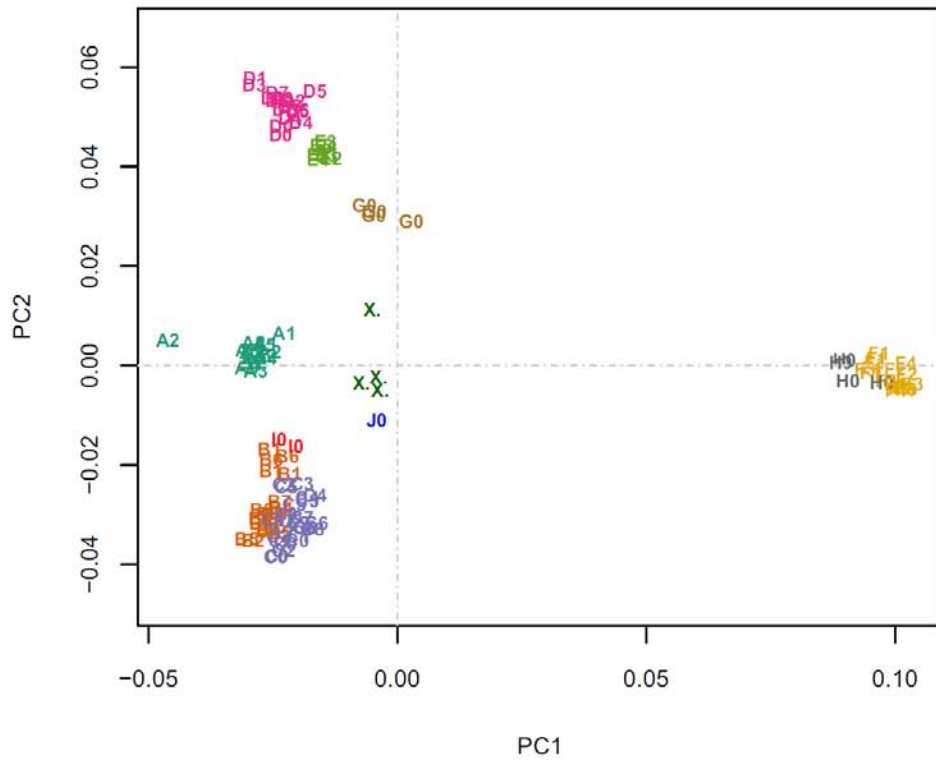
Smpl	$\Phi_A^2(k)$	$\Phi_B^2(k)$	$\Phi_C^2(k)$	$\Phi_D^2(k)$	$\Phi_E^2(k)$	$\Phi_F^2(k)$	$\Phi_G^2(k)$	$\Phi_H^2(k)$	DB Rule Type
G0_AF160501	0,011500	0,014100	0,013900	0,012200	0,011400	0,020300	0,000009	0,023300	G
G0_EF464098	0,012600	0,014900	0,014800	0,013300	0,012700	0,019500	0,000149	0,022300	G
G0_HE981172	0,012700	0,015500	0,015200	0,013400	0,012500	0,022000	0,000084	0,024700	G
G0_HE981176	0,011600	0,014100	0,013800	0,012100	0,011300	0,020200	0,000019	0,023100	G
H0_AB179747	0,020000	0,019300	0,018000	0,019300	0,019200	0,005360	0,022200	0,000044	H
H0_AB275308	0,020400	0,019700	0,018300	0,019600	0,019400	0,005540	0,022600	0,000091	H
H0_AB516395	0,024500	0,023900	0,021800	0,024100	0,023800	0,008190	0,026000	0,000817	H
H0_AP007261	0,020900	0,020100	0,018600	0,020500	0,020600	0,006010	0,023200	0,000240	H
I0_FJ023660	0,005270	0,006840	0,003590	0,008840	0,009220	0,017600	0,011900	0,019300	I
I0_FJ023664	0,005640	0,006990	0,003690	0,008810	0,009190	0,018400	0,011600	0,021200	I
J0_AB486012	0,014400	0,012000	0,011300	0,015100	0,013700	0,020700	0,017000	0,022500	C
X.chimpanzee AF222323	0,009060	0,009410	0,008670	0,009010	0,008040	0,016600	0,012000	0,019200	E
X.gibbon AB037927	0,011700	0,011700	0,010700	0,013300	0,012500	0,019800	0,016500	0,023600	C
X.orangutan AF193863	0,012600	0,011700	0,010200	0,012900	0,012300	0,019000	0,017700	0,022000	C
X.wmonkey AF046996	0,067800	0,063100	0,063600	0,065100	0,064200	0,071900	0,068000	0,075700	X

$$\hat{\Phi}_I^2(k) = \frac{1}{n_I} \sum_{i \in I} d_{i,k}^2 - \frac{1}{2n_I^2} \sum_{i,j \in I} d_{i,j}^2$$

UPGMA tree (raw): HDV whole genome



MDS map (K80): HBV WholeGenome



Discusión

5. DISCUSIÓN

El VHB es un miembro de la Familia *Hepadnaviridae* cuyo genoma consiste en una única molécula de ADN circular bicatenario parcial, dentro de la nucleocápside adopta una conformación relajada circular (ADNrc) y dentro del núcleo del hepatocito adopta una conformación circular covalentemente cerrada (ADNccc), que se pliega como un minicromosoma junto con histonas y proteínas no histonas y permanece durante toda la vida de la célula. Posee cuatro marcos de lectura abiertos (ORFs) altamente solapados; preS/Superficie (preS/S), Polimerasa (P), preCore/Core (PC/C) y X, cuya transcripción por la ARN polimerasa II del hospedador da lugar a cinco ARNm; preS1, preS2/S, pregenómico (ARNpg), preCore (ARNpc) y X, bajo el control de cuatro promotores virales (preS1p, Sp, Cp; que incluye el PBC y Xp), que se traducen en siete proteínas; tres tipos de proteínas de superficie (SHBs, MHBs y LHBs), la polimerasa viral (P), y los antígenos/proteínas e (HBeAg), core (HBcAg) y X (HBx) [31]. El ORF preCore/Core tiene dos codones de inicio en fase que sintetizan dos proteínas co-terminales desde diferentes transcritos, la proteína preCore/Core que es la precursora del HBeAg, usando el primer codón y el ARNpc como transcrito, y la proteína core desde el segundo y el ARNpg como transcrito. La proteína core (HBc) forma la nucleocápside y es muy inmunogénica y el gen core es la única parte no solapada del genoma y se considera la más tolerante a la variabilidad. HBeAg es una proteína soluble no estructural, con funciones inmunoregulatoras que facilita el establecimiento de la infección crónica y es clínicamente relevante, ya que es un índice de la replicación viral, inefectividad, inflamación, severidad de la enfermedad y respuesta al tratamiento y normalmente la seroconversión anti-HBe indica el control de la infección por el sistema inmune [31–34]. El genoma incluye también múltiples elementos reguladores; el PBC, el ENHI y II, cuatro cajas TATA-like y las DR1 y DR2, todos ellos agrupados en la región donde el gen X y el preCore se solapan y que hace que esta región tenga una gran relevancia clínica. El ARNpg, además de ser el transcrito para P y HBcAg, hace de molde para la transcripción reversa por la cual se replica el VHB [44], su señal de encapsidación ϵ es esencial para la unión de la ADN polimerasa viral [20], la cual carece de actividad correctora de errores exonucleasa 3'-5', lo que genera una alta tasa de mutación ($3,2 \times 10^{-5}$ – $7,9 \times 10^{-5}$ nt sustituidos/ciclo replicativo) que hace que el virus circule como una cuasiespecie (QS) [162].

La naturaleza de QS del VHB le permite adaptarse rápidamente a los cambios del ambiente [149], por tanto el estudio de la dinámica de QS es importante para entender la adaptabilidad, el poder patogénico y la persistencia de los virus [149]. Cómo de diferentes son las variantes genéticas que la componen, es decir la complejidad de la QS, se puede cuantificar

mediante los índices de diversidad, y se ha visto que puede influir en el potencial patogénico, la evolución clínica, la seroconversión a anti-HBe o la respuesta al tratamiento antiviral [165,167–170]. Los principales índices de diversidad son la entropía de Shannon normalizada (Sn), que mide la diversidad de la QS teniendo en cuenta la cantidad de variantes que se han encontrado y la frecuencia relativa entre ellas, pero es insensible al número de mutaciones, la frecuencia de mutación (Mf), que es la proporción de sitios mutados en una población de genomas virales con respecto a la secuencia dominante, y la diversidad nucleotídica (Pi), que mide el promedio de nucleótidos diferentes entre dos genomas de la QS.

Además de la variabilidad en forma de QS, el VHB se ha clasificado en ocho genotipos bien caracterizados (A-H) y dos más propuestos (I-J) cuando la divergencia en el genoma es >8% [134]. Su distribución geográfica y étnica está bien caracterizada, en España, el 90% de los casos corresponden a Gen/A y D, en proporciones similares, con un 6% de Gen/F [194]. Además se han descrito fenómenos de recombinación intergenotípica, deducida por las controversias en el análisis filogenético de diferentes zonas del genoma [148]. El mecanismo de recombinación no se conoce pero se considera que el primer paso es la co-infección con varios genotipos, de hecho los recombinantes tienden a tener la misma distribución geográfica que los genotipos parentales. En Europa Central y del Este se han descrito infecciones mixtas (Gen/A+D, Gen/A+D+F y Gen/D+F) en el 8% de los individuos infectados [146], en España se encontró una tasa considerable (22%) de infecciones mixtas (Gen/A+D y Gen/A+D+F) en pacientes tratados con antivirales orales análogos de nucleós(t)idos (NUCs) [194,198]. La clasificación genotípica es importante ya que cada vez existen más evidencias de que los genotipos, incluso las infecciones mixtas y los recombinantes, presentan diferente patogénesis y respuesta al tratamiento [135,136,150].

La infección crónica por el VHB es uno de los principales problemas de salud del mundo, se estima que 2.000 millones de personas han estado infectadas en algún momento de su vida y que 240 millones los están en la actualidad, causa el 30 y 53% de las cirrosis y HCC, respectivamente y 650.000 muertes al año [99,100]. Para su tratamiento existen cinco NUCs; Lamivudina (LAM), Telbivudina (LdT), Adefovir dipivoxil (ADV), Tenofovir disoproxil fumarato (TDF) y Entecavir (ETV), la principal desventaja de estos tratamientos es la selección de mutaciones en el gen de la polimerasa que confieren resistencias, aunque este fenómeno es poco frecuente con ETV y TDF, que son las opciones terapéuticas recomendadas actualmente. El objetivo de estos tratamientos es inhibir la replicación del virus de forma mantenida, ya que la infección crónica por el VHB no puede erradicarse completamente por la persistencia del ADNccc en el núcleo de los hepatocitos infectados [122]. Así pacientes con actividad viral nula, que incluso han negativizado el HBsAg y en los que se detecta anti-HBc, pueden presentar una

reactivación de la infección si son tratados con inmunosupresores [205]. Este hecho evidencia la persistencia de la infección incluso cuando los criterios serológicos y virológicos parecen indicar que esta estaba curada, y sugiere la presencia de ADNccc en los hepatocitos incluso décadas después de la supuesta curación de la infección. Para explicar esta situación no se puede descartar la existencia de una replicación de muy bajo nivel, que permita la infección de nuevos hepatocitos, pero que no pueda detectarse ni a nivel serológico ni virológico, similar a lo que ocurre en la IOB [104]. Pacientes que han estado en contacto con el VHB, unos 2.000 millones de personas según estimaciones de la OMS [99], lo seguirían estando aunque sus datos serológicos (HBsAg negativo, ADN-VHB indetectable, anti-HBc y anti-HBs positivos) indiquen que ha superado la infección.

Debido al grave problema de salud que supone la infección por el VHB, es importante el estudio de la complejidad y dinámica de la QS y las variantes a nivel de genotipos, ya que pueden dar información relevante sobre la patogénesis y la respuesta del virus al tratamiento, por lo que en el primer estudio de esta tesis se ha analizado la complejidad a nivel de la población individual (QS) en la región del genoma más susceptible a la variabilidad por la menor restricción por solapamiento (PC/Core/PBC). El estudio de la variabilidad a nivel de población viral general (genotipos) se ha abordado en el segundo estudio, para ello se han utilizado dos regiones; (1) X/PC, donde se encuentran muchos de los elementos reguladores y (2) P/S, donde el gen S se solapa con el dominio RT de la polimerasa y que es la utilizada normalmente para el genotipaje con el método de hibridación reversa INNO-LiPA en los laboratorios asistenciales. El hecho de analizar dos regiones diferentes en las mismas muestras pretendía detectar posibles diferencias en la distribución de genotipos entre ambas como evidencia del fenómeno de recombinación intergenotípica.

5.1 Técnicas para el estudio de la variabilidad

Tradicionalmente el estudio de QS se ha basado en el clonaje bacteriano y secuenciación, pero se requiere un elevado número de clones (más de 100) para el análisis, lo que la convierte en una técnica muy tediosa. Incluso llevando a cabo este procedimiento con centenares de clones (difícilmente asumible por el tiempo necesario y el coste económico), tendría poca sensibilidad para la detección de componentes minoritarios de una población viral, compuesta por billones de partículas virales, y el estudio de su complejidad mediante los índices de diversidad indicados previamente. Con la llegada de las técnicas de NGS esta limitación se ha superado y permiten obtener un número muy elevado de secuencias de una

molécula individual de ADN (desde cientos/miles a millones de *reads* en una misma muestra) [184]. No obstante estas tecnologías presentan una serie de limitaciones:

1. La posible recombinación artefactual durante el proceso de PCR. El elevado número de fragmentos amplificados y su elevada identidad favorecen que estos se entrecrucen e hibriden parcialmente, facilitando que la polimerasa produzca amplicones híbridos, es decir, recombinantes. Este fenómeno es inherente a cualquier metodología de PCR y debido a la sensibilidad actual de todas las plataformas de NGS para genomas virales se requieran etapas de PCR en la creación de las librerías, ya que la cantidad de material genético que se puede obtener en una infección viral es limitada, incluso en viremias muy elevadas como las que se pueden observar en algunas infecciones por el VHB. No obstante, este fenómeno es muy poco probable para componentes muy minoritarios de la QS e incluso prácticamente omisible cuando se pretende estudiar distribuciones genotípicas, como ha sido el caso de nuestro segundo estudio, sobre todo por el nivel de sensibilidad al que se ha realizado. Las nuevas tecnologías de NGS (conocidas como NNGS o NGS de tercera generación) permiten el estudio de fragmentos de ADN de longitudes significativamente mayores y no requieren de amplificaciones clonales como la emPCR, pero siguen necesitando de la amplificación de los genomas virales como fase previa, por lo que el riesgo de recombinación se mantiene.

2. La PCR puede sesgar la frecuencia relativa de las variantes y cualquier estudio de detección de variantes minoritarias, como podría ser el estudio de variantes que confieren resistencia al tratamiento con NUCs, debería tenerlo en cuenta. No obstante, aquí el objetivo no ha sido detectar variantes minoritarias sino describir la complejidad mediante índices poblacionales, que no se afectan por variantes o haplotipos en frecuencias bajas, por lo tanto, los estudios no se han visto afectados por esta limitación.

3. Imprecisión en regiones homopoliméricas, especialmente la UDPS usada en esta tesis presenta este inconveniente. Los resultados de los estudios no han tenido en cuenta la presencia de variantes específicas localizadas en dichas regiones por lo que esta limitación tampoco es valorable. No obstante, el desarrollo futuro de estos estudios (indicado al final de esta discusión) sí que requiere valorar los resultados en regiones homopoliméricas, como por ejemplo la región rica en timinas (TTTTT) localizada entre los nt 1821-1825. Para abordar esta situación se debe aumentar la cobertura (número de *reads* por muestra) y refinar el algoritmo bioinformático de análisis de los datos de secuenciación.

En general todas estas limitaciones se pueden detectar y controlar por el desarrollo de potentes algoritmos bioinformáticos. En este sentido, el desarrollo de herramientas propias para el filtraje de los datos ya había sido validado previamente en otros estudios [171,173,184,185,206,207].

En cuanto a las técnicas para el genotipaje, el *gold standard* es la secuenciación y análisis filogenético del genoma completo. Como esto no es factible en los laboratorios clínicos se ha aceptado secuenciar una región concreta del genoma que refleje lo mejor posible el estudio del genoma completo, como se hace por ejemplo con el VHC en el que se ha consensuado el estudio de un fragmento de la región NS5B para el subgenotipaje [207]. En el caso del VHB, no existe un consenso sobre la región a analizar, pero se considera que la región preS1/preS2 muestra un genotipaje similar al del genoma completo [135]. No obstante, la principal desventaja de la secuenciación directa (sin clonaje previo), tanto del genoma completo como parcial, es su incapacidad para detectar mezclas de genotipos y recombinantes.

Para la aplicación asistencial del genotipaje del VHB se han desarrollado múltiples tecnologías indirectas, una de las más ventajosas en términos coste-beneficio es la hibridación reversa INNO-LiPA, muy sensible, barata y además capaz de detectar mezclas. Sin embargo esta tecnología analiza la región P/S, que ni siquiera es la más precisa, este hecho se debe a que la casa comercial que la desarrolló aprovechó el método de hibridación reversa que tenía desarrollado para la detección de resistencias a antivirales en RT. Para ello diseñaron un conjunto de sondas, que fijadas a una matriz sólida (tiras), permitían la clasificación del genoma del VHB en los ocho genotipos conocidos en aquel momento con bastante buena aproximación a los estudios filogenéticos de genoma completo [200]. La gran difusión de este procedimiento, así como la gran actividad a nivel mundial de estudio de variantes de resistencia a tratamientos antivirales en la polimerasa viral, ha convertido esta tecnología, y la región P/S estudiada, en un método de referencia para clasificar el VHB. Además en el momento en el que se desarrolló esta metodología tampoco se disponía de mucha información sobre la complejidad de las poblaciones virales en una infección por el VHB, con lo que cualquier variante en el genoma puede afectar a los resultados de la hibridación con las sondas estándar para los genotipos. Por este motivo, se puede deducir que muchos de los estudios que han intentado asociar genotipo y patogénesis o respuesta al tratamiento pueden estar partiendo de un genotipaje incorrecto y de ahí que los resultados sean contradictorios.

5.2 Complejidad de la QS y seroconversión anti-HBe

En el estudio de la complejidad de QS en la región PC/Core/PBC mediante los índices de diversidad Sn, Mf y Pi en pacientes con HBC y fallo del tratamiento con NUCs, se ha encontrado una mayor variabilidad en pacientes HBeAg(-) que HBeAg(+), evidenciando que el aumento de la diversidad está asociado a la seroconversión y confirmando estudios previos [170]. En este sentido, debemos tener en cuenta que HBeAg altera la respuesta inmune del hospedador facilitando el establecimiento de la infección crónica. Esta proteína parece regular de manera negativa la respuesta de cT a los antígenos de la nucleocápside por una variedad de mecanismos, por ejemplo, se ha sugerido que los Th2 específicos contra HBeAg podrían regular de manera cruzada los Th1 específicos contra HBcAg o que HBeAg podría actuar preferentemente como un tolerógeno e inactivar las cT específicas contra HBcAg [208]. Así, parece lógico pensar que en aquellos pacientes en los que el efecto inmunomodulador de HBeAg está ausente se desarrollaría una mayor presión del sistema inmune. Debido a que la región PC/Core es la más susceptible a la variabilidad por la ausencia de restricción por solapamiento, esta presión del sistema inmune induciría la aparición de variantes creando un repertorio de genomas que hace que la QS aumente en heterogeneidad. Ante un cambio en el ambiente, el VHB tendría más opciones de seleccionar una variante capaz de sobrevivir en este nuevo ambiente replicativo.

A nivel de aa, se ha visto una mayor complejidad en pacientes con estatus HBeAg fluctuante en la región core cuando el VHB evoluciona en la dinámica natural de la QS bajo la presión del sistema inmune y en ausencia de presión externa, lo que sugiere que el sistema inmune ejercería una mayor presión sobre todo en el gen core en pacientes con estatus cambiante de HBeAg, la proteína HBc es mucho más inmunogénica que HBeAg y tiene un papel inmunoestimulador, induce una respuesta más fuerte de cB, cT y LCT y actúa como antígeno T dependiente o independiente [208]. Esto sumado a la pequeña tendencia que se ha observado a acumular variantes en las principales regiones epitópicas del core en estos pacientes y también en pacientes HBeAg(-), indicaría que el sistema inmune estaría actuando principalmente contra los epítomos del core, la QS aumentaría su complejidad en esa zona, dando lugar a variantes en HBc, sugiriendo un mecanismo potencial adicional del VHB para escapar de la respuesta inmune del hospedador (además de atenuar la respuesta con HBeAg) que favorecería la persistencia del virus.

5.3 Complejidad de la QS y actividad viral

De este estudio también se deduce que la actividad viral y complejidad son inversamente proporcionales, la complejidad en la región preCore a nivel de aa presentó una correlación negativa estadísticamente significativa con los niveles de ADN, también en la región core se observó que a mayor complejidad, menor carga viral pero sin llegar a la significación estadística, quizás debido a que la alta carga viral de las muestras esté actuando como factor de confusión.

Se puede asumir que cuando aumenta la respuesta inmune la carga viral disminuye por la eliminación de las células infectadas, que son reconocidas por el sistema inmune por expresar antígenos virales de las variantes mayoritarias de la QS, que han llegado a un cierto nivel de predominancia por tener un *fitness* elevado en el ambiente existente. Es decir, el sistema inmune ataca una QS con unas pocas variantes, poco compleja, pero con un alto *fitness*, que se refleja en una alta carga viral, de ahí la relación inversa entre complejidad de QS y carga viral observada en este estudio. Este reconocimiento y acción del sistema inmune disminuye la población de células hepáticas infectadas y, por lo tanto, al haber una menor producción de partículas virales, los niveles de ADN-VHB disminuyen. Pero la misma presión del sistema inmune lleva a cabo una selección de mutantes de escape que re-infectan nuevas células, generando diversidad viral. Estas variantes probablemente presenten menor *fitness*, es decir menor capacidad replicativa, que las variantes principales que infectaban las células que han sido eliminadas por la acción del sistema inmune, ya que de haber tenido un *fitness* mayor hubieran sido las variantes mayoritarias en la situación basal. En esta nueva situación, las variantes minoritarias presentan un *fitness* bajo y el número de hepatocitos infectados es menor, lo que resulta en una población viral pequeña (niveles de ADN-VHB bajos) pero muy compleja, reflejando de nuevo la relación inversa observada en nuestro estudio entre carga viral y complejidad de QS.

Por el contrario cuando la respuesta inmune es baja, hay poca presión selectiva con lo que se observan pocas mutaciones adaptativas, pero el virus puede replicarse mejor y el tamaño de la población aumenta, creándose un ambiente de baja complejidad en la población viral a pesar del enorme tamaño de la misma. Este sería el escenario de los pacientes HBeAg (+), o sea de los que no seroconvierten a anti-HBe, la carga viral es alta pero la diversidad de la QS es baja y típicamente tiene una respuesta inmune débil. Esta situación parece una consecuencia clara del efecto inmunomodulador de HBeAg, que atenúa la acción del sistema inmune, y su actividad como factor evolutivo generador de complejidad viral. Por el contrario, en pacientes que seroconvierten, lo que quiere decir que desaparece o se hace imperceptible

la actividad inmunomoduladora de HBeAg, la complejidad es alta y se produce una reducción de la carga viral, como se ha discutido en el párrafo anterior. Ocurriría la situación inversa al caso anterior de forma que la ausencia de HBeAg haría que la presión del sistema inmune aumentara generando una mayor heterogeneidad en la QS pero impidiendo que el virus se replique, por la eliminación de hepatocitos infectados, reduciendo la carga viral. En resumen, una carga viral elevada no significa necesariamente una mayor complejidad, sino que una variante especialmente eficaz para la replicación, con un mejor *fitness* en unas determinadas condiciones, puede convertirse en la predominante, llevando a una población de virus muy grande pero muy homogénea.

Las causas por las que la heterogeneidad en el preCore y en el core afectarían a la replicación podrían ser, por un lado el preCore se solapa con el PBC, que es un elemento regulador esencial y dirige la transcripción del ARNpg y ARNpc, y también en esta región se encuentra la señal de encapsidación ϵ , esencial para el inicio de la replicación del virus y, por otro lado, la proteína core, además de empaquetar el ARNpg para iniciar la retrotranscripción, es un componente estructural del ADNccc, su unión resulta en una reducción de espacio nucleosomal de 200 a 180 pb y aparentemente tiene un efecto regulador positivo en la transcripción y replicación, manteniendo un estado epigenético permisivo en él [89,209,210]. Por lo tanto las variantes en la región PC/Core podrían alterar sus funciones en la replicación, lo que llevaría a una reducción en los niveles de ADN-VHB, la población viral es pequeña pero muy compleja.

5.4 Descenso de la complejidad asociada al fallo del tratamiento

Al comparar los índices de diversidad obtenidos en las muestras basales, en las muestras tras un tiempo sin tratamiento, que reflejan la dinámica natural de la QS sin la presión de factores externos, y en las muestra obtenidas tras el fallo de LAM, que reflejan la evolución de la QS bajo la presión del tratamiento antiviral, se ha observado que eran significativamente más elevados en la primera muestra respecto a la tercera y en la segunda respecto a la tercera (aunque no significativamente) y similares entre la primera y segunda ($1^a \approx 2^a > 3^a$). Lo que en conjunto sugiere un descenso de la complejidad de la QS asociado al fallo del tratamiento, es decir, la población se hace menos compleja. Por un lado, se ha visto que cuando un tratamiento controla la infección por el VHB la respuesta inmune aumenta [196], pero aquí la situación sería la contraria, al fallar el tratamiento la respuesta del sistema inmune seguiría atenuada, esta menor presión de evolución provocaría una reducción de la complejidad. Por otro lado, si el aumento de la respuesta inmune por la acción del tratamiento

fuera significativo, los tratamientos con NUCs darían lugar a un descenso de la población de hepatocitos infectados evidenciado por una reducción del ADNccc, que permitiría prever un límite temporal a la aplicación de los mismos. Sin embargo, esta situación se produce con una baja intensidad por lo que se asume la duración ilimitada de los tratamientos y su escasa o nula capacidad curativa [122]. Como evidencia clínica de este hecho está la bajísima tasa de pérdida de HBsAg observada en pacientes tratados con NUCs, incluso con los más eficientes (ETV y TDF) [122]. Entonces durante el tratamiento con antivirales no parece previsible un proceso de aumento de complejidad asociado al descenso de la carga viral por la presión del sistema inmune, por un lado por lo expuesto en las líneas anteriores y por otro porque este descenso de la carga viral no es debido a la seroconversión como ocurre en la dinámica natural y que se ha visto en el apartado anterior, sino a la inhibición de la RT. La presión del tratamiento conlleva la selección de variantes virales resistentes o menos sensibles a dicho tratamiento por la presencia de mutaciones en la región RT de la polimerasa, por un proceso tipo “cuello de botella” asociado a una homogeneización de la complejidad de la QS. Este descenso de la complejidad se observaría en cualquier región del genoma que se analice, como se ha hecho en nuestro estudio en la región PC/Core/PBC estudiada.

La confirmación de este fenómeno requeriría del análisis por NGS de ambas regiones en una misma secuencia, actualmente no accesible de forma directa por la limitación de longitud de los fragmentos que se pueden secuenciar (unos 500 pb). De forma indirecta sería posible mediante la circularización de fragmentos de PCR de mayor tamaño como ya se hizo en una tesis anterior de nuestro grupo [206], no obstante este método incluye muchas manipulaciones que podrían afectar a la complejidad y solo analizaría unos 200 pb por región estudiada, con lo que el estudio de complejidad no sería fiable. El análisis por separado de ambas regiones se ha llevado a cabo en el segundo estudio pero no se ha valorado el efecto selector de una región sobre otra, precisamente porque se ha tratado de un análisis de las regiones por separado. Para estudiar adecuadamente cómo realmente la variación en una región afecta a la otra se debería analizar un fragmento que abarcara ambas a la vez y que se espera que sea posible con la nueva plataforma de NGS (SMRT-Avenio Z1), disponible en nuestra institución y para la que ya se están diseñando los próximos experimentos de continuación de esta tesis.

Resulta muy interesante que al comparar la complejidad de la QS en ambos períodos (período sin tratamiento y período de tratamiento con LAM) se encontró una correlación negativa, es decir, cuanto más compleja es la QS al final de la dinámica natural del virus, más homogénea es la población después del tratamiento. Este resultado sugiere que los cambios que ocurren durante la evolución natural podrían estar afectados por la respuesta inmune y

determinar la evolución de la misma región bajo el tratamiento con NUCs. Como se ha sugerido previamente, la causa de la simplificación la QS en el tratamiento podría implicar un proceso de tipo “cuello de botella” de selección de una o unas pocas variantes de la QS resistentes o menos sensibles a la acción del tratamiento. Durante la dinámica natural la presión de la respuesta inmune hace que la QS varíe y se haga más compleja, dotando a esta población viral de más recursos para sobrevivir ante otra presión externa, cuanto más amplio sea el número de secuencias disponibles, más probable es que exista entre ellas alguna(s) variante(s) que sean resistentes o menos sensibles al tratamiento. Durante el tratamiento con NUCs, LAM en este estudio, la QS se simplificaría porque una variante es seleccionada convirtiéndose en predominante, aquella que confiere un mejor *fitness* en presencia del antiviral, que por otro lado inhibe a las formas mayoritarias en la QS pre-tratamiento.

Debe tenerse en cuenta que en esta tesis el estudio de complejidad se ha efectuado en la región PC/Core/PBC, que no es donde el tratamiento ejercería la presión evolutiva, ya que actuaría sobre el dominio RT, pero ambas están en el mismo genoma del VHB de forma que la simplificación de la QS por variantes seleccionadas en la región RT se podría evidenciar en cualquier otra región del genoma, en nuestro caso la que hemos observado en la región PC/Core/PBC. Nuestros datos parecen confirmar esta hipótesis al detectar una relación inversa de complejidad de la QS; cuanto más alta es esta complejidad en el período de evolución natural, más baja es durante el período de tratamiento. No obstante, como se ha dicho anteriormente, se requiere el estudio del fragmento completo que abarque ambas regiones para dilucidar cómo la variación en el PC/Core/PBC influye en la presión que ejercen los NUCs en la región RT, analizando la posible selección de variantes que justifiquen el proceso selectivo de estos tratamientos.

En el segundo estudio no se observó un patrón de evolución claro de la complejidad de la QS, esto puede deberse a que las regiones estudiadas no son las mismas (X/preCore y P/S), a que la tercera muestra no corresponde al mismo momento del tratamiento (en el primer estudio es una muestra tomada en el momento del fallo del tratamiento y en el segundo en la respuesta parcial) y a que el fragmento X/preCore analizado en este segundo estudio presentaba una alta tasa de solapamiento, para poder ser comparado con el fragmento P/S, por lo que la complejidad está más restringida que en el primer estudio. Lo que si se observó es que los dos índices de diversidad calculados (M_f y P_i) eran significativamente mayores en la región X/preCore que en P/S, esto está de acuerdo con el mayor grado de conservación de RT y que es poco tolerante a los cambios, por su solapamiento con S y por los requerimientos de funcionalidad de ambas regiones. Por otro lado los posibles efectos selectores de variantes en la región P/S por la acción del tratamiento antiviral afectarían a posiciones muy concretas de

los dominios catalíticos en una proporción importante de los haplotipos posiblemente no afectando a la complejidad global de la QS mediante índices como M_f o P_i . En cualquier caso en el VHB, a diferencia de otros virus como el VIH o VHC, el alto grado de solapamiento del genoma lo hace menos susceptible a la variabilidad, ya que un cambio en un ORF podría afectar a otro negativamente o incluso a una región reguladora, que también están solapadas con los ORFs. Aun y todo, el VHB “ha conseguido manejar bien esta situación” y en algunos casos una misma mutación puede afectar a dos proteínas distintas como ocurre en la región X/PC, estas dos proteínas tolerarían los cambios porque supondrían una ventaja evolutiva como puede ser disminuir la expresión de HBeAg y favorecer el proceso de seroconversión. Este sería el caso de la doble mutación en el PBC: A1762T y G1764A que reduce la expresión de HBeAg y que al solapar con el gen X, causa el doble cambio en HBx: K130M y V131I, cuya posible ventaja o inocuidad para ciclo replicativo del VHB no se conoce todavía [47]. A pesar de que unos cambios muy concretos pueden no afectar a la complejidad de la QS, como se ha comentado en el caso de la región P/S, los resultados de la mayor complejidad de QS en X/PC que en P/S, parecen indicar que esta región X/PC tolera mejor un mayor número de cambios que P/S, probablemente porque esta mayor complejidad le confiere ventajas evolutivas. Estas ventajas podrían ser debidas a la generación de formas alternativas de HBx con diferencias funcionales que participarían en la compleja tarea reguladora de esta proteína. Esta posibilidad se está desarrollando en la continuación de esta tesis y se muestran algunos datos preliminares al final de esta discusión. Además, aunque no se expuso en el artículo correspondiente, se encontró una correlación negativa entre la complejidad en P/S y el ADN-VHB, lo que de nuevo pone de manifiesto que una población viral grande puede implicar unas pocas variantes pero con un gran *fitness* para la replicación. En este segundo estudio no se utilizó la S_n , porque gracias a un análisis más profundo de la utilidad de los índices de diversidad en Virología se ha visto que no tiene en cuenta la heterogeneidad de las diferentes variantes de la QS porque es insensible al número de mutaciones, con lo cual una población viral constituida por dos haplotipos al 50%, con sólo una sustitución de diferencia entre ellos, tendría la misma S_n que una población con 100 haplotipos al 1% cada uno y una media de 10 diferencias entre ellos [171], y obviamente no son igual de complejas.

5.5 Utilidad clínica de los índices de diversidad

La complejidad de la QS es un predictor de la progresión de la enfermedad o la respuesta al tratamiento, en el primer caso, el rango de complejidad de QS confiere una gran adaptabilidad al virus ante la respuesta del sistema inmune que tiene implicaciones en la

patogénesis viral, ya que ciertas variantes pueden permitir la persistencia de la replicación viral, en cierto modo su supervivencia, y en cierto modo también es responsable de las dificultades en la prevención y control de la enfermedad. En el segundo caso, refleja la barrera genética de resistencia, con lo que la heterogeneidad de la QS podría ser importante para determinar la eficacia del tratamiento. Por tanto adquirir un conocimiento preciso de la complejidad es un reto actual para el manejo de los pacientes y estudios recientes apoyan que la complejidad es un factor clínicamente relevante en el curso y pronóstico de la enfermedad y su respuesta al tratamiento [165,170,195] y nuestros resultados parecen confirmarlo. El aumento de la diversidad está asociado a la seroconversión y la complejidad presenta una correlación negativa antes del tratamiento y después del fallo del mismo, esto sugiere la importancia de estudiar la complejidad de QS antes de iniciar el tratamiento y así usarla como un posible factor predictivo de la evolución del virus en el caso de no respondedores a NUCs.

No obstante se debe asumir lo excepcional de que un paciente no responda a un tratamiento antiviral con los fármacos recomendados actualmente (ETV y TDF), sin embargo hay que tener en cuenta que la eficacia para disminuir la carga viral hasta límites indetectables a los 12 meses del inicio del tratamiento es del 70% en pacientes HBeAg(+) y del 90% en pacientes HBeAg(-) [122], por lo que existe una proporción importante de pacientes que se podrían beneficiar de los datos de complejidad para optimizar su abordaje terapéutico.

Este proyecto de tesis doctoral parece confirmar la utilidad de la secuenciación masiva por UDPS para el estudio de complejidad de QS y su utilidad clínica, con lo que gracias a la consolidación de los métodos de NGS en los laboratorios clínicos, se podría llevar a cabo un estudio sistemático y preciso de la QS para calcular los parámetros de complejidad y así aplicarlos al manejo clínico de la infección por el VHB. El uso asistencial de las técnicas de NGS requiere de validación clínica y analítica, monitorización y documentación del control de calidad y evaluación del método como cualquier otra técnica, con lo que la ACMG (*American College for Medical Genetics and Genomics*) ha publicado los requerimientos de documentación, validación y control de calidad tanto del trabajo de poyata como bioinformático para el uso de NGS en los laboratorios clínicos [211].

5.6 Complejidad del VHB: Mezclas de genotipos

Mediante UDPS se ha evidenciado una alta prevalencia de infecciones mixtas que implican los Gen/A, D y F, los tres principales que circulan en España [194]. En la región P/S los resultados globales son similares a los del genotipaje por INNO-LiPA, pero la UDPS muestra una mayor sensibilidad para la detección de infecciones mixtas. El genotipaje mediante INNO-

LiPA es un método indirecto de hibridación con sondas prefijadas, con lo que cualquier variante en el genoma puede alterar los resultados, en consecuencia las mezclas de genotipos podrían haber estado subestimadas en otros estudios por la baja sensibilidad de los métodos utilizados hasta ahora. En cuanto a la región X/PC todas las muestras menos una presentaron mezclas, la mayoría implicando a los tres genotipos, estos resultados, sumados a la alta complejidad de esta región, incluyendo la DR1 y el gap de la hebra (-), la convertirían en un *hot spot* para la recombinación y llevarían a la interpretación incorrecta de los resultados como una infección mixta. De hecho la región entre las posiciones nt 1600-2000 (cerca de la región DR1) ya ha sido descrita como una de las más comunes para los *breakpoints* implicados en la recombinación inter-genotípica [138], la mayor prevalencia de la reportada en España de Gen/F en X/PC podría indicar que esta región es un *hot spot* especialmente para la recombinación del Gen/F.

Al poder cuantificar estas mezclas de genotipos mediante UDPS, y no sólo detectarlas como ocurre con el método de hibridación reversa INNO-LiPA, ha sido posible analizar la evolución de las mismas y se ha visto que cambian a lo largo del tiempo, lo que ha puesto de manifiesto la compleja dinámica de la QS del VHB como un mecanismo adicional para adaptarse a nuevas situaciones, como son la respuesta del sistema inmune o el tratamiento con NUCs. En la región P/S, durante la dinámica natural de la QS, se han observado variaciones en las mezclas en seis de los pacientes, tres de ellas implicando un cambio de genotipo mayoritario (Gen/D principalmente). Durante el tratamiento se han observado cambios de genotipo principal en dos pacientes (cambio a Gen/A) y cambios de proporciones en tres. Los cambios de genotipos durante el tratamiento podrían deberse a una superinfección por otro genotipo durante el mismo, situación extremadamente improbable por las historias clínicas de estos pacientes, y parece más lógico pensar que pacientes con mezclas de genotipos al inicio del tratamiento, con diferentes sensibilidades al mismo, cambian de proporciones por la presión de selección que ejerce, llevando a la selección de otro genotipo mayoritario. En conjunto los resultados de este estudio sugieren que el Gen/A es menos sensible al tratamiento con LAM, lo que concuerda con nuestras observaciones anteriores [198] y que el tratamiento antiviral es un factor importante para inducir el cambio de genotipo. Otros estudios también han evidenciado cambios de mezclas de genotipos durante el tratamiento con NUCs, en un estudio en China todos los pacientes antes del inicio del tratamiento con ADV presentaban una infección mixta Gen/B+C, después del tratamiento la proporción de Gen/B aumentaba, sugiriendo que el Gen/C era más sensible que el Gen/B a ADV [12]. En la región X/PC, por el contrario, no se han observado cambios de genotipo principal, pero sí cambios en las proporciones en todos los pacientes menos en uno, lo que sugiere que esta región es más

permissiva a los cambios que P/S cuando está bajo la presión del sistema inmune o el tratamiento, y esto concuerda con los resultados de los índices de diversidad significativamente mayores en la región X/PC que en P/S, como ya se ha comentado anteriormente. Es importante el uso de una tecnología lo suficientemente sensible para detectar las mezclas de genotipos ya que las consecuencias de la co-infección son importantes e incluyen el cambio de genotipo cuando la presión selectiva cambia y la posible formación de virus recombinantes, como se explica en el siguiente apartado.

5.7 Posible recombinación inter-genotípica

Dos resultados del segundo estudio sugieren la recombinación entre genotipos, por una lado la alta prevalencia de infecciones mixtas en la región X/PC comparado con la región P/S, en este sentido, como ya se ha dicho en el punto anterior, la alta complejidad de esta región X/PC la convertirían en un *hot spot* para la recombinación, pero sobre todo las discrepancias encontradas entre genotipos obtenidos en ambas regiones de la misma muestra del mismo paciente. Está claro que estos resultados podrían deberse a la propia variabilidad inherente a la UDPS, que utiliza un paso de PCR previo, pero la comparación de los resultados en ambas regiones en algunos pacientes muestra porcentajes claramente muy diferentes que hacen extremadamente improbable que el error en la PCR sea la causa, no obstante se deberían confirmar mediante UDPS de un fragmento más largo que abarcara ambas regiones en la misma secuencia.

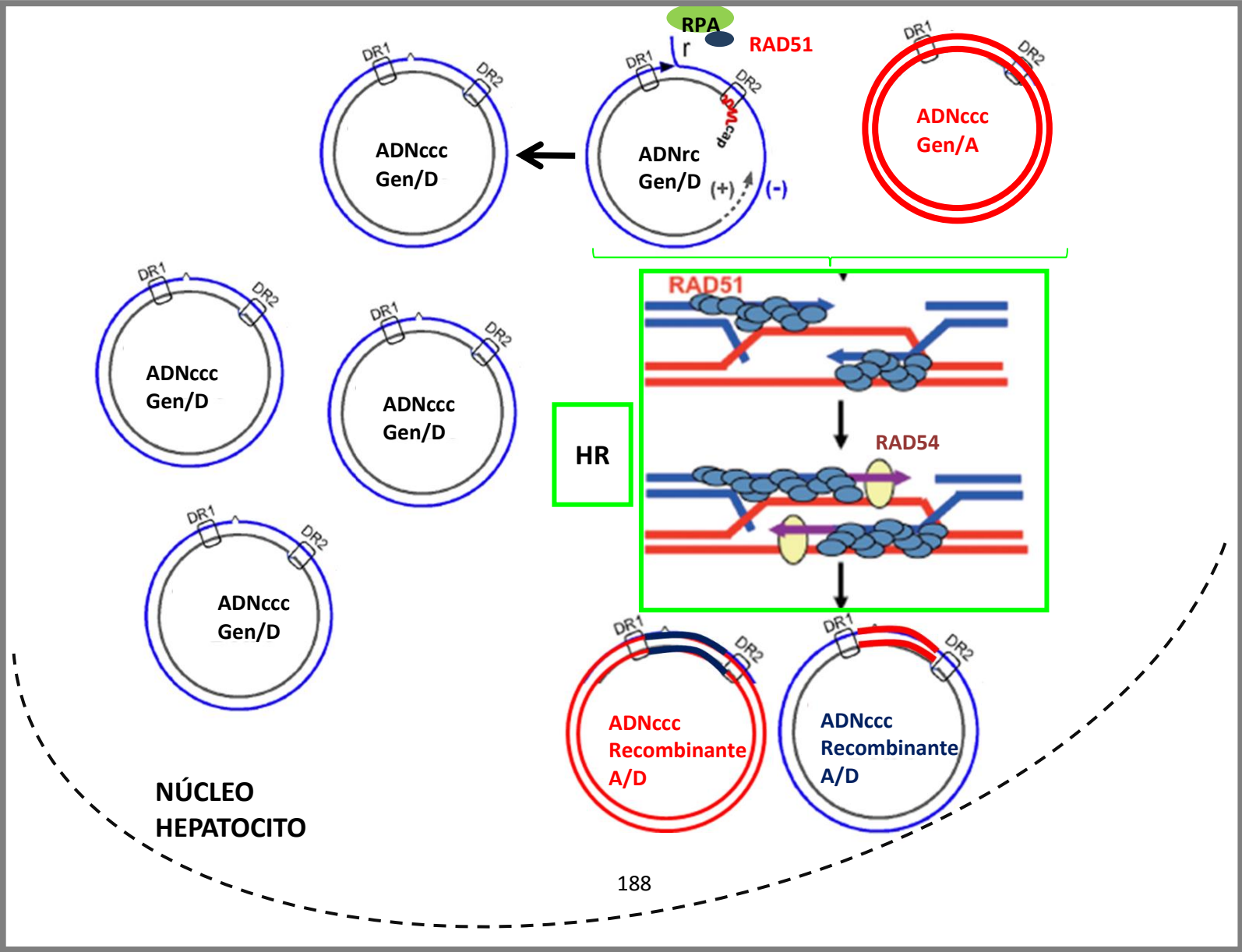
El mecanismo de recombinación no se conoce pero la circulación de los Gen/A y D en España en proporciones similares y la vida media larga de los hepatocitos infectados (de 2,4 a más de 70 días durante el tratamiento con LAM [204]) podría facilitar la infección de un único hepatocito con diferentes genotipos y favorecer el intercambio de material genético entre ambos. Parece poco probable que la recombinación ocurra durante la retrotranscripción, ya que el proceso tiene lugar en la nucleocápside implicando sólo una molécula de ARN_{pg}, con lo que probablemente ocurre en el núcleo entre moléculas de ADN_{ccc} de diferentes genotipos co-infectantes, durante el proceso de síntesis del ADN_{ccc} y por la maquinaria de reparación de ADN celular. Al igual que ocurre con otros virus como los papilomavirus o adenovirus, el VHB reclutaría esta maquinaria para usarla a su favor [212].

El ADN_{rc} podría reclutar la maquinaria de reparación debido a sus peculiaridades, la hebra(-) contiene la proteína P unida por una Tyr del dominio PT, una mella y un fragmento libre monocatenario (la redundancia terminal “r”) y la hebra(+) contiene un gap y el primer de ARN en el extremo 5'. En la célula la reparación de estos aductos implica al menos dos

mecanismos diferentes o el corte con endonucleasas o el corte del enlace fosfotirosil por tirosil-DNA-fosfodiesterasas (TDPs) específicas, esta misma actividad sería la que se llevaría a cabo durante la conversión del ADNrc a ADNccc para liberar la proteína P del ADNrc, imprescindible para el inicio de la síntesis del ADNccc, y de hecho hay evidencias de la liberación de P del ADNrc por TDP2 [84]. Una vez liberado de P, el ADNrc sería reconocido como ADN dañado y reparado por un mecanismo semejante al de la recombinación homóloga (HR), que utiliza una doble hebra homóloga para reparar las roturas bicatenarias del ADN celular DSB (Double-Strand Breaks) [147].

El mecanismo propuesto incluiría los siguientes pasos que se reflejan en la figura 37. La HR se inicia por la presencia de ADN de cadena simple como el que resulta de la rotura del ADN bicatenario o, en este caso, se iniciaría por la redundancia terminal “r” en el extremo 5’ de la hebra(-) del ADNrc. Los fragmentos de ADN monocatenarios serían reconocidos por la proteína nuclear RPA y se uniría a Rad51, esta proteína forma filamentos sobre el extremo monocatenario del ADN y entonces se produce la invasión de la cadena libre a la doble cadena homóloga, que existiría en el núcleo del hepatocito infectado debido a la presencia de más de una molécula de ADNccc. Se pueden acumular entre 1-50 moléculas de ADNccc en un misma célula [213–215], las cuales provienen mayoritariamente de moléculas de ADNrc ya que el ADNccc no se replica semiconservativamente como el ADN celular [97]. La invasión de la doble hebra homóloga por parte del ADN monocatenario recubierto de Rad51 generaría un loop al que se uniría Rad54, una translocasa dependiente de ATP. Rad 54 hace posible que el extremo 3’ “invasor” del ADNmc esté disponible para la ADNpol, entonces ocurriría la síntesis de nuevo ADN usando la cadena homóloga como molde, ligación y resolución de este loop intermedio para dar lugar a la conversión génica con entrecruzamiento [216].

Figura 37: Posible mecanismo de recombinación entre moléculas de ADNccc de diferentes genotipos que co-infectan un mismo hepatocito por un mecanismo similar al mecanismo de HR implicado en la reparación de roturas bicatenarias del ADN celular DSB (*Double-Strand Breaks*) en la que se utiliza una doble hebra homóloga como molde para la reparación.



Aunque no hay evidencias experimentales que confirmen el papel de los mecanismos de reparación en la formación del ADNccc (y por tanto en la recombinación) el hecho de que la mayoría (más del 60 %) de los genomas recombinantes descritos tienen un punto de corte entre los nt 1640 y 1900 del genoma del VHB, una región entre DR 1 y 2 que se solapa con las redundancias terminales y el punto de unión de la polimerasa viral con el ADNrc [151], apoyan esta hipótesis.

Otro mecanismo podría ser que la propia polimerasa viral llevara a cabo la recombinación, por ejemplo para la RT del VIH, con gran homología con la del VHB, se ha propuesto un modelo de elección de copia dinámico para la recombinación del VIH durante la replicación que propone un balance entre la actividad polimerasa y ARNasa H durante la retrotranscripción que influenciaría el intercambio de molde. Cuando la actividad de la polimerasa es baja (y tiene alta actividad ARNasa H) la disociación de RT de la hebra naciente de ADN aumenta y como consecuencia la probabilidad de intercambio entre hebras del molde. Por el contrario cuando la actividad ARNasaH es baja (manteniendo una alta actividad RT) las oportunidades de recombinación son bajas porque la RT está unida a la hebra en extensión [143].

Independientemente de cuál sea el mecanismo de recombinación, es un fenómeno relativamente frecuente tal y como se ha observado en el análisis filogenético a gran escala de secuencias de genoma completas del VHB de diferentes genotipos disponibles en las bases de datos públicas [148,217] y que puede tener importantes consecuencias a la hora de clasificar el VHB. Además cada vez existen más estudios que muestran la asociación entre recombinantes y patogénesis o respuesta al tratamiento [138], la recombinación es un importante elemento de variabilidad genética con posibles implicaciones clínicas ya que un solo fenómeno de recombinación el virus puede transferir varias mutaciones en una determinada región de un ADNccc a otro ADNccc que ya tenía mutaciones en otra región, creando recombinantes con una virulencia aumentada.

5.8 Importancia de una correcta clasificación genotípica

Para poder investigar el impacto de los genotipos en diferentes aspectos como la profilaxis, el diagnóstico, la progresión de la enfermedad, la respuesta al tratamiento antiviral y el pronóstico, es fundamental una clasificación única, de hecho muchos expertos creen que se debería rectificar la clasificación actual, ya que podría ser incorrecta a causa, fundamentalmente, de los fenómenos de recombinación [218]. Las asociaciones genotipo-patogénesis/respuesta al tratamiento podrían estar artefactadas porque la clasificación en

algunos estudios ha sido incorrecta llevando a conclusiones contradictorias. De hecho, hasta ahora la asociación genotipo-respuesta al tratamiento sólo se ha confirmado para el IFN [139] y se ha sugerido para los NUCs, pero es difícil poder establecer una asociación genotipo-respuesta al tratamiento si partimos de la base de que quizás no se esté genotipando bien.

Primeramente se debería consensuar una región para el genotipaje del VHB, ya que este es altamente dependiente de la región seleccionada para ello, el análisis de fragmentos de 400 pb deslizados 40 pb hasta cubrir todo el genoma (*Sliding Windows*), realizado en el segundo estudio, ha mostrado que la región preS es la de mayor poder de discriminación, en concordancia con los estudios previos en los que el análisis filogenético de la región preS1/preS2 (nt 3025-80) muestra un genotipaje similar al del genoma completo [135], además también se ha encontrado que la región P/S entre las posiciones nt 615-969 es capaz de discriminar entre los ocho genotipos principales (Gen/A-H). Por el contrario la región X/PC (nt 1596-1912) muestra una clara mezcla entre los Gen/D y E y Gen/C e I, la causa de unas secuencias tan homóloga en esta parte del genoma en las poblaciones donde son mayoritarios estos genotipos podría ser la posible ventaja evolutiva, en cierto modo, a estos genotipos “les habría convenido parecerse en esta región y no divergir”. La consecuencia en términos prácticos es que esta región no es buena para la clasificación del VHB, no obstante se han valorado los resultados obtenidos en la misma por la bajísima frecuencia de Gen/E, C o I en nuestra población, lo que garantiza que un resultado de Gen/D o E en nuestras muestras en la región X/PC corresponde a un Gen/D.

Una limitación de la filogénesis de un fragmento determinado es que puede enmascarar los fenómenos de recombinación, de los cuales se han identificado los *breakpoints* específicos dentro del genoma, concretamente cuatro fragmentos son los que agrupan los puntos más susceptibles a la recombinación. La región usada para la secuenciación y filogénesis por tener la mayor potencia preS1/preS2 (nt 3025-80) está incluida dentro de uno de estos fragmentos situado en la región preS/S (nt 3000-300), pero existen otros tres en las posiciones nt 1600-2000 (cerca de la región DR1), nt 2100-2500 (extremo 3' del gen core) y nt 700-1000 (extremo 3' del gen S) [138], con lo cual muchos de los resultados de genotipaje en diferentes estudios podrían estar enmascarando genotipos recombinantes y las asociaciones deducidas en dichos estudios podrían no ser correctas. Además el hecho de que la secuenciación no sea capaz de detectar infecciones mixtas, supone otro sesgo a los estudios de asociación, la hibridación reversa INNO-LiPA puede detectarlas, pero no con la sensibilidad de la UDPS ni de forma cuantitativa, impidiendo el estudio dinámico de estas mezclas o la posible relevancia clínica de sus proporciones, y tampoco detecta fenómenos de recombinación, con lo que estudios con esta metodología estarían en la misma situación que la secuenciación.

El análisis filogenético del genoma completo sigue siendo el método más concluyente para el genotipaje y se considera el *gold standard*, pero actualmente es un proceso laborioso que consume mucho tiempo. Gracias a las nuevas técnicas de NNGS, que permiten la secuenciación de fragmentos más largos, sería posible la amplificación del genoma completo de 3,2 kb del VHB, con el aumento de la rapidez y abaratamiento, se podrían llegar a usar con propósitos clínicos de manera rutinaria.

5.9 Perspectivas de futuro

Debido a que la región X/PC tiene mucho interés por su extrema multifuncionalidad, incluye muchos elementos reguladores, y por las proteínas que codifica, HBx es la principal proteína asociada a la carcinogénesis, se ha continuado el trabajo presentado en esta tesis estudiando la variabilidad en dicha región. Se han analizado muestras basales de pacientes con HBC en un período libre de tratamiento para evaluar la alta complejidad de esta zona, en la que además se han reportado, no solo variantes por sustitución, sino también variantes con inserciones y/o deleciones (indel), asociadas a la severidad de la infección [157–161]. El análisis de los datos preliminares en muestras de 50 pacientes (960.921 secuencias en total) ha evidenciado más de 100 variantes con indel en el ORF X, que dan lugar a versiones truncadas de HBx en el extremo Ct, que ya han sido asociado con HCC, e incluso a versiones más largas de la proteína, sugiriendo que estas variantes representan un posible mecanismo de multicodificación del virus como forma de generar formas alternativas de la proteína con diferentes funciones. Esta continuación del presente proyecto de tesis doctoral está siendo posible gracias a la mejora de los algoritmos para procesar los datos de la UDPS, que están permitiendo detectar inserciones y/o deleciones, incluso de una única pb, en regiones altamente homopoliméricas de la región X/PC. De hecho estos datos preliminares han sido expuestos en congresos nacionales e internacionales en comunicaciones tipos póster que se resumen en la tabla 22.

Tabla 22: Comunicaciones tipo póster a congresos nacionales e internacionales describiendo variantes por indel en la región X/preCore.

CONGRESO	TÍTULO PÓSTER
The 54th Annual Meeting of the American Association for the Study of Liver Disease (Noviembre 2013, Washington)	Dynamic study of Hepatitis B Virus (HBV) quasispecies in X Gene/Basal Core Promoter/preCore region by Ultradeep Pyrosequencing (UDPS) technology: high rates of deletions, <i>de novo</i> stop codons, and viral genotype mixtures
	Dynamic of quasispecies complexity and main regulatory motifs of Hepatitis B virus (HBV) X gene/Basal Core Promoter/PreCore region analyzed by Ultradeep Pyrosequencing (UDPS) technology
39 Congreso anual de la Asociación Española para el estudio del hígado (Febrero 2014, Madrid)	Dinámica de la complejidad de cuasiespecies y principales motivos reguladores del Virus de la Hepatitis B (VHB) en la región X/PROMOTOR BÁSICO DEL CORE/PRECORE analizada mediante secuenciación masiva (<i>UDPS-Ultradeep Pyrosequencing</i>)
	Estudio dinámico de cuasiespecies del virus de la hepatitis B (VHB) en la región del gen X, promotor básico del core (PBC) y preCore (PC) mediante <i>Ultradeep Pyrosequencing</i> (UDPS): altas tasas de deleciones y codones stop <i>de novo</i>
40 Congreso anual de la Asociación Española para el estudio del hígado (Febrero 2015, Madrid)	La presencia de genomas con inserciones y/o deleciones en la principal región reguladora del Virus de la Hepatitis B sugiere una multicodificación de la proteína X
The International Liver Congress, EASL (Abril 2015, Viena)	Insertions and/or deletions in the main regulatory region of Hepatitis B Virus suggest multicoding of the X protein

Conclusiones

6. CONCLUSIONES

1. Las técnicas de NGS basadas en UDPS permiten estudiar las variantes genéticas que forman una QS y su evolución a lo largo del tiempo bajo diferentes presiones selectivas, con un nivel de sensibilidad y capacidad cuantitativa mucho mayor que las técnicas convencionales de secuenciación directa con un paso previo de clonaje. No obstante tienen una serie de limitaciones a tener en cuenta como la posible recombinación artefactual durante el proceso de PCR, la modificación de las frecuencias relativas de las variantes o la imprecisión en regiones homopoliméricas, por lo que es necesario un procesamiento previo de los datos mediante algoritmos bioinformáticos.

2. La mayor variabilidad de la QS en pacientes HBeAg(-) que HBeAg(+) en la región preCore/Core, evidencia que el aumento de la diversidad está asociado a la seroconversión. En aquellos pacientes en los que el efecto inmunomodulador de HBeAg está ausente se desarrollaría una mayor presión del sistema inmune en esta región, induciendo la aparición de variantes que hacen que la QS aumente en heterogeneidad.

3. La selección positiva de variantes en el core en pacientes HBeAg(-) y HBeAg(+/-) durante la dinámica natural de la QS sugiere un mecanismo potencial de escape del sistema inmune por cambios en la secuencia de la nucleocápside. El sistema inmune estaría actuando principalmente contra los epítomos del core, la QS aumentaría su complejidad en esa zona dando lugar a variantes en HBc, sugiriendo un mecanismo potencial adicional del VHB para escapar de la respuesta inmune del hospedador, además de atenuar la respuesta con HBeAg.

4. La alta complejidad en la región preCore está asociada con una baja replicación viral, esta región se solapa con el PBC y la señal de encapsidación ϵ con lo que sus variantes podrían afectar a dicho proceso, y sugiere una respuesta inmune aumentada en pacientes HBeAg(-), probablemente debido a la falta de actividad inmunomoduladora de HBeAg. La ausencia de HBeAg provocaría un aumento de la presión del sistema inmune generando una mayor heterogeneidad en la QS pero impidiendo que el virus se replique, por la eliminación de hepatocitos infectados, reduciendo la carga viral.

5. Una carga viral elevada no significa necesariamente una mayor complejidad, sino que una variante especialmente eficaz para la replicación, con un mejor *fitness* en unas determinadas condiciones, puede convertirse en la predominante, llevando a una población de virus muy grande pero muy homogénea.

6. La correlación negativa entre la complejidad durante la evolución de la QS en el período sin tratamiento y tras el fallo de este en la región PC/Core/PBC, sugiere que los cambios que ocurren durante la evolución natural de la QS podrían estar afectados por la respuesta inmune y determinar la evolución de la misma región bajo el tratamiento con NUCs.

7. Se requiere el estudio de la complejidad de QS en un fragmento que abarque la región donde los NUCs ejercen la presión selectiva (RT) y la región PC/Core/PBC simultáneamente para dilucidar cómo la variación en el PC/Core/PBC influye en la presión que ejercen los NUCs en esta región RT.

8. La región X/preCore presenta una mayor complejidad que la región P/S, esto está de acuerdo con el mayor grado de conservación de RT y parece indicar que la región X/preCore tolera mejor un mayor número de cambios que P/S.

9. El aumento de la diversidad está asociado a la seroconversión y la complejidad presenta una correlación negativa antes del tratamiento y después del fallo del mismo, lo que sugiere la importancia de estudiar la complejidad de QS antes de iniciar el tratamiento y usarla como un posible factor predictivo de la evolución del virus en el caso de no respondedores a NUCs. Con la consolidación de los métodos de NGS se podría llevar a cabo el estudio de haplotipos de manera reproducible y así usar los parámetros de complejidad de QS para el manejo clínico de la infección por el VHB.

10. Mediante UDPS se evidencia la existencia de mezclas de genotipos que cambian a lo largo del tiempo, lo que pone de manifiesto la compleja dinámica de la QS del VHB como un mecanismo adicional para adaptarse a nuevas situaciones, como son la respuesta del sistema inmune o el tratamiento con NUCs. En este sentido los resultados sugieren que el Gen/A es menos sensible al tratamiento con LAM y que el tratamiento antiviral es un factor importante para inducir el cambio de genotipo.

11. La alta prevalencia de infecciones mixtas en la región X/PC sumado a su elevada complejidad sugieren que esta región podría ser un *hot spot* para la recombinación. Las discrepancias encontradas en los genotipos obtenidos en la región X/PC y P/S de la misma muestra del mismo paciente también sugieren fenómenos de recombinación intergenotípica, que debería confirmarse mediante técnicas de NNGS de un fragmento más largo que abarcara ambas regiones en la misma secuencia.

12. El resultado del genotipaje es altamente dependiente de la región seleccionada para ello, esto sumado a la alta prevalencia de infecciones mixtas y de posibles genotipos recombinantes, ponen de manifiesto la necesidad de consensuar tanto la región como la tecnología para el genotipaje del VHB y así poder establecer el significado clínico de la clasificación genotípica.

13. Con las nuevas técnicas NNGS, que permiten la secuenciación de fragmentos más largos, sería posible la amplificación del genoma completo de 3,2 kb del VHB y la posterior filogénesis (técnica *gold standard* para el genotipaje) y además permitiría confirmar la recombinación intergenotípica, con el aumento de la rapidez y abaratamiento de estas técnicas, se podrían llegar a usar con propósitos clínicos de manera rutinaria.

Bibliografía

7. BIBLIOGRAFÍA

1. Blumberg BS, Alter HJ ViS. A “new” antigen in leukemia sera. *Jama*. 1965;191: 541–546.
2. Almeida JD, Rubenstein D SE. New antigen-antibody system in Australia-antigen-positive hepatitis. *Lancet (London, England)*. 1971;2: 1225–1227.
3. Dane DS, Cameron CH BM. Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *Lancet (London, England)*. 1970;1: 695–698.
4. Galibert F, Mandart E, Fitoussi F, Tiollais P CP. Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in *E. coli*. *Nature*. 1979;281: 646–650.
5. Gust ID, Burrell CJ, Coulepis AG, Robinson WS ZA. Taxonomic classification of human hepatitis B virus. *Intervirology*. 1986;25: 14–29.
6. Heermann KH. Large Surface Proteins of Hepatitis B Virus Containing the Pre-s Sequence. *J Virol*. 1984;52: 396–402.
7. Wynne SA, Crowther RA, Leslie AGW. The Crystal Structure of the Human Hepatitis B Virus Capsid. 1999;3: 771–780.
8. Huovila AP, Eder AM FS. Hepatitis B Surface Antigen Assembles in a Post-ER , Pre-Golgi Compartment. *J Cell Biol*. 1992;118: 1305–1320.
9. Luangsay S ZF. Structure and molecular virology. Thomas HC, Lok AS, Locarnini SA, Zuckerman AJ, editors *Viral Hepatitis Fourth edition*. 2013. pp. 65–80. doi:10.1017/CBO9781107415324.004
10. Rodríguez-Frías F, Jardi R. Virología molecular del virus de la hepatitis B. *Enferm Infecc Microbiol Clin*. 2008;26: 2–10. doi:10.1016/S0213-005X(08)76514-5
11. Locarnini S BS. Hepatitis B surface antigen quantification: Not what it seems on the surface. *Hepatology*. 2012;56: 411–414.
12. Bruss V, Lu X, Thomssen R, Gerlich WH. Post-translational alterations in transmembrane topology of the hepatitis B virus large envelope protein. *EMBO J*. 1994;13: 2273–9.
13. Yan H, Zhong G, Xu G, He W, Jing Z, Gao Z, et al. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *Elife*. 2012;13. doi:10.7554/eLife.00049
14. Schädler S, Hildt E. HBV life cycle: Entry and morphogenesis. *Viruses*. 2009;1: 185–209. doi:10.3390/v1020185
15. Radziwill G, Tucker W, Schaller H. Mutational analysis of the hepatitis B virus P gene product: domain structure and RNase H activity. *J Virol*. 1990;64: 613–20.

16. Bartenschlager R, Schaller H. The amino-terminal domain of the hepadnaviral P-gene encodes the terminal protein (genome-linked protein) believed to prime reverse transcription. *EMBO J.* 1988;7: 4185–4192.
17. Wang Y-X, Luo C, Zhao D, Beck J, Nassal M. Extensive Mutagenesis of the Conserved Box E Motif in Duck Hepatitis B Virus P Protein Reveals Multiple Functions in Replication and a Common Structure with the Primer Grip in HIV-1 Reverse Transcriptase. *J Virol.* 2012;86: 6394–6407. doi:10.1128/JVI.00011-12
18. Li MD, Bronson DL, Lemke TD, Faras AJ. Phylogenetic analyses of 55 retroelements on the basis of the nucleotide and product amino acid sequences of the pol gene. *Mol Biol Evol.* 1995;12: 657–70.
19. Poch O, Sauvaget I, Delarue M, Tordo N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* 1989;8: 3867–74.
20. Jones S a, Hu J. Hepatitis B virus reverse transcriptase: diverse functions as classical and emerging targets for antiviral intervention. *Emerg Microbes Infect.* 2013;2: e56. doi:10.1038/emi.2013.56
21. 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
22. van Hemert FJ, van de Klundert MAA, Lukashov V V., Kootstra NA, Berkhout B, Zaaijer HL. Protein X of hepatitis B virus: Origin and structure similarity with the central domain of DNA glycosylase. *PLoS One.* 2011;6. doi:10.1371/journal.pone.0023392
23. Cao F, Jones S, Li W, Cheng X, Hu Y, Hu J, et al. Sequences in the terminal protein and reverse transcriptase domains of the hepatitis B virus polymerase contribute to RNA binding and encapsidation. *J Viral Hepat.* 2014;21: 882–893. doi:10.1111/jvh.12225
24. Li J, Du Y, Liu X, Shen Q, Huang A, Zheng M, et al. Binding sensitivity of adefovir to the polymerase from different genotypes of HBV: molecular modeling, docking and dynamics simulation studies. *Acta Pharmacol Sin.* Nature Publishing Group; 2013;34: 319–28. doi:10.1038/aps.2012.146
25. Steitz TA. DNA Polymerases: Structural Diversity and Common Mechanisms. *J Biol Chem.* 1999;274: 17395–17398. doi:10.1074/jbc.274.25.17395
26. Stuyver LJ, Locarnini SA, Lok A, Richman DD, Carman WF, Dienstag JL, et al. Nomenclature for antiviral-resistant human hepatitis B virus mutations in the polymerase region. *Hepatology.* 2001;33: 751–7. doi:10.1053/jhep.2001.22166
27. Huang H, Chopra R, Verdine GL, Harrison SC. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science.* 1998;282: 1669–75. doi:10.1126/science.282.5394.1669

28. Jacobo-Molina A, Ding J, Nanni RG, Clark AD, Lu X, Tantillo C, et al. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc Natl Acad Sci U S A*. 1993;90: 6320–4.
29. Tantillo C, Ding J, Jacobo-Molina A, Nanni RG, Boyer PL, Hughes SH, et al. Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase. Implications for mechanisms of drug inhibition and resistance. *J Mol Biol*. 1994;243: 369–87. doi:10.1006/jmbi.1994.1665
30. Bartholomeusz A, Tehan BG, Chalmers DK. Comparisons of the HBV and HIV polymerase, and antiviral resistance mutations. *Antivir Ther*. 2004;9: 149–60.
31. Rodríguez-Frias F, Jardi R. Virología molecular del virus de la hepatitis B. *Enferm Infecc Microbiol Clin*. 2008;26: 2–10. doi:10.1016/S0213-005X(08)76514-5
32. Bruss V, Gerlich WH. Formation of transmembraneous hepatitis B e-antigen by cotranslational in vitro processing of the viral precore protein. *Virology*. 1988;163: 268–75.
33. Messageot F, Salhi S, Eon P, Rossignol J-M. Proteolytic processing of the hepatitis B virus e antigen precursor. Cleavage at two furin consensus sequences. *J Biol Chem*. 2003;278: 891–5. doi:10.1074/jbc.M207634200
34. Alexopoulou A. HBeAg negative variants and their role in the natural history of chronic hepatitis B virus infection. *World J Gastroenterol*. 2014;20: 7644. doi:10.3748/wjg.v20.i24.7644
35. Kramvis A. The clinical implications of hepatitis B virus genotypes and HBeAg in pediatrics. *Rev Med Virol*. 2016;26: 285–303. doi:10.1002/rmv.1885
36. Ito K, Kim K-H, Lok AS-F, Tong S. Characterization of Genotype-Specific Carboxyl-Terminal Cleavage Sites of Hepatitis B Virus e Antigen Precursor and Identification of Furin as the Candidate Enzyme. *J Virol*. 2009;83: 3507–3517. doi:10.1128/JVI.02348-08
37. Pumpens P, Grens E. HBV Core Particles as a Carrier for B Cell / T Cell Epitopes. 2001;1067: 98–114.
38. Seitz S, Urban S, Antoni C, Böttcher B. Cryo-electron microscopy of hepatitis B virions reveals variability in envelope capsid interactions. *EMBO J*. 2007;26: 4160–7. doi:10.1038/sj.emboj.7601841
39. Roseman AM, Berriman J a, Wynne S a, Butler PJG, Crowther RA. A structural model for maturation of the hepatitis B virus core. *Proc Natl Acad Sci U S A*. 2005;102: 15821–15826. doi:10.1073/pnas.0504874102

40. Dryden KA, Wieland SF, Whitten-Bauer C, Gerin JL, Chisari F V., Yeager M. Native Hepatitis B Virions and Capsids Visualized by Electron Cryomicroscopy. *Mol Cell*. 2006;22: 843–850. doi:10.1016/j.molcel.2006.04.025
41. Bock CT, Schwinn S, Locarnini S, Fyfe J, Manns MP, Trautwein C, et al. Structural organization of the hepatitis B virus minichromosome. *J Mol Biol*. 2001;307: 183–96. doi:10.1006/jmbi.2000.4481
42. Yeh C, Wong SW, Fung Y, Ou J. Cell cycle regulation of nuclear localization of hepatitis B virus core protein. *PNAS*. 1993;90: 6459–6463.
43. Geng M, Xin X, Bi L-Q, Zhou L-T, Liu X-H. Molecular mechanism of hepatitis B virus X protein function in hepatocarcinogenesis. *World J Gastroenterol*. 2015;21: 10732–10738. doi:10.3748/wjg.v21.i38.10732
44. 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
45. Minor MM, Slagle BL. Hepatitis B virus HBx protein interactions with the ubiquitin proteasome system. *Viruses*. 2014;6: 4683–4702. doi:10.3390/v6114683
46. Bouchard MJ, Schneider RJ. The enigmatic X gene of hepatitis B virus. *J Virol*. 2004;78: 12725–12734. doi:10.1128/JVI.78.23.12725
47. Kim H, Lee S, Kim B, Kim H, Lee S, Kim B. X region mutations of hepatitis B virus related to clinical severity. *World J Gastroenterol*. 2016;22: 5467–5478. doi:10.3748/wjg.v22.i24.5467
48. Kidd-Ijunggren K, Bla J. Mutations Within the Hepatitis B Virus Genome Among Chronic Hepatitis B Patients With Hepatocellular Carcinoma. *J Med Virol*. 2003;71: 18–23. doi:10.1002/jmv.10458
49. Slagle BL, Bouchard MJ. Hepatitis B virus X and regulation of viral gene expression. *Cold Spring Harb Perspect Med*. 2016;6. doi:10.1101/cshperspect.a021402
50. Arii M, Takada S KK. Identification of three essential regions of hepatitis B virus X protein for trans-activation funtion. *Oncogene*. 1992;7: 397–403.
51. Feitelson MA, Bonamassa B, Arzumanyan A. The roles of hepatitis B virus-encoded X protein in virus replication and the pathogenesis of chronic liver disease. *Expert Opin Ther Targets*. 2014;18: 293–306. doi:10.1517/14728222.2014.867947
52. Tu H, Bonura C, Giannini C, Bre C, Kremsdorf D. Biological Impact of Natural COOH-Terminal Deletions of Hepatitis B Virus X Protein in Hepatocellular Carcinoma Tissues Biological Impact of Natural COOH-Terminal Deletions of Hepatitis B Virus X Protein in Hepatocellular Carcinoma Tissues 1. *Cancer Res*. 2001;61: 7803–7810.

53. Wang Q, Zhang W, Ye L, Zhang X. A mutant of HBx (HBx Δ 127) promotes hepatoma cell growth via sterol regulatory element binding protein 1c involving 5-lipoxygenase. *Nat Publ Gr. Nature Publishing Group*; 2010;31: 367–374. doi:10.1038/aps.2010.5
54. Fomby P, Cherlin AJ. Hepatitis B virus regulatory HBx protein binding to DDB1 is required but is not sufficient for maximal HBV replication. *Virology*. 2012;426: 73–82. doi:10.1016/j.virol.2012.01.021
55. Zhang Z, Torii N, Furusaka A, Malayaman N, Hu Z, Liang TJ. Structural and Functional Characterization of Interaction between Hepatitis B Virus X Protein and the Proteasome Complex*. *J Biol Chem*. 2000;275: 15157–15165. doi:10.1074/jbc.M910378199
56. Wang M, Wu H, Huang S, Zhang H, Qin C, Zhao L-H, et al. HBx regulates fatty acid oxidation to promote hepatocellular carcinoma survival during metabolic stress. *Oncotarget*. 2016;7: 6711–26. doi:10.18632/oncotarget.6817
57. Yue D, Zhang Y, Cheng L, Ma J, Xi Y, Yang L, et al. Hepatitis B virus X protein (HBx) - induced abnormalities of nucleic acid metabolism revealed by 1H-NMR-based metabonomics. *Sci Rep. Nature Publishing Group*; 2016;14: 24430. doi:10.1038/srep24430
58. Samuel Martín-Vílchez, Paloma Sanz-Cameno, Yolanda Rodríguez-Munoz PLM, Francisca Molina-Jimenez, Manuel Lopez-Cabrera, Ricardo Moreno-Otero and EL-P. The Hepatitis B Virus X Protein Induces Paracrine Activation of Human Hepatic Stellate Cells. *Hepatology*. 2008;47: 1872–1883. doi:10.1002/hep.22265
59. Ng K, Chai S, Tong M, Guan X, Lin C, Ching Y-P, et al. C-terminal truncated hepatitis B virus X protein promotes hepatocellular carcinogenesis through induction of cancer and stem cell-like properties. *Oncotarget*. 2016;7: 24005–17. doi:10.18632/oncotarget.8209
60. Zhang H, Shan CL, Li N, Zhang X, Zhang XZ, Xu FQ, et al. Identification of a natural mutant of HBV X protein truncated 27 amino acids at the COOH terminal and its effect on liver cell proliferation. *Acta Pharmacol Sin*. 2008;29: 473–480. doi:10.1111/j.1745-7254.2008.00764.x
61. Wang Q, Zhang W, Ye L, Zhang X. A mutant of HBx (HBx Δ 127) promotes hepatoma cell growth via sterol regulatory element binding protein 1c involving 5-lipoxygenase. *Nat Publ Gr. Nature Publishing Group*; 2010;31: 367–374. doi:10.1038/aps.2010.5
62. Tian Y, Yang W, Song J, Wu Y, Ni B. Hepatitis B virus X protein-induced aberrant epigenetic modifications contributing to human hepatocellular carcinoma pathogenesis. *Mol Cell Biol*. 2013;33: 2810–2816. doi:10.1128/MCB.00205-13

63. Zhang Y, Liu H, Yi R, Yan T, He Y, Zhao Y, et al. Hepatitis B virus whole-X and X protein play distinct roles in HBV-related hepatocellular carcinoma progression. *J Exp Clin Cancer Res. Journal of Experimental & Clinical Cancer Research*; 2016; 1–10. doi:10.1186/s13046-016-0366-3
64. Moolla N, Kew M AP. Regulatory elements of hepatitis B virus transcription. *J Viral Hepat.* 2002;9: 323–331.
65. Zheng Y, Li J, Ou J. Regulation of Hepatitis B Virus Core Promoter by Transcription Factors HNF1 and HNF4 and the Viral X Protein Regulation of Hepatitis B Virus Core Promoter by Transcription Factors HNF1 and HNF4 and the Viral X Protein. 2004;78: 6908–6914. doi:10.1128/JVI.78.13.6908
66. Nassal M. Hepatitis B viruses: Reverse transcription a different way. *Virus Res.* 2008;134: 235–249. doi:10.1016/j.virusres.2007.12.024
67. Yuh CH, Ting LP. The genome of hepatitis B virus contains a second enhancer: cooperation of two elements within this enhancer is required for its function. *J Virol.* 1990;64: 4281–4287.
68. Kidd AH, Kidd-Ljunggren K. A revised secondary structure model for the 3'-end of hepatitis B virus pregenomic RNA. *Nucleic Acids Res.* 1996;24: 3295–3301.
69. Pollack JR GD. Site-Specific RNA Binding by a Hepatitis B Virus Reverse Transcriptase Initiates Two Distinct Reactions: RNA Packaging and DNA Synthesis. *J Virol.* 1994;68: 5579–5587.
70. Knaus T, Nassal M, Biologie M, Heidelberg U, Feld IN. The encapsidation signal on the hepatitis B virus RNA pregenome forms a stem-loop structure that is critical for its function. 1993;21: 3967–3975.
71. Schleucher È, Cromsigt J, Ippel H, Kidd-ljunggren K, Flodell S. The apical stem - loop of the hepatitis B virus encapsidation signal folds into a stable tri - loop with two underlying pyrimidine bulges. 2002;30.
72. Lok ASF, Akarca U, Greene S. Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal. *PNAS.* 1994;91: 4077–4081.
73. Wang G, Seeger C. Novel Mechanism for Reverse Transcription in Hepatitis B Viruses. *J Virol.* 1993;67: 6507–6512.
74. Hoofnagle JH. Reactivation of Hepatitis B. *Hepatology.* 2009;49: S156–165. doi:10.1002/hep.22945

75. Guan Huei Lee, Shanthi Wasser SGL. Hepatitis B pregenomic RNA splicing—The products, the regulatory mechanisms and its biological significance. *Virus Res.* 2008;136: 1–7.
76. Yu X, Mertz JE. Differential regulation of the pre-C and pregenomic promoters of human hepatitis B virus by members of the nuclear receptor superfamily. *J Virol.* 1997;71: 9366–74.
77. Doitsh G SY. Enhancer I predominance in hepatitis B virus gene expression. *Mol Cell Biol.* 2004;24: 1799–1808.
78. Glebe D, Bremer CM. The Molecular Virology of Hepatitis B Virus. *Semin Liver Dis.* 2013;33: 103–112.
79. Stieger B, Hagenbuch B, Landmann L, Höchli M, Schroeder A MP. In situ localization of the hepatocytic Na⁺/Taurocholate cotransporting polypeptide in rat liver. *Gastroenterology.* 1994;107: 1781–1787.
80. Stoeckl L, Funk A, Kopitzki A, Brandenburg B, Oess S, Will H, et al. Identification of a structural motif crucial for infectivity of hepatitis B viruses. *PNAS.* 2006;103: 6730–6734.
81. Berting A, Fischer C, Schaefer S, Garten W, Klenk HD GW. Hemifusion activity of a chimeric influenza virus hemagglutinin with a putative fusion peptide from hepatitis B virus. *Virus Res.* 2000;68: 35–49.
82. KannM, Sodeik B, Vlachou A, Gerlich WH, Helenius A. Phosphorylation-dependent binding of hepatitis B virus core particles to the nuclear pore complex. *J Cell Biol.* 1999;145: 45–55.
83. Schimtz A, Schwarz A, Foss M, Zhou L, Rabe B, Hoellenriegel J, et al. Nucleoporin 153 Arrests the Nuclear Import of Hepatitis B Virus Capsids in the Nuclear Basket. *Plos Pathog.* 2010;6: e1000741. doi:10.1371/journal.ppat.1000741
84. Nassal M. HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. *Gut.* 2015;64: 1972–1984. doi:10.1136/gutjnl-2015-309809
85. Gao W, Hu J. Formation of hepatitis B virus covalently closed circular DNA: removal of genome-linked protein. *J Virol.* 2007;81: 6164–74. doi:10.1128/JVI.02721-06
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 U S A.* 2014;111: E4244–53. doi:10.1073/pnas.1409986111

87. Cui X, McAllister R, Boregowda R, Sohn JA, Cortes Ledesma F, Caldecott KW, et al. Does Tyrosyl DNA Phosphodiesterase-2 Play a Role in Hepatitis B Virus Genome Repair PLoS One. 2015;10: e0128401. doi:10.1371/journal.pone.0128401
88. Y Murakami, K Saigo, H Takashima, M Minami, T Okanou, C Bréchet PP-B. Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. Gut. 2005;54: 1162–1168. doi:10.1136/gut.2004.054452
89. Levrero M, Pollicino T, Petersen J, Belloni L, Raimondo G, Dandri M. Control of cccDNA function in hepatitis B virus infection. J Hepatol. European Association for the Study of the Liver; 2009;51: 581–592. doi:10.1016/j.jhep.2009.05.022
90. Bruss V. Hepatitis B virus morphogenesis. World J Gastroenterol. 2007;13: 65–73. doi:10.3748/wjg.v13.i1.65
91. 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
92. Abraham TM, Loeb DD. The topology of hepatitis B virus pregenomic RNA promotes its replication. J Virol. 2007;81: 11577–84. doi:10.1128/JVI.01414-07
93. Yang W, Summers J. Illegitimate replication of linear hepadnavirus DNA through nonhomologous recombination. J Virol. 1995;69: 4029–36.
94. Juergen B. Hepatitis B virus replication. World J Gastroenterol. 2007;7: 48–64.
95. Watanabe T, Sorensen EME, Naito A, Schott M, Kim S, Ahlquist P. Involvement of host cellular multivesicular body functions in hepatitis B virus budding. Proc 2007;104: 10205–10. doi:10.1073/pnas.0704000104
96. Lambert C, Döring T, Prange R. Hepatitis B virus maturation is sensitive to functional inhibition of ESCRT-III, Vps4, and gamma 2-adaptin. J Virol. 2007;81: 9050–60. doi:10.1128/JVI.00479-07
97. Tuttleman JS, Pourcel C, Summers J. Formation of the pool of covalently closed circular viral DNA in hepadnavirus-infected cells. Cell. Elsevier; 1986;47: 451–460. doi:10.1016/0092-8674(86)90602-1
98. Wong DKH, Yuen MF, Yuan H, Sum SSM, Hui GK, Hall J, et al. Quantitation of covalently closed circular hepatitis B virus DNA in chronic B patients. Hepatology. 2004;40: 727–737. doi:10.1002/hep.20353
99. WHO media centre. Fact sheet N°204 [Internet]. 2015. Available: <http://www.who.int/mediacentre/factsheets/fs204/en/>

100. &, Vasilios Papastergiou, Rosa Lombardi¹, Douglas MacDonald EAT. Global Epidemiology of Hepatitis B Virus (HBV) Infection. *Curr Hepatol Rep.* 2015;14: 171–178. doi:10.1016/S1473-3099(05)70216-4
101. Trépo C, Chan HLY, Lok A. Hepatitis B virus infection. *Lancet.* 2014;384: 2053–2063. doi:10.1016/S0140-6736(14)60220-8
102. CDC. Center for Disease Control and Prevention. In: Preventing Hepatitis B [Internet]. 2014. Available: http://www.cdc.gov/globalhealth/immunization/othervpds/preventing_hepatitisb.html
103. Samal J, Kandpal M, Vivekanandan P. Molecular mechanisms underlying occult hepatitis B virus infection. *Clin Microbiol Rev.* 2012;25: 142–163. doi:10.1128/CMR.00018-11
104. Buti M, García-samaniego J, Prieto M, Rodríguez M, Sánchez-tapias JM, Suárez E, et al. Documento de consenso de la AEEH sobre el tratamiento de la infección por el virus de la hepatitis B (2012). *Gastroenterol Hepatol. SEGO;* 2012;35: 512–528. doi:10.1016/j.gastrohep.2012.04.006
105. Hadziyannis SJ. Natural history of chronic hepatitis B in Euro-Mediterranean and African Countries. *J Hepatol. European Association for the Study of the Liver;* 2011;55: 183–191. doi:10.1016/j.jhep.2010.12.030
106. Knöll A, Pietrzyk M, Loss M, Goetz WA, Jilg W. Solid-organ transplantation in HBsAg-negative patients with antibodies to HBV core antigen: low risk of HBV reactivation. *Transplantation.* 2005;79: 1631–3.
107. Chisari F V., Ferrari C. Hepatitis B Virus Immunopathogenesis. *Annu Rev Immunol.* 1995;13: 29–60. doi:10.1146/annurev.iy.13.040195.000333
108. You CR, Lee SW, Jang JW, Yoon SK. Update on hepatitis B virus infection. *World J Gastroenterol.* 2014;20: 13293–13305. doi:10.3748/wjg.v20.i37.13293
109. Song IH. Emerging therapeutics and relevant targets for chronic Hepatitis B. *Turk J Gastroenterol.* 2016;27: 210–5. doi:10.5152/tjg.2016.15521
110. Tong S, Revill P. Overview of hepatitis B viral replication and genetic variability. *J Hepatol. European Association for the Study of the Liver;* 2016;64: S4–S16. doi:10.1016/j.jhep.2016.01.027
111. Lang T, Lo C, Skinner N, Locarnini S, Visvanathan K, Mansell A. The hepatitis B e antigen (HBeAg) targets and suppresses activation of the toll-like receptor signaling pathway. *J Hepatol.* 2011;55: 762–9. doi:10.1016/j.jhep.2010.12.042
112. Visvanathan K, Skinner NA, Thompson AJ V, Riordan SM, Sozzi V, Edwards R, et al. Regulation of Toll-like receptor-2 expression in chronic hepatitis B by the precore protein. *Hepatology.* 2007;45: 102–110. doi:10.1002/hep.21482

113. Milich DR, Schödel F, Hughes JL, Jones JE, Peterson DL. The hepatitis B virus core and e antigens elicit different Th cell subsets: antigen structure can affect Th cell phenotype. *J Virol.* 1997;71: 2192–201.
114. Vanlandschoot P, Cao T, Leroux-Roels G. The nucleocapsid of the hepatitis B virus: A remarkable immunogenic structure. *Antiviral Res.* 2003;60: 67–74. doi:10.1016/j.antiviral.2003.08.011
115. Diepolder HM, Ries G, Jung MC, Ju H. Differential Antigen-Processing Pathways of the Hepatitis B Virus e and Core Proteins. *Gastroenterology.* 1999; 650–657.
116. Guerrieri F, Belloni L, Pediconi N, Levrero M. Molecular Mechanisms of HBV-Associated Hepatocarcinogenesis. *Semin Liver Dis.* 2013;33: 147–156. doi:10.1055/s-0033-1345721
117. Ding D, Lou X, Hua D, Yu W, Li L, Wang J, et al. Recurrent Targeted Genes of Hepatitis B Virus in the Liver Cancer Genomes Identified by a Next-Generation Sequencing-Based Approach. *PLoS Genet.* 2012;8. doi:10.1371/journal.pgen.1003065
118. Toh ST, Jin Y, Liu L, Wang J, Babrzadeh F, Gharizadeh B, et al. Deep sequencing of the hepatitis B virus in hepatocellular carcinoma patients reveals enriched integration events, structural alterations and sequence variations. *Carcinogenesis.* 2013;34: 787–798. doi:10.1093/carcin/bgs406
119. Minami M, Daimon Y, Mori K, Takashima H, Nakajima T, Itoh Y, et al. Hepatitis B virus-related insertional mutagenesis in chronic hepatitis B patients as an early drastic genetic change leading to hepatocarcinogenesis. *Oncogene.* 2005;24: 4340–4348. doi:10.1038/sj.onc.1208628
120. Zhang Z, Wu C, Chen X, Li X, Li J, Lu M. Genetic variation of hepatitis B virus and its significance for pathogenesis. *World J Gastroenterol.* 2016;22: 126–144. doi:10.3748/wjg.v22.i1.126
121. Betz-Stablein BD, Töpfer A, Littlejohn M, Yuen L, Colledge D, Sozzi V, Angus P, Thompson A, Revill P, Beerwinkler N, Warner N LF. Single-molecule sequencing reveals complex genomic variation of hepatitis B virus during 15 years of chronic infection following liver transplantation. *J Virol.* 2016; doi:10.1128/JVI.00243-16
122. Papatheodoridis G, Buti M, Cornberg M, Janssen H, Mutimer D, Pol S, et al. Guía de práctica clínica de la EASL: Tratamiento de la infección crónica por el virus de la hepatitis B. *J Hepatol.* 2012;57: 167–185.
123. Zoulim F, Lebossé F, Levrero M. Current treatments for chronic hepatitis B virus infections. *Curr Opin Virol.* 2016;18: 109–116. doi:10.1016/j.coviro.2016.06.004

124. Gish R, Jia JD, Locarnini S, Zoulim F. Selection of chronic hepatitis B therapy with high barrier to resistance. *Lancet Infect Dis.* Elsevier Ltd; 2012;12: 341–353. doi:10.1016/S1473-3099(11)70314-0
125. Terrault NA, Bzowej NH, Chang K-M, Hwang JP, Jonas MM, Murad MH. AASLD guidelines for treatment of chronic hepatitis B. *Hepatology.* 2016;63: 261–283. doi:10.1002/hep.28156
126. Levrero M, Testoni B, Zoulim F. HBV cure: why, how, when? *Curr Opin Virol.* Elsevier B.V.; 2016;18: 135–143. doi:10.1016/j.coviro.2016.06.003
127. Domingo E. Rna Virus Mutations. *Annu Rev Microbiol.* 1997;51: 151–178.
128. Caligiuri P, Cerruti R, Icardi G, Bruzzone B. Overview of hepatitis B virus mutations and their implications in the management of infection. *World J Gastroenterol.* 2016;22: 145–154. doi:10.3748/wjg.v22.i1.145
129. Ghany M, Liang TJ. 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
130. Zoulim F, Locarnini S. Hepatitis B Virus Resistance to Nucleos(t)ide Analogues. *Gastroenterology.* Elsevier Inc.; 2009;137: 1593–1608. doi:10.1053/j.gastro.2009.08.063
131. Tenney DJ, Levine SM, Rose RE, Walsh AW, Weinheimer SP, Discotto L, et al. Clinical emergence of entecavir-resistant hepatitis B virus requires additional substitutions in virus already resistant to lamivudine. *Antimicrob Agents Chemother.* 2004;48: 3498–3507. doi:10.1128/AAC.48.9.3498-3507.2004
132. Amini-Bavil-Olyae 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: 1158–65. doi:10.1002/hep.22790
133. Lin C-L, Kao J-H. Review article: novel therapies for hepatitis B virus cure - advances and perspectives. *Aliment Pharmacol Ther.* 2016;44: 213–222. doi:10.1111/apt.13694
134. Kramvis A, Kew M, François G. Hepatitis B virus genotypes. *Vaccine.* 2005;23: 2409–2423. doi:10.1016/j.vaccine.2004.10.045
135. Kramvis A, Arakawa K, Yu MC, Nogueira R, Stram DO, Kew MC. Relationship of Serological Subtype , Basic Core Promoter and Precore Mutations to Genotypes / Subgenotypes of Hepatitis B Virus. *J Med Virol.* 2008;80: 27–46. doi:10.1002/jmv

136. Tian Q, Jia J. Hepatitis B virus genotypes: epidemiological and clinical relevance in Asia. *Hepatol Int*. Springer India; 2016; [Epub ahead of print]. doi:10.1007/s12072-016-9745-2
137. Shi W, Zhang Z, Ling C, Zheng W, Zhu C, Carr MJ, et al. Hepatitis B virus subgenotyping: History, effects of recombination, misclassifications, and corrections. *Infect Genet Evol*. Elsevier B.V.; 2013;16: 355–361. doi:10.1016/j.meegid.2013.03.021
138. Araujo NM. Hepatitis B virus intergenotypic recombinants worldwide: An overview. *Infect Genet Evol*. Elsevier B.V.; 2015;36: 500–510. doi:10.1016/j.meegid.2015.08.024
139. Kao J-H. Molecular epidemiology of hepatitis B virus. *Korean J Intern Med*. 2011;26: 255–61. doi:10.3904/kjim.2011.26.3.255
140. Sugauchi F, Orito E, Ichida T, Kato H, Sakugawa H, Kakumu S, et al. Hepatitis B virus of genotype B with or without recombination with genotype C over the precore region plus the core gene. *J Virol*. 2002;76: 5985–92. doi:10.1128/JVI.76.12.5985
141. Cui C, Shi J, Hui L, Xi H, Zhuoma A, Quni A, et al. The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. *J Gen Virol*. 2002;83: 2773–2777.
142. Chekaraou MA, Brichtler S, Mansour W, Gal F Le, Garba A, Dény P, et al. A novel hepatitis B virus (HBV) subgenotype D (D8) strain, resulting from recombination between genotypes D and E, is circulating in Niger along with HBV/E strains. *J Gen Virol*. 2010;91: 1609–1620. doi:10.1099/vir.0.018127-0
143. González-candelas F. Recombination in viruses: Mechanisms , methods of study , and evolutionary consequences. *Infect, Genet Evol*. 2015;30: 296–307. doi:10.1016/j.meegid.2014.12.022
144. Wang Y, Shan X, Liang Z, Shan Y, Huang W, Zhang D, et al. Deep sequencing analysis of HBV genotype shift and correlation with antiviral efficiency during adefovir dipivoxil therapy. *PLoS One*. 2015;10: 1–15. doi:10.1371/journal.pone.0131337
145. Zhang Q, Cao G. Genotypes, mutations, and viral load of hepatitis B virus and the risk of hepatocellular carcinoma: HBV properties and hepatocarcinogenesis. *Hepat Mon*. 2011;11: 86–91.
146. Deterding K, Constantinescu I, Nedelcu FD, Gervain J, Nemecek V, Srtunecy O, et al. Prevalence of HBV genotypes in Central and Eastern Europe. *J Med Virol*. 2008;80: 1707–11. doi:10.1002/jmv.21294
147. Lafrance-Vanasse J, Williams GJ, Tainer J a. Envisioning the dynamics and flexibility of Mre11-Rad50-Nbs1 complex to decipher its roles in DNA replication and repair. *Prog Biophys Mol Biol*. Elsevier Ltd; 2015;117: 182–193. doi:10.1016/j.pbiomolbio.2014.12.004

148. Shi W, Carr MJ, Dunford L, Zhu C, Hall WW, Higgins DG. Identification of novel inter-genotypic recombinants of human hepatitis B viruses by large-scale phylogenetic analysis. *Virology*. 2012;427: 51–59. doi:10.1016/j.virol.2012.01.030
149. 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
150. Luca AS, Ursu RG, Teu T, Luca CM, Pleşca CE. The Need for HBV Genotyping : a Cost-Efficient Approach. *Rev Med Chir Soc Med Nat Iasi*. 2015;119: 982–987.
151. Pourkarim MR, Amini-Bavil-Olyae S, Kurbanov F, Van Ranst M, Tacke F. Molecular identification of hepatitis B virus genotypes/ subgenotypes: Revised classification hurdles and updated resolutions. *World J Gastroenterol*. 2014;20: 7152–7168. doi:10.3748/wjg.v20.i23.7152
152. Li F, Zhang D, Li Y, Jiang D, Luo S, Du N, et al. Whole genome characterization of hepatitis B virus quasispecies with massively parallel pyrosequencing. *Clin Microbiol Infect*. 2015;21: 280–287. doi:10.1016/j.cmi.2014.10.007
153. Park YM, Jang JW, Yoo SH, Kim SH, Oh IM, Park SJ, et al. Combinations of eight key mutations in the X / preC region and genomic activity of hepatitis B virus are associated with hepatocellular carcinoma. *J Viral Hepat*. 2014;21: 171–177. doi:10.1111/jvh.12134
154. Shi Y, Wang J, Wang Y, Wang A, Guo H, Wei F, et al. A novel mutant 10Ala/Arg together with mutant 144Ser/Arg of hepatitis B virus X protein involved in hepatitis B virus-related hepatocarcinogenesis in HepG2 cell lines. *Cancer Lett*. Elsevier Ireland Ltd; 2016;371: 285–291. doi:10.1016/j.canlet.2015.12.008
155. Sung F, Lan C, Huang C, Lin C, Liu C, Chen P, et al. Progressive accumulation of mutations in the hepatitis B virus genome and its impact for time to diagnosis of hepatocellular carcinoma. *Hepatology*. 2016; doi:10.1002/hep.28654
156. Kim H, Park J, Jee Y, Lee S, Kim H, Song B, et al. Hepatitis B Virus X Mutations Occurring Naturally Associated With Clinical Severity of Liver Disease Among Korean Patients With Chronic Genotype C Infection. *J Med Virol*. 2008;1343: 1337–1343. doi:10.1002/jmv
157. Günther S, Piwon N, Iwanska a, Schilling R, Meisel H, Will H. Type, prevalence, and significance of core promoter/enhancer II mutations in hepatitis B viruses from immunosuppressed patients with severe liver disease. *J Virol*. 1996;70: 8318–8331.
158. Kidd-Ljunggren K, Öberg M, Kidd AH. Hepatitis B virus X gene 1751 to 1764 mutations: Implications for HBeAg status and disease. *J Gen Virol*. 1997;78: 1469–1478.

159. Lee S-A, Mun H-S, Kim H, Lee H-K, Kim B-J, Hwang E-S, et al. Naturally occurring hepatitis B virus X deletions and insertions among Korean chronic patients. *J Med Virol*. 2011;83: 65–70. doi:10.1002/jmv.21938
160. Günther S, Baginski S, Kissel H, Reinke P, Krüger DH, Will H, et al. Accumulation and persistence of hepatitis B virus core gene deletion mutants in renal transplant patients are associated with end-stage liver disease. *Hepatology*. 1996;24: 751–8. doi:10.1002/hep.510240401
161. Peng Y, Liu B, Hou J, Sun J, Hao R, Xiang K, et al. Naturally occurring deletions/insertions in HBV core promoter tend to decrease in HBeAg-positive chronic hepatitis B patients during antiviral therapy. *Antivir Ther*. 2015;20: 623–632. doi:10.3851/IMP2955
162. Domingo E, Escarmís C, Lázaro E, Manrubia SC. Quasispecies dynamics and RNA virus extinction. *Virus Res*. 2005;107: 129–139. doi:10.1016/j.virusres.2004.11.003
163. Olsper A, Chung BY, Atkins JF, Carr JP, Firth AE. Transcriptional slippage in the positive-sense RNA virus family Potyviridae. *EMBO Rep*. 2015;16: 995–1004.
164. Barzon L, Lavezzo E, Militello V, Toppo S, Palù G. Applications of next-generation sequencing technologies to diagnostic virology. *Int J Mol Sci*. 2011;12: 7861–7884. doi:10.3390/ijms12117861
165. Liu F, Chen L, Yu D-M, Deng L, Chen R, Jiang Y, et al. Evolutionary patterns of hepatitis B virus quasispecies under different selective pressures: correlation with antiviral efficacy. *Gut*. 2011;60: 1269–77. doi:10.1136/gut.2010.226225
166. Domingo E, Gomez J. Quasispecies and its impact on viral hepatitis. *Virus Res*. 2007;127: 131–150. doi:10.1016/j.virusres.2007.02.001
167. Domingo E, Sheldon J, Perales C. Viral Quasispecies Evolution. *Microbiol Mol Biol Rev*. 2012;76: 159–216. doi:10.1128/MMBR.05023-11
168. Chen L, Zhang Q, Yu D, Wan M, Zhang X. Early changes of hepatitis B virus quasispecies during lamivudine treatment and the correlation with antiviral efficacy q. *J Hepatol*. European Association for the Study of the Liver; 2009;50: 895–905. doi:10.1016/j.jhep.2008.12.018
169. Lim SG, Cheng Y, Guindon S, Seet BL, Lee LY, Hu P, et al. Viral Quasi-Species Evolution During Hepatitis Be Antigen Seroconversion. *Gastroenterology*. 2007;133: 951–958. doi:10.1053/j.gastro.2007.06.011
170. Cheng Y, Guindon S, Rodrigo A, Wee LY, Inoue M, Thompson AJ V, et al. Cumulative viral evolutionary changes in chronic hepatitis B virus infection precedes hepatitis B e antigen seroconversion. *Gut*. 2013;69: 1347–1355. doi:10.1136/gutjnl-2012-302408

171. Gregori J, Perales C, Rodriguez-Frias F, Esteban JI, Quer J, Domingo E. Viral quasispecies complexity measures. *Virology*. Elsevier; 2016;493: 227–237. doi:10.1016/j.virol.2016.03.017
172. Yang Z, Huang S, Chen L, Liu F, Cai X, Guo Y, et al. Characterization of Full-Length Genomes of Hepatitis B Virus Quasispecies in Sera of Patients at Different Phases of Infection. 2015;53: 2203–2214. doi:10.1128/JCM.00068-15
173. Gregori J, Salicrú M, Domingo E, Sanchez A, Esteban JI, Rodríguez-Frías F, et al. Inference with viral quasispecies diversity indices: Clonal and NGS approaches. *Bioinformatics*. 2014;30: 1104–1111. doi:10.1093/bioinformatics/btt768
174. Osiowy C, Giles E. Evaluation of the INNO-LiPA HBV Genotyping Assay for Determination of Hepatitis B Virus Genotype Evaluation of the INNO-LiPA HBV Genotyping Assay for Determination of Hepatitis B Virus Genotype. *J Clin Microbiol*. 2003;41: 5473–5477. doi:10.1128/JCM.41.12.5473
175. Hannachi N, Fredj N Ben, Bahri O, Thibault V, Ferjani A, Gharbi J, et al. Molecular analysis of HBV genotypes and subgenotypes in the Central-East region of Tunisia. *Virology*. 2010;7: 302. doi:10.1186/1743-422X-7-302
176. Lowe CF, Merrick L, Harrigan PR, Mazzulli T, Sherlock CH, Ritchie G. Implementation of next-generation sequencing for hepatitis B resistance and genotyping in a clinical microbiology laboratory. *J Clin Microbiol*. 2015;54: 127–133. doi:10.1128/JCM.02229-15
177. Hussain M, Chu CJ, Sablon E, Lok ASF. Rapid and sensitive assays for determination of hepatitis B virus (HBV) genotypes and detection of HBV precore and core promoter variants. *J Clin Microbiol*. 2003;41: 3699–3705. doi:10.1128/JCM.41.8.3699-3705.2003
178. Chayama K, Suzuki Y, Kobayashi M, Tsubota A, Himoto M, Miyano H, Koike H, Kobayashi M, Koida I, Arase Y, Saitoh S, Murashima N, Ikeda K KH. Emergence and Takeover of YMDD Motif Mutant Hepatitis B Virus During Long-Term Lamivudine Therapy and Retakeover by Wild Type After Cessation of Therapy. *Hepatology*. 1998;27: 1711–1716.
179. Cane PA, Cook P, Ratcliffe D, Mutimer D. Use of Real-Time PCR and Fluorimetry To Detect Lamivudine Resistance-Associated Mutations in Hepatitis B Virus. *Antimicrob Agents Chemother*. 1999;43: 1600–1608.
180. Kim HS, Han KH, Ahn SH, Kim EO, Chang HY MM, Chung HJ, Yoo W, Kim SO HS. Evaluation of methods for monitoring drug resistance in chronic hepatitis B patients during lamivudine therapy based on mass spectrometry and reverse hybridization. *Antivir Ther*. 2005;10: 441–449.

181. Tran N, Berne R, Chann R, Gauthier M, Martin D, Armand M, et al. European Multicenter Evaluation of High-Density DNA Probe Arrays for Detection of Hepatitis B Virus Resistance Mutations and Identification of Genotypes. *J Clin Microbiol.* 2006;44: 2792–2800. doi:10.1128/JCM.00295-06
182. Rong-Sheng Wang, Hua Zhang, Yu-Fen Zhu, Bei Han Z-JY. Detection of YMDD mutants using universal template real-time PCR. *World J Gastroenterol.* 2006;12: 1308–1311.
183. Mignardi M, Nilsson M. Fourth-generation sequencing in the cell and the clinic. *Genome Med.* 2014;6: 1–4. doi:10.1186/gm548
184. Domingo E, Gregori J, Perales C, Rodriguez-Frias F, Esteban JI, Quer J, et al. Deep Sequencing : Becoming a Critical Tool in Clinical Virology. *J Mol Diagnostics.* Elsevier B.V.; 2015;4: 265–276. doi:10.3390/ijms12117861
185. Ramírez C, Gregori J, Buti M, Tabernero D, Camós S, Casillas R, et al. A comparative study of ultra-deep pyrosequencing and cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. *Antiviral Res.* 2013;98: 273–283. doi:10.1016/j.antiviral.2013.03.007
186. Vincent AT, Derome N, Boyle B, Culley AI, Charette SJ. Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money. *J Microbiol Methods.* Elsevier B.V.; 2015; doi:10.1016/j.mimet.2016.02.016
187. Datta S, Budhaliya R, Das B, Chatterjee S, Vanlalhmuka, Veer V. Next-generation sequencing in clinical virology: Discovery of new viruses. *World J Virol.* 2015;4: 265–76. doi:10.5501/wjv.v4.i3.265
188. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* Nature Publishing Group; 2016;17: 333–351. doi:10.1038/nrg.2016.49
189. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nat Biotechnol.* 2006;437: 376–380. doi:10.1038/nature03959.
190. Metzker ML. Sequencing technologies — the next generation. *Nat Rev Genet.* Nature Publishing Group; 2009;11: 31–46. doi:10.1038/nrg2626
191. Roche Applied Science. GS FLX/Junior Titanium Technology. www.My454.Com. 2016.
192. Roche. 454 Sequencing System Software Manual Version 2.9. 454 Life Sciences. Signal Processing. 2013. pp. 14–33.

193. Lefterova MI, Suarez CJ, Banaei N, Pinsky BA. Next-Generation Sequencing for Infectious Disease Diagnosis and Management. *J Mol Diagnostics*. American Society for Investigative Pathology and the Association for Molecular Pathology; 2015;17: 623–634. doi:10.1016/j.jmoldx.2015.07.004
194. Buti M, Elefsiniotis I, Jardi R, Vargas V, Rodriguez-Frias F, Schapper M, et al. Viral genotype and baseline load predict the response to adefovir treatment in lamivudine-resistant chronic hepatitis B patients. *J Hepatol*. 2007;47: 366–72. doi:10.1016/j.jhep.2007.04.011
195. Cheng Y, Guindon S, Rodrigo A, Lim SG. Increased viral quasispecies evolution in HBeAg seroconverter patients treated with oral nucleoside therapy. *J Hepatol*. European Association for the Study of the Liver; 2013;58: 217–224. doi:10.1016/j.jhep.2012.09.017
196. Boni C, Laccabue D, Lampertico P, Giuberti T, Viganò M, Schivazappa S, et al. Restored Function of HBV-Specific T Cells After Long-term Effective Therapy With Nucleos(t)ide Analogues. *Gastroenterology*. 2012;143: 963–973.e9. doi:10.1053/j.gastro.2012.07.014
197. Tatematsu K, Tanaka Y, Kurbanov F, Sugauchi F, Mano S, Maeshiro T, et al. A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. *Journal of virology*. 2009. pp. 10538–10547. doi:10.1128/JVI.00462-09
198. Jardi R, Rodriguez-Frias F, Schaper M, Giggi E, Tabernero D, Homs M, et al. Analysis of hepatitis B genotype changes in chronic hepatitis B infection: Influence of antiviral therapy. *J Hepatol*. European Association for the Study of the Liver; 2008;49: 695–701. doi:10.1016/j.jhep.2008.03.030
199. Sablon E, Shapiro F. Advances in molecular diagnosis of HBV infection and drug resistance. *Int J Med Sci*. 2005;2: 8–16. doi:10.7150/ijms.2.8
200. Osiowy C, Giles E. Evaluation of the INNO-LiPA HBV genotyping assay for determination of hepatitis B virus genotype. *J Clin Microbiol*. 2003;41: 5473–7.
201. Simmonds P, Midgley S. Recombination in the Genesis and Evolution of Hepatitis B Virus Genotypes Recombination in the Genesis and Evolution of Hepatitis B Virus Genotypes. *J Virol*. 2005;79: 15467–15476. doi:10.1128/JVI.79.24.15467
202. Mese S, Arıkan M, Cakiris A, Abaci N, Gumus E, Kursun O, et al. Role of the line probe assay INNO-LiPA HBV DR and ultradeep pyrosequencing in detecting resistance mutations to nucleoside/nucleotide analogues in viral samples isolated from chronic hepatitis B patients. *J Gen Virol*. 2013;94: 2729–38. doi:10.1099/vir.0.053041-0

203. Yang J, Xing K, Deng R, Wang J, Wang X. Identification of Hepatitis B virus putative intergenotype recombinants by using fragment typing. *J Gen Virol*. 2006;87: 2203–2215. doi:10.1099/vir.0.81752-0
204. Neumann AU. Hepatitis B Viral Kinetics : A Dynamic Puzzle Still to Be Resolved. *Hepatology*. 2005;42: 249–254. doi:10.1002/hep.20831
205. Sanchez MJ, Buti M, Homs M, Palacios A, Rodriguez-Frias F, Esteban R. Successful use of entecavir for a severe case of reactivation of hepatitis B virus following polychemotherapy containing rituximab. *J Hepatol*. 2009;51: 1091–6. doi:10.1016/j.jhep.2009.07.012
206. Homs M, Buti M, Quer J, Jardí R, Schaper M, Tabernero D, et al. Ultra-deep pyrosequencing analysis of the hepatitis B virus preCore region and main catalytic motif of the viral polymerase in the same viral genome. *Nucleic Acids Res*. 2011;39: 8457–8471. doi:10.1093/nar/gkr451
207. Quer J, Gregori J, Rodríguez-Frias F, Buti M, Madejon A, Perez-del-Pulgar S, et al. High-resolution hepatitis C virus subtyping using NS5B deep sequencing and phylogeny, an alternative to current methods. *J Clin Microbiol*. 2015;53: 219–26. doi:10.1128/JCM.02093-14
208. Chen MT, Billaud J-N, Sällberg M, Guidotti LG, Chisari F V, Jones J, et al. A function of the hepatitis B virus precore protein is to regulate the immune response to the core antigen. *Proc Natl Acad Sci U S A*. 2004;101: 14913–8. doi:10.1073/pnas.0406282101
209. Belloni L, Pollicino T, De Nicola F, Guerrieri F, Raffa G, Fanciulli M, et al. Nuclear HBx binds the HBV minichromosome and modifies the epigenetic regulation of cccDNA function. *Proc Natl Acad Sci U S A*. 2009;106: 19975–9. doi:10.1073/pnas.0908365106
210. Guo YH, Li YN, Zhao JR, Zhang J YZ. HBc binds to the CpG islands of HBV cccDNA and promotes an epigenetic permissive state. *Epigenetics*. 2011;6: 720–726.
211. Rehm HL, Bale SJ, Bayrak-toydemir P, Jonathan S, Brown KK, Deignan JL, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med*. 2013;15: 733–747. doi:10.1038/gim.2013.92.ACMG
212. Lilley CE, Schwartz RA, Weitzman MD. Using or abusing : viruses and the cellular DNA damage response. *Trends Microbiol*. 2007;15: 119–26. doi:10.1016/j.tim.2007.01.003
213. Wong DK-H, Yuen M-F, Yuan H, Sum SS-M, Hui C-K, Hall J, et al. Quantitation of covalently closed circular hepatitis B virus DNA in chronic hepatitis B patients. *Hepatology*. 2004;40: 727–37. doi:10.1002/hep.20353

214. Dandri M, Burda MR, Will H, Petersen J. Increased hepatocyte turnover and inhibition of woodchuck hepatitis B virus replication by adefovir in vitro do not lead to reduction of the closed circular DNA. *Hepatology*. 2000;32: 139–46. doi:10.1053/jhep.2000.8701
215. Moraleda G, Saputelli J, Aldrich CE, Averett D, Condreay L, Mason WS. Lack of effect of antiviral therapy in nondividing hepatocyte cultures on the closed circular DNA of woodchuck hepatitis virus. *J Virol*. 1997;71: 9392–9.
216. Lisby M, Rothstein R. Cell biology of mitotic recombination. *Cold Spring Harb Perspect Biol*. 2015;7: a016535. doi:10.1101/cshperspect.a016535
217. Simmonds P, Midgley S. Recombination in the genesis and evolution of hepatitis B virus genotypes. *J Virol*. 2005;79: 15467–76. doi:10.1128/JVI.79.24.15467-15476.2005
218. Shi W, Zhu C, Zheng W, Carr MJ, Higgins DG, Zhang Z. Subgenotype reclassification of genotype B hepatitis B virus. *BMC Gastroenterol. BMC Gastroenterology*; 2012;12: 116. doi:10.1186/1471-230X-12-116
219. Magnus LO, Norder H. Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. *Intervirology*. 1995;38: 24–34.