

Aus dem Institut für Immunologie

Laboratoire National de Santé, Luxemburg

Die weltweite Verbreitung und Evolution des Hepatitis B  
Virus: Rekombinationen, neue Genotypen und Subtypen

*The worldwide distribution and evolution of the hepatitis B  
virus: recombinations, new genotypes and subtypes*

DISSERTATION

der medizinischen Fakultät

der UNIVERSITÄT DES SAARLANDES

zur Erlangung des Grades eines Doktors der Naturwissenschaften

vorgelegt von

Christophe M. Olinger

2008

## Table of Contents

Zusammenfassung .....	1
Abstract .....	3

## Chapter 1: Introduction ..... 5

1. State of the art and objectives .....	5
2. The hepatitis B virus .....	7
2.1. General information.....	7
2.2. Structure of <i>hepadnaviridae</i> .....	9
2.3. Non structural proteins .....	10
2.4. The HBV genome.....	12
2.5. Life cycle of <i>hepadnaviridae</i> .....	13
3. Clinical features of hepatitis B infection.....	19
4. Epidemiology .....	23
4.1. Global distribution of HBV infection.....	24
4.2. HBV serotypes and genotypes.....	24
4.3. Transmission of HBV.....	26
4.4. Prevention of HBV infection.....	27
4.5. Treatment of chronic HBV infection.....	27
4.6. Co-infection with other viruses .....	28

## Chapter 2: Material..... 29

1. Chemicals .....	29
2. Buffers and Solutions .....	30
3. Enzymes .....	31
4. DNA markers .....	31
5. Bacterial strain.....	31

6.	Bacterial media.....	31
7.	Kits .....	32
8.	Vectors .....	33
8.1.	pCR <sup>®</sup> 4-TOPO <sup>®</sup> .....	33
8.2.	pCR <sup>®</sup> -Blunt II-TOPO <sup>®</sup> .....	34
9.	PCR primers .....	35
9.1.	Complete genome PCR .....	35
9.2.	M13 cloning verification PCR.....	35
10.	Bioinformatics.....	36
11.	Machinery used .....	36

## Chapter 3: Methods ..... 38

1.	HBsAg detection .....	38
2.	DNA extraction .....	38
3.	RNA extraction .....	39
4.	DNA quantification .....	39
5.	Polymerase chain reaction.....	40
5.1.	Amplification of the HBV complete genome .....	41
5.2.	M13 PCR.....	42
5.3.	Semi-quantitative detection of HBV DNA by real time PCR .....	43
6.	Agarose gel electrophoresis .....	43
7.	Cloning .....	44
7.1.	TOPO TA Cloning <sup>®</sup> kit .....	45
7.2.	Zero Blunt <sup>®</sup> TOPO <sup>®</sup> PCR Cloning kit.....	45
8.	Sequencing .....	46
9.	Phylogenetic analysis .....	48
9.1.	Electropherogram analysis .....	48
9.2.	Sequence alignment.....	48

9.3.	Nucleotide distance calculation .....	49
9.4.	Phylogenetic reconstruction .....	50
9.5.	Recombination analysis.....	51

## Chapter 4: Results and Discussions .. 52

### Part I..... 52

Hepatitis B virus genotype E HBsAg detection with different immunoassays and diagnostic impact of mutations in the preS/S gene

1.	Clinical samples .....	53
2.	Serological assays .....	53
3.	Results .....	54
3.1.	HBV DNA detection .....	54
3.2.	HBsAg detection .....	54
3.3.	Quantitative impact of preS/S gene mutations on HBsAg detection.....	56
4.	Discussion .....	58

### Part II..... 60

The Central African Republic is part of the West-African hepatitis B virus genotype E crescent

1.	Clinical samples .....	61
2.	Results .....	62
2.1.	Serology .....	62
2.2.	Sequence analysis and genotyping .....	65
2.3.	PreS/S gene analysis.....	65
2.4.	PreC/C gene analysis.....	67
2.5.	Mixed infections and recombination .....	67
3.	Discussion .....	68

## Part III ..... 71

### Phylogenetic analysis of the preC/C gene of hepatitis B genotypes E and A in West-Africa: New subtypes, mixed infections and recombinations

1.	Clinical samples .....	72
2.	Amplification of the preC/C region .....	72
3.	Results .....	74
3.1.	Screening for mixed infection .....	74
3.2.	Genotype prevalence and mixed infections .....	78
3.3.	Genetic variability of genotypes and subtypes .....	79
3.4.	Mutations in the preC/C gene .....	81
3.5.	Insertions and deletions in the preC/C gene .....	82
3.6.	HBeAg status and mutations .....	84
4.	Discussion .....	84

## Part IV ..... 88

### Multiple genotypes and subtypes of HBV and HCV in Belarus: Similarities with Russia and West-European influences

1.	Clinical samples .....	89
2.	Amplification and cloning of HBV and HCV .....	89
3.	Sequencing and phylogenetic analysis of HCV .....	90
4.	Results and discussion .....	90
4.1.	Hepatitis B virus .....	90
4.2.	Hepatitis C virus .....	93

## Part V ..... 96

### Identification and molecular characteristics of a new hepatitis B genotype I in Lao People's Democratic Republic

1.	Clinical samples .....	97
2.	Alternative complete genome amplification .....	97
3.	Phylogenetic analysis .....	97
4.	Results .....	98
4.1.	Detection and sequencing of HBV DNA.....	98
4.2.	Distribution of genotypes .....	98
4.3.	Definition of a new genotype I.....	99
4.4.	Nucleotide sequence analysis.....	99
4.5.	Amino acid sequence analysis.....	103
4.6.	Mixed infections and recombinations.....	103
5.	Discussion .....	107

## Chapter 5: Conclusions and Perspectives . 109

References .....	113
Annexes .....	133
Curriculum.....	1

## Index of Figures

Figure 1 : Electron-microscopy image of HBV particles.....	8
Figure 2 : Schematic representation of hepatitis B particles.....	9
Figure 3 : The human HBV genome .....	12
Figure 4 : Schematic view of the life cycle of HBV .....	14
Figure 5 : The mRNAs of the hepatitis B virus.....	18
Figure 6 : Serological profiles of acute and chronic HBV infection.....	20
Figure 7 : Global prevalence of HBsAg in 2005.....	23
Figure 8 : Worldwide distribution of HBV genotypes.....	25
Figure 9 : Map of the pCR4-TOPO cloning vector.....	33
Figure 10 : Map of the pCR-Blunt II-TOPO vector.....	34
Figure 11 : Example of a sequence electropherogram.....	48
Figure 12 : Example of a sequence alignment .....	49
Figure 13 : Example of a phylogenetic tree .....	50
Figure 14 : Example of bootscan analysis.....	51
Figure 15 : Phylogenetic tree based on the complete genome of HBV strains isolated from patients from the Central African Republic .....	63
Figure 16 : Phylogenetic tree based on the S gene of HBV strains from the Central African Republic .....	64
Figure 17 : Phylogenetic tree based on the preC/C gene of HBV strains from the Central African Republic .....	66

Figure 18 : Similarity plots comparing the S, X and preC/C genes of CAR177 .....	69
Figure 19 : PCR reactions used to detect mixed infections and for genotype specific PCR ...	74
Figure 20 : Phylogenetic tree of sequences clustering with genotype A or E reference strains .....	77
Figure 21 : Phylogenetic tree of genotype A complete genome sequences and preS fragment sequences.....	79
Figure 22 : Phylogenetic analysis of the NIE24072 outlier based on the first 200 bp, the complete preC/C gene and the last 317 bp .....	81
Figure 23 : Phylogenetic reconstruction of the preS fragment and complete genome of HBV strains found in Belarus and reference sequences .....	94
Figure 24 : Phylogenetic reconstruction of the core/E1 region of the HCV strains found in Belarus and reference sequences.....	94
Figure 25 : Relative distribution of genotypes and subtypes identified in 386 DNA positive rejected blood donors from Lao PDR .....	98
Figure 26 : Phylogenetic comparison of genotype I complete genomes to a set of reference sequences.....	100
Figure 27 : Bootscan analysis of genotype I .....	101
Figure 28 : Phylogenetic comparison of genotype I with other genotypes on the distinct sequence fragments .....	103
Figure 29 : Evolutionary history of genotype I and three recombinant strains.....	104
Figure 30 : Diagrams of recombinants and genotype I .....	106

## Index of Tables

Table 1: Single letter amino acid code .....	XII
Table 2 : Primers and conditions for the amplification and sequencing of the complete genome using 4 semi-nested PCR reactions .....	35
Table 3 : Primers used in the M13 cloning verification PCR .....	35
Table 4 : Constant parameters for the complete genome PCR reactions .....	42
Table 5 : Conditions of the M13 PCR reaction .....	43
Table 6 : Quantity of DNA necessary for sequencing .....	47
Table 7 : Nucleotide substitution models .....	49
Table 8 : Comparison of DNA status with the HBsAg status obtained with 3 different assays .....	54
Table 9 : Comparison of DNA status and HBsAg status of patients with discordant results in the 3 HBsAg assays .....	55
Table 10 : Classification of DNA positive serum samples into 3 groups according to the HBsAg detection signals .....	57
Table 11: Characteristics of the patients cohort and their HBV serology status.....	62
Table 12 : Comparison of the results of the genotype A (rvA) and nonA (rvnonA) specific PCR, with the genotype of the sequence of the PCR fragments as well as the genotype-independent PCR.....	75
Table 13 : Prevalence of single and mixed genotype infections .....	76
Table 14 : Specific mutations in the preC/C gene of 29 genotype A and 89 genotype E sequences.....	83
Table 15 : Comparison of PCR results with HBsAg and anti-HCV ELISA results .....	91

Table 16 : Genotype distribution of HBV and HCV in Belarus .....	92
Table 17 : Distances within and between the different genotypes and subtypes present in Belarus.....	93
Table 18 : Genotype, accession number and genetic distances and estimated years of evolution of the three sequence domains identified in genotype I.....	102
Table 19 : Genotype, accession number and genetic distances and estimated years of evolution of three strains showing signs of recombination involving genotype I. ....	104

## List of abbreviations

aa	amino acid
ALT	alanine aminotransferase
bp	base pair
CAR	Central African Republic
DHBV	duck hepatitis virus
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide-triphosphate
DR	direct repeat
ELISA	enzyme-linked immunosorbant assay
ER	endoplasmic reticulum
GSHV	ground squirrel hepatitis virus
HBV	hepatitis B virus
HCC	hepatocellular carcinoma
HCV	hepatitis C virus
HIV	human immunodeficiency virus
kb	kilobases
Lao PDR	Lao People's Democratic Republic
MgCl <sub>2</sub>	magnesium chloride
mRNA	messenger RNA

Neg	negative
NLS	nuclear localization signal
nt	nucleotides
ORF	open reading frame
PCR	polymerase chain reaction
Pg	pregenomic
Pos	positive
RNA	ribonucleic acid
RPM	rounds per minute
SNP	single nucleotide polymorphism
WHV	woodchuck hepatitis virus

**Table 1: Single letter amino acid code**

<b>Single letter code</b>	<b>Amino acid</b>
G	Glycine (Gly)
P	Proline (Pro)
A	Alanine (Ala)
V	Valine (Val)
L	Leucine (Leu)
I	Isoleucine (Ile)
M	Methionine (Met)
C	Cysteine (Cys)
F	Phenylalanine (Phe)
Y	Tyrosine (Tyr)
W	Tryptophan (Trp)
H	Histidine (His)
K	Lysine (Lys)
R	Arginine (Arg)
Q	Glutamine (Gln)
N	Asparagine (Asn)
E	Glutamic Acid (Glu)
D	Aspartic Acid (Asp)
S	Serine (Ser)
T	Threonine (Thr)

# Zusammenfassung

Das Hepatitis B Virus wurde in acht Genotypen und vor Kurzem auch mehrere Subtypen unterteilt, basierend auf der Akkumulation von Mutationen in seinem DNS Genom. Die Mehrzahl dieser Genotypen weist eine spezifische geographische Verbreitung auf. Die vorliegende Arbeit stellt zusätzliche Daten über Genotyp E vor, der nur in den Ländern südlich der Sahara vorkommt und in dieser Region auch als fast einziger Genotyp. Vorherige Studien haben belegt, dass Genotyp E Stämme eine niedrigere Diversität aufweisen, als Stämme anderer Genotypen, was in Anbetracht ihrer Verbreitung über fast ein Drittel des Afrikanischen Kontinents überrascht. Zusätzlich beinhaltet diese Arbeit eine phylogenetische Studie über Genotyp D in Weißrussland, sowie die Definition und Beschreibung eines neuen Genotyps I, das von uns in Laos gefunden wurde.

In einem ersten Ansatz stellen wir dar, dass kommerziell erhältliche Testkits für das Detektieren von HBsAg, dem primären Marker für Hepatitis B Infektionen, unterschiedliche Resultate aufzeigen können und spezifische Genotyp E Mutationen sich positiv auf die Detektion auswirken. Folglich besteht die Möglichkeit, dass diagnostische Tests weltweit unterschiedlich reagieren und besondere Vorsicht ihrer Auswahl und Entwicklung nötig ist.

In einer Studie von Stämmen, die wir in der Zentralafrikanischen Republik und schon vorher untersuchten Ländern gefunden haben, zeigen wir, dass Genotyp E auch weiter im Osten Afrikas vorkommt und konnten die niedrige Genotyp E Diversität bestätigen. Während vorherige Studien auf Sequenzen des S Genes basierten, stützen sich unsere Resultate auf eine Vielzahl von C-Gen Sequenzen. Die einzig anderen von uns gefunden Genotypen waren Genotyp A und D. Die Genotyp A Stämme konnten in 3 verschiedene Subtypen eingeteilt werden, von denen zwei (A4 und A5) zum ersten Mal von uns beschrieben werden. Jeder dieser Genotyp A Subtypen wurde in einem anderen afrikanischen Land gefunden (A3 in Kamerun, A4 in Mali und A5 in Nigeria). Diese Resultate sind umso überraschender angesichts der niedrigen Diversität die in Genotyp E Stämmen aus den gleichen Ländern beobachtet wurde. Ausserdem zeigen wir, dass in Kamerun, dem einzigen Land in dieser Region in dem die beiden Genotypen A und E mit ähnlichen Prevalenzen zirkulieren, 37% der Patienten gleichzeitig mit Stämmen beider Genotypen infiziert sind. Vier Sequenzen von

Stämmen aus Kamerun, Nigeria und der Zentralafrikanischen Republik weisen Zeichen von Rekombinationen zwischen den Genotypen A und E, E und D und in einem Fall sogar E, A und G auf. Diese Resultate zeigen, dass Rekombinationen häufiger vorkommen als bisher angenommen.

In einer phylogenetischen Studie von 69 Hepatitis B Stämmen aus Weißrussland weisen wir die dortige Präsenz der Genotypen A und D nach, wobei die Genotyp D Stämme in vier erst kürzlich definierte unterschiedliche Subtypen eingeteilt werden konnten. Diese Sequenzen zeigen, dass die Hepatitis B Verbreitung in Weißrussland der Russlands ähnlich ist, allerdings mit sichtbaren West-Europäischen Einflüssen.

Letztendlich beschreiben wir zusätzlich zu den schon gefundenen Rekombinationen in Afrika, 19 verwandte Stämme aus Laos, die das Resultat von wahrscheinlich mehreren Rekombinationen zwischen mindestens 3 Genotypen sind. Diese Stämme gruppierten mit keinem der bisher bekannten Genotypen und erfüllen die Kriterien für die Definition eines neuen Genotyps I. Weitere Analysen zeigten, dass dieser Genotyp ein Resultat von Rekombinationen zwischen den Genotypen G, C (C3), A (A3) und sogar E ist, die wahrscheinlich vor etwa 500 Jahren stattgefunden haben. Ausserdem zeigen wir, dass Genotyp I höchstens vor 200 Jahren in Laos eingeführt wurde. Die Identifizierung und Analyse dieses neuen Genotyps ist ein zusätzlicher Beweis dafür, dass Rekombinationen eine wichtige Antriebsfeder der Evolution des Hepatitis B Virus sind.

## Abstract

The hepatitis B virus has been divided into 8 genotypes and recently also multiple subtypes, based on the accumulation of mutations in its DNA genome. Most of these genotypes present distinct geographic distributions. Here we further characterize genotype E which only circulates and is dominant in Sub-Saharan Africa. It was previously shown that genotype E strains present a lower diversity than other HBV genotypes, which, considering its presence on over a third of the African continent is unexpected. In addition, we present a phylogenetic analysis of genotype D in Belarus and describe a new genotype I found by us in Lao People's Democratic Republic.

In a first approach, we demonstrate that currently available commercial assays for the detection of HBsAg, the primary hepatitis B marker, perform differently when compared to each other and present a positive detection bias caused by genotype E specific mutations. Thus, diagnostic assays may not perform the same worldwide and careful attention is warranted in their development and selection.

We also demonstrate that the genotype E prevalence in sub-Saharan Africa extends as far as the Central African Republic and confirm the low genotype E diversity in this sub-Saharan Africa, by obtaining and analysing sequences covering a different region of the hepatitis B genome than the one used in previous studies. The only other genotypes we found were genotypes A and D. The genotype A strains could be divided into three distinct subtypes of which two had not been described before (A4, A5). Each subtype of genotype A was found in a specific country (A3 in Cameroon, A4 in Mali and A5 in Nigeria), in contrast to genotype E of which all sequences identified from these countries presented a conspicuously low diversity. We also show that in Cameroon, the only country in sub-Saharan Africa where genotypes A and E co-circulate with equal prevalences, 37% of patients were co-infected with both genotypes. Four sequences, obtained from strains in Nigeria, Cameroon and the Central African Republic, showed signs of recombination events between genotypes A and E, E and D and even E, A and G and indicate that recombination events may be more frequent than previously suspected.

In a phylogenetic study of 69 hepatitis B strains in Belarus, we show that these strains belong to genotypes A and D, with strains of genotype D being dividable into 4 distinct recently described subtypes. The sequences obtained here, further add to the current definition of subtypes and indicate a genotype distribution in Belarus similar to Russia but with West-European influences.

Finally, adding to the recombination events identified during our studies, we also show that 19 related strains identified in Lao PDR, are the result of recombination events between at least three genotypes. These strains, which did not cluster with any known HBV genotype, fulfil the formal criteria for the definition of a new genotype (I). Further analysis indicated that this new genotype is a result of multiple recombination events involving genotypes G, C (C3), A (A3) and even E which probably occurred around 500 years ago. We also show that genotype I was only recently introduced in Lao PDR, probably not more than 200 years ago. The identification and analysis of the new genotype I provided further evidence of the importance of recombination in the evolution of hepatitis B.

# Chapter 1: Introduction

## 1. State of the art and objectives

The hepatitis B virus (HBV) is a DNA virus infecting humans, non-human primates (*e.g.* gibbon, chimpanzee and woolly monkey) and other animals (*e.g.* duck and woodchuck). Currently it is estimated that a third of the world population has come in contact with HBV during their lifetime and that as a consequence, 400 million people are suffering from chronic hepatitis B infection. A majority will eventually die of liver complications or hepatocellular carcinoma (HCC). Although most of the infected people live in areas such as sub-Saharan Africa, Asia and the Pacific, in which HBV is endemic, the virus is circulating globally with highly variable prevalences (Moyer and Mast 1994).

Based on the accumulation of single nucleotide polymorphisms in the genome and phylogenetic reconstruction, the circulating strains have been classified into 8 genotypes (A – H) (Okamoto *et al.* 1988; Norder *et al.* 1992) with some having been classified into additional distinct subtypes. While most of these genotypes present specific geographical distributions (B, C, E, F and H) some can be found worldwide (A and D) while genotype G appears only sporadically. A particularly noticeable geographic distribution has been observed for genotype E, a genotype found only in sub-Saharan Africa and of which until recently only few genomic sequences had been reported (Norder *et al.* 1994; Odemuyiwa *et al.* 2001). For several sub-Saharan countries HBV markers were found in as much as 50% of participants in unselected cohort studies and were even more frequent in human immunodeficiency virus (HIV) positive patients (Mulders *et al.* 2004). In 2004, a phylogenetic study covering hepatitis B strains from seven sub-Saharan countries (Mali, Burkina Faso, Togo, Benin, Nigeria, Cameroon and Democratic Republic of the Congo), representing a third of the African continent, revealed that a majority of circulating strains were of genotype E (Mulders *et al.* 2004) with one exception: in Cameroon the prevalence of genotype E strains was equal to the prevalence of strains of genotype A. In addition, sequence analysis of the circulating genotype E strains found during this study, revealed a surprisingly low variability, less than half of what would normally be expected. This is even more remarkable when considering that these strains were recovered from an area covering more than 6 million km<sup>2</sup> and in which more than 40% of the

African population lives. This suggests that genotype E has a short evolutionary history in humans and that the virus was introduced relatively late in sub-Saharan Africa. This late introduction would also be supported by the conspicuous absence of this genotype in Afro-Americans but it would be incompatible with an evolution from the closest known human HBV genotype, genotype D. The evolution from the latter would have taken about 700 years with an estimated evolutionary rate of  $4.2 \times 10^{-5}$  SNPs per site and year (Orito *et al.* 1989; Fares and Holmes 2002). Therefore, the relatively recent introduction from an animal reservoir must be seriously considered. If genotype E was introduced by a single introductory event it would have taken about 200 years to develop the 1.67% observed diversity. In contrast, introduction of HBV from other host-species into primates (including humans) is thought to have occurred 6000 years ago (Fares and Holmes 2002). Cross-species infections from humans to monkeys have been shown experimentally but there is no evidence of natural transmission of HBV from primates to humans. Genotype E has been found once in chimpanzee, but the direction of transmission could not be established (Takahashi *et al.* 2000). The possibility of transmission by blood-feeding arthropods has been considered but never been proven.

The present study aims to further characterize HBV genotype E and to advance our understanding of its widespread circulation but conspicuously low diversity. In addition, phylogenetic studies of HBV in Belarus and Lao PDR aim at gaining further insights in the mechanisms of distribution and evolution of HBV genotypes and subtypes. During these studies, special interest will be placed in detecting evidence of recombinations between genotypes, and to evaluate their importance in the history of HBV evolution.

In a first approach, data will be provided on the sensitivity and specificity of currently available hepatitis B diagnostic kits in light of the exceptionally high prevalences of hepatitis B found in sub-Saharan Africa and to evaluate a possible bias in regard to genotype E (Chapter 4, Part I). Phylogenetic classification of sequences obtained from strains circulating in the Central African Republic will give us valuable insights into the extent of the genotype E dominant crescent towards East Africa and the influence of genotypes D and A found in North and South-East Africa respectively (Chapter 4, Part II). Furthermore, the low diversity of genotype E will be confirmed on additional strains from the region with particular interest on the surprising co-circulation of genotype E and A in Cameroon and its consequences for

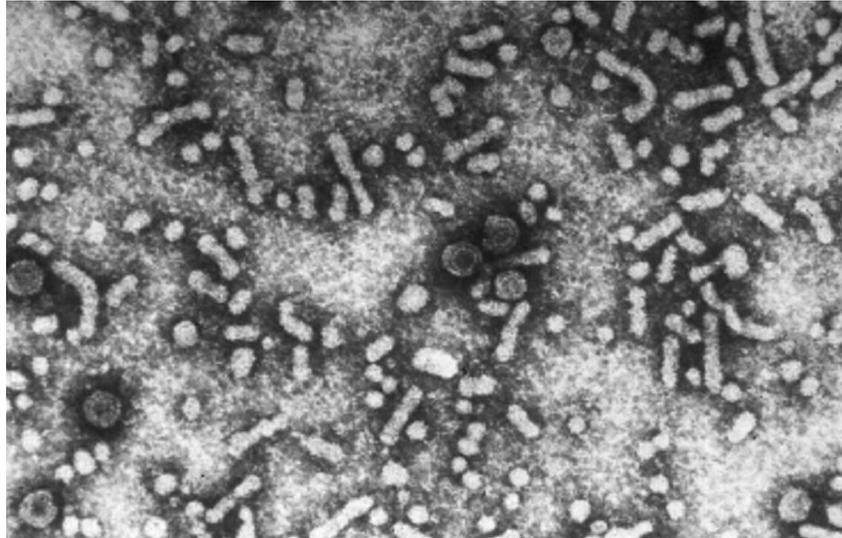
the evolution of hepatitis B, such as recombination events and infection with multiple HBV genotypes and subtypes (Chapter 4, Part III).

During this study additional information on the distribution of the subtypes of hepatitis B genotype D, were obtained from strains found in patients co-infected with HIV and/or HCV in Belarus (Chapter 4, Part IV). Finally, the analysis of a high number of hepatitis B strains found in Lao People's Democratic Republic (PDR) revealed the circulation of several strains not attributable to any known HBV genotype. These strains are clearly the results of multiple recombination events involving also genotype E or one of its ancestors. The tentative creation of a new genotype I was proposed (Chapter 4, Part V) and new insights in the evolutionary history of HBV were gained.

## 2. The hepatitis B virus

### 2.1. General information

The hepatitis B virus was the first human hepatitis virus from which the proteins as well as the genome were identified and characterized. Prior to this, hepatitis infections were mainly described by their routes of transmission: type A was mostly transmitted by the faecal-oral route while type B was transmitted from mother to child. In 1967 the type B infection was related to the appearance of a specific serum protein designated as the Australia antigen which was first observed in the serum of an Australian Aborigine (Blumberg *et al.* 1967). Using electron-microscopy this antigen was later linked to the presence of virus-like particles found in hepatitis type B patients which marked the discovery of the hepatitis B virus (Figure 1) (Dane *et al.* 1970). The name was retained, although other viruses causing hepatitis like symptoms that were also transmitted parenterally, were later described (hepatitis C, D and G viruses). In 1973 the viral nature of the particles was confirmed by the detection of an endogenous DNA polymerase activity leading to the characterization of the HBV genome as a small, circular DNA based molecule (Kaplan *et al.* 1973; Robinson *et al.* 1974; Robinson and Greenman 1974).



**Figure 1 : Electron-microscopy image of HBV particles**

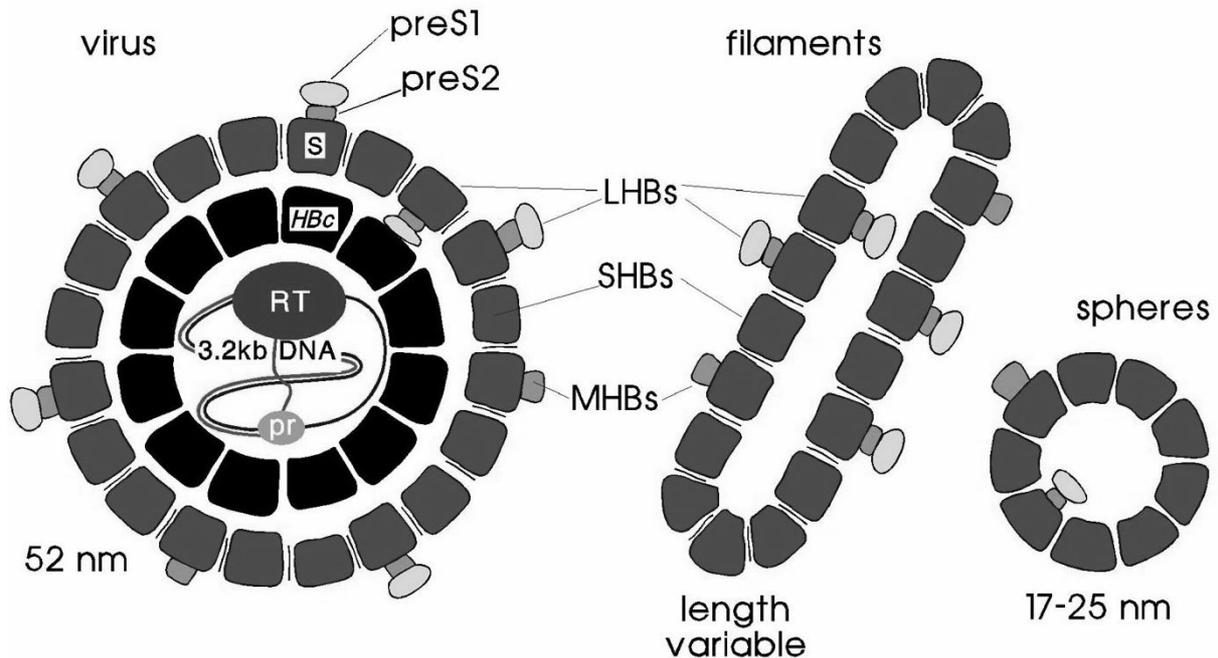
Hepatitis B particles found in the serum of an infected patient. The three forms: filaments, small spheres and complete particles are visible (Figure from [www.cdc.gov](http://www.cdc.gov)).

The hepatitis B virus and its relatives are classified into the family of the *Hepadnaviridae*, the name derived from their hepatotropism and the molecular nature of their DNA genome (Howard 1995). All other members of the family infect non human hosts but are not commonly of interest to veterinary medicine. On the other hand these viruses serve as models in the study of human HBV. The woodchuck hepatitis virus (WHV) for instance was discovered in woodchucks which had developed liver cancer (Summers *et al.* 1978). This ultimately led to the association of human HBV and an often occurrence of liver cancer in HBV infected patients. A similar virus was found in healthy ground squirrels (GSHV) (Marion *et al.* 1980) and arctic ground squirrels (Testut *et al.* 1996). Among non human primates, hepadnaviruses could be isolated from chimpanzees, orang-utans, gorillas, gibbons and woolly monkeys (Robertson and Margolis 2002). The virus strains found in the chimpanzees were very similar to the human HBV virus but the direction of transmission remains unclear (Takahashi *et al.* 2000; Makuwa *et al.* 2007). More distantly related hepadnaviruses were found in ducks (Yokosuka *et al.* 1985), grey herons (Sprengel *et al.* 1988), Ross' goose, snow goose (Chang *et al.* 1999), white storks (Pult *et al.* 2001) and cranes (Prassolov *et al.* 2003). Most important of the latter is the model of the duck hepatitis B virus (DHBV) in which a reverse transcription step from RNA to DNA was detected in core particles (Summers and Mason 1982), later also found in all other hepadnaviruses. Thus

hepadnaviruses belong to the group of retroviruses and more specifically, the pararetroviruses, to differentiate between other retroviruses such as the human immunodeficiency virus (HIV), which encode their genome as an RNA molecule and are designated as orthoretroviruses.

## 2.2. Structure of *hepadnaviridae*

Figures 1 and 2 show the three types of HBV virus associated structures found in the blood of infected persons. Similar particles can also be found in the blood of hepadnavirus infected wood chucks or ground squirrels (Summers *et al.* 1978; Mason *et al.* 1980). Only the larger structure, with a diameter of 42 to 45 nm is an actual viral particle, since the smaller spheres and filaments do not contain any genomic material. The outer protein shell or envelope is formed by the HBs protein or HBsAg (Dane *et al.* 1970) which forms an icosaedrical (spherical) structure. This HBs protein exists in three different forms depending on the start codon used during translation. The small HBs protein (SHBs) constitutes the majority of the viral particles, filaments and spheres, while the middle HBs protein (MHBs) and the small



**Figure 2 : Schematic representation of hepatitis B particles**

Apart from the infectious viral particle (left), the sera of highly viremic carriers contain non infectious particles formed of excessive HBs proteins and taking the shapes of filaments of variable lengths as well as small spheres with a diameter of 17 to 25 nm. Infected hepatocytes secrete 100 to 10000-fold more small spheres than

infectious particles. Additional components of the viral particle are the HBc proteins, the DNA genome and the viral polymerase. Between the SHBs units of particles and filaments small traces of ER derived lipids can be found (Figure from Thomas *et al.* 2005).

HBs protein, which have additional domains referred to as the preS1 and preS2 domain, are lesser constituents. The proportion of these forms determines the morphology (Marquardt *et al.* 1987). The inner shell, only found in virus particles, is formed by 240 HBc protein subunits and constitutes the capsid of the virus. The capsid surrounds the 3.2 kb long DNA genome which is covalently linked to the viral polymerase by a primase domain. Finally, each viral particle contains a cellular protein kinase and the heat shock proteins 90 and 70 (hsp90 and hsp70).

### 2.3. Non structural proteins

Three additional proteins encoded by the HBV genome are the HBe protein or HBeAg, the X protein and the polymerase. While the polymerase is fixed covalently to the genome inside the viral particle, the HBe protein and the X protein are not included in the viral particle. The HBe protein is encoded by the same open reading frame (ORF) than the HBc protein but uses an alternative upstream start codon which adds a small additional domain referred to as the preC domain. The preC sequence encodes a hydrophobic alpha helix, which is a secretion signal and allows for translocation of the HBe protein into the lumen of the ER (Bruss and Gerlich 1988; Standring *et al.* 1988). During that process, a signal peptidase cuts off 19 of the 29 amino acids of the preC region. The remaining 10 amino acids are sufficient to prevent the HBe protein from assembling into a capsid structure (Wasenauer *et al.* 1992). Thus, although the HBe and the HBc protein share 90% of their sequence, they follow completely different pathways in the viral life cycle. Following the first cleavage, parts of the HBe protein undergo additional cleavages and can be found either in the plasma membrane, in the nucleus or as secreted protein (Standring *et al.* 1988; Schlicht and Schaller 1989; Yeh *et al.* 1990; Wang *et al.* 1991; Nassal and Rieger 1993). Uncleaved HBe precursor protein accumulates as a phosphoprotein (Yang *et al.* 1992). The HBe protein is not essential for the viral life cycle and variants not producing HBe protein are often found during acute and chronic HBV infection. Nevertheless, the genome of all hepadnaviruses encodes the HBe protein. It is thought that the expression of the HBe protein, also called the HBeAg, causes immune tolerance and represents a viral strategy to evade the immune system. In fact, high levels of secreted HBeAg are found in low symptomatic but high viremic carriers and elimination of HBeAg is often accompanied by a flare-up of pathogenesis caused by the activation of the adaptive immune system and mass death of infected hepatocytes. Further prove of immune evasion was found

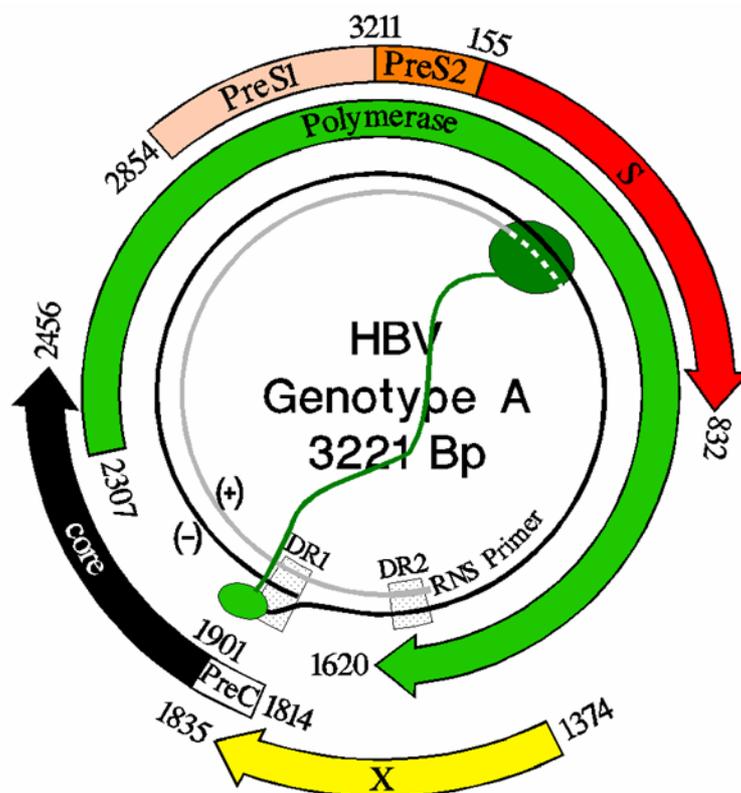
in the woodchuck hepatitis model in which the risk of chronic infection was higher in newborn of hepatitis B infected HBeAg positive mothers, while the infection was transient when mothers were HBeAg negative (Chen *et al.* 1992).

The HBV polymerase has four distinguishable domains and is essential for the viral lifecycle (Schlicht *et al.* 1991). The amino-terminal domain, also called the primase, is covalently linked to the 5' end of the minus-strand of virion DNA (Bartenschlager and Schaller 1988). The next domain functions as a spacer and is followed by the RNA- or DNA dependant DNA polymerase, i.e. reverse transcriptase. It has similarities to the reverse transcriptase of retroviruses (Toh *et al.* 1983; Beck *et al.* 2002) and is a target of current antiviral drugs against HBV (Torresi *et al.* 2002). The final domain encodes an RNase H activity which degrades the RNA component of RNA/DNA hybrid molecules. This activity, similar to the DNA dependant DNA polymerase activity, is only functional when incorporated into a fully assembled viral capsid (Nassal 1992; Gong *et al.* 2001). It is thought that only one polymerase protein is packaged in each viral particle (Gong *et al.* 2001).

Mutational studies suggest that the HBV X protein is not essential for virus particle production in an *in vitro* cell culture transfection system (Blum *et al.* 1992). However, transfection of WHV DNA lacking the X ORF into liver cells of woodchucks did not yield an infection (Chen *et al.* 1993; Zoulim *et al.* 1994). There is no clear indication that the X protein is a structural component of the virus particle. Within an infected cell, the protein is mainly found in the cytoplasm and a minor fraction in the nucleus (Doria *et al.* 1995). Although not all functions of the X protein are currently clear, it has been well documented that it presents pro-apoptotic activities (Schuster *et al.* 2002). Other studies have demonstrated that the presence of the X protein activates the description of several genes in a more or less unspecific manner, but whether it functions as a direct transcription factor or interacts with cellular pathways remains unclear (Rossner 1992; Kekule *et al.* 1993; Natoli *et al.* 1995; Qadri *et al.* 1995; Chirillo *et al.* 1996; Haviv *et al.* 1996; Henkler and Koshy 1996; Su and Schneider 1996; Xu *et al.* 2002). The most significant side-effect of the HBV X protein is its tumorigenicity in mouse hepatocytes (Su and Schneider 1996) and transgenic mice (Kim *et al.* 1991) which could however not be confirmed in all systems (Schaefer and Gerlich 1995).

## 2.4. The HBV genome

The genome of HBV is a circular, partially double stranded DNA molecule of approximately 3200 nucleotides, depending on the genotype (Robinson *et al.* 1974) (Figure 3). In general, the starting point is defined as the unique *EcoR* I restriction site or at a homologous site if the specific sequence is not present. While the 5' end of the plus strand is fix, the 3' end is “floating” with single stranded regions of 300 to 2000 nucleotides having been observed (Summers *et al.* 1975; Landers *et al.* 1977). Within this gap lodges the HBV DNA polymerase with the exception of the primase domain which is covalently linked to the 5' end of the minus strand (Weber *et al.* 1994). The minus strand has a terminal redundancy of 8 to 9 bases, resulting in a triple stranded region (Will *et al.* 1987). The 5' end of the plus strand consist of an mRNA-like, capped oligoribonucleotide of 18 bases (Seeger *et al.* 1986). The genome contains two directly repeated sequences of 10 or 11 bases, DR1 and



**Figure 3 : The human HBV genome**

The human HBV genome is a partially double stranded DNA molecule of approximately 3200 nucleotides which encodes 7 viral proteins. After entry into the cell the complete genome is reverse transcribed into a large RNA molecule referred to

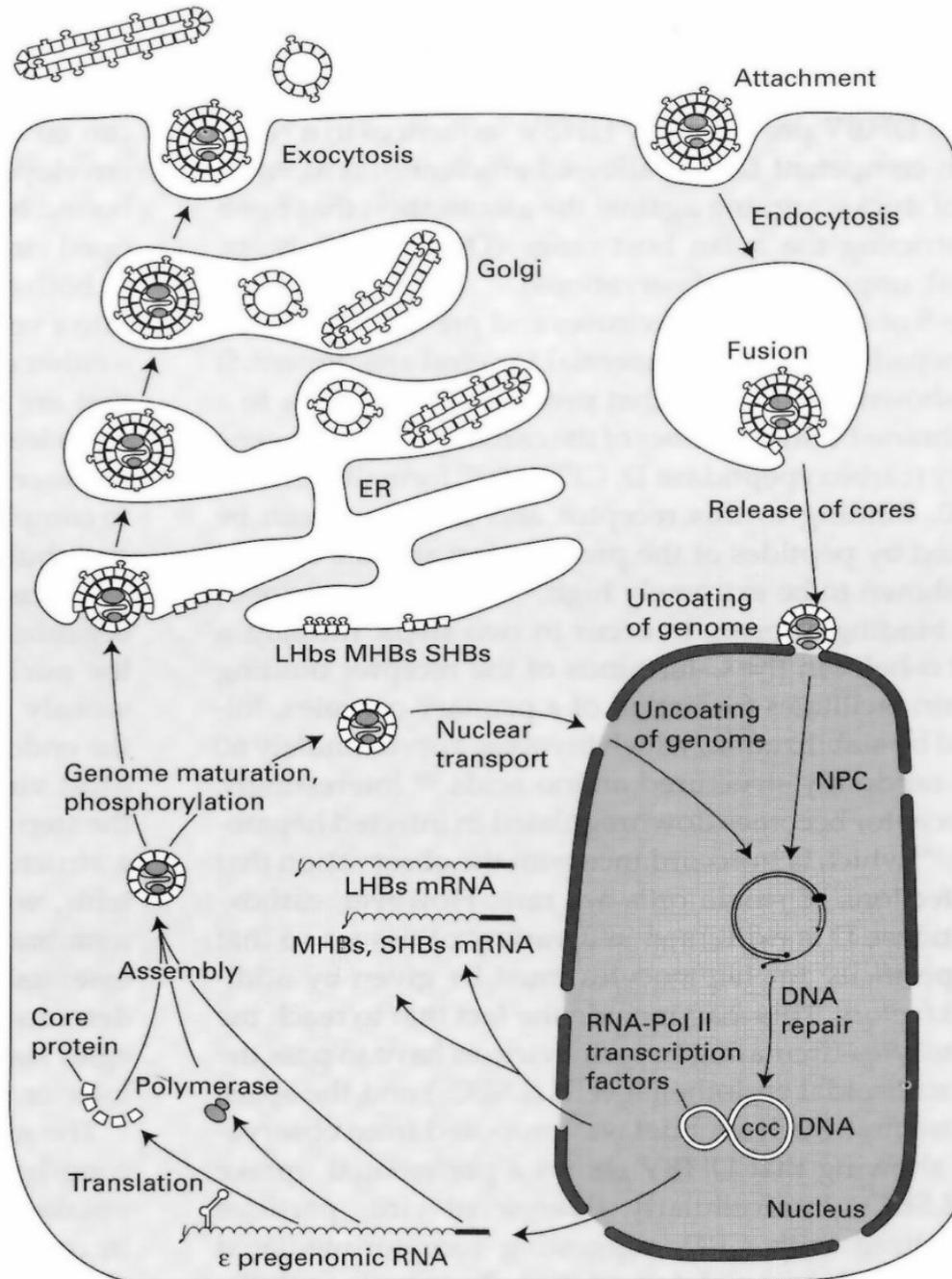
as the pregenomic RNA which is encapsidated during virus assembly, followed by reverse transcription into a DNA molecule by the viral polymerase. Nucleotide positions are based on strain AM180623.

DR2 (Dejean *et al.* 1984). The complex structure of an RNA/DNA mix with single-, double- and triple stranded regions is a uniqueness of *hepadnaviridae* and directly results from its rolling circle replication mechanism.

Using the typical search criteria, the HBV genome contains four open reading frames (ORF) referred to as the ORF P, X, S and C. These ORFs do not follow each other linearly but are embedded within the genome in a complex arrangement. In fact, the genome is not substantially longer than the longest ORF P and not one base is non-coding. ORF S is completely located within ORF P but uses a different reading frame. ORFs C and X overlap partially with ORF P. Furthermore, the usage of internal start codons allows one ORF to encode for several proteins, resulting in proteins with identical carboxy-terminal ends but amino-terminal ends of various lengths. Thus, ORF S codes for three co-terminal HBs proteins (45). ORF C encodes the HBe and HBc protein and ORF X encodes the X protein with additional products suspected (47). The regulatory elements necessary for RNA transcription and protein translation are contained within coding regions. While numerous frame shifts and overlapping ORFs allow for a minimal size of the genome, they also increase the risk of mutations causing a non functional protein and interrupting the viral life cycle.

### 2.5. Life cycle of *hepadnaviridae*

The life-cycle of *hepadnaviridae* can be divided into several steps: (1) attachment of the virus particle to the host cell, (2) penetration into the cell, (3) transport within the cell, (4) release of the viral genome, (5) transcription and translation of viral genes, (6) replication of the viral genome, (7) assembly of the viral particle, and (8) release of the virus. Unlike most other viruses, the life cycle of HBV contains an additional step consisting of the entry of newly synthesized viral DNA into the nucleus. These steps are further detailed in the following paragraphs and summarized in Figure 4.



**Figure 4 : Schematic view of the life cycle of HBV**

In many details, this model is still speculative. It is assumed that the virus is endocytosed after attachment; the nucleocapsid is released to the cytosol and transported into the nucleus. In the nucleus the HBV genome is converted into a covalently closed circular DNA and transcribed into three essential classes of mRNA. For the sake of simplicity, the HBe and HBx mRNA classes are

omitted. Translation of the core/pol transcript in the cytosol allows for assembly of core particles that contain the pregenome. The three HBs proteins are translated at the ER and inserted in that membrane. While budding, they envelop the core particles, and the HBV viral particles as well as the HBs structures are secreted by exocytosis (Figure from Thomas *et al.* 2005).

### *Attachment and entry into the host cell*

The attachment step is crucial in determining host specificity and organ tropism and represents a first attack point for protective immunity. The basis of many vaccines is the production of neutralizing antibodies against surface epitopes of viral particles. The study of attachment of HBV is still hampered by the lack of cell systems allowing infection. Several *in vitro* systems show replication and secretion of infectious particles but only after transfection of the HBV genome into the cell. An example of such a system is the human HepG2 hepatoma cell line (Sureau *et al.* 1986; Sells *et al.* 1987). This lack of uptake of HBV by cell lines can be explained by the lack of differentiation of such cell lines when compared to hepatocytes in a natural environment. In fact, complete viral life cycles including attachment and entry of viral particles could be studied only using susceptible hepatocytes from newly hatched ducklings (Tuttleman *et al.* 1986) as well as primary woodchuck hepatocytes cultures (Aldrich *et al.* 1989) and primary hepatocytes of human origin (Galle *et al.* 1994). Unfortunately such hepatocyte cultures retain their susceptibility for infection only for a few days and quickly undergo changes in differentiation. In addition, human hepatocytes are not readily available. This problem was overcome only recently by a newly established human hepatocyte cell line (Gripon *et al.* 2002) although the number of viral particles secreted remains lower than the number of viral particles necessary for infection (Gripon *et al.* 1988; Gripon *et al.* 1993; Galle *et al.* 1994; Gripon *et al.* 2002).

Several studies have linked the attachment itself to the preS1 sequence 21-47 (Neurath *et al.* 1986; Pontisso *et al.* 1989). The binding could be blocked by antibodies against this epitope while competed for by the peptide sequence itself. Furthermore, antibodies to peptide preS1(21-47) were able to neutralize *in vitro* HBV inocula, which were no longer infectious (Neurath *et al.* 1986). In the DHBV model, the preS sequence binds to a membrane-bound member of the carboxypeptidase gene family (carboxypeptidase D, CPD) also referred to as gp180 (Kuroki *et al.* 1995; Tong *et al.* 1995). Interestingly, the receptor is down-regulated in infected hepatocytes (Breiner *et al.* 2001), which would explain why co-infections of a single cell are rare. However, it is thought that since this receptor family occurs not only in hepatocytes, the cell tropism is determined by additional factors.

The mechanism of HBV entry into hepatocytes is not completely understood. Although HBV particles are enveloped, they use endocytosis to enter the cell (Kock *et al.* 1996; Breiner and Schaller 2000). Independent of acidification of the endosome, the capsid is then released into the cytoplasm by fusion of the viral and endosomal envelopes. This fusion process is similar to the one used by influenza and swapping experiments have demonstrated that the fusion peptide of *hepadnaviridae* functions when inserted into influenza hemagglutinin (Berting *et al.* 2000).

### *Transport and release of viral genome*

Following the release of the capsid into the cytoplasm, the capsids are transported towards the nucleus via the cellular microtubule network. The interaction is made possible after the core protein has undergone a change in conformation caused by phosphorylation (Kann *et al.* 1999). The new conformation also exposes a nuclear localization signal (NLS) allowing an interaction with the cellular importin  $\alpha/\beta$  pathway and an active transport of the complete viral capsid to the nuclear pores (Pante and Kann 2002). Here, a not completely understood mechanism allows for the entry of the uncoated viral genome and core proteins into the nucleus (Rabe *et al.* 2003). In the DHBV model, DNA appears in the nucleus only 24 h after infection (Tuttleman *et al.* 1986).

### *Transcription and translation of viral genes and replication of the viral genome*

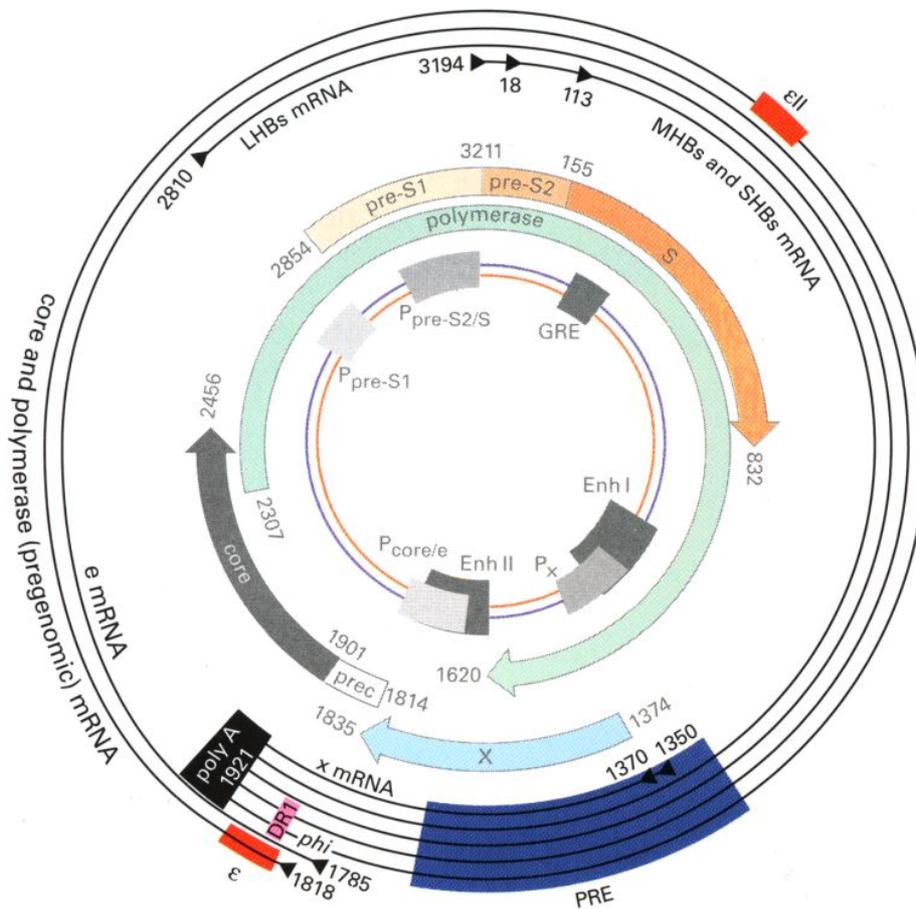
The template for transcription and synthesis of the pregenomic RNA is not the encapsidated partially double-stranded DNA genome but a circular covalently closed DNA molecule (cccDNA) obtained after repair of the viral genome. This involves removing the covalently bound viral polymerase, the conversion of the triple and single stranded regions to double stranded DNA, and the closing of the gaps at the 5' and 3' ends of both strands. After this conversion, the plasmid-like genome binds to histones to form a mini-chromosome (Bock *et al.* 1994). Based on *in vitro* studies the conversion is performed by both the viral and cellular polymerase as well as other cellular enzymes (Kock and Schlicht 1993; Pourquier *et al.* 1999; Kock *et al.* 2003). HBV DNA has been found to occasionally integrate into the host DNA but contrary to *retroviridae*, this is not an essential step. Integrated HBV DNA often contains fusions between HBV genes and host cell genes resulting in potential oncogenic proteins.

Although very compact, the genome of *hepadnaviridae* encodes multiple regulatory sequences crucial for correct transcription of viral proteins (Figure 5). These sequences bind both cellular and viral transcription factors and present themselves as two types: The HBV *promoters*, of which at least four have been identified in the HBV genome (Schaller and Fischer 1991), initiate the transcription of multiple mRNA molecules, thus presenting heterogeneous 5' ends. For instance the HBc/e promoter allows for the transcription of mRNAs encoding either the HBe protein or the HBc protein. A similar phenomenon occurs at the M/SHBs promoter and the X promoter. Only the LHBs promoter presents a single entry site. The pregenomic RNA, which is later encapsidated, starts at the preC sequence and is the second longest RNA molecule (Figure 5). *Enhancers*, of which two have been identified in the hepadnaviral genome, up- or downregulate the activity of the promoters up to 50 times. The viral X protein, as well as numerous cellular proteins, can bind to these enhancers but most are liver-specific (Dikstein *et al.* 1990; Hu and Siddiqui 1991) and contribute to the hepatotropism of *hepadnaviridae*. In addition to promoters and enhancers, three other regulatory elements were identified in hepadnaviral genomes. A glucocorticoid response element (GRE) is able to enhance transcription 2 to 5 fold after binding of the glucocorticoid receptor (Tur-Kaspa *et al.* 1986). The negative regulating element (NRE) inhibits only the transcription from the core/precore promoters (Tur-Kaspa *et al.* 1988), while the so-called CCAAT element inhibits transcription of the preS1 promoter and enhances transcription of the S mRNA in the presence of a binding protein (Lopez-Cabrera *et al.* 1991). For most viruses, the first gene products to be expressed are transcription factors helping to regulate the expression of the other viral proteins in a specific order by binding to viral promoters and enhancers. These genes are called immediate early genes. In the case of HBV, this crucial role is played by the HBx protein, which is the first protein expressed (Wu *et al.* 1991) and acts as a potent transactivator on viral but also cellular proteins. Detailed analysis of this regulation is currently hampered by the availability of cell systems.

Similar to eukaryotic messenger RNAs, the mRNAs of *hepadnaviridae* are 5' capped and 3' polyadenylated before being exported from the nucleus, but contrary to the former, they do not undergo splicing. Since the export from the nucleus is tightly coupled to the splicing mechanism, the HBV mRNAs have to use a different export route which is currently not fully understood although it is known that it involves a splicing inhibitory RNA sequence contained in the X gene (Huang and Liang 1993; Huang and Yen 1994; Huang and Yen 1995). The

common termination signal for all mRNAs is the TATAAA box at the beginning of the C gene.

Translation occurs either in the cytosol or the ER and its efficiency is determined by the ability of ribosomes to start translation as well as the speed of translation. Apart from regular translation, the translation of several HBV proteins uses mechanisms normally not found in eukaryotic cells, such as internal ribosome entry sites (IRES), leaky scanning for start codons, usage of atypical start codons, frame shifting and reinitiation.



**Figure 5 : The mRNAs of the hepatitis B virus**

The initiation sites of mRNA synthesis for the various HBV proteins are shown as triangles. All HBV mRNAs are polyadenylated and comprise the post-transcriptional regulatory element that prevents splicing. Other elements such as ε I, ε II, phi and the DR1, are only functional on the

pregenomic mRNA and are required for packaging and subsequent minus-strand DNA synthesis. Light grey boxes indicate promoter regions while dark grey boxes indicate enhancers and binding sites (Figure from Thomas *et al.* 2005).

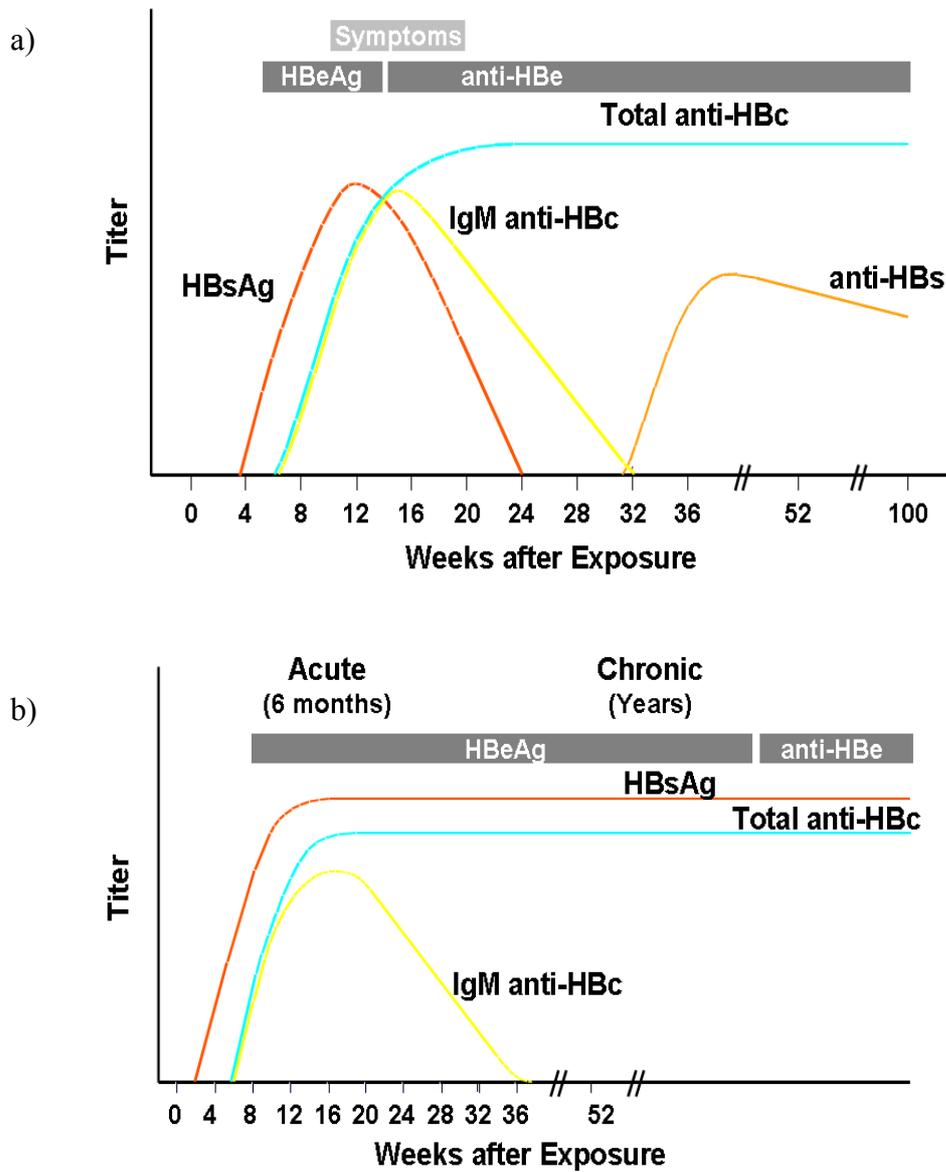
The viral genome is amplified by the transcription of the pregenomic mRNA in multiple copies from the circular DNA genome. The new molecules are synthesized by the cellular RNA polymerase II which transcribes the DNA genome to more than genome-length mRNA with redundant ends. These pregenomic (pg) RNA molecules are then exported to the cytosol.

### *Assembly and release of viral particles*

In the cytosol, the pgRNA binds to the viral polymerase and to the core protein with the help of cellular Hsp90 and other factors (Hu *et al.* 1997; Hu and Anselmo 2000; Park and Jung 2001; Hu *et al.* 2002; Beck and Nassal 2003). The binding and packaging is ensured by the  $\epsilon$  encapsidation signal and necessitates a high enough concentration of HBc protein (Beck and Nassal 1998). After packaging and formation of the capsid, the pgRNA is reverse transcribed during a complex series of events into the partially double stranded DNA genome. The reverse transcription involves several priming and translocation events as well as direct repeat sequences on the genome. It has been shown that capsids can at this point re-enter the nucleus but preferentially follow the excretion pathway (Tuttleman *et al.* 1986). The assembly of the core particle increases the affinity of the HBc protein towards membrane inserted LHBs proteins (Bruss and Ganem 1991). In the case of an excess of SHBs, the LHBs, MHBs and SHBs proteins aggregate and the virion as well as the HBs proteins move from the ER via the Golgi apparatus to the cell surface (Masuda *et al.* 1990). At this stage, the HBs proteins are glycosylated and disulfide bridges are formed between the HBc and HBs proteins. The assembled viral particles exit the cell by following the secretion pathway.

### 3. Clinical features of hepatitis B infection

It is estimated that more than 400 million people are chronically infected with HBV throughout the world. The carrier rate of the primary marker of HBV infection, the HBsAg, varies worldwide from 0.1% to 0.2% in Britain, the United States and Scandinavia to more than 3% in Greece and Southern Italy, and 10% to 15% and more in Africa and the Far East. Infections can either be acute or become chronic which depends mostly on the age of the infected individual. The chronicity rate for infections before the age of 1 year is 80% to 90%, for infections in early childhood it is 20% to 50%, whereas infection in adults may lead to chronic infection in only 1% to 2%.



**Figure 6 : Serological profiles of acute and chronic HBV infection**

The differentiation between acute (a) and chronic (b) HBV infection is based on the serology of the main viral markers: HBsAg and HBeAg as well as

the main immunological markers: total antibodies against HBc, HBe and HBs and IgM antibodies against HBc (Figure adapted from [www.cdc.com](http://www.cdc.com)).

The incubation period ranges from 1 to 5 months but may be shorter when exposed to a high viral load (Barker and Murray 1972). Typical symptoms are pains and inflammations of joints in shoulders, arms and hands (polyarthralgia and polyarthritis) as well as red skin patches (urticaria) that are caused by deposition of HBeAg complexes in blood vessels (Dienstag *et al.* 1978). The majority of infections evolve without additional symptoms although some infections are accompanied by a yellowing of eyes and mucosal membranes (jaundice). It is known that patients that do not show signs of jaundice have a higher risk of developing a chronic infection. In patients who recover, the serum liver enzyme levels (amino transferase, ALT) return to normal levels after 1 to 4 months. Persistence of high ALT levels for more than 6 months usually suggests chronic infection (Tassopoulos *et al.* 1987).

HBV infection is diagnosed, in addition to ALT levels and symptoms, by the appearance of viral proteins in the serum (Figure 6a). The HBsAg usually appears 1 to 10 weeks after HBV infection and 2 to 8 weeks before the onset of symptoms and can be present with up to  $10^{13}$  particles per ml (Frosner *et al.* 1982). The second viral protein that is secreted is the HBeAg and usually appears in parallel to HBsAg. First signs of immune reaction are the appearance of IgM and IgG antibodies against the viral core protein (IgM and IgG anti-HBc). Diagnostic assays usually measure the IgG antibodies together with the IgM antibodies and the results are referred to as total anti-HBc. During acute infection, the IgM anti-HBc appears shortly after HBsAg and is detectable for 6 months, while IgG antibodies appear later and persist for many years. Resolution of acute infection is marked by the decrease of circulating HBsAg and the appearance of anti-HBs and anti-HBe.

Chronic infections are identified by the persistence of HBsAg and HBeAg as well as a lack of anti-HBs and anti-HBe appearance after 6 months of infection (Figure 6b). The presence of HBeAg signifies the active replicative phase of chronic HBV infection which tends to persist for months to years and thus is an important marker for chronic infection assessment. Most patients will ultimately undergo seroconversion with loss of HBeAg and appearance of anti-HBe, coincident with a decline in viral replication, a return of ALT levels to normal and a decrease in serum HBV DNA below detection threshold of most assays ( $<10^3$  copies per ml). These patients are referred to as *inactive carriers*. In some patients the absence of HBeAg is caused by mutations on the viral genome resulting in the translation of a non-functional HBeAg protein. In general these patients have a detectable serum HBV DNA titre, usually above  $10^5$  copies per ml, and fluctuating ALT levels (Lok *et al.* 2001). The most significant of

these mutations is the appearance of a stop codon in the precore sequence at position 1896 (Okamoto *et al.* 1994) which is found in up to 27% of persons with chronic HBV infection in the USA, 50% in Asia and 92% in the Mediterranean (Funk *et al.* 2002; Chu *et al.* 2003).

Chronically infected patients with replicating HBV, display various degrees of liver damage, from benign forms of chronic lobular hepatitis to more severe forms of active cirrhosis and HCC. The pathogenesis of hepadnaviruses is not due to a high cytotoxicity caused by replication and release of viral particles but due to the destruction of HBV infected liver cells by the host immune system, which also marks the entry into an inflammatory phase. In most cases, with the appearance of anti-HBe and thus the decline of viral replication, the destruction of liver cells subsides, and the patients become inactive carriers. However, HBV replication can reactivate and lead to a deterioration of the underlying disease from chronic active hepatitis to active cirrhosis. The 5-year survival rate of patients with cirrhosis is 66% to 78%. Reactivation of a latent HBV infection is a frequent event in immunocompromised patients infected with HIV (Vento *et al.* 1989), homosexuals (Davis *et al.* 1984), and patients treated with immunosuppressive drugs (Lok *et al.* 1991). In immunocompetent patients, the probability of clearing HBeAg within 5 to 10 years of diagnosis is about 50% (Bortolotti *et al.* 1990; Yuen *et al.* 2000; McMahon *et al.* 2001).

Chronic hepatitis B accounts for more than 75% of all HCC cases and is the predominant cause of cancer mortality in Africa and China. It has been suggested that a nutrition containing high amounts of aflatoxin, secreted by a certain species of *Aspergillus*, is a risk factor for HBV-related HCC in these geographical areas (Sun *et al.* 1999). There is also evidence that HCV co-infection and high alcohol consumption contribute to the risk. On a molecular level, the increase of risk to develop HCC in chronic HBV infected patients is linked to the frequent liver damage and repair which can lead to the accumulation of oncogeneous mutations. However, the possible integration in the cellular genome of parts of the viral DNA is an even greater risk and can cause dysregulation of important cell-cycle and cell-death related genes.

#### 4. Epidemiology

More than 2 billion ( $2 \times 10^9$ ) people have been infected worldwide with HBV. According to WHO estimates, there were over 5.2 million cases of acute HBV infection in 2000, while 400 million people are currently estimated to suffer from chronic infection. Acute hepatitis B usually runs a self-limited course in adult subjects, with most patients recovering completely. Fulminant hepatitis occurs in 1% to 2% of acute infections, resulting in the death of most of these patients. About 15% to 40% of chronically infected subjects will develop complications, leading to an estimated 520 000 to 1 200 000 deaths each year due to acute and chronic hepatitis, cirrhosis and liver cancer. Chronic HBV infection is the major cause of hepatocellular carcinoma (HCC) worldwide (WHO 2000), although its prevalence is known to vary widely among the world population, and those areas with higher prevalence of viral infection (Figure 7) present the highest HCC rates (Brechot *et al.* 1998; Blakely *et al.* 1999; Yu *et al.* 2000). HCC is one of the major three causes of death in Africa, Asia and the Pacific Rim (Lemon *et al.* 2000). HCC itself constitutes the sixth most frequent cancer, representing around 5% of all cancers worldwide (Lee 1997; Parkin *et al.* 2001; Lok 2002).

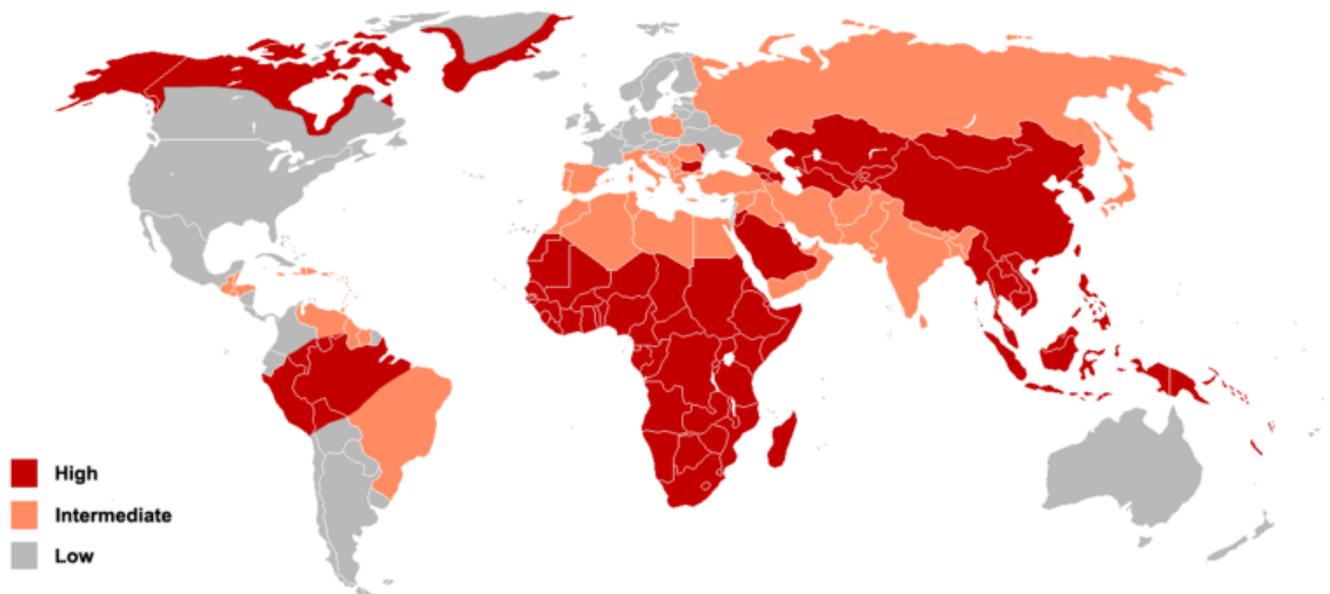


Figure 7 : Global prevalence of HBsAg in 2005 (Figure from [www.cdc.com](http://www.cdc.com))

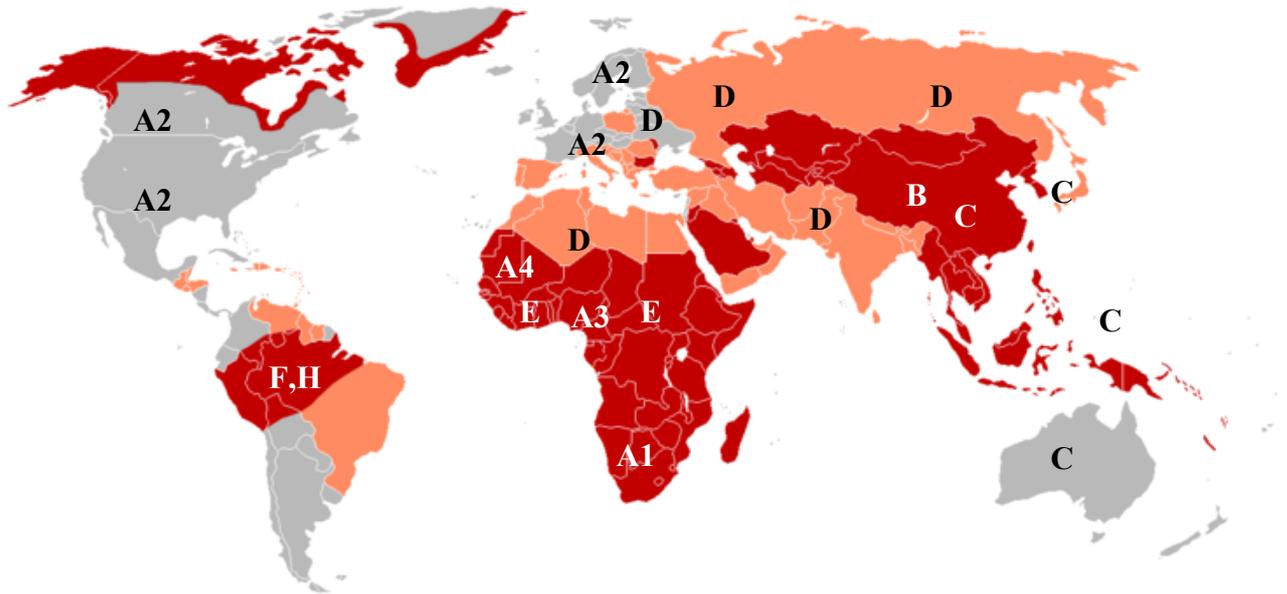
### 4.1. Global distribution of HBV infection

The incidence of infection and patterns of transmission vary greatly in different population subgroups throughout the world. It is influenced primarily by the age at which infection occurs. Endemicity of infections is considered high in those parts of the world where at least 8% of the population is HBsAg positive. In these areas, 70% to 90% of the population generally have evidence of previous HBV infection (Figure 7).

In high endemic areas such as Asia, sub-Saharan Africa and the Pacific, carrier rates for HBsAg range from 8% to 25% and anti-HBs prevalences from 60% to 85%. Thus, exposure to HBV in high endemic areas, measured serologically, may approach 100% (Moyer and Mast 1994). In areas of the world with an intermediate pattern of HBV infection (eastern and central Europe, the Middle East, the Indian subcontinent and the Amazonian basin), the prevalence of HBsAg positivity ranges from 1% to 8%, serological evidence of past infection is found in 10% to 60% of the population and the lifetime risk of becoming infected with HBV is estimated to be 20% to 60% (FitzSimons and Van Damme 1999; Lok *et al.* 2001). In most developed parts of the world, the prevalence of chronic HBV infection is less than 1% and the overall infection rate is 5% to 7% (Lok *et al.* 2001). Overall, approximately 45% of the global population live in areas of high chronic HBV prevalence (Mahoney 1999).

### 4.2. HBV serotypes and genotypes

HBV shows substantial genetic heterogeneity. Four serotypes of the HBsAg referred to as adw, ayw, adr and ayr have been defined based on two mutually exclusive determinant pairs, d/y and w/r and a common determinant *a*. Sequencing of viral genomes has now become the major tool used for descriptive virology, and sequence data are now used to reconstruct the phylogenetic history of viruses and to delimit genetic subtypes. Genetic analysis has allowed for the classification of HBV into 8 distinct genotypes (A to H) that have different geographical distributions (Figure 8) and associations with different risk groups for infection (Okamoto *et al.* 1988; Norder *et al.* 1992). Genotypes were first defined by an 8% nucleotide difference on the complete genome or a 4.1% nucleotide difference on the S gene (Okamoto *et al.* 1988). Recently the necessary nucleotide difference on the complete genome level was revised downwards to 7.5% (Kramvis and Kew 2007; Kramvis and Kew 2007). Subtypes are defined by a complete genome nucleotide difference of 4%.



**Figure 8 : Worldwide distribution of HBV genotypes (Figure modified from [www.cdc.com](http://www.cdc.com))**

Genotype A is predominantly found in north-western Europe, North America and east and south Africa (Norder *et al.* 1993; Bowyer *et al.* 1997), genotypes B and C are found predominantly in eastern and south-eastern Asia including China and Japan but also in Australia and the Pacific (Okamoto *et al.* 1988; Theamboonlers *et al.* 1999), genotype D circulates worldwide with highest prevalences in the Mediterranean area and Russia (Norder *et al.* 1993), genotype E is predominant in western Africa (Odemuyiwa *et al.* 2001; Mulders *et al.* 2004; Olinger *et al.* 2006) and the most divergent genotypes F and H are found exclusively among indigenous populations in Central and South America (Arauz-Ruiz *et al.* 1997; Arauz-Ruiz *et al.* 2002). Genotype G has been identified in a few samples in the USA and France (Stuyver *et al.* 2000; Vieth *et al.* 2002) (Figure 8). More recently (Mulders *et al.* 2004; Norder *et al.* 2004), HBV genotypes have been divided into subtypes with the clearest subtype distribution apparent for genotype A. Subtype A1 is present in South Africa while subtype A2 circulates in Europe and North Africa. At least 5 subtypes have also been identified for genotypes B, C and D and more studies on their geographic distributions are warranted.

Worldwide, genotypes B, C and D are dominant, with an estimated 240 million individuals infected with genotypes B or C and 40 million infected with genotype D. Genotype A has infected approximately 3 million, and genotype E 20 million individuals. The clinical significance of genotypes has not yet been fully evaluated but some recent data suggests that associations between certain genotypes and HBeAg seroconversions, viral mutations, severity of liver disease and response to treatment may exist.

Interestingly, most current genotypes of HBV seem to be the result of one or several recombination events (Bollyky *et al.* 1996; Simmonds and Midgley 2005; Szmargd and Balloux 2007). In particular, this is evident for the B/C recombinant which has spread in mainland Asia (Sugauchi *et al.* 2004) and has been defined as genotype B1 but also for genotypes B and C themselves which show similarities to genotype A. Genotype E is thought to be a recombinant between genotype D and another unknown or extinct genotype (Bowyer and Sim 2000) and genotype G is highly divergent from all currently known genotypes in most regions of the genome but shows similarities to genotype E in the end of the S gene. The importance of recombination events in the evolution of HBV, as well as their molecular mechanisms remains controversial.

### 4.3. Transmission of HBV

Although HBsAg has been detected in a wide variety of body fluids, only serum, semen and saliva have been demonstrated to be infectious (Alter *et al.* 1977; Scott *et al.* 1980). Breast milk, tears, faces and urine remain controversial, although HBsAg or HBV particles have been detected in these fluids (Shimoda *et al.* 1981; Blum *et al.* 1983). Person to person spread include blood transfusions, contaminated equipment used for therapeutic or drug injections and needle sticks or injuries with other contaminated sharp instruments Sexual transmission remains an important route. The risk of perinatal transmission is greatest for infants born to HBeAg positive mothers and ranges from 70% to 90% at 6 months of age. About 90% of these children remain chronically infected (Stevens *et al.* 1979). The risk of perinatal infection among infants born to HBeAg negative mothers ranges from 10% to 40%, with 40% to 70% of these infected infants remaining chronically infected (Stevens *et al.* 1979; Xu *et al.* 1985). The transmission by blood-feeding arthropods has been investigated but never proven.

### 4.4. Prevention of HBV infection

Hepatitis B is preventable by immunization. First-generation vaccines were prepared from HBsAg particles purified and inactivated from plasma donations from chronic asymptomatic HBV carriers. These preparations were safe and immunogenic but have been superseded by recombinant DNA vaccines produced by expression of HBsAg in yeast cells. These vaccines induce protection against all known genotypes and subtypes of HBV. Around 5% to 10% of vaccine recipients present no anti-HBs antibody production or their levels remain under the recommended 10 IU/L. These non-responders remain susceptible to infection with HBV. It is thought that the risk of non-responsiveness is linked to the ethnic group, but also to the site and route of injection as well as gender, age, body mass index, immunosuppression and immunodeficiency. Currently promising third-generation vaccines are developed that aim to reduce the frequency of such cases. There is also evidence that amino acid substitutions within the a-determinant of the surface antigen can allow replication of HBV in vaccinated persons, as antibodies produced do not recognize the changes induced by such mutations. Reports indicated that the frequency of such mutations is increasing (Hsu *et al.* 1999; Nainan *et al.* 2002; Ni *et al.* 2007)

Before the availability of HBV vaccines and still today, passive immunization with hepatitis B immunoglobulin (HBIG) is used for prophylaxis under certain conditions. HBIG is prepared from pools of plasma with high titres of anti-HBs and may confer temporary passive immunity. The major indication for the administration of HBIG is a single acute exposure to HBV. Results following the use of HBIG for prophylaxis in babies born to HBV infected mothers indicate encouraging prevention of infection in the newborn if the immunoglobulin is administered immediately. The risk of the baby developing chronic infection is reduced by up to 70%. Combined passive and active immunisation, have an efficiency approaching 90%.

### 4.5. Treatment of chronic HBV infection

Three agents are currently approved for the treatment of chronic HBV: interferon-alpha (IFN- $\alpha$ ), lamivudine and adefovir. Each agent has inherent limitations for use in the treatment of chronic HBV (Lok *et al.* 2001; Conjeevaram and Lok 2003; Fung and Lok 2004). IFN- $\alpha$  is effective in a minority of patients and has frequent side effects that limit its tolerability (Hoofnagle and di Bisceglie 1997). The efficacy of lamivudine is limited by the emergence of

drug-resistant HBV mutants (Yuen *et al.* 2007), restricting its usability as a long-term therapy while adefovir is well tolerated and is associated with low incidence of resistance, but its antiviral effect is sub-optimal. The first objective of a treatment is to decrease HBV replication in order to decrease inflammation in the liver and thus prevent the progression of fibrosis. Interrupting the fibrosis process prevents progression to cirrhosis and its complications, including HCC. If the antiviral effect is sufficient (less than 100 000 copies of HBV DNA per ml) and is maintained, the chances of an effective immune response are high and seroconversion to HBeAg may occur.

IFN- $\alpha$  has a direct antiviral effect by inhibiting the synthesis of viral DNA and activating antiviral enzymes, in addition, it also increases the cellular immune response against infected hepatocytes. Lamivudine is a nucleoside analogue which directly inhibits the HBV DNA polymerase but long-term administration may lead to drug-resistance induced by mutations on the viral polymerase. Adefovir works similar to lamivudine and also blocks the DNA polymerase activity.

The activity of these antivirals especially in combination with each other, although already in use, still needs further long-term studies. New drugs, such as entecavir, PEG-IFN and others, are under consideration.

#### 4.6. Co-infection with other viruses

Evidence has been generated that HCV co-infection and / or co-infection with HIV are clinically and virologically important factors. HBV-positive patients co-infected with HIV tend to have milder liver disease, but the overall morbidity and mortality seem to be higher than in patients infected with HIV alone. Patients co-infected with HCV tend to have more severe liver disease and more commonly occurring HCC in comparison to patients having either infection alone (Colin *et al.* 1999; Di Martino *et al.* 2002; Puoti *et al.* 2002). The infection with hepatitis delta virus (HDV) requires prior HBV infection since HDV uses HBsAg to form its envelope and a HBV/HDV co-infection may result in a more severe acute disease and a higher risk (2%-20%) of developing acute liver failure.

# Chapter 2: Material

## 1. Chemicals

Agarose	Invitrogen
Bromophenol blue	Invitrogen
Ethanol 100%	Merck
Ethidium bromide	Invitrogen
Ethylenediaminetetraacetic acid (EDTA)	Biorad
Glycerol	Sigma
LB agar	Invitrogen
Luria broth base (LB)	Invitrogen
MgCl <sub>2</sub> 50 mM	Invitrogen
Nucleotides (dNTPs)	Invitrogen
Oligonucleotides/primers	Eurogentec
PCR buffer without MgCl <sub>2</sub> 10x	Invitrogen
Picogreen <sup>®</sup> 10 000x	Molecular Probes
Sodium acetate	Merck
Sulphuric acid (concentrated)	Sigma
SYBR <sup>®</sup> Green I nucleic acid stain 10 000x	Molecular Probes
Tris	Sigma

## 2. Buffers and Solutions

DNA loading buffer	Bromophenol blue	0.25% (w/v)
	Glycerol	50% (v/v)
	EDTA	50 mM
Sodium acetate 3M	Sodium acetate	26.409 g
	ddH <sub>2</sub> O	100 ml
	Adjust pH 5.2	
	Autoclave	
TAE-buffer (50x)	Tris	2 M
	Sodium acetate	25 mM
	EDTA	0.5 M
	Adjust pH 7.8	
TE-buffer	Tris	10 mM
	EDTA	1 mM
	Adjust pH 7.6	

### 3. Enzymes

Phusion™ High Fidelity DNA polymerase	Finnzyme
Proteinase K	Qiagen
Platinum® Taq DNA polymerase	Invitrogen

### 4. DNA markers

1 kb plus DNA ladder™	Life Technologies
-----------------------	-------------------

To determine the length of DNA fragments during agarose gel electrophoresis the 1 kb plus DNA ladder™, containing DNA fragments of the following sizes was used:

12000, 11000, 10000, 9000, 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1650, 1000, 850, 650, 500, 400, 300, 200, 100 bp.

### 5. Bacterial strain

E. coli One Shot® TOP10: This electrocompetent bacterial strain was provided with the TOPO TA Cloning® kit (Invitrogen) and was used for transfection with PCR product containing vectors by electroporation.

### 6. Bacterial media

Bacteria were grown in liquid LB medium containing 25 mg/l of Luria Broth base (Invitrogen) that was autoclaved at 121°C for 15 minutes. Growth plates for spreading bacteria were prepared with an autoclaved 32 mg/l LB agar (Invitrogen) medium. Transformed bacteria were selected by adding ampicillin (100 µg/ml) or kanamycin (30 µg/ml) to the liquid or solid medium.

## 7. Kits

- **HBsAg detection**

Murex HBsAg Version 3<sup>®</sup> (ELISA test)                      ABBOTT Diagnostics

- **DNA extraction**

QIAamp<sup>®</sup> DNA Blood Mini kit                                      Qiagen

- **RNA extraction**

QIAamp<sup>®</sup> Viral RNA Mini kit                                      Qiagen

- **Cloning**

TOPO TA Cloning<sup>®</sup> kit    Invitrogen

Zero Blunt<sup>®</sup> TOPO<sup>®</sup> PCR cloning kit                              Invitrogen

- **Sequencing**

Jet Quick PCR Purification Spin<sup>®</sup> kit                              Genomed

Big Dye Terminator v3.1 Cycle Sequencing<sup>®</sup> kit              Applied Biosystems

## 8. Vectors

### 8.1. pCR<sup>®</sup>4-TOPO<sup>®</sup>

The pCR<sup>®</sup>4-TOPO<sup>®</sup> vector, included as a linear molecule in the TOPO TA Cloning<sup>®</sup> kit (Invitrogen), has 3' thymidine overhangs at the insertion site as well as several restriction and primer binding sites up- and downstream of the insertion site. The M13 sequences can be used to amplify the insert while the T3 and T7 sequences can be used for transcription by a T7 or T3 RNA polymerase. In addition the vector encodes resistance genes for ampicillin and kanamycin as well as a P<sub>lac</sub>/LacZ mechanism for insertion verification (Figure 9).

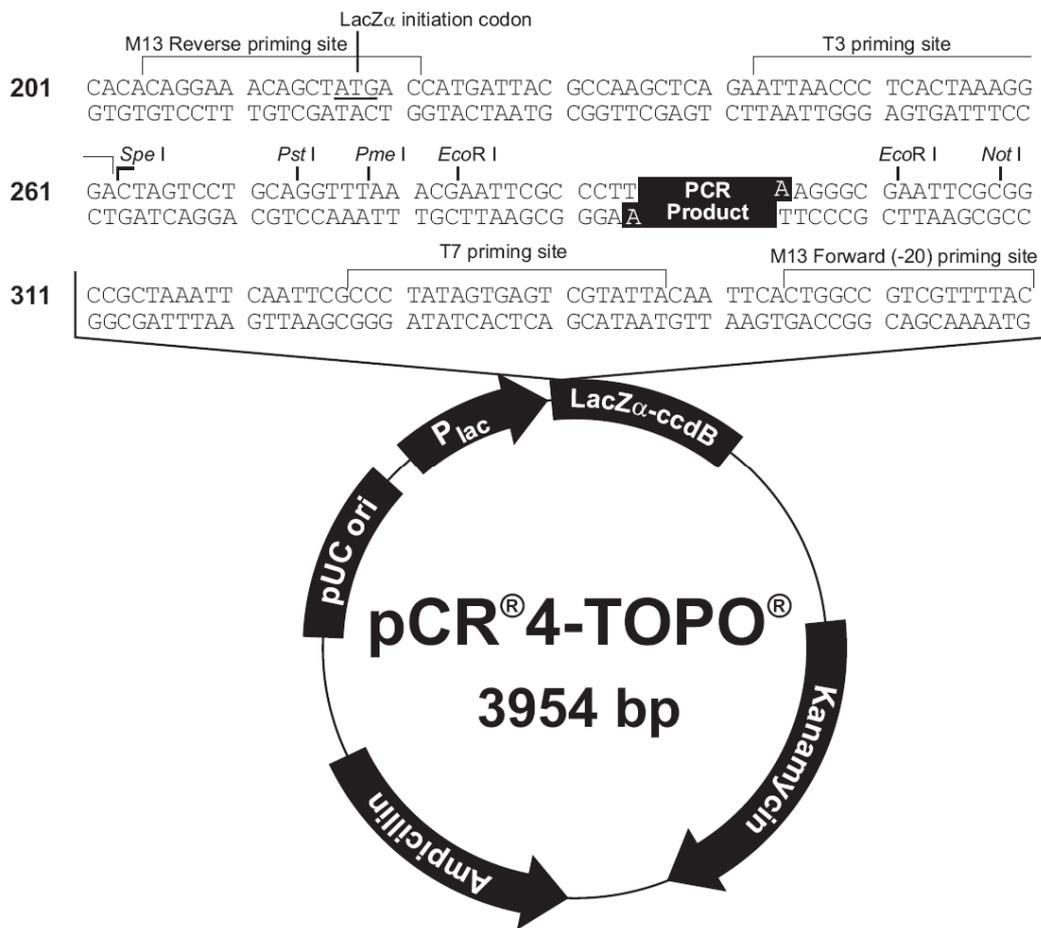


Figure 9 : Map of the pCR<sup>®</sup>4-TOPO<sup>®</sup> cloning vector (Figure from Invitrogen)

8.2. pCR<sup>®</sup>-Blunt II-TOPO<sup>®</sup>

The pCR<sup>®</sup>-Blunt II-TOPO<sup>®</sup> vector, included as a linear molecule in the Zero Blunt<sup>®</sup> TOPO<sup>®</sup> PCR cloning (Invitrogen), does not have 3' thymidine overhangs at the insertion site. Several restriction and primer binding sites are located up- and downstream of the insertion site. The M13 sequences can be used to amplify the insert while the T3 and T7 sequences can be used for transcription by a T7 or T3 RNA polymerase. In addition the vector encodes resistance genes for zeocin and kanamycin as well as a P<sub>lac</sub>/LacZ mechanism for insertion verification (Figure 10).

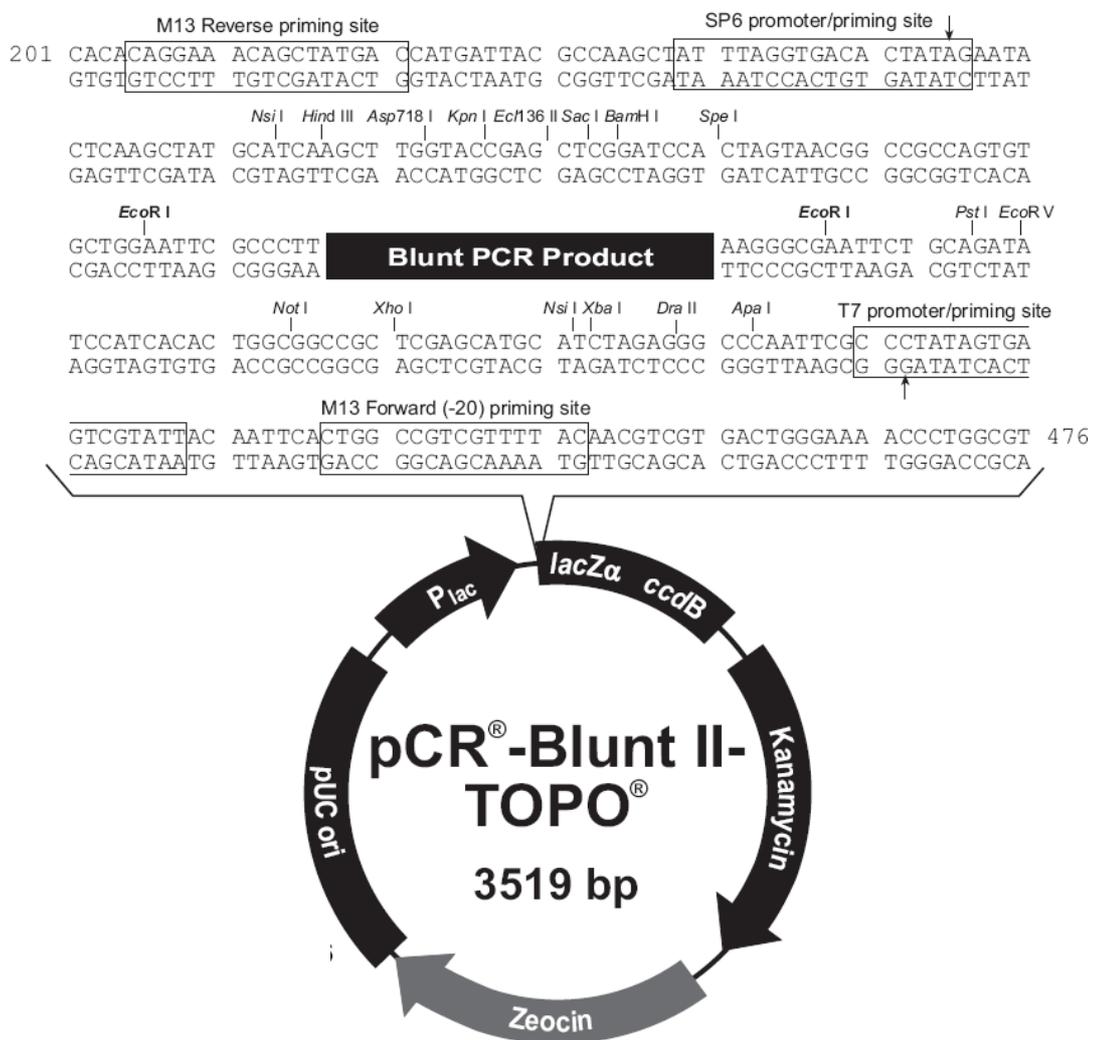


Figure 10 : Map of the pCR<sup>®</sup>-Blunt II-TOPO<sup>®</sup> vector (Figure from Invitrogen)

## 9. PCR primers

### 9.1. Complete genome PCR

**Table 2 : Primers and conditions for the amplification and sequencing of the complete genome of HBV using 4 semi-nested PCR reactions**

Fragment	Primer* (sense)	5'-3' sequence	Position <sup>†</sup>	MgCl <sub>2</sub> mM	Annealing °C
<b>preS</b>	fw2422c (fw)	AGAACTCCCTCGCCTCGCAGAC	2375-2396	1.65	63
	preS-R (rv)	ACAGGCGGKGTTTTTCTTGTTGA	199-221		
	fw2451 (fw)	TCAATCGCCGCGTCGCAGAA	2404-2423	1.5	65
<b>S</b>	P2f (fw)	CCTGCTGGTGGCTCCAGTTC	56-75	1.65	63
	979 (rv)	ATTGGAAAGTATGTCAAAGAATTGTGGGTCTTTTG	977-1011		
	Mc2r (rv)	TGGAAGTTGGGGATCATTGCC	891-911	1.5	60
<b>X</b>	455(fw)	CAAGGTATGTTGCCCGTTTG	455-473	1.5	62
	1800 (rv)	AGACCAATTTATGCCTACAGCCTCCTA	1774-1800		
	fw696 (fw)	TCAGTGGTTCGTAGGGCTTTCC	694-715	1.5	63
<b>C</b>	fw1608X (fw)	GCATGGAGACCACCGTGAACG	1606-1626	2	60
	rv2661 (rv)	TCATTTACAGTGAGAGGGCCACAAATTG	2586-2614		
	fw1644Xmm (fw)	TGCCCAAGGTCTTACATAATAGGACTCTTG	1639-1668	1.8	60

\* For each fragment the first two lines describe primers and conditions for the first round PCR; the third line describes the semi-nested primer and conditions for the second round (semi-nested) PCR.

<sup>†</sup> According to the HBV/E reference strain accession number X75657.

### 9.2. M13 cloning verification PCR

**Table 3 : Primers used in the M13 cloning verification PCR**

Fragment	Primer	5'-3' sequence
<b>M13 PCR</b>	Forward	GTAAAACGACGGCCAG
	Reverse	CAGGAAACAGCTATGAC

## 10. Bioinformatics

The specific usage of these applications, in particular for phylogenetic analysis, is explained in Chapter 3, Section 9.

**Sequence manipulation:** Bioedit v7.0.5.3, <http://www.mbio.ncsu.edu/BioEdit/>

**Distance calculations and neighbour-joining tree construction:** MEGA3.1, (Kumar *et al.* 2004)

**Maximum likelihood tree calculation:** PAUP\* v4.0, (Swofford 2003)

**Electropherogram analysis:** SeqScape<sup>®</sup> v2.5, Applied Biosystems

**Sequence acquisition:** Data Collection Software v3.0, Applied Biosystems

**Similarity plot and bootscan analysis:** SimPlot v3.5.1, (Lole *et al.* 1999)

**Sequence alignment:** Clustal W v1.4, (Higgins *et al.* 1996)

**Substitution model determination:** Modeltest, (Posada and Crandall 1998)

**Real time PCR fluorescence acquisition and analysis:** Opticon Monitor<sup>™</sup> v3.1, Biorad

**ELISA software:** Softmax PRO v5, Molecular Devices

## 11. Machinery used

**ELISA reader** SpectraMax<sup>®</sup> Plus 384, Molecular Devices

**ELISA washer** 96PW, SerColab

**Incubator** HERAcell<sup>®</sup> 150, Heraeus

<b>Centrifuges</b>	Pico 17, Heraeus; Megafuge <sup>®</sup> 1.0R, Heraeus, UNIVAP 150H, UniEquip
<b>Heating block</b>	Thermomixer Comfort, Eppendorf
<b>Fluorescence reader</b>	GENios Plus, Tecan
<b>Electroporation apparatus</b>	Pulse Controller Plus, Capacity Extender Plus, Gene Pulser II Plus, Biorad
<b>Vortex</b>	Vortex-Genie <sup>®</sup> 2, Scientific Industries
<b>Shaker</b>	Multitron 2, INFORS-HT
<b>Sequencer</b>	ABI PRISM <sup>®</sup> 3130x/ Genetic Analyzer, Applied Biosystems
<b>Gel tank and casting form</b>	Biozyme
<b>Electrophoresis power supply</b>	E835, Consort
<b>Gel documentation system</b>	InGenius, Syngene
<b>PCR machine</b>	Mastercycler <sup>®</sup> Gradient, Eppendorf
<b>Real time PCR machines</b>	Opticon <sup>®</sup> 2 DNA Engine, Biorad; Chromo4 <sup>™</sup> /PTC200, Biorad; LightCycler <sup>®</sup> <sup>1</sup> , Roche Diagnostics

---

<sup>1</sup> Other company and product denominations mentioned in this document, such as: Invitrogen, Merck, Biorad, Sigma, Eurogentec, Molecular Probes, Finnzyme, Qiagen, Life Technologies, Abbott Diagnostics, Genomed, Applied Biosystems, Molecular Devices, SerColab, Heraeus, UniEquip, Eppendorf, Tecan, Scientific Industries, INFORS-HT, Biozyme, Consort, Syngene, Roche Diagnostics, Biomérieux, may be trademarks or registered trademarks of their respective trademark owners.

## Chapter 3: Methods

The following are the general methods applied during the creation and collection of data and results applicable to all parts in Chapter 4. Specific methods, such as the description of clinical material or certain PCR reactions or serological assays are described in the respective Part of Chapter 4.

### 1. HBsAg detection

The detection of HBsAg in the serum of voluntary blood donors or liver patients was performed using the Murex HBsAg kit version 3<sup>®</sup> (ABBOTT Diagnostics) and following the manufacturer's protocol. The assay is based on the enzyme-linked immunosorbant assay principle (ELISA). Briefly, 25 µl of sample diluent were added to the wells of a 96 well microtitre plate followed by 75 µl of serum sample. On each plate, instead of serum sample, two wells were filled with a negative control and one well with a positive control (both included in the kit). The plate was then covered by a lid and incubated at 37°C for 1 h. After incubation, 50 µl of conjugate were mixed to each well and the plate was incubated again at 37°C for 30 minutes. At the end of the incubation time, the plate was washed 5 times with wash fluid using an ELISA washer and 100 µl of substrate solution were added to each well. After incubation for 30 minutes at 37°C, 50 µl of stop solution (0.5 M sulphuric acid) were added, again to each well. Development of a purple colour, signifying a positive result, was detected by an ELISA reader at a wavelength of 450 nm and using 690 nm as a reference wavelength.

### 2. DNA extraction

DNA was extracted from the serum of voluntary blood donors or liver patients using the QIAamp<sup>®</sup> DNA Blood Mini kit (Qiagen) and following the manufacturer's protocol. Briefly, 200 µl of serum sample were added to 20 µl of proteinase K in a 1.5 ml microcentrifuge tube. To this, 200 µl of lysis buffer were added (buffer AL, included in the kit), and after vortexing, the resulting solution was incubated for 10 minutes at 56 °C. After a brief centrifugation to remove drops from the lid of the tube, 200 µl of 100% ethanol were added and the solution

was again vortexed and centrifuged. Then, the mixture was applied to a QIAamp<sup>®</sup> spin column (included) in a 2 ml centrifugation tube. The column was centrifuged at 6000 g for 1 minute and the flowthrough was discarded. After two washing steps (500 µl of Buffers AW1 and AW2, included), the DNA was eluted by applying 100 µl of distilled deionised water on the column membrane and by centrifugation at 6000 g for 1 minute and collecting the eluate in a 1.5 ml microcentrifuge tube. Eluted DNA was stored at -20 °C.

### 3. RNA extraction

RNA was extracted from the serum of voluntary blood donors or liver patients using the QIAamp<sup>®</sup> Viral RNA Mini kit (Qiagen) and following the manufacturer's protocol. Briefly: 560 µl of lysis buffer containing carrier RNA (Buffer AVL, included in the kit) were added to 140 µl of serum, mixed by vortexing and incubated at room temperature for 10 minutes. To this, 560 µl of ethanol 100% were added and, after mixing, the solution was centrifuged briefly to remove drops from the inside of the lid. The mixture was applied to a QIAamp<sup>®</sup> Mini spin column in a 2 ml centrifugation tube which was then centrifuged for 1 minute at 6000 g. The flowthrough was discarded and the column was washed twice with Buffer AW1 and AW2 (included). RNA was then eluted from the column by adding 60 µl of distilled deionised water by centrifuging at 6000 g for 1 minute and collecting the eluate in a 1.5 ml microcentrifuge tube. Eluted RNA was stored at -20 °C.

### 4. DNA quantification

The amount of total DNA in a solution was determined by using a molecule called Picogreen<sup>®</sup> that is able to bind to DNA and emits fluorescence only in a bound state. First, the sample to be analysed was diluted 20 times by adding 5 µl of sample to 95 µl of TE buffer and this volume was added to 100 µl of a Picogreen<sup>®</sup> solution (200 times dilution of concentrated Picogreen<sup>®</sup> in TE buffer). A dilution series of a DNA solution of known quantity, mixed with Picogreen<sup>®</sup> was used as a quantification standard. After mixing and incubation for 5 minutes at room temperature in the dark, the fluorescence was measured at a wavelength of 480 nm using the GENios<sup>®</sup> fluorescence reader (Tecan) with a 520 nm wavelength filter. The fluorescence values of the standards were used to plot a standard curve and a trend line ( $y=ax+b$ ) was inferred. Based on the equation of this trend line, the quantity of DNA in the sample was determined.

## 5. Polymerase chain reaction

The polymerase chain reaction (PCR) is used to amplify double stranded PCR products, to quantify the presence of a specific DNA fragment, or to obtain sufficient DNA product for downstream applications. The specificity of amplification is determined by a set of oligonucleotides (referred to as primers), which specifically bind to a given sequence. A DNA dependant DNA polymerase will then amplify a sequence by starting at the forward primer and finishing at the reverse primer during multiple PCR cycles. Each cycle consist of three steps with each step requiring a specific temperature. The first step, the denaturation, is usually run at a temperature between 92 °C to 95 °C and separates the double stranded DNA into single stranded molecules. During the following step, the annealing, usually at a temperature of 50 °C to 65 °C, the primers will bind specifically to these single stranded molecules. During the final step, the elongation, the DNA polymerase will copy the single stranded sequence between the primers to a double stranded molecule, usually at a temperature of 72 °C. In certain applications, the annealing and the elongation steps are merged. The result of PCR reactions is verified using agarose gel electrophoresis (*cf* Section 6).

A semi-nested PCR reaction consists of using the PCR product of a first PCR as template in a second PCR but with one primer shifted (downstream for the forward primer or upstream for the reverse primer) with the second primer unchanged. This approach increases the sensitivity of the amplification and, by shifting one primer, reduces the risk of amplifying unspecific products that could have appeared in the first PCR. The two PCR reactions are referred to as first round and second round PCRs. Nested PCRs, which consist of changing both primers from first to second round PCR, have a higher specificity but need more primers to be designed. For virus detection applications this often poses a problem in primer design since different strains of the same virus can have high sequence variability.

In addition to primers, template and the DNA polymerase, PCR mixtures also contain dNTPs, the building blocks of DNA, Magnesium Chloride (MgCl<sub>2</sub>), necessary for enzyme activity and primer binding, as well as a buffer, ensuring the correct ionic strength and pH for the PCR reaction to work.

In recent years, the usage of a double stranded DNA binding, fluorescent molecule (Sybergreen) has added the possibility to follow the DNA amplification during each cycle of a PCR reaction. These real time PCRs allow for an easier quantification of template DNA and have made consequent verification steps unnecessary. Another type of real time PCR, referred to as 5' nuclease assay (or TaqMan<sup>®</sup> PCR) adds additional specificity by using a third oligonucleotide (the probe), which is coupled to a fluorescent molecule and binds between the forward and reverse primers. During the elongation step, the DNA polymerase separates the initially non-fluorescent molecule from the probe, which then becomes fluorescent. Thus, while Sybergreen fluoresces after binding to any double stranded DNA molecule, the fluorescence detected in a TaqMan<sup>®</sup> assay is linked to the amplification of a specific DNA sequence, thus effectively increasing the specificity of the PCR reaction. TaqMan<sup>®</sup> assays are generally used for diagnostic and quantification purposes.

#### 5.1. Amplification of the HBV complete genome

In order to amplify the complete 3200 nucleotides circular genome of HBV, four semi-nested PCR reactions were developed yielding four overlapping fragments of approximately 1000 bp. These fragments are herein referred to as preS, S, C and X fragments and are obtained by using a total of 12 different primers in 8 PCR reactions. First round and second round reactions were run for 40 or 30 cycles respectively. The initial denaturation and final elongation steps consisted of 5 minutes. Table 4 lists constant parameters of each of the 8 PCR reactions while Table 2 (Chapter 2) lists variable parameters as well as primer names and sequences. Real time PCR reactions were performed on the Opticon<sup>®</sup> 2 DNA engine (Biorad) or the Chromo4<sup>®</sup>/PTC200 (Biorad), while non real time PCR reactions were performed on a Mastercycler<sup>®</sup> Gradient (Eppendorf).

**Table 4 : Constant parameters for the complete genome PCR reactions**

<b>Reagent</b>	<b>End concentration</b>	
PCR Buffer	1x	
MgCl <sub>2</sub>	Variable, <i>cf</i> Table 2	
dNTP	200 nM	
Forward primer	200 nM	
Reverse primer	200 nM	
(SYBR <sup>®</sup> Green	1x)	
DNA polymerase	1 unit	
Template	1st round: 5 µl, 2nd round: 2 µl of 1st round	
Volume	25 µl	

<b>Step</b>	<b>Temperature (°C)</b>	<b>Time (s)</b>
Denaturation	95	20
Annealing	Variable, <i>cf</i> Table 2	20
Elongation	72	60

## 5.2. M13 PCR

To verify the cloning of PCR products into plasmids, the inserted fragments were amplified by M13 PCR. The M13 primer binding sequences are located at the borders of the cloning site in the vectors and allow for the amplification of the inserted fragment, independent on its sequence. The template consists of a single bacterial colony picked from a growth plate and inserted directly into the PCR mixture using a sterile wooden tooth pick. During the first denaturation step of the PCR, the bacterial cell wall is denaturated and plasmids are freed. The time of the elongation step is dependant on the length of the inserted fragment and generally is 1 minute for each 1000 bp. M13 PCR reactions were run for 35 cycles in a Mastercycler<sup>®</sup> Gradient (Eppendorf). Primers are listed in Table 3 and PCR conditions in Table 5.

**Table 5 : Conditions of the M13 PCR reaction**

<b>Reagent</b>	<b>End concentration</b>	
PCR Buffer	1x	
MgCl <sub>2</sub>	2.5 mM	
dNTP	200 nM	
Forward primer	800 nM	
Reverse primer	800 nM	
DNA polymerase	1 unit	
Template	Bacterial colony	
Volume	25 µl	

<b>Step</b>	<b>Temperature (°C)</b>	<b>Time (s)</b>
Denaturation	95	20
Annealing	55	20
Elongation	72	variable

### 5.3. Semi-quantitative detection of HBV DNA by real time PCR

Real time TaqMan<sup>®</sup> HBV DNA amplification was performed on a LightCycler<sup>®</sup> (Roche Diagnostics) with two sets of PCR primers and probes, corresponding to the HBV core gene (Jardi *et al.* 2001). Both sets were universally conserved among 44 known sequences obtained from GenBank corresponding to genotypes A-H. As an external standard for quantification a serial dilution of a 10<sup>-6</sup> pre-diluted Eurohep HBV standard reference 1 serum, (genotype A, subtype adw2), diluted in HCV RNA negative, HAV RNA negative, parvovirus B19 DNA negative and HIV RNA negative human cryosupernatant (National Institute for Biological Standards and Control code 98/780, South Mimms, England) was used. The detection limit was 5 IU/ml.

## 6. Agarose gel electrophoresis

Agarose gel electrophoresis allows for the visualisation and separation of DNA products. Due to their high content of phosphates, DNA molecules are negatively charged and when exposed to an electrical current, they migrate to the positively charged cathode, a process referred to as

electrophoresis. During gel electrophoresis the DNA is pipetted into pockets, or slots, in an agarose gel which is placed into an ion containing buffer (TAE buffer). When applying an electrical current to this buffer, ions, including the DNA, migrate through the agarose gel. Since agarose forms a tightly meshed structure, the DNA molecules, while migrating through the pores of this mesh, are more or less slowed down depending on their size. The migration is visually followed by mixing the DNA with an ionic marker (Bromophenol blue) which migrates as similar to a DNA molecule of 200 bp. The migration is stopped when this marker has passed through 2/3 of the gel. By adding a molecule to the gel which fluoresces only when bound to double stranded DNA (ethidium bromide or Sybergreen), the DNA products can be visualized under ultraviolet light (UV) after migration. By letting a DNA marker or ladder (1 kb plus DNA ladder™, Life Technologies) run in parallel with the PCR products, the size of the latter can be evaluated by comparison to the known sizes of DNA fragments in the marker.

One percent agarose gels were prepared by adding 1 g of agarose to 100 ml of 1x TAE buffer. By heating the suspension, the agarose was dissolved in the buffer and formed a solid gel while cooling down. While still liquid, 1.5 µl of ethidium bromide were added to the agarose solution which was then poured into a rectangular casting form of 14x12 cm. An added comb with a thickness of 1 mm formed the slots into which the DNA was pipetted. After 30 minutes of solidification, the gel was completely submerged into 1x TAE buffer inside a tank, perpendicular to the electrical current. 5 µl of PCR product were mixed to 2 µl of 6x loading buffer and pipetted into a slot. The glycerol contained in the loading buffer prevents the PCR product from dissolving in the TAE buffer. The gel was exposed to an electrical current of 140 V until the end of migration and pictures were taken under UV light in an InGenius Gel documentation system (Syngene). PCR products that presented single bands of the correct size were selected for downstream applications.

## 7. Cloning

Cloning can be used to identify the presence of multiple DNA populations in a single PCR product. The PCR product is mixed with vectors and an enzyme which ligates each DNA molecule into a different vector. While transfecting these vectors into bacteria, an antibiotics resistance is introduced and allows for the selection of successfully transfected bacteria. Since each bacterium can only maintain one vector, it is possible to separate the vectors, and thus

the DNA inserts, by spreading the bacterial mix on a growth plate and, after an incubation period, analysing appeared colonies that each was initially based on a single bacterial cell.

### 7.1. TOPO TA Cloning<sup>®</sup> kit

The TOPO TA Cloning<sup>®</sup> kit (Invitrogen) uses linear pCR<sup>®</sup>4-TOPO<sup>®</sup> vector with 3' thymidine (T) overhangs and a covalently bound topoisomerase. The Platinum<sup>®</sup> Taq DNA polymerase used in PCR reactions, non-template specifically adds 5' adenosine overhangs to every PCR product, a property used by the topoisomerase to introduce the PCR product into the vector and form a circular closed plasmid. Specific sequences upstream and downstream of the inserted fragment are used in downstream applications such as sequencing, additional PCR reactions and restriction assays. In addition, vectors encode resistance genes to ampicillin and kanamycin as well as an origin of replication site (pUC-ori) necessary for maintenance in the bacterial cell (Figure 9).

The following protocol was followed: 4 µl of fresh PCR product were mixed to 1 µl of diluted salt solution and 1 µl of TOPO<sup>®</sup> vector (included in the kit). The reaction mix was then incubated for 5 minutes at room temperature and put on ice. Electrocompetent One Shot<sup>®</sup> TOP10 bacteria (Invitrogen) were diluted 1:1 with distilled water and 100 µl were mixed with 4 µl of the TOPO<sup>®</sup> cloning reaction in a 0.1 cm electroporation cuvette. The bacteria/vector mix was then electroporated using the following conditions: tension: 2.25 kV, resistance: 200 Ohm, capacity: 25 µF, and immediately mixed to 250 µl of room temperature S.O.C. medium (Invitrogen). In order to allow for the expression of the resistance genes, the bacterial suspension was incubated for 1 h at 37 °C before 60 µl were spread on a LB solid agar growth plate containing kanamycin. After 24 h incubation at 37°C bacterial colonies were picked using sterile wooden toothpicks and analysed by M13 PCR.

### 7.2. Zero Blunt<sup>®</sup> TOPO<sup>®</sup> PCR Cloning kit

Complete HBV genomes amplified using a single PCR reaction giving a DNA fragment of 3200 bp (used only in Chapter 4, Part.5) were cloned into the pCR<sup>®</sup>-Blunt II-TOPO<sup>®</sup> vector (Figure 10) using the Zero Blunt<sup>®</sup> TOPO<sup>®</sup> PCR Cloning kit (Invitrogen). The main difference to the pCR<sup>®</sup>4 vector is the lack of 3' thymidine overhangs and that it can incorporate longer fragments of DNA. This is necessary since the polymerase used in this PCR reaction

(Phusion™ High Fidelity DNA polymerase, Finnzyme) does not add 3' adenosine overhangs to PCR products. The bacterial strain used and the experimental conditions are identical to the TOPO TA Cloning® kit protocol.

## 8. Sequencing

The sequencing method used is based on the dye terminator method (Sanger sequencing). In dye terminator sequencing, extension is initiated at a specific site on the template DNA by using a short oligonucleotide (or primer) complementary to the template at that region. The primer is extended using a DNA polymerase, very similar to a typical PCR reaction. Included with the primer and DNA polymerase are the four dNTPs, along with a low concentration of fluorescently labelled, chain terminating di-deoxynucleotides. Random incorporation of the chain terminating nucleotide by the DNA polymerase results in a series of related DNA fragments that are terminated only at positions where the terminating nucleotide is used. Each fragment terminates either with a ddATP, ddTTP, ddGTP or ddCTP, each labelled with a different fluorophore. The fragments are then size-separated by capillary electrophoresis in a polyacrylamide gel with a laser reading the fluorescence at the end of the capillary. Sequences are assembled by comparing the size dependant order of appearance of fragments and the nucleotide specific fluorescence peaks, referred to as sequence electropherograms.

The following protocol was used: Before running a sequencing PCR, the sample has to be separated from primers and none incorporated nucleotides, which would otherwise interfere with the sequencing PCR reaction. This is performed using a DNA binding spin column. Small fragments, such as primer and non-incorporated nucleotides pass through the column while larger PCR fragments (80 bp - 20 kbp) are bound. The PCR purification was performed using the Jet Quick PCR Purification Spin® kit (Genomed) as follows: 20 µl of PCR product were mixed with 400 µl of buffer H1 (included in the kit) in a 1.5 ml microcentrifuge tube. The mixture was then loaded in a Jet Quick® Spin column sitting in a 2 ml centrifugation tube and centrifuged for 1 minute at 12 000 g. The flowthrough was discarded and 500 µl of buffer H2 (included) were added to the column, followed by centrifugation as before. The spin column was transferred into a clean 1.5 ml microcentrifugation tube and 30 µl of preheated 70°C TE buffer were added. After 2 minutes of incubation at room temperature and centrifugation as before, the eluted DNA was quantified as described in Section 4. The quantity of DNA required for sequencing is listed in Table 6.

**Table 6 : Quantity of DNA used for sequencing**

<b>Template</b>	<b>Quantity</b>
100-200 bp	1 - 3 ng
200-500 bp	3 -10 ng
500-1000 bp	5 - 20 ng
1000-2000bp	10 - 40 ng
>2000 bp	20 - 50 ng

To 1  $\mu$ l of BigDye Terminator<sup>®</sup> mix (included) were added: 1.5  $\mu$ l of 5x TE buffer, 1  $\mu$ l of 5  $\mu$ M primer and the necessary quantity of DNA, diluted in deionised water (max 5  $\mu$ l). The mixture was brought to a volume of 10  $\mu$ l with deionised water. Depending on the number of samples to be sequenced this was done in a 96 well plate. The PCR conditions were as follows: 25 cycles of 96 °C for 10 seconds, 50 °C for 5 seconds, 60 °C for 2 min with an initial denaturation step of 96°C for 1 minute.

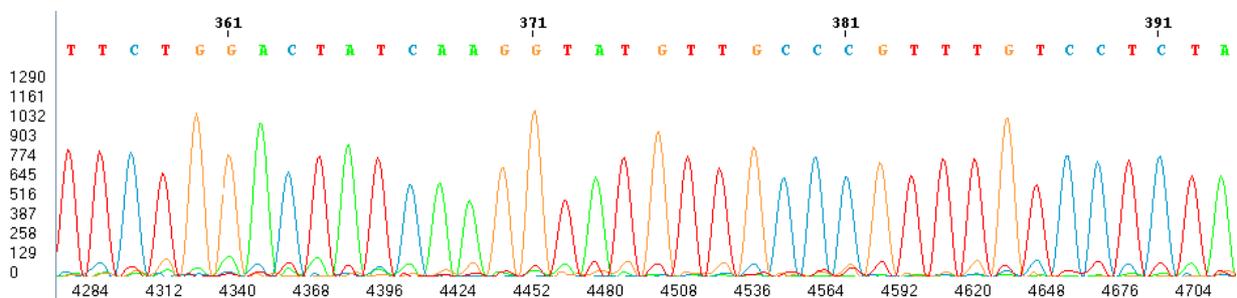
Non-incorporated dye needs to be removed before sequencing and was done as follows: To each sample, 5  $\mu$ l of 125 mM EDTA and 10 mM of deionised water were added. After mixing, 60  $\mu$ l of 100% ethanol were added to each well, the plate was vortexed and incubated for 15 minutes at room temperature in the dark and centrifuged at 4 °C, 3000 RPM for 30 minutes in a Megafuge<sup>®</sup> 1.0R (Heraeus). Immediately afterwards, the ethanol was removed by inverting the plate on tissue paper with subsequent centrifugation at 1000 RPM for 60 seconds. The previous steps were repeated with a 70% ethanol solution and a centrifugation at 4°C for 15 minutes. The plate was then dried for 15 minutes in a UNIVAP 150 H (UniEquip) and stored at 4 °C until use.

Preparation for sequencing consisted of heating the samples for 5 minutes to 95°C, adding 10  $\mu$ l of HI-DI (Applied Biosystems) and heating again at 95°C for 5 minutes. The plate was then loaded on the capillary sequencer (ABI PRISM<sup>®</sup>, 3130xl Genetic Analyzer, Applied Biosystems). All applications used capillaries with a length of 80 cm.

## 9. Phylogenetic analysis

### 9.1. Electropherogram analysis

Sequences were extracted from electropherograms by using the SeqScape<sup>®</sup> software (Applied Biosystems) and visually inspected for inconsistencies. The application also allows assembly of individual sequences to complete genomes by alignment to a known reference sequence. Sequences were then imported into the Bioedit sequence manipulation software and aligned to a set of reference sequences using its internal Clustal W algorithm. Figure 11 shows an example of an electropherogram.



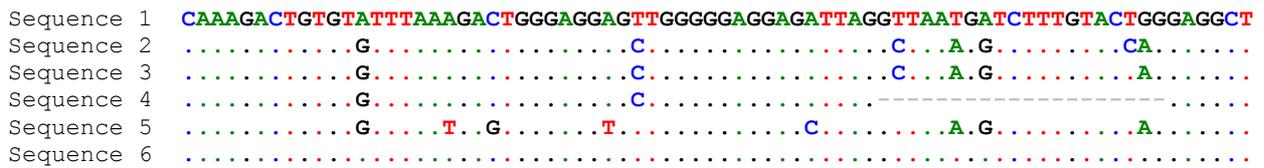
**Figure 11 : Example of a sequence electropherogram**

During capillary sequencing, fragments are separated by size (horizontal) and the type and intensity (vertical) of fluorescence is read. Depending on their colour, peaks indicate thymidine, cytidine, adenine and guanine. Usually

low background fluorescence is visible although when reaching at least 50% of the intensity of single peaks, could indicate quasi-species or a mixed DNA population.

### 9.2. Sequence alignment

Sequence alignments consist of arranging sequences, known to be linked by function or genome location, in such a way as to minimize differences between them. Gaps are introduced when they maximize the overall similarity. Sequences are said to be similar when at least 50% of nucleotide positions are identical after alignment. Identical nucleotides are usually indicated by a dot. In the case of HBV, a minimal similarity of 85% is generally observed, while the similarity between hepatitis C strains can descend to as low as 60%. An example alignment is shown in Figure 12.



**Figure 12 : Example of a sequence alignment**

Identical positions compared to a reference are indicated as dots. Gaps are shown as dashes.

### 9.3. Nucleotide distance calculation

Nucleotide distance calculations, as well as phylogenetic links between sequences, were deduced using the MEGA3.1 software and the Kimura 2 parameter and neighbour-joining models. While calculation of the number of single nucleotide polymorphisms between two or more sequences is sufficient for basic analysis and phylogenetic reconstruction, they do not take into account evolutionary mechanisms such as reversion or back mutations. Reversions are a series of mutations that restore the original nucleotide at a position. A simple example would be a T to A mutation which is followed by an A to T mutation at the same nucleotide position. It is important to take into account such events, especially if the time of evolution between sequences is long. In the last three decades, several substitution models have been developed which in general take into account the probability of mutation from one nucleotide to another, as well as the frequency of a given nucleotide. A simple model would be to assign the same probability to each possible mutation and ignore nucleotide frequencies. This model is referred to as the one parameter Jukes-Cantor model (Table 7a). A very complicated model would be to consider a different probability for each possible mutation (12 possibilities, *i.e.* A to T, A to C, A to G, T to A, etc.) and to take into account the frequency of all 4 nucleotides. This model is referred to as the 16 parameter model (Table 7c). The most widely used model however, for basic time-independent phylogeny is the Kimura 2 parameter model, which ignores nucleotide frequencies but differentiates between the probability of transitions and transversions (Table 7b).

**Table 7 : Nucleotide substitution models**

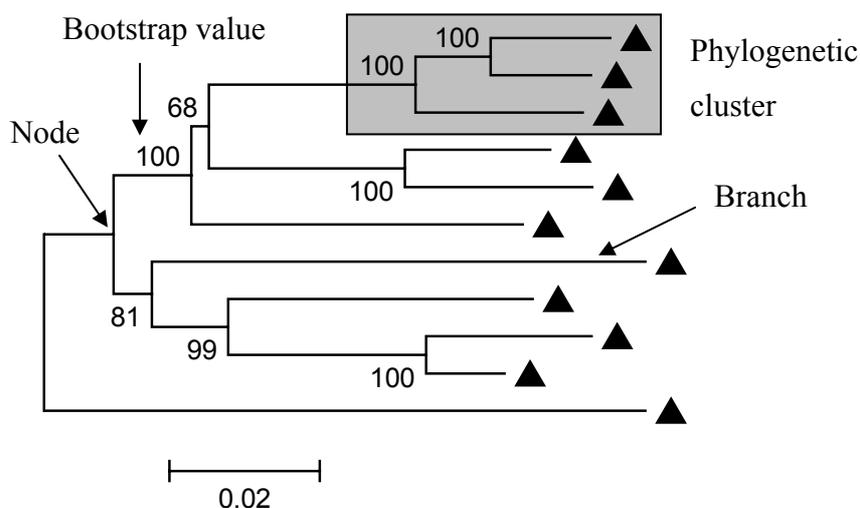
a)	A	T	C	G	b)	A	T	C	G	c)	A	T	C	G
A	-	a	a	a	A	-	b	b	a	A	-	dx <sub>T</sub>	gx <sub>C</sub>	jx <sub>G</sub>
T	a	-	a	a	T	b	-	a	b	T	ax <sub>A</sub>	-	hx <sub>C</sub>	kx <sub>G</sub>
C	a	a	-	a	C	b	a	-	b	C	bx <sub>A</sub>	ex <sub>T</sub>	-	lx <sub>G</sub>
G	a	a	a	-	G	a	b	b	-	G	cx <sub>A</sub>	fx <sub>T</sub>	ix <sub>C</sub>	-

Substitution matrices of the Jukes-Cantor one parameter (a), Kimura 2 parameter (b) and 16 parameter (c) models. Mutation probabilities are indicated by the letters *a* to *l* while nucleotide frequencies are indicated by the letter *x*.

## 9.4. Phylogenetic reconstruction

Distance based phylogenetic reconstruction uses the distances calculated between sequences to assemble a phylogenetic tree. On this tree, sequences are linked by nodes which represent hypothetical ancestors or points of deviation between two sequences and connected by horizontal lines (branches) of lengths proportional to the actual genetic distance (Figure 13). Sequences are said to be of the same phylogenetic group when they cluster on the same node. Depending on the organism analysed, phylogenetic groups are referred to as lineages and sub-lineages (*e.g.* avian influenza) genotypes and subtypes (*e.g.* hepatitis B, measles, etc.) or families and species (*e.g.* mammals, insects, etc.). The most commonly used tree construction method is the neighbour-joining model which assumes variable substitution rates across lineages and uses few computational resources.

In the case of more complex phylogenies, maximum likelihood or Bayesian methods, which are very computationally intensive, are applied. The latter are not distance based methods but use the actual sequences to construct trees. Briefly, every possible tree that can be constructed from each nucleotide position of a set of sequences is evaluated by calculating its probability of being the correct tree, estimated by applying a nucleotide substitution model. The final “correct” tree would be the one with the highest probability or likelihood. Maximum likelihood trees were calculated using the PAUP\* 4 software package and the Modeltest application to determine the substitution model which best describes the sequence data.



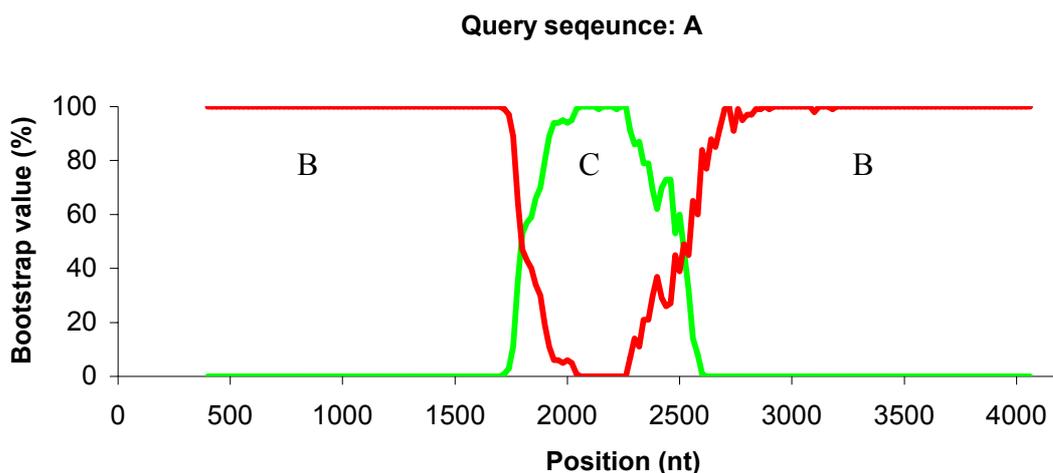
**Figure 13 : Example of a phylogenetic tree**

Sequences (indicated by triangles) are linked to each other by horizontal lines (branches) with lengths proportional to the genetic distance. Nodes indicate hypothetical ancestors. Numbers indicate bootstrap values of nodes significance.

In order to evaluate the significance of nodes, bootstrap calculation is applied: Sequence positions of a given alignment are reshuffled to construct alignments of the same length than the original. Not all sites are considered and sites can be considered more than once. Thus an alignment of 7 positions: 1 2 3 4 5 6 7 could be reshuffled into 1 1 3 4 4 5 6 or 2 2 4 5 6 6 7. For each alignment a phylogenetic tree is constructed and for each node of the phylogenetic tree of the original alignment, the number of times it appears in the phylogenetic trees of the reshuffled alignments is counted. These values are referred to as bootstrap values and are usually expressed as percentages. High bootstrap values indicate significant nodes. Bootstrap values below a certain cut-off mean that a node can be ignored.

### 9.5. Recombination analysis

Recombination between strains can be detected by comparing phylogenetic reconstructions based on different sequence regions. When a given sequence can be found in different phylogenetic clusters a recombination event has occurred. The software Simplot calculates phylogenetic trees based on sequence fragment windows. The size of these fragments is referred to as windows size. Windows can overlap. The bootstrap value of the node connecting the analysed sequence with the most similar sequence is plotted on a graph, referred to as bootscan analysis. Recombination events have occurred when different most similar sequences appear during the analysis (Figure 14).



**Figure 14 : Example of bootscan analysis**

Bootstrap values of the node connecting the query sequence (A) with its most similar sequence on a given region. This example shows that the query sequence is a recombination between two different sequences B and C.

## Chapter 4: Results and Discussions

### Part I

Results of part I were published as:

Olinger CM, Weber B, Otegbayo JA, Ammerlaan W, van der Taelen-Brulé N, Muller CP. (2007) Hepatitis B virus genotype E surface antigen detection with different immunoassays and diagnostic impact of mutations in the preS/S gene. *Medical Microbiology and Immunology*. 196:247-252.

#### Hepatitis B virus genotype E HBsAg detection with different immunoassays and diagnostic impact of mutations in the preS/S gene

The influence of genetic variability on the sensitivity of serological and molecular assays has so far received little attention. However, it seems that the analytical sensitivity of HBsAg assays is dependent on HBV genotypes and subtypes (Weber *et al.* 2003). By testing dilution series of different HBV subtypes, up to 10-fold differences in the sensitivity of three commercial assays were observed (Weber 2005). HBsAg detection with monoclonal antibody (mab)-based diagnostic assays may be unreliable in populations where circulating subtypes/genotypes or variants are distinct from the virus strain used for the production of mabs. The major neutralising epitope, (or a-determinant, residues 110-164), on the hydrophilic portion of the surface antigen of the prevalent genotype E strains is the most diverging from that of genotype A, with a difference of eight distinct amino acids (Norder *et al.* 2004). Thus, an important question is whether diagnostic reagents initially developed for genotype A strains are suitable also for the detection of genotype E infections. In the present study the performance of three commercial assays for the detection of surface antigen was evaluated in sera of chronic carriers of genotype E virus.

## 1. Clinical samples

Sera were collected at the University College Hospital (Ibadan, Nigeria) from 93 patients with liver disease (2003) and 107 HIV patients (1995 and 1998) and stored at -80°C. All sera were diluted 1/10 in HBV pooled negative serum and were tested with three different HBsAg assays (A, B and C). Real time quantification PCR and nested PCR of the preS1, preS2 and S gene were performed on DNA extracted from undiluted sera and sequenced as described in Chapter 3.

## 2. Serological assays

Diluted serum samples were tested using three different commercial HBsAg diagnostic tests. The AxSYM<sup>®</sup> HBsAg v2 assay (Assay A, Abbott Diagnostics) is a microparticle enzyme immunoassay that uses a monoclonal capture antibody directed against a conformational epitope of the first loop of the a-determinant (aa 121 to 124) and a polyclonal tracer antibody. The Elecsys<sup>®</sup> HBsAg (Assay B, Roche Diagnostics) is an electrochemoluminescent assay based on a monoclonal capture antibody directed against the second loop of the a-determinant (aa 143-145) and a monoclonal antibody with the same specificity as the tracer antibody. The VIDAS<sup>®</sup> HBsAg Ultra (Assay C, Biomérieux) is an enzyme linked fluorescent assay and uses two monoclonal antibodies for capture that recognize two non-overlapping conformational epitopes outside the a-determinant (aa 101-105 and aa 199-208) and a polyclonal tracer antibody.

A test result was interpreted as true negative or true positive if it was negative or positive in all the three HBsAg assays. Discrepant samples were retested with the assays that were initially reactive. If the sample was again positive it was submitted to confirmatory testing in a neutralisation assay of the same manufacturer. Anti-HBc and anti-HBs (Elecsys<sup>®</sup> HBsAg and anti-HBc) testing was performed in all sera with discordant results in the first test as long as enough sample was available. A test result of a discrepant sample was interpreted as true positive if it was confirmed positive in the neutralisation assay of the same manufacturer(s) independently of the anti-HBc or HBV DNA PCR result. A discrepant sample was interpreted as initially false positive if it was negative after repeated testing and as repetitively false positive if it was not confirmed in the neutralisation assay of the same manufacturer.

### 3. Results

#### 3.1. HBV DNA detection

HBV DNA was detected in 84 of the 200 patients (Table 8) both by real time PCR and nested PCR and these were considered as true DNA positive. Six sera with low viral loads (< 50 IU/mL) and two sera with 229 and 1320 IU/ml of HBV DNA were only positive by real time PCR.

**Table 8 : Comparison of DNA status with the HBsAg status obtained with 3 different assays**

	DNA neg	DNA pos	Total
<b>3 assays HBsAg pos</b>	6	55	<b>61</b>
<b>3 assays HBsAg neg</b>	87	16	<b>103</b>
<b>HBsAg discordant</b>	23	13	<b>36</b>
<b>Total</b>	<b>116</b>	<b>84</b>	<b>200</b>

#### 3.2. HBsAg detection

*Concordant HBsAg detection.* One hundred and three samples were tested negative with the 3 HBsAg assays (Table 8), 16 of these had low viral loads ranging from 1.0 to  $7 \times 10^3$  IU/ml. Sixty-one of the 200 samples gave congruent positive results (Table 8) in the 3 HBsAg assays, 90.2% (55 of 61) were HBV DNA positive with viral loads ranging from 1.2 to  $7 \times 10^9$  IU/ml by real time PCR. Two HBsAg positive samples gave very low index values, close to the cut-off with all three assays but were confirmed as true HBsAg positives by neutralisation assay. Viral loads of these two samples were 1320 and 5224 IU/ml, both were anti-HBc and anti-HBs negative.

*Discordant HBsAg detection.* The remaining 36 serum samples gave discordant results between HBsAg assays (Table 9), 30, 1 and 2 sera were only initially reactive in one of the assays A, B or C respectively. Two and one sample were initially reactive in assays A and C and in assays A and B, respectively. Of the discrepant samples, 13 were tested HBV DNA

positive with viral loads ranging from 10 to 432 IU/ml. The HBsAg index values of the discrepant sera were relatively low and ranged from 1.6 to 14.1 (assay A), 1.0 to 3.1 (assay B) and 1.3 to 1.4 (assay B). Of the 36 discrepant sera, 20 were anti-HBc positive. For 20 sera there was enough material left for additional anti-HBs testing and 3 samples were anti-HBs positive (10.5 to 15.5 IU/l).

**Table 9 : Comparison of DNA status and HBsAg status of patients with discordant results in the 3 HBsAg assays**

<b>Positive result with:</b>	<b>DNA neg</b>	<b>DNA pos</b>	<b>Total</b>	<b>positive after neutralisation</b>	<b>Interpretation</b>
<b>Assay C only</b>	0	2	<b>2 (0)*</b>	0	False positive
<b>A + C</b>	1	1	<b>2 (0)</b>	0	False positive
<b>Assay A only</b>	21	9	<b>30 (29)</b>	1 (DNA +)	False positive (n =29) True positive (n =1)
<b>A + B</b>	1	0	<b>1 (1)</b>	0	False positive
<b>Assay B only</b>	0	1	<b>1 (1)</b>	0	False positive
<b>Total</b>	<b>23</b>	<b>13</b>	<b>36(31)</b>	1	

\*Numbers in brackets indicate samples repeatedly positive in a given assay

Sixty-eight of the 84 samples positive in both DNA detection assays were positive in at least one HBsAg assay (55 where triple HBsAg positive, 16 were triple negative and 13 presented discordant results). The resolution of the discrepancies by repeated testing and by neutralisation assays showed a relatively high number of false positive results (Table 9). Of the 30 assay A-only positive samples, 29 were repeatedly reactive, but only one serum with a viral load of 101 IU/ml was confirmed positive by neutralisation assay; this serum was anti-HBc and anti-HBs negative. Thus 29 were false positive among the 30 assay A-only positive samples. All the samples that were initially reactive in either assay B or C or double positive in assays A and B or A and C (DNA positive or negative; *cf* Table 9) were not confirmed by neutralisation assay and thus considered false positive. Only one of the 36 discrepant samples was considered true positive (confirmed by neutralisation assay) and HBV DNA positive.

Sequencing of 48 HBV DNA positive samples revealed the presence of genotype E in 45 and genotype A1 in 3 individuals. Amino acid substitutions within the pre S and S gene from genotype E strains are summarised in Table 10. Mutations in the preS1, preS2 and S gene were classified in four groups depending on their effect of HBsAg detection: (1) no effect on HBsAg detection, (2) reduced signal in one or more HBsAg assays, (3) no HBsAg detection and (4) mutations that were present in both groups 2 and 3. A reduced HBsAg signal was defined as a decrease of at least 25% in comparison to the mean value obtained within the same assay for the 61 HBsAg positive sera. Only 2 mutations within the a-determinant were observed that were associated with a reduced HBsAg signal (L127P) or a negative HBsAg result (S143T). The highest number of mutations that were associated with impaired HBsAg detection were located in the S gene (n=15), 13 of which were outside of the a-determinant region. Only 6 mutations of each the preS1 and preS2 region were associated with a reduced or a negative HBsAg signal. The A184V mutation of the S gene was the most frequent amino acid substitution with reduced (n = 3) or negative HBsAg signal (n=3). In 5 of these 6 sera A184V was the only mutation present. However there was one sample that was HBsAg positive in assay A only (and confirmed) and that harboured the A184V mutation. One sample with reduced signal intensities in the three HBsAg assays (serum 173, data not shown) and 5 HBsAg negative sera (sera 58, 72, 83, 114 and 162; Table 10) did not show any amino acid substitution which may be associated with an expected effect on the HBsAg detection.

### 3.3. Quantitative impact of preS/S gene mutations on HBsAg detection

The detection of HBsAg was most frequently impaired in assay B, since 13 samples showed reduced signal intensities ranging between 0.1% and 68.5% of the mean value for HBsAg positive sera in this assay. Assays A and C showed decreased HBsAg signals in 9 (2% - 71%) and 5 (2% - 74.8%) serum samples, respectively.

**Table 10 : Classification of DNA positive serum samples into 3 groups according to the HBsAg detection signals**

Sample assignment	Test*			Mutations						
	A	B	C	preS1			preS2		S	
10	130.8	173	127.1	H86A	P94L					
13	226.1	183	123.1	R35K						
16	189.4	169	128.6	E54D	K86Q					
18	99.7	112	122.5	T19S					P203Q	
20	117.6	100	126.5	H16Q	R35K	K86T				
23	199	138	127.2	H16Q	T19S					
27	179.2	85	131.9	F67S			P52L		R122K	
32	200.7	96	120.6	T19S	T53K		A39V P52L			
34	159.5	122	120.3	S96F						
49	91.2	102	121.1	K86T					P52L	
60	100.1	116	121.6	P94L						
78	209.7	133	117.3	S77W						
14	<b>16.6</b>	<b>67</b>	128.6	<b><i>N48T</i></b>	E54Q					
21	<b>43.7</b>	112	127.2	H16Q			<b><i>D51G</i></b>		P52L	
25	<b>55.3</b>	<b>69</b>	131	H16Q	<b><i>V14A</i></b>	<b><i>L49R</i></b>	<b><i>F46S</i></b>			
37	<b>13.3</b>	<b>51</b>	130.6	W77T	P94L	S96A		P203Q		<b><i>Y206H</i></b>
47	<b>23.6</b>	77	127.2	R35K	P94L		A39V	A53V	V224A	
53	<b>0.4</b>	<b>4</b>	<b>29.4</b>	<b><i>N48T</i></b>	<b><i>Q82P</i></b>	P94L				A184V
92	<b>48.6</b>	95	122.2	R35K	P94L		A39V	A53V	F83C	V224A
131	<b>47</b>	113	83.9	F67L			<b><i>P54L</i></b>			
141	<b>68.5</b>	<b>52</b>	<b>8.4</b>							A184V
146	<b>0</b>	<b>2</b>	<b>2</b>							A184V
164	<b>55.7</b>	104	<b>74.8</b>	<b><i>L49R</i></b>			<b><i>L127P T189I S204N</i></b>			
171	<b>0</b>	<b>2</b>	<b>2</b>	E54D						
187	<b>29.8</b>	<b>71</b>	87.8	T19S	T53K	K86T	A39V	<b><i>F46S</i></b>	<b><i>I49T</i></b>	
33							A39V	A53V	C76F	F83C S143T V224A
54							<b>T31A</b>			
58				W77S						
64							<b>P70T I82T</b>			
72							P52L			
83				P94L						
86							<b>T31A</b>			
114				P94L						
126				<b>D91N</b>						
135 <sup>†</sup>							A184V			
138							A184V			
140							A184V			
162				H16Q						
165							<b>F85L</b>			
169				K86Q	<b>T97N</b>					
189				K86T	P94L	<b>P93Q</b>	<b>I213T</b>			

Mutations in the preS/S gene that occur: only in samples with reduced HBsAg detection signals in at least one assay are shown in ***bold and italic***; in samples that are HBsAg negative in the three assays on black background; in samples with reduced or

negative HBsAg signal on grey background; in samples with no effect on HBsAg signal, no formatting. \*Reduced HBsAg signals are marked in **bold**; †only AxSYM HBsAg positive

## 4. Discussion

HBsAg detection assays rely on the capture of the surface antigen by monoclonal antibodies. Most of these antibodies are directed towards immunogenic epitopes, mainly in the a-determinant of genotype A surface antigen. The surface antigen of genotype E differs by 5.7% from genotype A on a nucleotide level and 8.5% on the amino acid level. Within the a-determinant (110 – 164) 90% of genotype E strains differ by at least 8 amino acids (14.8%) from a prototype A sequence. Here the characteristics of three different HBsAg assays were tested with respect to their ability to detect genotype E. All three assays were similar in sensitivity (Assay A: 100% assays B and C: 98.4%). Assay A however showed a poor specificity (84%) in comparison with assay B (99%) and C (98%). It is our unpublished experience that repetitive freezing and thawing of samples generates low positive signals in assay A that are not confirmed by neutralisation assay. The apparent false positivity in assay A in comparison to B and C was already previously observed by one of us in non-E genotypes (Weber *et al.* 2006). Our results are also in agreement with the results obtained by Mizuochi *et al.* (Mizuochi *et al.* 2006) who could not demonstrate any failure of 10 commercial diagnostic kits to detect recombinant HBsAg of genotypes E even at very low levels of HBsAg (0.2 IU/ml).

Despite the overall low genetic variability of genotype E strains (Odemuyiwa *et al.* 2001; Mulders *et al.* 2004; Huy *et al.* 2006), several new mutations throughout the preS/S gene seem to be associated with impaired HBsAg detection. The comparison of DNA positivity with the intensity of HBsAg detection signals revealed a set of mutations that were only found in HBsAg impaired samples (reduction of > 25% of average value or HBsAg negative) suggesting that these mutations interfere with HBsAg detection. Two of these new mutations, L127P and S143T, are within the otherwise conserved a-determinant. One strain with the mutation S143T was not detected, by assay A only however mutations outside the a-determinant (preS2: A53V and S: C76F, F83C and V224A) may be responsible for this “failure”, although also explainable by low HBsAg titres (Alhababi *et al.* 2003; Jeantet *et al.* 2004).

In addition, our study revealed several mutations outside of the a-determinant that reduced HBsAg detection such as an A184V substitution present exclusively in 5 of 6 samples with impaired HBsAg detection. Although the A184V mutation impaired the HBsAg detection also

in assay A, only this assay detected a strain (sample 135, confirmed by neutralization, Table 9 and 10) harbouring this mutation. Several of the above mutations with a negative influence on HBsAg detection are found in positions discordant between genotype E and A, such as L127, S143 and A184 in genotype E which in genotype A correspond to P127, T143 and V184. Interestingly these mutations that seem to reduce HBsAg detection are normally found in genotype A and only exceptionally in genotype E. Thus the presence of the genotype E wild-type amino acids could cause a positive detection bias for genotype E in comparison to genotype A. The association of these mutations with the failure to detect HBsAg needs to be further confirmed by epitope mapping with monoclonal antibodies and recombinant HBsAg.

HBsAg was the only serological marker in two HBV DNA positive samples. Anti-HBc antibodies may have been below the detection limit of the competitive assay since the sera were diluted 1:10. However systematic dilution experiments tend to exclude this possibility (Melchior and Kirch; unpublished data). Alternatively, samples may have been drawn before anti-HBc seroconversion at the end of the incubation period. Decreased sensitivity for anti-HBc detection in genotype E infected patients, in-frame deletions or other mutations in the precore/core gene may also (partially) account for the isolated HBsAg reactivity (Echevarria *et al.* 1991; Valliammai *et al.* 1995). In a recent study, the isolated HBsAg positive result was described in 24 patients (Echevarria and Leon 2004) and seems more frequent during pregnancy and in spring (Echevarria JM, personal communication). In studies from France and China (Ni *et al.* 1993; Laperche *et al.* 2001) immune tolerance of HBV during perinatal transmission was thought to be the most likely explanation for the absence of anti-HBc conversion. In sub-Saharan Africa, early childhood transmission is frequent but isolated HBsAg seems to be less than 3%.

The results of our study show that the three assays have a similar sensitivity for the detection of genotype E HBsAg. Failure to detect HBsAg and differences in levels of HBsAg detection signals are probably due to mutations in the preS/S gene most of which were found outside of the a-determinant and suggest a positive detection bias for HBV genotype E compared to genotype A.

## Part II

Results of part II were published as:

Bekondi C, Olinger CM, Boua N, Talarmin A, Muller CP, Le Faou A, Venard V. (2007) Central African Republic is part of the West-African hepatitis B virus genotype E crescent. *Journal of Clinical Virology*. 40:31-37.

The Central African Republic is part of the West-African hepatitis B virus genotype E crescent

In sub-Saharan Africa, the seroprevalence of anti-HBV antibodies and the prevalence of chronic carriers are excessively high (Pawlotsky *et al.* 1995; Olubuyide *et al.* 1997; Mulders *et al.* 2004; Kurbanov *et al.* 2005; Makuwa *et al.* 2007). HBV has been classified into eight genotypes A to H (Okamoto *et al.* 1987; Norder *et al.* 1993; Stuyver *et al.* 2000), most of which have a more or less distinct geographic distribution (Kimbi *et al.* 2004; Hannoun *et al.* 2005; Kurbanov *et al.* 2005; Huy *et al.* 2006; Olinger *et al.* 2006; Makuwa *et al.* 2007). In South Africa, in south-eastern (Tanzania, Malawi) and eastern Africa (Somalia) genotype A dominates (Kramvis *et al.* 2002; Sugauchi *et al.* 2003; Kimbi *et al.* 2004; Hannoun *et al.* 2005). This genotype was also frequently found in Cameroon, in particular among HIV carriers, whereas in healthy Bantu and Pygmies population of this country genotype E and A were both prevalent (Kurbanov *et al.* 2005). Only sporadic strains of genotype A were found in Mali, Burkina-Faso (Mulders *et al.* 2004) and Nigeria (Olinger *et al.* 2006). In West-Africa genotype E predominates in a vast crescent spanning from Senegal to Angola, including Ivory Coast (Suzuki *et al.* 2003) Ghana (Huy *et al.* 2006), Nigeria, Togo, Benin, Mali, and at least the western part of The Democratic Republic of Congo (Mulders *et al.* 2004). However it is not clear how far the genotype E crescent extends to the east. This genotype was not found in any other part of the world except for some sporadic cases in African Black in Europe (Ganne-Carrie *et al.* 2006) and South America (Ganne-Carrie *et al.* 2006; Mathet *et al.* 2006).

Here we report the first complete genome of HBV from the Central African Republic and show that genotype E also dominates in this country, further extending the genotype E crescent to the east.

### 1. Clinical samples

Samples were randomly collected in 2004 from 112 male and 84 female patients admitted with symptoms of acute or chronic hepatitis to the Central Hospital in Bangui, the most populated part of the country. Most patients came from within 50 km from Bangui an area populated by a highly mobile population originating from most part of the country. Some other patients were directly referred from provinces in the Eastern and Western part of the country. The patients or next of kin gave their written informed consent. HBsAg and total anti-HBc and anti-HBs antibodies were determined by commercial tests (Architect<sup>®</sup>; Abbott Laboratories). The viral load was measured by a TaqMan<sup>®</sup> real time PCR kit (Abbott Laboratories) (detection threshold = 29 ( $\pm 3$ ) copies/ml; range from 29 to  $2.9 \times 10^{10}$  copies/ml). Weak signals below the calibration range were interpreted as positive, but not quantifiable.

## 2. Results

### 2.1. Serology

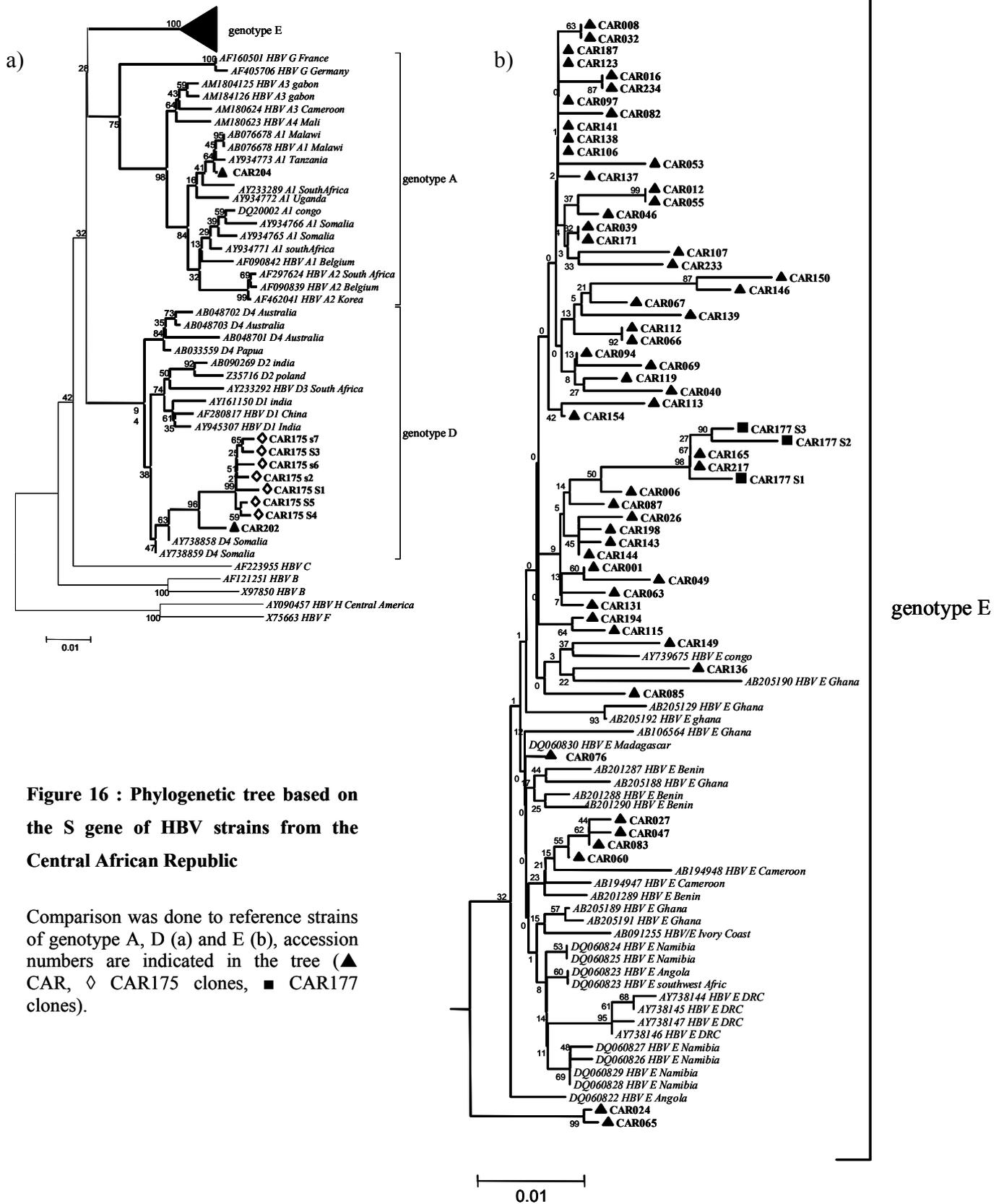
Among the 196 patients admitted for liver disease 183 were positive for at least one HBV marker (Table 11). One was positive for anti-HBs only and was excluded from the following analysis. Among the 120 that were HBsAg and anti-HBc positive, 79 (66%) were also HBV DNA positive by real time PCR (range 29 to  $5.6 \times 10^7$  copies/ml) and few were anti-HBs positive. Only 2 were HBsAg positive but anti-HBc negative; these two patients were also HBV DNA positive and anti-HBs negative. Conversely one third (60 of 182; 33%) were anti-HBc positive and HBsAg negative. However, 3 of these 60 patients contained detectable HBV DNA but not enough to be quantified (<29 copies/ml).

**Table 11: Characteristics of the patients cohort and their HBV serology status**

	HBsAg(+)/ Anti-HBc(+)	HBsAg(+)/ Anti-HBc(-)	HBsAg(-)/ Anti-HBc(+)	HBsAg(-)/ Anti-HBc(-)
Number	120	2	60	14
Mean age (range)	29.8 (10 - 70)	21.5 (18 - 25)	33.4 (17 - 72)	24.3 (15 - 50)
Sex ratio (male/female)	1.7	NA	0.9	0.8
Anti-HBs-positive	9	0	26	1
HBV-DNA-positive	79	2	3	0
Range (copies/ml)	29 – 5.6 10 <sup>7</sup>	37 - 105	< 29	–
ALT (UI/L)	1158	365	153.5	176.4
Range	22 - 8130	179 - 551	11 - 616	19 - 769

NA: not applicable (two males).





**Figure 16 : Phylogenetic tree based on the S gene of HBV strains from the Central African Republic**

Comparison was done to reference strains of genotype A, D (a) and E (b), accession numbers are indicated in the tree (▲ CAR, ◇ CAR175 clones, ■ CAR177 clones).

## 2.2. Sequence analysis and genotyping

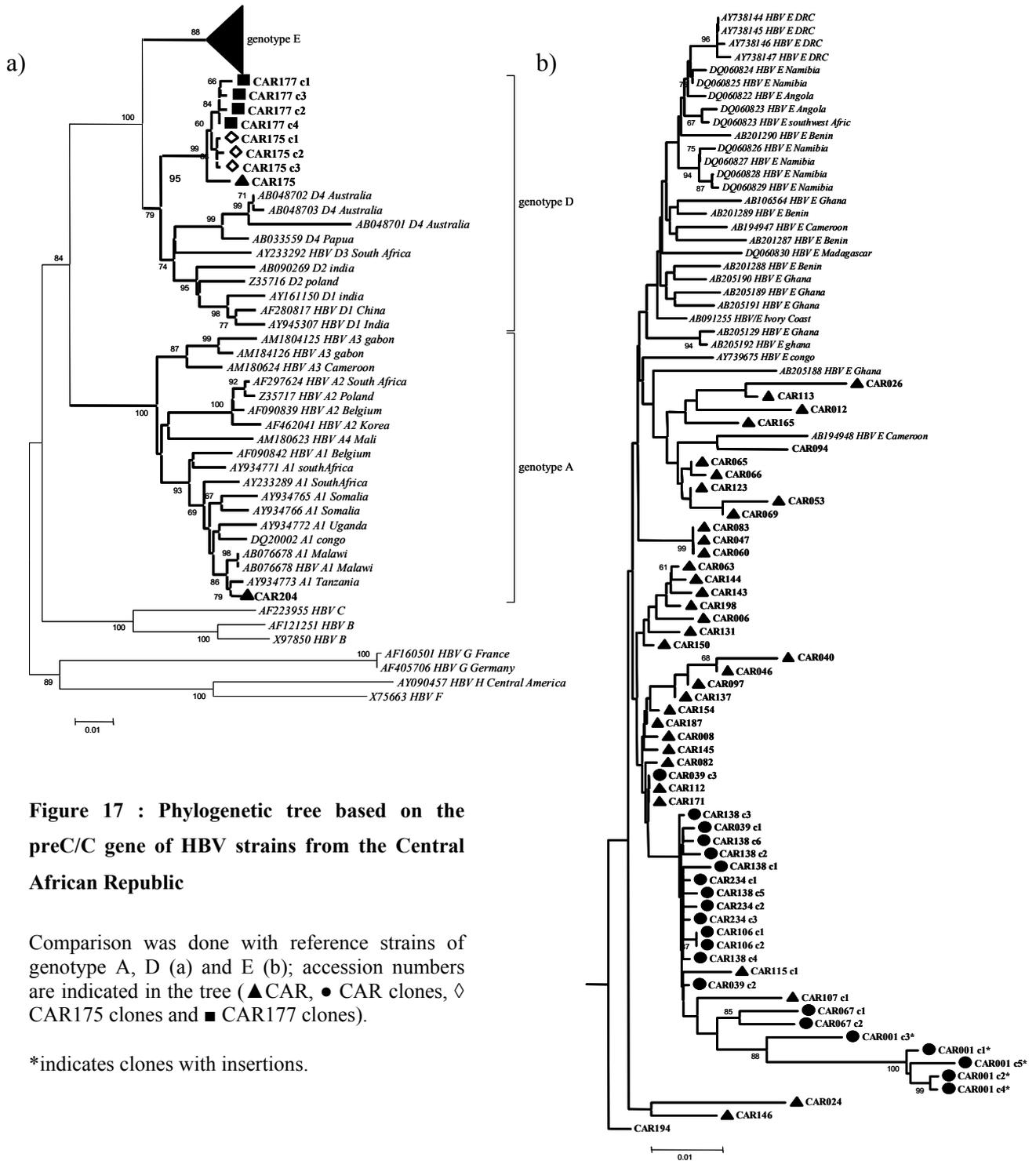
Sequences of the complete genome were obtained and analysed for 30 selected HBV strains from patients that had HBV DNA levels above 1000 copies/ml. Mostly because of low viral load (<100 copies/ml), only partial sequences were obtained for another 36 HBV strains (15 preS, 14X, 27S and 6 preC/C fragments). Phylogenetic comparison of all sequences with reference strains assigned 62 of the 66 strains to genotype E (94%), 3 (3.5%; CAR175, CAR177, CAR202) sequences to genotype D and one sequence (CAR204) to an A1 subgenotype, most closely related to A1 strains from Tanzania (accession number AY934773) and Malawi (AB076678) and some other East African strains (Figures 15, 16a and 16b).

S regions of two of the genotype D strains (CAR175, CAR202) were more closely related to genotype D4 from Somalia (Norder *et al.* 2004), (mean distance of 0.6% for CAR175; and 1.3% for CAR202) than to genotype D3 from South Africa (3% mean genetic distance). CAR177 for which only a partial sequence was available, was an outlier to genotype D3 based on the *PreC/C* region (mean distance of 4.2%) (Figure 17a) and was further analysed by cloning (see below).

The intra-group variability of the genotype E strains showed a diversity of 1.37% for the full-length genome, 1.08% for the preS/S, 1.59% for the preC/C and 1.77% for the X gene. When compared to genotype E isolates from Benin, Togo, Mali, Burkina Faso, Nigeria and Cameroon, the average distance at the nucleotide level was of 0.7% (range 0% to 1.8%) excluding the CAR001 (see below) (Figure 17b).

## 2.3. PreS/S gene analysis

Among the 45 available preS/S sequences, 10 strains of genotype E presented deletions of 3 to 33 nucleotides corresponding to 1 to 11 amino acids, in the preS2 gene. All of these deletions were in frame. Some mutations were also found in the a-determinant of the S gene. The most important mutations were G588C and C546A leading to amino acids substitutions G145A and T131N already described (Odemuyiwa *et al.* 2001). Other non-synonymous nucleotide substitutions were found in amino acids positions M103I, S143L/T



**Figure 17 : Phylogenetic tree based on the preC/C gene of HBV strains from the Central African Republic**

Comparison was done with reference strains of genotype A, D (a) and E (b); accession numbers are indicated in the tree (▲CAR, ●CAR clones, ◇CAR175 clones and ■CAR177 clones).

\*indicates clones with insertions.

and D144Q as described before (Weinberger *et al.* 2000; Weber 2005). These mutations could interfere with the detection by commercial HBsAg kits. An additional 27 redundant S fragments did not contribute any further information.

#### 2.4. PreC/C gene analysis

Nine of 36 strains (25%), all of genotype E, displayed a mutation in the start codon of the *PreC* ORF (G1896A). This mutation induces a stop codon generating a non-functional HBeAg. Thirty-six X genes (overlapping with the core promoter) were obtained and the A1762T and G1764A substitutions leading to the amino acid changes K130M and V131I previously described as hot spot mutation sites were found in 8 of 36 strains, all of genotype E.

Some clones of the CAR001 strain contained a rare replacement mutation in the core promoter involving the hepatocyte nuclear factor binding site described by (Kurbanov *et al.* 2005). This mutation may be specific to HBV genotype E and results in a lower HBeAg titre, a high HBV DNA level and progression towards liver fibrosis.

#### 2.5. Mixed infections and recombination

Twelve serum samples presenting mixed nucleotides in the sequencing electropherogram were screened for mixed infections by cloning either the product of the preS, S, or the C PCR fragment into a vector (see Figures 16a, 16b and 17a, 17b). The cloning showed no evidence of mixed infections except for the CAR001 mentioned above.

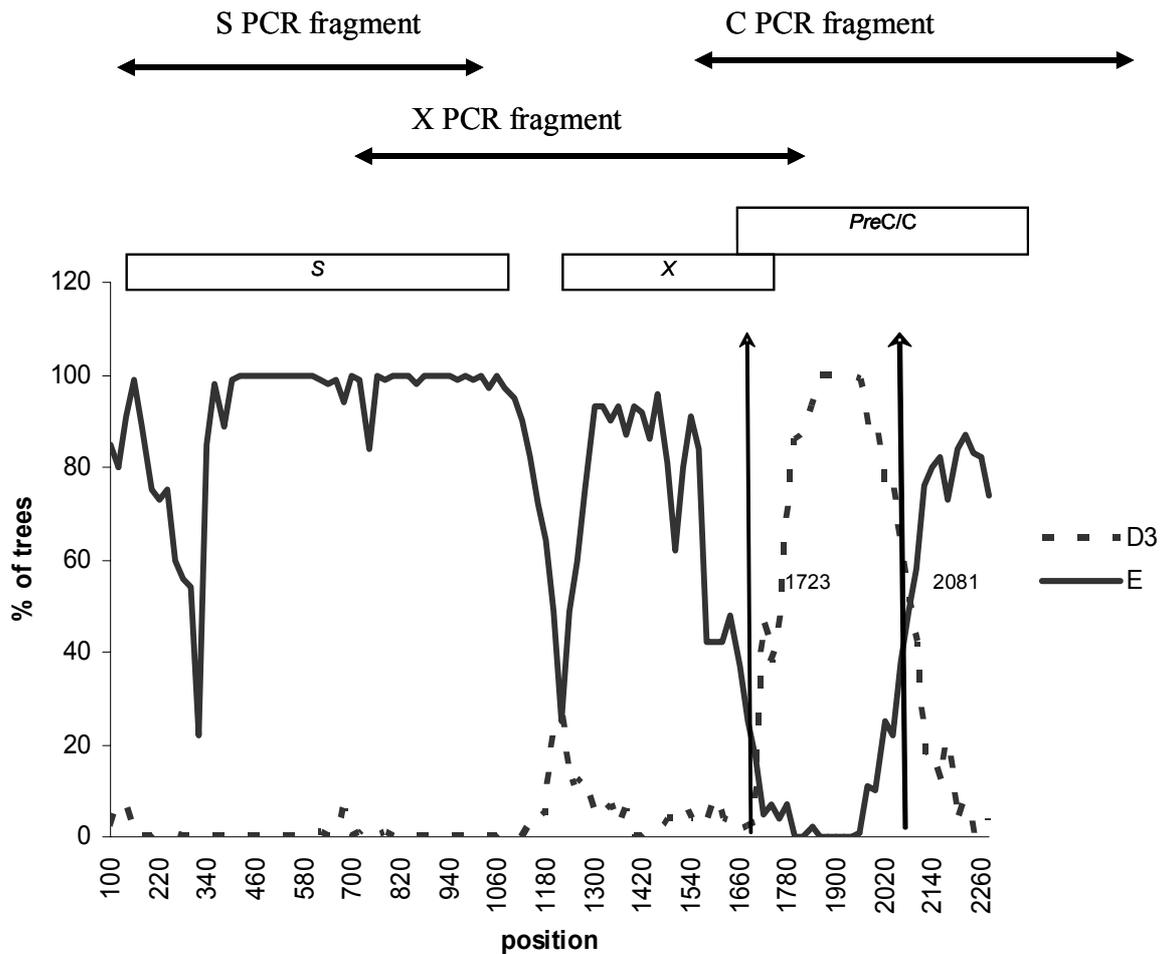
Phylogenetic analyses and distance plots of clones of the CAR177 assigned all S and X PCR fragments to genotype E. Within the preC/C ORF genotype E was partially replaced by a genotype D fragment (358bp). Simplot analysis mapped the E/D and D/E recombination sites to position 1723 and 2081 respectively (nt positions 200 and 558 in the C fragment) corresponding essentially to first third of the 1067 bp long C fragment (Figure 18).

### 3. Discussion

The present study shows that the genotype E crescent previously described to span from Senegal in the Northwest to Angola in the South, extends far enough to the East to include at least the most populated areas of CAR. Based on 30 full-length sequences and 36 partial sequences a prevalence of 94% of genotype E was found. More than 98% of sequences were obtained from HBsAg and anti-HBc positive patients with chronic liver disease. Thus this high prevalence may be biased if genotype E strains are more likely to cause chronic infections or if they have a higher perinatal infectivity. However, similar studies in unselected populations in other sub-Saharan countries such as Benin, Togo, Mali, Gabon and Cameroon, in chronic liver disease patients in Nigeria and Ivory Coast (Odemuyiwa *et al.* 2001; Suzuki *et al.* 2003; Mulders *et al.* 2004; Makuwa *et al.* 2006; Olinger *et al.* 2006) also found very high prevalence of genotype E 43% (Cameroon) to 87.4%; (liver patients in Ivory Coast), although less high than in CAR. The diversity of the genotype E CAR strains is only 1.37% for the full-length genome and these do not increase the overall diversity of all genotype E sequences in the database (1.67%).

Careful analysis of the branching pattern shows that most CAR strains are similar to strains from West African countries. Because of the low genetic diversity and its exclusive endemicity in sub-Saharan Africa we have speculated that genotype E strains emerged from a virus introduced less than 200 years ago (Mulders *et al.* 2004). However, the low diversity and the low bootstrap values do not allow delineating geographic regions from where the virus may have emerged. In addition to genotype E, three genotype D strains were found which according to their branching pattern and genetic distances seem to share a common ancestor with Somalian strains, classified as D4 (Norder *et al.* 2004).

Recombinations between HBV viruses may be a more frequent than initially anticipated (Bowyer and Sim 2000; Hannoun *et al.* 2000; Morozov *et al.* 2000, Sugauchi, 2001 #51; Sugauchi *et al.* 2002). Several hybrid genotypes have been described such as C/D in Tibet (Cui *et al.* 2002) recombination between genotype A and C in Vietnam (Hannoun *et al.* 2000).



**Figure 18 : Similarity plots comparing the S, X and preC/C genes of CAR177**

Comparison was done to the sequence of genotype E (line) and genotype D (interrupted line). Only relevant genotypes are shown. Breakpoints are indicated.

In Nigeria, genotype A (Mulders *et al.* 2004; Olinger *et al.* 2006) recombined with an E/D sequence to a triple recombination. Here a new HBV hybrid of genotype E and D, of the S, X and preC/C gene has been identified. The recombination site was located in the C gene, but it is unclear whether the parent strain belongs to D3 or D4 because of an incomplete sequence of the Somalia D4 strains. This would at present seem like the most probably donor strain of the D/E recombination. As discussed by Bowyer *et al.* (2000) the X/core genes in which we detected the recombination of genotype E and D are very similar suggesting a common precursor. Similar event might have occurred earlier in the history of genotype E now the dominant genotype in West Africa. Interestingly only a single A1 strain was detected that resembled by Kimura distance and branching to those that were most frequently found in

Eastern and South-eastern Africa (Kramvis *et al.* 2002; Sugauchi *et al.* 2002; Hannoun *et al.* 2005). Although only few sequences are available from the latter regions, these seem to be dominated by A1 and D3/D4 strains, while the genotype E extends from West Africa to Central Africa with little overlap of genotypes in CAR.

## Part III

Results of part III were published as:

Olinger CM, Venard V, Njyou M, Oyefolu AO, Maïga I, Kemp AJ, Omilabu SA, le Faou A, Muller CP. (2006) Phylogenetic analysis of the precore/core gene of hepatitis B virus genotypes E and A in West Africa: new subtypes, mixed infections and recombinations. *Journal of General Virology*. 87:1163-1173.

### Phylogenetic analysis of the preC/C gene of hepatitis B genotypes E and A in West-Africa: New subtypes, mixed infections and recombinations

In different cohorts of otherwise healthy children or adults from 7 West-African countries, the Institute of Immunology observed in an earlier study, between 9% and 65% of chronic carriers and up to 100% in HIV patients from Cameroon (Olubuyide *et al.* 1997; Mulders *et al.* 2004). Most of these studies were based on the genetic analysis of the preS/S gene. An intriguing finding, confirmed since also by others, was the conspicuously low sequence diversity of HBV/E of 1.67% in the preS/S gene (Odemuyiwa *et al.* 2001; Mulders *et al.* 2004); (Kramvis *et al.* 2005). Only relatively few preC/C gene sequences of genotype E have been analyzed and genotype E did not seem to separate from genotype D in the X and C ORFs (Bowyer and Sim 2000), raising some initial questions about its existence (Kidd-Ljunggren *et al.* 1995; Kramvis *et al.* 2005). Therefore further studies and sequencing of larger numbers of in particular the C gene have been recommended (Kramvis *et al.* 2005). Here, we report 122 new preC/C sequences from 3 West African countries in order to determine whether the low sequence diversity found in the preS/S gene can also be confirmed for this gene and how these sequences compare phylogenetically with the other genotypes. Cameroon was included in this study to investigate whether the high genotype A prevalence within the genotype E endemic crescent may have led to mixed infections and recombinations. To distinguish between genotype A and nonA, a 6 nt insertion in the former

was exploited (Hannoun *et al.* 2002). Considering the very high prevalence of chronic carriers of HBV/E throughout West-Africa (Ahmed *et al.* 1998; Allain *et al.* 2003), genetic analysis of the preC and C gene may also provide clues to explain this excessive rate of chronicity associated with this genotype.

### 1. Clinical samples

Serum samples were collected from 110 HBsAg positive donors between 1998 and 2004 in three West African countries: Nigeria (Lagos, Ibadan), Mali (Bamako) and Cameroon (North, East, West, South and Central Province). The adults from Nigeria were patients admitted to local hospitals, many of them with liver disease. The donors from Mali were otherwise healthy students. About two thirds of patients were known to be HIV positive or became known during hospitalisation. Blood was drawn after informed consent of donors or their parents or guardians in the case of children. The larger part of the described cohorts had already been included in an earlier study (Mulders *et al.* 2004).

Sera were tested for HBs and HBe antigen using the Murex<sup>®</sup> kits (Abbott Laboratories). HIV infections were confirmed using the Murex<sup>®</sup> HIV-1.2.0 kit (Abbott Laboratories). Serum samples were stored at -80 °C until use.

### 2. Amplification of the preC/C region

A first PCR served as detection PCR for mixed infections (Figure 19b). Whenever both genotype specific reverse primers amplified a PCR product, a mixed infection was assumed. To confirm this and to obtain a larger PCR fragment for sequencing, the second round of the detection PCR was repeated with C1 as the forward primer (Figure 19c). The PCR products obtained (C1/rvA and C1/rvnonA) were purified in a 1% agarose gel and sequenced using the same primers. When after the detection PCR only one of both genotype specific reactions was positive, the sequencing was done with the corresponding reverse primer and the C1 forward primer. In parallel, a genotype-insensitive PCR (Figure 19a) was run on all samples. Technical details of the different PCR reactions are described below.

HBV DNA was isolated from 200 µl of serum using the QIAamp<sup>®</sup> DNA Blood Mini Kit (Qiagen) and eluted in 200 µl volume. Genomic amplification of the preC/C region was

performed by polymerase chain reaction in a semi-nested format using primers PC1 (5'-GGAGACCACCGTGAACGC-3', pos 1610-1627) and C2 for the first round and C1 (5'-CTGGGAGGAGTTGGGGGA-3', pos 1730-1747) and C2 (5'-GTAGAAGAATAAAGCCC-3', pos 2487-2503) (Kao *et al.* 2002) for the second round (Figure 19a). Nucleotides are numbered according to the HBV/E reference strain GenBank X75657 (Norder *et al.* 1994). The first round was performed for 40 cycles (95 °C for 1 min; 50 °C, 1 min; 72 °C, 2 min and a final extension step at 72 °C for 10 min) in a 50 µl reaction volume containing 5 µl of extracted DNA, 1X PCR buffer, 0.2 mM dNTPs, 1.8 mM MgCl<sub>2</sub>, 0.3 µM of each primer and 2 U of Platinum® Taq DNA polymerase (Invitrogen). After the first amplification, 1 µl of the PCR products was reamplified with primers C1 and C2 for another 40 cycles and using the same PCR conditions (Figure 19a).

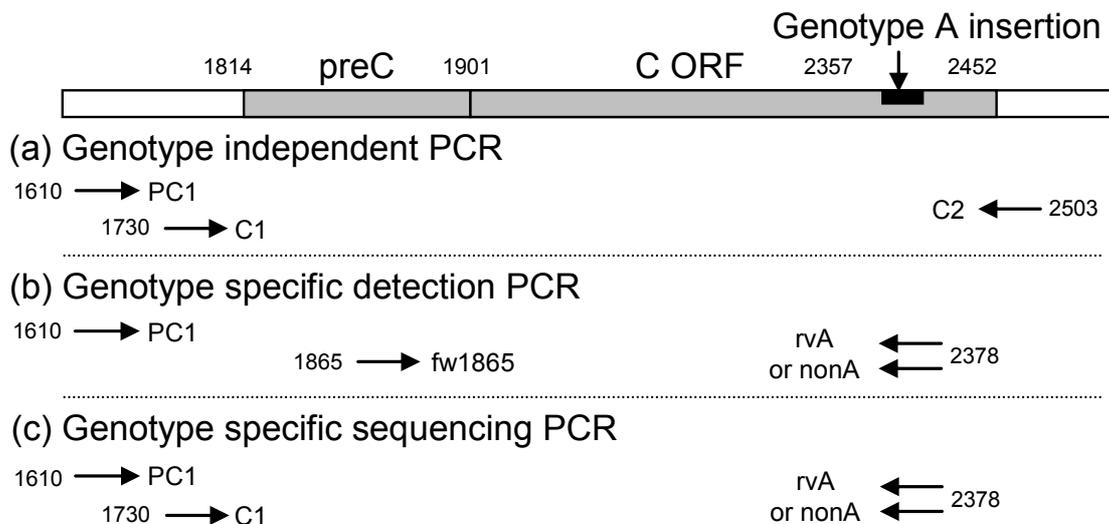
Genotype mixtures were analysed by amplification of the C gene with A-specific and non-A specific primers as previously described (Hannoun *et al.* 2002). In order to increase PCR sensitivity, each sample was tested first in parallel using the forward primer PC1 in combination with either rvA (5'-TTCTTCTTCTAGGGGACCTGCCTCAGTCC-3', pos 2356-2384) or rvnonA (5'- TTCTTCTTCTAGGGGACCTGCCTCATCGT-3', pos 2350-2378). In the second round, PC1 was replaced by fw1865 (5'-CAAGCCTCCAAGCTGTGCCTTGGGTGGCCTT-3', pos 1865-1895) giving a fragment of a maximum of 560 bp (Figure 19b). PCR conditions were as described above except for 30 cycles and an annealing temperature of 58 °C in the first round and 62 °C in the second round. The amplified products were separated in a 1% agarose gel, stained with ethidium bromide. In the case of a positive reaction, the second round was repeated under the same PCR conditions, using C1 as forward primer (Figure 19c).

Complete genome sequencing using four semi-nested PCR reactions, cloning and phylogenetic analysis were done as described in Chapter 3. 180 sequences were obtained either with primers C1C2 (Figure 19a), or C1rvA or C1rvnonA (Figure 19c) and included the entire preC/C gene with a total length of 517-584 bp, depending on the genotype (genome position 1814-2331, numbering according to X75657). In addition, 3 complete genomes and 4 preS fragments (pos 2455-159) were sequenced. Sequences were submitted to the EMBL/GenBank/DDBJ database under accession no.: AM110794-AM110915 for the preC/C gene and AM180623-AM180628 for the complete genome and preS fragment sequences.

### 3. Results

#### 3.1. Screening for mixed infection

All serum samples were screened for mixed infections by genotype A and nonA strains, using two genotype-specific reverse primers (rvA; rvnonA), exploiting the 6-nucleotides insertion in genotype A (Figure 19b) (Hannoun *et al.* 2002). All samples were sequenced with the C1 forward primer and either the reverse primer rvA or rvnonA depending on which reverse primer gave a positive fragment in the *genotype-specific PCR* (Figure 19c). A total of 49 samples were only positive in either the rvA (20 of 49) or the rvnonA detection PCR reaction (29 of 49). Phylogenetic analysis showed that 12 rvA products obtained were of genotype A and 28 rvnonA products of genotype E. Surprisingly 8 rvA products were of genotype E and 1 rvnonA product was of genotype A. The C1C2 sequences showed that in the latter 9 samples binding of the rvA and rvnonA primers was non-specific (Table 12).



**Figure 19 : PCR reactions used to detect mixed infections and for genotype specific PCR**

are semi-nested with different forward primers. Positions are given according to the HBV/E reference strain accession no. X75657.

PCR reactions used to detect mixed infections (b) and to generate genotype-specific (c) and genotype unspecific (a) products for sequencing. All PCRs

Another 33 samples were positive in both genotype-specific detection PCR reactions after two rounds (Figure 19b) giving 66 fragments. Sequencing assigned 13 of 33 rvA products to

genotype A and 31 of 33 rvnonA to genotype E. 22 of the 66 sequences did not correspond to the expected genotype: 2 of the 33 rvnonA products were of genotype A and 20 of the 33 rvA products were identified as genotype E. Thus only 84 of 115 sequences that were obtained by genotype-specific detection PCR were confirmed by sequencing (Table 12). In 11 patients among the above 110, both rvA and rvnonA sequences were of the predicted genotype and for one, the C1C2 sequence was of genotype A, while the rvnonA sequence of genotype E, thus giving a total of 12 mixed infected samples with two distinct sequences available.

One sample (NIE24072) was further analysed by *cloning* the product of the S fragment PCR into a vector. 5 colonies were randomly selected and the plasmid was sequenced using M13 primers. One clone did not contain an insert, 3 clones contained sequences identified as genotype A3 with a divergence of 2.5% among the 3 sequences and one clone contained a genotype E sequence.

**Table 12 : Comparison of the results of the genotype A (rvA) and nonA (rvnonA) specific PCR, with the genotype of the sequence of the PCR fragments as well as the genotype-independent PCR**

Genotype	Genotype-specific PCR				Total
	rvA sequence		rvnonA sequence		
	A	E	A	E	
rvA+ rvnonA-	12	7+1 <sup>*, †</sup>	NA	NA	19+1
rvA- rvnonA+	NA	NA	1 <sup>†</sup>	26 (1) <sup>‡</sup> +2	27+2
rvA+ rvnonA+	13 (11)	20	2	27 (11)+4	62+4
<b>Total</b>	25	27+1	3	53+6	108+7

Genotype	Genotype-independent PCR		Total
	A	E	
Genotype-specific PCR correct	4	23	27
Genotype-specific PCR negative	3	25	28
Genotype-specific PCR false	1 <sup>†</sup>	8 <sup>†</sup>	9
C1C2 of NIE24072	1 (1)		1

\* Numbers after '+' indicate the number of samples identified as mixed infected based on both >5 ambiguous nucleotides found in genotype specific positions in the sequencing electropherograms and an unreadable reverse fragment.

† For these patients, the C1C2 sequence confirmed binding of the irrelevant reverse primer in the genotype-specific PCR.

‡ Numbers in brackets indicate sequences of this category identified from patients with mixed infections from whom both a distinct genotype A and E sequence were obtained.

**Table 13 : Prevalence of single and mixed genotype infections**

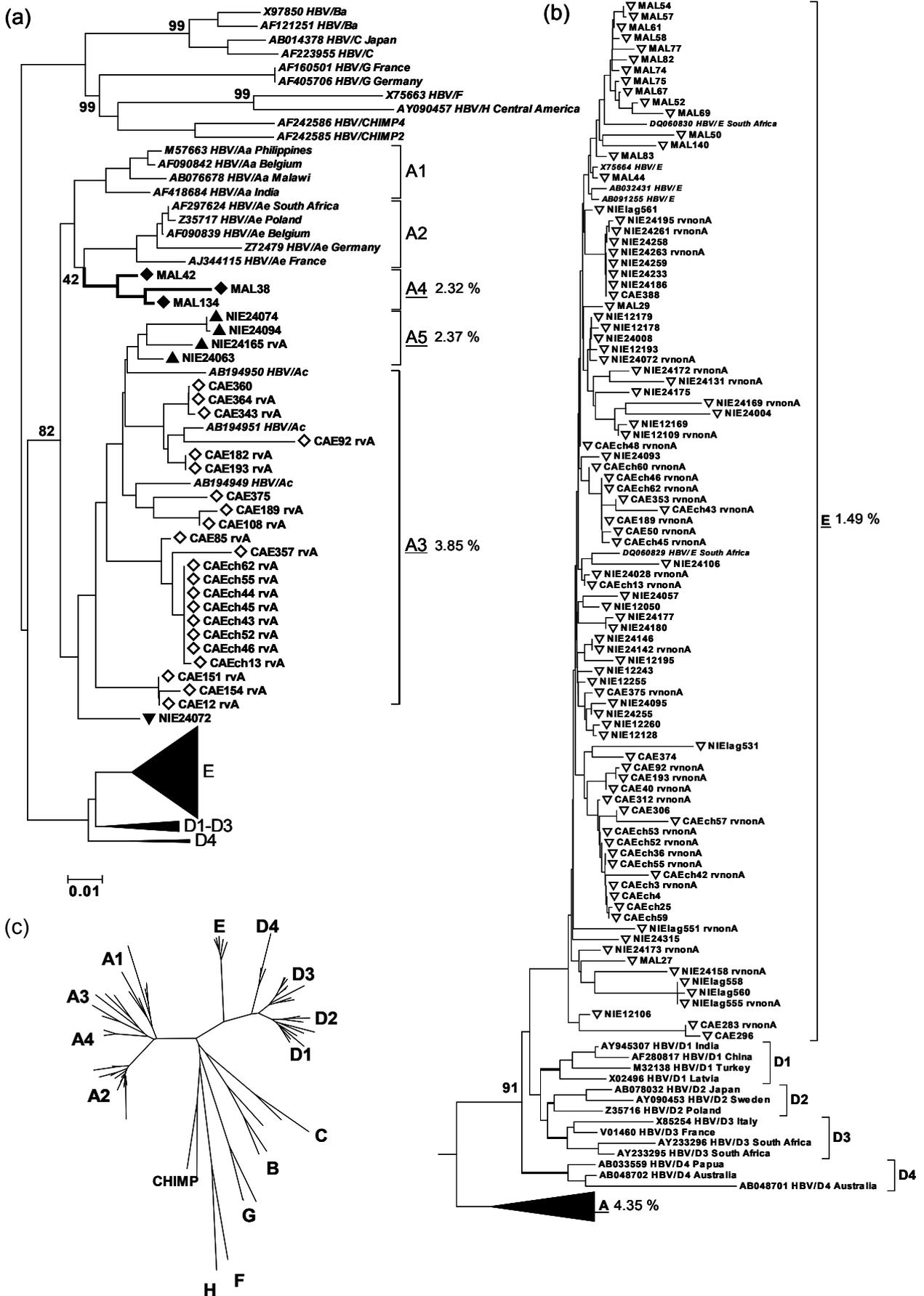
Country	Cohort	Age (yrs)		A	E	A/E*	Total
Cameroon	Adults	18-48	HIV+	10	9	4	23
			HIV-	0	0	0	0
	Children	0.6-5	HIV+	1	6	6+4	17
			HIV-	0	0	1	1
Mali	Adults	22-28	HIV+	1	1	0	2
			HIV-	2	16	0	18
Nigeria	Adults	12-57	HIV+	4	26	1+3	34
			HIV-	0	3	0	3
	Children	6-14	HIV+	0	5	0	5
			HIV-	0	7	0	7
Subtotals			HIV+	16	47	11+7	81
			HIV-	2	26	1	29
<b>Total</b>			HIV+/-	18	73	12+7	110

\* Numbers after '+' indicate the number of samples identified as mixed infected based on both >5 ambiguous nucleotides found in genotype specific positions in the sequencing electropherograms and an unreadable reverse fragment.

The *genotype-independent PCR* (Figure 19a) was positive for 65 samples. For 27, the sequence of this PCR confirmed the sequence obtained by the genotype-specific PCR. For 28 patients, sequences were obtained by the genotype-independent PCR only, while the genotype-specific PCR was negative for both fragments probably because of the lower sensitivity of the latter. For 9 samples, the genotype-independent PCR gave the same sequence as the genotype-specific PCR. However, in these cases the genotype specific sequence was obtained by binding of the false reverse primer (Table 12).

The robustness of the genotype-sensitive PCR was tested using 2 plasmids containing the C fragment (pos 1644-2661) of either genotype A or E. The plasmid concentration of one genotype was kept constant ( $10^3$  copies/ $\mu$ l) and increasing concentrations of the other genotype ( $10^3$  to  $10^9$  copies/ $\mu$ l, ratios of 1:1 to 1000:1) were added. When the irrelevant genotype was present in a sufficient excess (>100 times) over the specific genotype, binding and amplification of the false template occurred, explaining most of the false PCR products.

Figure 20 : Phylogenetic tree of sequences clustering with genotype A (a) or E (b) reference strains



**Legend to Figure 20 (previous page) : Phylogenetic tree of sequences clustering with genotype A (a) or E (b) reference strains**

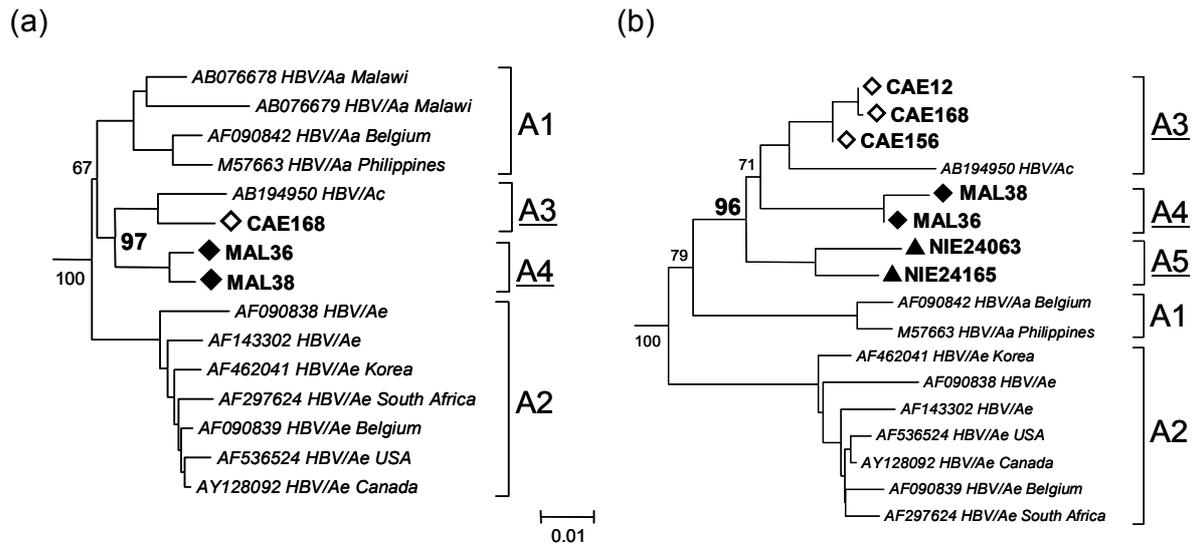
Sequences clustering with subtype A3 are marked as  $\diamond$ , the Mali subtype A4 as  $\blacklozenge$ , the Nigerian subtype A5 as  $\blacktriangle$  and the Nigerian recombinant as  $\blacktriangledown$ . The genotype E sequences are labelled with  $\nabla$ . Bootstrap values of important nodes are indicated as well as the diversities of subgroups. The following abbreviations have been used: MAL: Mali adults; CAE: Cameroon adults; CAEch: Cameroon children; NIE24xxx and NIElagxxx: Nigerian adults and NIE12xxx: Nigerian children. The insert (c) shows a tree calculated using complete genome sequences of references.

---

The *electropherograms* of seven of the forward sequences obtained in the C1/rvA/rvnonA (Figure 19c) and the C1/C2 (Figure 19a) PCR reactions showed distinct peaks in at least 5 genotype specific positions, supposedly because of the presence of both genotypes. The reverse reaction of these seven sequences showed a high background or was not readable, an expected result of a frame shift between the two reverse fragments of A and nonA (i.e. E) starting at the genotype A insertion. These 7 patients were therefore also considered as mixed infected. Thus, the sole detection of both sequences may lead to an underestimation of the occurrence of mixed infections (Table 12).

### 3.2. Genotype prevalence and mixed infections

122 sequences from 110 patients from HIV-negative and HIV-positive children and adults were further analysed (Table 13). Ninety-two of 110 patients were (co)-infected with genotype E and this genotype was also the most prevalent in each of the three countries. Genotype A was the only other genotype detected in the three countries (Nigeria 8/49, Mali 3/20 and Cameroon 26/41 patients). This genotype was most dominant in HIV-positive donors from Cameroon. Mixed infections of genotype E with genotype A were essentially limited to Cameroon and Nigerian HIV-positive donors (18 of 19).



**Figure 21 : Phylogenetic tree of genotype A complete genome sequences (a) and preS fragment sequences (pos. 2455-159)**

Sequences clustering with subtype A3 are marked as ◇, the Mali subtype A4 as ◆ and the Nigerian subtype A5 as ▲. Bootstrap values of important nodes are indicated. The following abbreviations have been used: MAL: Mali adults; CAE: Cameroon adults and NIE24xxx Nigerian adults.

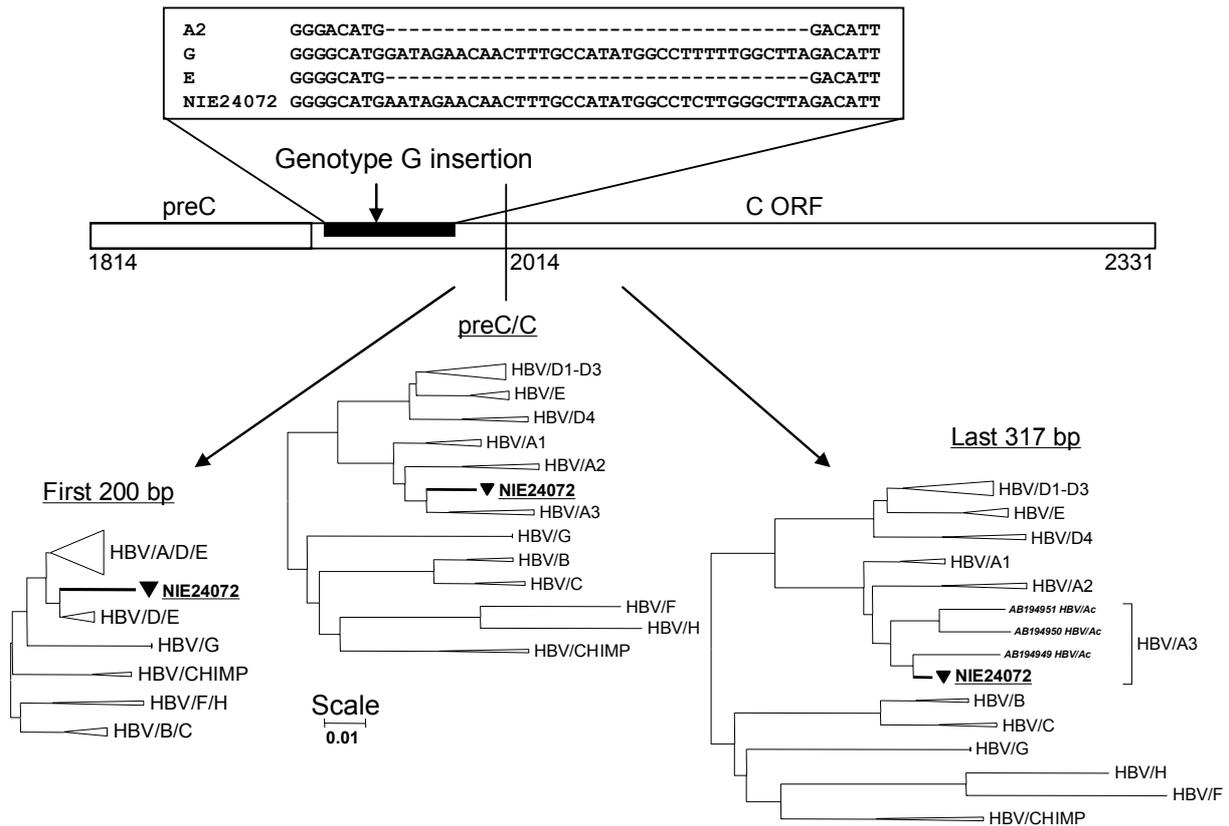
### 3.3. Genetic variability of genotypes and subtypes

With a maximal diversity of 9.43%, genotype A sequences found in the three countries were highly diverse. Phylogenetic analysis differentiated 3 distinct subtypes. The A sequences from Cameroon clustered with the recently defined subgroup A<sup>\*/</sup>Ac/A3 (Figure 20a). The bootstrap value separating this A3 lineage from the A1 and A2 subtypes was 82%. The genetic distance between our A3 strains and the European/American A2 subtype (5.88%) is smaller than the distance to the African/Asian A1 subtype (5.97%) although this was not the case for the complete genome sequences (Figure 21a). The 3 genotype A sequences from Mali did not cluster with the subtypes A1 to A3 but formed a cluster of their own tentatively designated as subtype A4, with an average distance of 5.41% (4.93% to A2 and 5.86% to A3) to the other A subtypes, although the bootstrap value was relatively low (42% of calculated trees). The diversity among these Mali sequences was large (2.32%) in comparison to the Mali E sequences (1.47%). For two A4 strains from Mali (AM110795 and AM180623) and one A3 strain from Cameroon (AM180624) complete genome sequences were obtained (Figure 21a). The node separating the A3 complete genome

to the A3 genome was 6% (divergence of 0.7%). Four Nigerian genotype A sequences formed another group of their own, (tentatively designated as A5), with an average diversity of 2.37% and an average distance to the other A subtypes of 5.11% (4.47% to A3 and 5.32% to A2). A4 and A5 were separated by an average genetic distance of 5.05%. For two A5 (AM180625 and AM180626) and two A3 strains (AM180627 and AM180628), only the preS fragment (in addition to the preC/C fragment) was obtained. The bootstrap value separating the A5 from the A3 group was 96% (Figure 21a). For the A5 preS sequences a distance to the A3 sequences of 4.3% was calculated. The diversity between A4 and A5 was 5.2%. The divergence within A4 and A5 was 0.7% and 2.4%.

The diversity of the genotype E sequences was 1.41%, which was significantly lower ( $P < 0.001$ ) than the average diversity of 2.85% found within the genotype A subtypes A3 to A5. The topology of the tree as well as the diversity calculations showed that the HBV/E preC/C sequences are closest to HBV/D1 and D2 with an average distance between E and D1-D4 of 4.8-5.9% versus 9.2% to the A3 subtype (Figure 20b).

Identical genotype E sequences were found in patients from distant provinces in Nigeria and Cameroon (e.g. Abia, NIE24072; Ogun, NIE24008; Imo NIE24146; Kogi NIE24142; Northern Province, CAE12 and Central Province, CAE151). Identical sequences were also found in the different countries such as East-Cameroon (CAE388) and the province of Oyo in Western Nigeria (NIE24233). In addition, 7 genotype A strains obtained from the Cameroon children cohort showed 100% identical sequences. When the patients were divided into three age groups (<14, 14-30 and >30 years), the ratio of E to A sequences in the three groups decreased: 3.3; 3 and 2.4, at least partially explaining the significant ( $P < 0.001$ ) increase in genetic diversity, when all sequences were combined (0.67%, 2.1% and 3.2%). Also, within both genotypes the diversity tended to increase with age, but significance was lost.



**Figure 22 : Phylogenetic analysis of the NIE24072 outlier based on the first 200 bp (left), the complete preC/C gene (centre) and the last 317 bp (right)**

The scale is the same as for the trees of Figure 20. Genotypes A, D and E group on the same node when the phylogenetic tree is constructed using only the first 200 bp.

### 3.4. Mutations in the preC/C gene

A summary of specific mutations is shown in Table 14. Four patients, all from Cameroon were infected with genotype A strains displaying a mutation in the start codon (A1814C or A1814T) of the preC ORF. Mutations in positions G1896A or G1897A were identified in 34 patients, primarily in genotype E sequences (32 of 34). In 18 cases both the wild-type and the mutation were found to coexist in position 1896 (10 quasi-species and 8 mixed infected patients). The latter mutation was frequently found (7 of 14) together with G1899A. These mutations introduce a new stop codon (TGG to TAG or TGA), which causes a premature non-functional HBeAg. Whenever G1896A is associated with the double mutation A1850T and C1858T, it causes an increase in stability of the pregenomic RNA encapsidation signal. This was the case in 31 of 88 sequences, all of genotype E. In two genotype A sequences a C1857T

mutation was found and the same 2 sequences had G1897A. The A1850T mutation, normally restricted to nonA genotypes, was also found in a genotype A strain from Mali. Interestingly, the G1896A mutation was 3 times more frequent in HIV infected patients (34.2% versus 11%). While in Nigeria, this mutation was restricted to adults; in Cameroon it was 4 times more frequent in children than adults (48% to 11%). In two patients, additional rare stop codons were identified: C1817T in the preC and G2262T in the core region. Finally, all identified A genotypes except two (NIE24063, MAL42) presented a C1862, found to be very rare in any of the genotypes A-H. No significant difference in the mutation pattern was found between sequences obtained from mixed or single infected patients.

### 3.5. Insertions and deletions in the preC/C gene

The *rvA* sequence of a mixed infected patient from Nigeria (AM110794), had an insertion of 36 bp at the beginning of the C ORF (position 1903) which differed only by 3 non-synonymous nucleotides from a 36 bp long C gene fragment of genotype G located in the same nucleotide positions in both the genotype G reference sequence and NIE24072 (Figure 22). A BLAST search of this insert showed that a similar sequence was not found in any other sequence in GenBank and that the closest other match has only 19 of 36 bp identity. The first 200 bp of this strain clustered with genotype E as well as genotypes A and D whereas the last 356 bp clustered with the HBV/A3 sequences and all other genotypes were grouped correctly (Figure 22). However, the first 200 bp upstream of the insert contain two mutations specific to nonA genotypes (positions: 1850 and 1858), while further downstream of the insertion, many typical genotype A mutations were identified. Five patients presented deletions in the C ORF (Table 14): one from Mali (2008-2054), two from Nigeria (2121 and 2172-2307), the latter being mixed infected and two with single nucleotide deletions from Cameroon (nt 2170; nt 2250). All these deletions cause a disruption of the reading frame of both the HBeAg and the core protein.

Table 14 : Specific mutations in the preC/C gene of 29 genotype A and 89 genotype E sequences

<b>Mutation</b>	<b>A</b>	<b>E</b>
<b>HBeAg start codon mutation</b>		
A1814C/A1814T	4	/
<b>HBeAg stop codon apparition pgRNA encapsidation signal</b>		
C1817T	/	1
A1850T	1	88
C1857T	2	/
C1858T	/	89
G1896A (A+G)	/	14 (18)
G1897A	2	/
C1858T and G1896A(A+G)	/	13 (18)
A1850T and C1858T	/	88
<b>Characteristic mutation</b>		
G1862C	26	/
G1862T	1	/
<b>Frequent mutation</b>		
G1899A(A+G)	3 (1)	6 (4)
G1896A and G1899A (A+G)	/	3 (4)
<b>Deletions in core protein</b>		
46 bp (2008-2054)	/	1
1 bp (2121)	/	1
1 bp (2170)	1	/
135 bp (2172-2307)	1	/
1 bp (2250)	1	/
<b>Core/HBeAg stop codon apparition</b>		
G2262T	/	1

The preC sequences of one genotype A and 3 genotype E strains were not entirely readable and were therefore excluded from this table.

### 3.6. HBeAg status and mutations

Forty-nine patients, for which there was enough serum, were tested for HBeAg, including 3 mixed infected with 2 sequences and 2 with only one sequence available. One of two sequences from a mixed infected HBeAg negative patient had the 1814 start codon mutation whereas the other sequence had no mutation affecting the HBeAg expression. 5 sequences from single infected patients had mutations at 1817, 1896 or 1897 (quasi-species in these positions was considered as wildtype). Two of the 5 patients with mutations at 1896/1897 and the patient with the C1817T mutation were HBeAg negative. Of the patients with wild-type nucleotides in these 3 positions, 46 were HBeAg positive (including 4 mixed infections), 6 HBeAg negative (including 1 mixed infection). In all the mixed infected cases, the second sequence, when available, was also of wildtype. Three patients had deletions: The one with the single nucleotide deletion at nt 2170 was HBeAg negative, the patients with the deletion at 2121 and 2172-2307 were associated with an HBeAg positive status.

## 4. Discussion

The present survey presents 122 new preC/C sequences from three major countries in West-Africa as well as 3 complete genome sequences and 4 preS regions of HBV. As expected from previous studies from ourselves and others (Norder *et al.* 1994; Odemuyiwa *et al.* 2001; Suzuki *et al.* 2003); (Liu *et al.* 2001; van Steenberg *et al.* 2002); (Baptista *et al.* 1999; Owiredu *et al.* 2001; Suzuki *et al.* 2003), the majority of the sequences from this region were of genotype E. Genotype A was the only other genotype found in the region. Although in our cohorts HIV-positive donors dominated the high prevalence of genotype A in Cameroon does not seem to be limited to HIV patients (Mulders *et al.* 2004; Kurbanov *et al.* 2005). Despite the large distances between the different collection sites (up to 3000 km) the 92 genotype E preC/C sequences showed very little genetic diversity irrespective of origin and other patients' characteristics. Similar to the genetic diversity reported earlier for the preS/S gene (1.67%, (Mulders *et al.* 2004) or 1.54 % all preS/S sequences included) the diversity of our 92 sequences was 1.41% or 1.59% when all preC/C sequences from GenBank were included. When the preS/S and preC/C genes of the same 45 strains were compared, the C gene sequences were 1.5 and 2.4 times more diverse than the S genes for genotypes A and E respectively. Similarity plots showed that similar to the preS/S gene, the preC/C gene was

most closely related to genotype D in particular the subgenotype D1 and D2 (about 5% compared to >9% for the other genotypes). Two single-infected samples seemed to be the result of an A/E recombination as the preS/S sequences was of genotype A (Mulders *et al.* 2004) and the preC/C of genotype E (e.g. AJ605037 and AM11089).

All genotype A sequences from Cameroon were identified as A3, and none grouped with the African/Asian A1 or European/American A2 subtype (Bowyer *et al.* 1997; Hasegawa *et al.* 2004). In our earlier study, the preS/S sequences of several of these strains had been assigned to a group A'' (Mulders *et al.* 2004), which has now been renamed as A3 (Hannoun *et al.* 2005; Kurbanov *et al.* 2005). Our results confirm that subtype A3 is somewhat closer to the European A2 subtype (average genetic distance: 5.88%) than to the African/Asian A1 (average genetic distance: 5.97% (Hannoun *et al.* 2005)). Instead of the G1862T mutation in the precore gene, thought to be characteristic for A3 strains (Hannoun *et al.* 2005) G1862C was predominant in our sequences. All strains were of the A2-type in position 1888. The genetic diversity of subtype A3 is 3.85% compared to the 2.19% of subtype A2. Thus A2 may have a shorter evolutionary history than A3 and may be a more recent progeny of A3.

The Mali genotype A strains are phylogenetically distinct from the other A subtypes (average distance of 5.4%). With a genetic distance of 4.93%, they are closest to the A2/Ae subtype but warrant a new subgenotype tentatively designated A4 according to the recommendation of >4% genetic distance between subgenotypes. Two complete genome sequences of this subtype corroborated the phylogenetic difference with the other subtypes with a bootstrap value of 97%.

The A sequences from Nigeria formed a cluster of their own with an average genetic distance of 4.47% and a maximum distance of 6.5% from the Cameroon A3 strains, complying with the proposed criteria (genetic distance >4%) for a new genotype subtype, tentatively designated A5. Phylogenetic analysis of two preS sequences further corroborated the definition of this new subtype. Assuming an estimated mutation rate of  $4.2 \times 10^{-5}$  per nt and per year (Okamoto *et al.* 1987; Orito *et al.* 1989; Hannoun *et al.* 2000) the two strains would have taken 500 years to evolve from a common ancestor. The Nigerian strain NIE24072 shows evidence of a triple recombination of a nonA (E/D) sequence and an A3 sequence separated by a G specific insert (Figure 22). Although genotype G has so far not been reported from

Africa, the insertion in some of the quasi-species sequences suggests that the patient must have been in contact with a genotype G strain.

Considering the co-existence of genotype A and E in West Africa, the frequency of co-infections is of interest. Mixed infections were identified by genotype-specific PCR in a number of patients, but sequences of both genotypes were confirmed only in 12 cases. 7 additional cases were identified by >5 ambiguous nucleotides in genotype-specific positions, paired with an unreadable reverse sequence (caused by the genotype A specific insertion). Despite these additional criteria, false positive mixed infections are unlikely and the 17.3% mixed infections may still be a slight underestimation. 37% of patients from Cameroon were co-infected with both genotypes and 79% of these were children. As expected a lower rate of co-infection was detected in Nigeria. All but one mixed-infected patient were HIV-positive, but this may be biased by the large number of HIV-positive donors from Cameroon. Thus mixed infections seemed to be frequent when (these) two genotypes co-circulate. Together with earlier studies by us and others in Cameroon (Mulders *et al.* 2004; Kurbanov *et al.* 2005), this suggests that both genotypes are present in the population independent of HIV status and that mixed infections are not restricted to HIV positive patients.

Considering the low genetic diversity of the most prevalent genotype E and its virtual absence in the Americas, we suggested a short evolutionary history and a recent introduction into humans (Mulders *et al.* 2004). This, however, is in contrast with the excessively high endemicity of acute and chronic hepatitis B infection throughout West-Africa. Perinatal transmission is thought to be the most frequent cause of chronic infection in African children (Edmunds *et al.* 1996). The risk of becoming a chronic carrier is especially high for children born to HBeAg positive mothers (Beasley *et al.* 1981; Beasley *et al.* 1981; Thomas 1982; Chu *et al.* 1985). 80% of the donors tested were HBeAg positive with a clear genotype bias: 35 of 39 (89.8%) genotype E carriers but only 2 of 6 (33.3%) genotype A carriers were HBeAg positive. If this can be confirmed in a larger study, it could partially explain the high(er) prevalence of genotype E. The HBeAg negativity associated with A and E sequences correspond to a number of mutations in preC/C sequences believed to affect HBeAg expression (1814, start codon mutation; 1896 and 1897, encoding a new stop codon), but in more than half of cases both wild-type and mutated nucleotides co-existed as quasi-species or mixed infections, explaining the expression of HBeAg. The 3 HBeAg positive cases, where

only mutant strains are detectable, wild-type strains may have been missed. HBeAg negative cases without the above mutations warrant the analysis of the core promoter region.

In all cases but one, the G1896A mutation was associated with the A1850T/C1858T double mutation. Only two sequences presented G1897A encoding a stop codon and for both a C1857T mutation was detected instead of the double mutation. This might be related to the stem loop formation in this region and the pregenomic RNA encapsidation signal, but the clinical significance of this is not clear.

This study confirms that genotype E is dominant throughout West-Africa and presents a low but 2 – 3 times higher diversity on the preC/C gene than on the preS/S gene supporting our earlier speculation of a short evolutionary history of this genotype and a recent introduction into humans of this genotype. The high prevalence of genotype E may be partially explained by the 2 – 3 fold higher rate of HBeAg positivity in comparison to genotype A. In contrast to genotype E, the diversity of genotype A suggests several new subtypes. In Cameroon, where genotypes A and E co-circulate, the frequency of mixed infections is high and A/E recombinations seem to have occurred. Recombinations with other genotypes or subtypes could partially explain the current genetic diversity.

## Part IV

Results of part IV were published as:

Olinger CM, Lazouskaya NV, Eremin VF and Muller CP. (2008) Multiple genotypes and subtypes of hepatitis B and C viruses in Belarus: similarities with Russia and West European influences. *Clinical Microbiology and Immunology*. In press.

### Multiple genotypes and subtypes of HBV and HCV in Belarus: Similarities with Russia and West-European influences

With an estimated 400 million people chronically infected with hepatitis B virus (HBV) and 170 million with hepatitis C virus (HCV) (Madhava *et al.* 2002; Norder *et al.* 2004), these infections represent major public health problems. There is increasing evidence that the risk of developing severe liver injury and the response to antiviral treatment is influenced by the genotypes and subtypes involved (Nolte 2001; Fung and Lok 2004; Yoo *et al.* 2005).

The hepatitis B virus has been phylogenetically classified into eight genotypes (A-H) with different geographic distributions (Okamoto *et al.* 1988; Norder *et al.* 1992). At least six major genotypes of HCV (1-6), with more than 60 subtypes, have been identified worldwide (Bukh *et al.* 1995; Schreier *et al.* 1996; Zein and Persing 1996). HCV subtypes 1a, 1b, 2a, 2b, 2c, and 3a are responsible for more than 90% of infections in North and South America, Europe and Japan. In the United States, genotypes 1a and 1b account for approximately 35% each of the infections (Zein *et al.* 1996) while in Japan, 73% are caused by the latter subtype (Takada *et al.* 1993). Subtype 3a is dominant in intravenous drug users in Europe and the United States (Pawlotsky *et al.* 1995). Genotype 4 is prevalent in North Africa and the Middle East (Chamberlain *et al.* 1997; Abdulkarim *et al.* 1998), and genotypes 5 and 6 are found in South Africa and Hong Kong, respectively (Cha *et al.* 1992; Simmonds *et al.* 1993).

Until now, little information on the genotype distribution of HBV and HCV is available for some parts of the world, especially for developing countries as well as for some of the former Soviet Republics. The Republic of Belarus, reports a seroprevalence of 4.8% for the hepatitis B and 1.26% for the hepatitis C virus but little is known about the molecular characteristics of the circulating viruses. Here, we report the molecular characterization of HBV and HCV strains from Belarus.

## 1. Clinical samples

A total of 157 serum samples were obtained in 2004-2005 from patients with known HBsAg and anti-HCV status from the Infectious Disease Hospital in Minsk. The residents of Minsk correspond to about 25% of the country's population and are representative of the whole population. In addition, the infectious disease hospital is the largest hospital in Belarus which as a central referral hospital attends to patients from both the capital and the rest of the country. HBV and HCV were confirmed with the Murex<sup>®</sup> HBsAg kit version 3 (Abbott Diagnostics) and the Ortho<sup>®</sup> HCV 3.0 ELISA test system (Ortho-Diagnostics). HBV DNA and HCV RNA were detected by RT-PCR and/or PCR (Table 15). HIV co-infection was diagnosed in 51% of all donors or 8.7% of HBsAg carriers and 69.9% of anti-HCV positive donors. 64.6% of the latter were HCV RNA positive.

## 2. Amplification and cloning of HBV and HCV

Nucleic acid extraction and complete HBV genome amplification were done as described in Chapter 3. For amplification of the core/E1 region of HCV a semi-nested PCR was carried out in a 25 µl reaction volume consisting of 0.5 µl cDNA, 2.5 mM of MgCl<sub>2</sub>, 200 nM of dNTPs, 50 nM of each primer (fw290utr(+), TGCCTGATAGGGTGCTTGCGAG, pos: 290-311; 1321e1(-), ACCAGTTCATCATCATATCCCATGCCAT, pos: 1293-1320), and 1U of Platinum<sup>®</sup> Taq DNA polymerase with 1x PCR buffer. After an initial denaturation step at 95 °C for 5 min, 40 cycles of PCR at 95 °C for 1 min, 63 °C for 1 min and 72 °C for 1 min were carried out. Nested PCR was performed with 5 µl of the first-round product diluted 1:100 using the same conditions but with a different forward primer (fw480c(+), CGCGCGACTAGGAAGACTTC, pos: 480-499; rv1321e1, 0.10 µM of each) a MgCl<sub>2</sub> concentration of 2 µM and an annealing temperature of 62°C. The product of the first round S

fragment PCR was cloned using the pCR<sup>®</sup>4-TOPO<sup>®</sup> Cloning kit (Invitrogen) as described in Chapter 3.

### 3. Sequencing and phylogenetic analysis of HCV

The nested and M13 PCR products were purified and sequenced as described in Chapter 3. Phylogenetic analysis of HCV sequences included the core to E1 region with a total length of about 844bp (pos: 480-1323, according to HCV H77). For each HBV strain phylogenetic analysis was done for the combined preS and S fragment sequences (pos: 56-2423, according to X75664) as well as the full length genome if available. All sequences were submitted to EMBL/GenBank/DDBJ under accession numbers: EU414031-EU414184.

## 4. Results and discussion

### 4.1. Hepatitis B virus

Forty-four patients (28%) were positive for HBsAg only, 88 (56%) for anti-HCV antibodies only and 25 (15.9%) for both serological markers. For HBV, less than 10% are HIV positive. Similar values have been found in other countries (Mathur *et al.* 2002). 37/44 HBsAg positive samples (84.1%), 1/88 anti-HCV positive samples (1.2%) and 5/25 double-positive samples (20%) were positive in at least one of the 4 nested PCRs covering the preS, S, X and C genes of HBV and were sufficient for genotyping (Table 15). One patient was HBsAg negative but HBV DNA positive. For a total of 12 strains complete genomic sequences were obtained. Phylogenetic analysis based on the preS or on the complete genome sequences (Figure 23) showed that 38 patients were infected by genotype D (88.4%) and 5 by genotype A2 (11.6%) (Figure 23). Despite the co-circulation of multiple subtypes in Belarus, cloning experiments revealed no evidence of mixed infections or recombinations. Only recently subtypes of genotype D were recognized in phylogenetic studies (Norder *et al.* 2004). Our strains from Belarus segregated into 4 subtypes (Table 16): D1 (n=5), D2 (n=25), D3 (n=7) and D4 (n=1) (Figure 23). There was no significant difference in genotype or subtype distribution between HIV positive and negative patients. Three complete genome sequences were obtained of each of the identified subtypes except for D4, for which the fragments from nt 198 to 2418 were missing.

**Table 15 : Comparison of PCR results with HBsAg and anti-HCV ELISA results**

		<b>pos</b>	<b>neg</b>	<b>Total</b>
<b>HBV DNA</b>				
<b>HBsAg</b>	<b>pos</b>	42 (26.7)*	27 (17.2)	<b>69 (43.9)</b>
	<b>neg</b>	1 (0.6)	87 (52.2)	<b>88 (56.1)</b>
	<b>Total</b>	<b>43 (27.4)</b>	<b>114 (72.6)</b>	<b>157</b>
<b>HCV RNA</b>				
<b>anti-HCV</b>	<b>pos</b>	78 (49.7)	35 (22.3)	<b>113 (72)</b>
	<b>neg</b>	0	44 (28)	<b>44 (28)</b>
	<b>Total</b>	<b>78 (49.7)</b>	<b>79 (50.3)</b>	<b>157</b>

\*Numbers in brackets indicate percentages.

Analysis of sequences available from GenBank/DDBJ/EMBL revealed that countries North (Estonia, Latvia; (Tallo *et al.* 2004)) and South (Hungary, Serbia; (Lazarevic *et al.* 2007; Szomor *et al.* 2007)) of Belarus display a similar genotypic distribution of genotypes A and D with prevalences of genotype D ranging from 71% to 82% and genotype A ranging from 18% to 28%. In Estonia, Latvia and Hungary D2 was the most prevalent subtype (50.8% to 66.6%) followed by D3 and D1. Further South in Serbia, subtype D3 was dominant (43.2%) followed by D2 (32.9%). To the West of Belarus, such as in Poland (86.5% versus 13.5%; (Dzierzanowska-Fangrat *et al.* 2006)) and the Czech Republic (73% versus 27%; (Krekulova *et al.* 2003)) genotype A becomes dominant over genotype D. No sequence data was available from the latter countries. In contrast, in Russia genotype D is dominant (Flodgren *et al.* 2000) with more than 90% prevalence consisting mainly of subtype D2 (80.6%) followed by D3 (12.9%) and D1 (3.2%). Thus, similar to the Baltic States, the genotype distribution found in Belarus resembles the one in Russia although not without clear influences from Western Europe. From Lithuania, Romania, Ukraine and Bulgaria no HBV genotype data was available.

Table 16 : Genotype distribution of HBV and HCV in Belarus

HBV genotype	Patients	D1	D2	D3	D4	A2	Total
HBsAg+	43	4 (11.1)*	21 (58.3)	6 (16.7)	1 (2.8)	4 (11.1)	36
HBsAg/anti-HCV+	20	0	1 (100)	0	0	0	1
HBsAg/HIV+	1	0	1 (100)	0	0	0	1
HBsAg/ anti-HCV/HIV+	5	1 (25)	2 (50)	0	0	1 (25)	4
anti-HCV+/HIV+ and -	88	0	0	1 (100)	0	0	1
<b>Total</b>	<b>157</b>	<b>5 (11.6)</b>	<b>25 (58.1)</b>	<b>7 (16.28)</b>	<b>1 (2.3)</b>	<b>5 (11.6)</b>	<b>43</b>

HCV genotype		1a	1b	3a	4a	4d	Total
anti-HCV+	14	0	9 (81.8)	2 (18.2)	0	0	11
anti-HCV/HBsAg+	20	0	10 (62.5)	6 (37.5)	0	0	16
anti-HCV/HIV+	74	4 (8.1)	21 (42.9)	22 (44.9)	1 (2.0)	1 (2.0)	49
anti-HCV/HBsAg/HIV+	5	0	2 (100)	0	0	0	2
HBsAg+/HIV+ and -	44	0	0	0	0	0	0
<b>Total</b>	<b>157</b>	<b>4 (5.1)</b>	<b>42 (53.8)</b>	<b>30 (38.5)</b>	<b>1 (1.3)</b>	<b>1 (1.3)</b>	<b>78</b>

\* Numbers in brackets indicate percentages.

Since D subtypes have so far been described only rarely (Norder *et al.* 2004; Bozdayi *et al.* 2005; Sunbul and Leblebicioglu 2005) we tested whether the Belorussian strains would comply with the proposed criteria for the definition of new subtypes. Sequence analysis of complete genomes revealed an average intrasubtypic genetic distance of 1.5% within the genotype D subtypes (D4 excluded), while the maximal intersubtypic genetic distance (D4 excluded) was 2.9% with D3 being the most divergent from the other two subtypes (2.7% from D1 and 2.9% from D2). The average distance between D1 and D2 was 2.2% (Table 17). The intersubtypic distances of all available strains are actually lower than the minimal 4% distance proposed to define subtypes (Norder *et al.* 2004) but phylogenetic reconstructions allow nevertheless an unambiguous separation with bootstrap values of 100% at the nodes separating the different subtypes on the phylogenetic tree calculated tree using complete genome sequences (Figure 23). Subtypes D1 and D2 are phylogenetically close to each other and are thought to have evolved in Europe and Western Russia (Norder *et al.* 2004). Subtype D3 originated probably in Russia and spread to Japan (Norder *et al.* 2004; Michitaka *et al.*

2006). Subtype D4 is predominantly found in Australia and Papua New Guinea and is also the genetically most distant from the other D subtypes. Blast searches of complete genomes, identified the following closest known relatives with known origins: D1: AB222711 from Uzbekistan; D2: Z35716 from Poland; D3: DQ111987 from Mongolia and A2: AB116079 from Japan. A blast search based on the S gene of D4 returned AB033559 from Papua New Guinea.

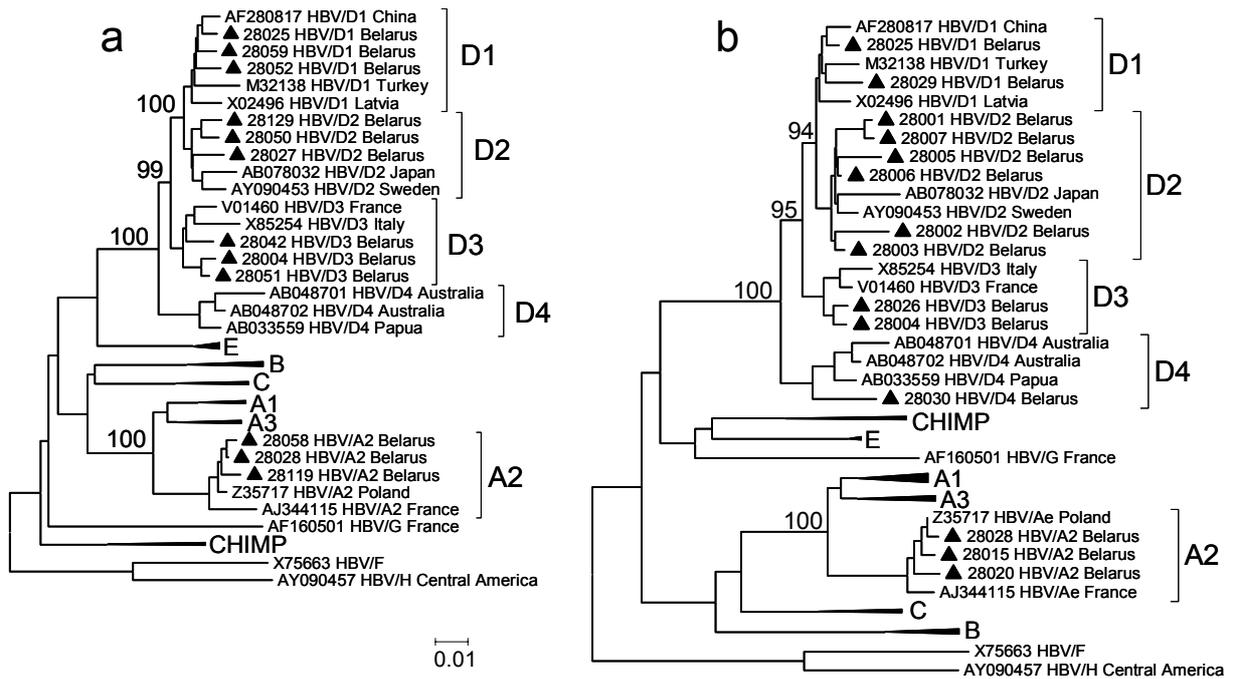
**Table 17 : Distances (in %) within (bold) and between the different genotypes and subtypes present in Belarus**

preS	A2	D1	D2	D3	D4
A2	<b>0.9 (1.04)</b>				
D1	10.5 (17.52)	<b>1.3 (1.41)</b>			
D2	10.6 (17.52)	2.2 (2.78)	<b>1.7 (1.98)</b>		
D3	10.5 (17.73)	2.7 (3.18)	2.9 (3.75)	<b>1.5 (1.48)</b>	
D4	na (17.68)	na (5.09)	na (5.77)	na (5.14)	<b>na (na)</b>

Calculations are based on complete genome or preS fragment (number in brackets) sequences.

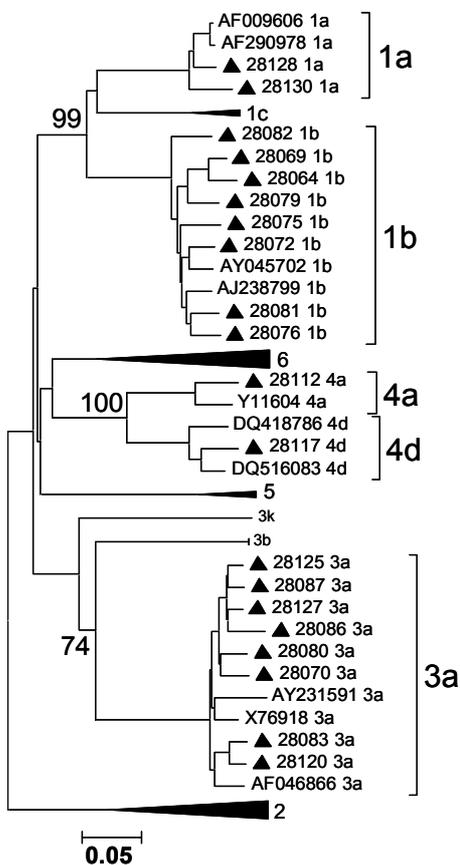
#### 4.2. Hepatitis C virus

Almost 70% of HCV patients were coinfecting with HIV, a number which is similar to the one found in other countries (Ramos *et al.* 2007). In 60 of the 88 anti-HCV positive only and 18 of the 25 HBsAg/anti-HCV double positive patients (Table 16), HCV RNA was detected and sequenced in the core/E1 region (Tables 15 and 16). Phylogenetically the 78 strains belonged to hepatitis C subtypes 1b (53.8%), 3a (38.5%), 1a (5.1%), 4a (1.3%) and 4d (1.3%) (Figure 24 and Table 16). A study from 1997 (Viazov *et al.* 1997) reported subtype 1b to be the most prevalent subtype (76%) followed by subtype 3a (19%) and single cases of 1a and 2a which suggests that the prevalence of subtype 3a is increasing and less subtype 1b strains are circulating. Genotypes 1b and 3a are also dominant over other genotypes in Russia, Estonia and Western Europe (Hraber *et al.* 2007). While in Russia both genotypes have a similar



**Figure 23 : Phylogenetic reconstruction of the preS fragment (a) and complete genome (b) of HBV strains found in Belarus (indicated by ▲) and reference sequences**

Bootstrap values of relevant nodes are indicated. Trees have been calculated using the neighbour-joining method and the Kimura 2 parameter model.



**Figure 24 : Phylogenetic reconstruction of the core/E1 region of the HCV strains found in Belarus (indicated by ▲) and reference sequences**

Bootstrap values of relevant nodes are indicated. Trees have been calculated using the neighbour-joining method and the Kimura 2 parameter model.

prevalence (45%), the relative prevalence of genotype 1b over 3a is 3 times higher in Estonia (71% versus 24%; (Tallo *et al.* 2004)) and 6.5 times higher in Western Europe (58% versus 9%). In Russia and Estonia the prevalence of all other genotypes is below 2% and below 1%, while in Europe genotype 1a represents 20%. Thus with a 1.5 fold dominance of genotype 1b over 3a and a prevalence of 5.1% the genotype pattern of HCV in Belarus is similar to the one in Russia although not without a clear Western European influence. Subtypes 1a, 4a and 4d were only found in HIV positive patients while the prevalence of subtypes 1b and 3a was 70.4 and 29.6% in HIV negative patients and 45.1% and 43.1% in HIV positive patients (Table 16). Similar studies for example in Spain (Ramos *et al.* 2007) also showed that among HIV positive donors subtype 1a and 3a are more prevalent indicating separate transmission networks of the HCV subtypes such as intravenous drug users (Mathei *et al.* 2005).

In the HCV strains found in Belarus, the average distances on the nucleotide level ranged from 5.7% to 7.1% within subtypes and from 22.4% to 33.8% between subtypes with an average distance between all Belorussian strains of 21.8%. The diversity within subtype 3a in Belarus was 1.7 times lower than the diversity of all worldwide strains belonging to the same subtype while for subtype 1b it was 1.2 times lower. Despite the higher prevalence of 1b, this may indicate that 3a has been circulating longer in Belarus or that it has been introduced multiple times from various sources and/or countries. BLAST searches of the different HCV strains did not reveal a closer relationship of Belorussian subtypes with other strains from the same geographic region but indicated for both subtypes 1b and 3a similarities with strains found worldwide. For subtype 1b, this could possibly be explained by the world-wide use of contaminated blood products such as in anti-D immunoglobulin in 1977 (Power *et al.* 1995). For subtype 3a however, circumstances similar to those in Egypt may have caused its spread (Frank *et al.* 2000). Also on a protein level, the HCV strains did not reveal amino acid substitutions specific to strains or subtypes found in Belarus confirming the multiple introductions from abroad.

In conclusion, HBV genotype D strains in Belarus form phylogenetic clusters (D1 to D4) that are compatible with the 4 subtypes recently proposed, although the intersubtypic distances may be lower than required. The relative prevalence of genotypes of both HBV and HCV in Belarus reflect Russian levels although with clear European influences, possibly explained by the socio-political context of the country. Surprisingly however, the virus variants do not seem to be clearly related to those from these neighbouring countries.

## Part V

Results of part V were submitted as:

Olinger CM, Jutavijittum P, Hübschen J, Yousukh A, Samountry B, Thammavong T, Toriyama K and Muller CP. Identification and molecular characteristics of a new Hepatitis B genotype I in Lao People's Democratic Republic.

### Identification and molecular characteristics of a new hepatitis B genotype I in Lao PDR

Mutations and recombinations between strains seem to drive a genetic evolution of HBV, which has resulted so far into 8 known genotypes (Bartholomeusz and Schaefer 2004; Norder *et al.* 2004). Since recombination events are not taken into account by current substitution model based molecular phylogeny, a careful approach in the classification of strains is warranted. Some of the current genotypes are already the result of multiple recombinations between existing and/or extinct genotypes (Bollyky *et al.* 1996; Bowyer and Sim 2000; Simmonds and Midgley 2005; Szmargd and Balloux 2007). Some genotypes have been associated with distinct clinical patterns (Tai *et al.* 1997; Torre and Naoumov 1998; Tsubota *et al.* 1998; Akuta and Kumada 2005; Kramvis and Kew 2005; Schaefer 2005; Kramvis and Kew 2007) and their detection and identification is important for virus and disease surveillance as well as for the understanding of HBV evolution.

Here we present the phylogenetic analysis of 19 related strains found in rejected blood donors from Lao PDR which do not cluster with any known HBV genotype.

## 1. Clinical samples

In 2004 and 2005, serum was collected from 498 HBsAg-positive first-time blood donors in several blood donation centres in Vientiane City and Central provinces, including from 5 donors from Northern provinces of Lao PDR. All donors were tested HIV, HCV and syphilis negative.

## 2. Alternative complete genome amplification

Extraction of DNA, amplification and sequencing of preS, S, X and C fragments were done as described in Chapter 3. Complete genome fragments were obtained using a single PCR reaction adapted from a PCR reaction based on two overlapping primers as described before (Gunther *et al.* 1995) and cloned using the pCR<sup>®</sup>-Blunt II-TOPO<sup>®</sup> Cloning kit (Invitrogen) as described in Chapter 3.

## 3. Phylogenetic analysis

Sequences were analysed as described in Chapter 3. All sequences were submitted to EMBL/GenBank/DDBJ. Similarity plots and bootscan analysis were done using windows of 800 or 400 nt, respectively. The time required for genetic diversity to develop within genotype I from the closest hypothetical common ancestor was calculated using a substitution rate of  $4.2 \times 10^{-5}$  substitutions per site and per year (Fares and Holmes 2002). Similarly the time of evolution of sequences or sequence fragments from a common ancestor was calculated by considering the genetic distance to the hypothetical common ancestor to be half of the genetic distance between the sequence of interest and its closest known relative.

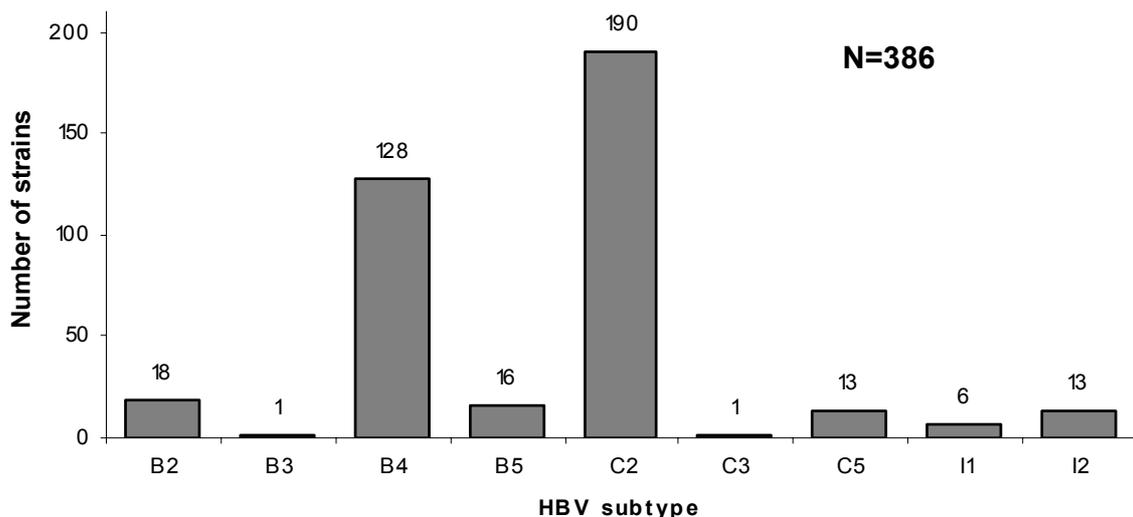
## 4. Results

### 4.1. Detection and sequencing of HBV DNA

Of 498 HBsAg positive rejected blood donors, 453 (90.8%) were PCR positive in at least one of the four HBV PCR reactions. Fifty-three strains were sequenced completely, whereas from 27 strains only 2 or 3 fragments could be sequenced. From the remaining samples, 369 S, 1 preS, 1 C and 2 X fragment sequences were obtained.

### 4.2. Distribution of genotypes

Phylogenetic analysis of all available sequences revealed that 163 strains (42.2%) were of genotype B and 204 (55.4%) of genotype C. Subtypes included B2 (18), B3 (1), B4 (128) and B5 (16) as well as C2 (190), C3 (1) and C5 (13) (Figures 25 and 26) (nomenclature according to Norder *et al.* 2004). 67 strains showed signs of mixed infections or recombinations between genotypes B and C and were not further characterized in this study. A genotype could not be attributed to 19 strains that clustered on the same node and separated with a 100% bootstrap value into two subgroups (Figure 26). From 15 of the latter strains, complete genome sequences were obtained.



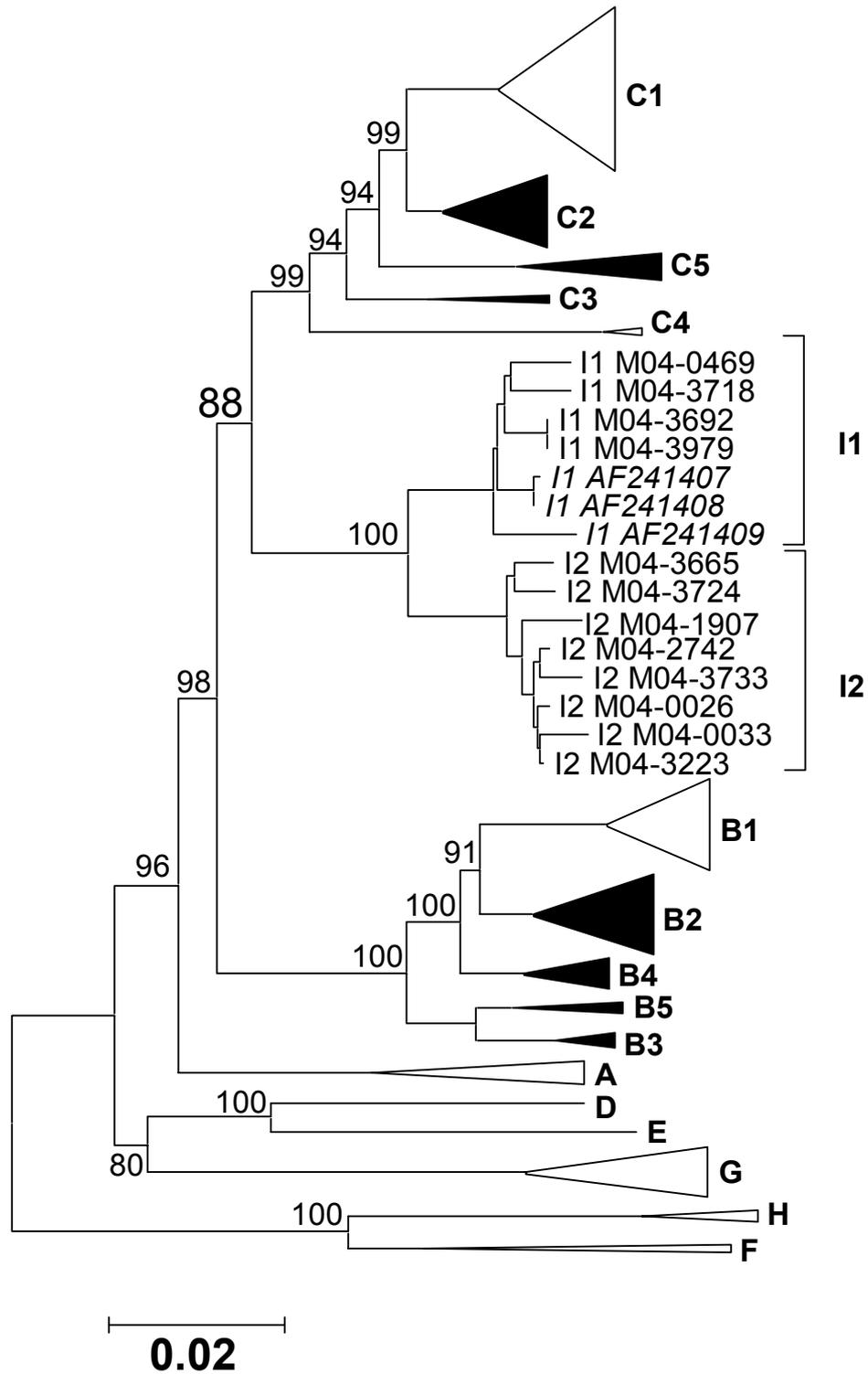
**Figure 25 : Relative distribution of genotypes and subtypes identified in 386 DNA positive rejected blood donors from Lao PDR**

#### 4.3. Definition of a new genotype I

When the complete genome sequences of the unclassifiable strains were compared to a representative set of full length HBV sequences of the known genotypes, genotype C (subtype C3) was the closest relative with an average Kimura distance of 7.89% (7.0% to C3). The bootstrap value of the separating node was 88% similar to the G/DE and B1/B2 nodes. On the S gene level the distance was 4.23% with a bootstrap value of 96% at the node separating genotypes I and C. Within the unclassifiable strains two phylogenetic subgroups were identified with a bootstrap value of 100%. The Kimura distances to genotype C of these two groups were 7.85% and 7.93%. Within the two groups an average diversity of 1.19% and 0.94% was calculated, which increased to 2.33% when all unclassifiable strains were considered as a single group. The maximal genetic distance between two genotype I strains was 4.3%. All clusterings were verified by Maximum Likelihood tree construction (data not shown). According to published criteria (Okamoto *et al.* 1988; Kramvis and Kew 2007), these values warrant the definition of a 9<sup>th</sup> HBV genotype, designated tentatively as genotype I. Accordingly, the two subgroups will be designated here as subtypes I1 and I2.

#### 4.4. Nucleotide sequence analysis

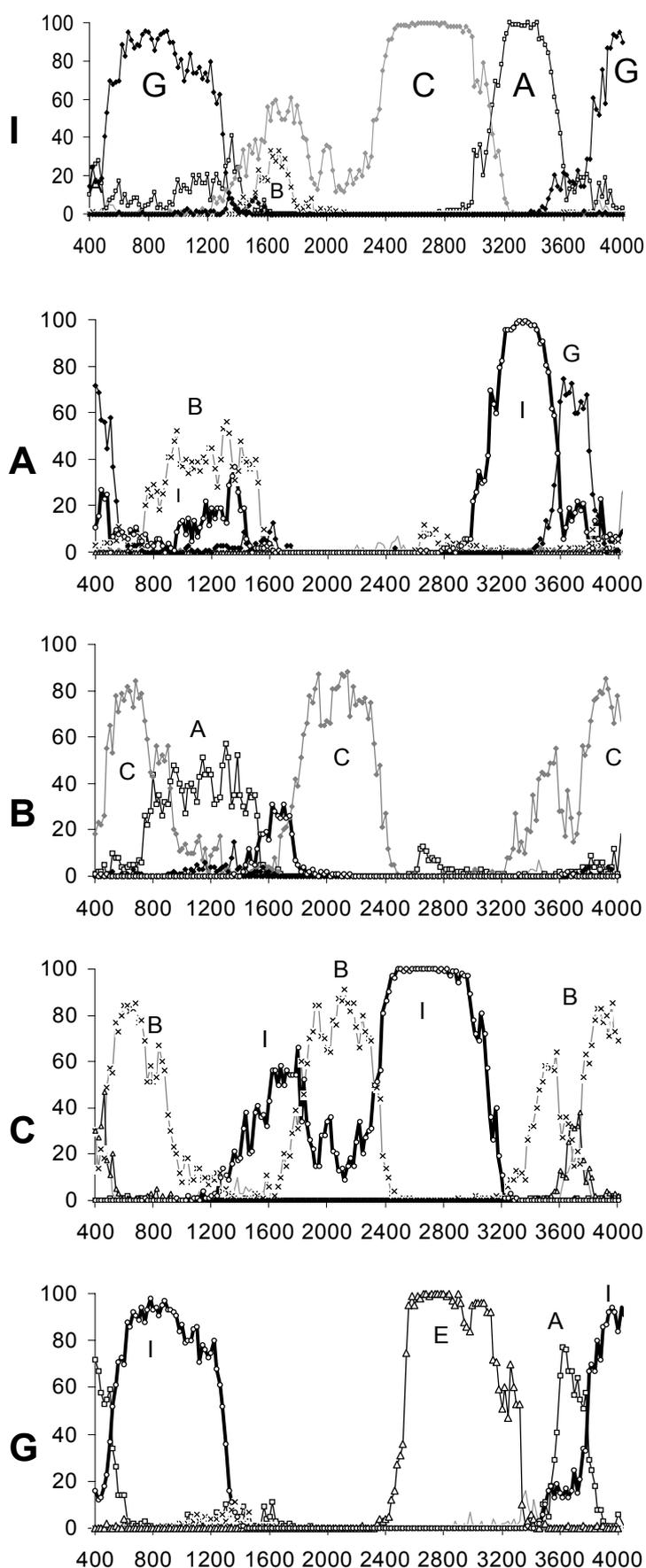
Bootscan analysis and phylogenetic reconstruction of genotype I strains (Figures 27 and 28) revealed some distant similarities with several different genotypes depending on the sequence region compared, indicating multiple recombination events. While most of the sequence of genotype I strains (nt 1400 – 3000) were most closely related to genotype C (6.29%), the rest of the circular genome was most closely related to genotype A (6.23%) (nt 3000-400) and genotype G (6.29%) (nt 400-1400). The next most similar genotypes for these fragments were genotypes B, C and A respectively, with genetic distances of 9.14%, 8.61% and 6.84%. For all segments, BLAST searches retrieved sequences AF241407 to AF241409 or sequences AB231908 and AB231909 as being the most similar. Phylogenetic analysis attributed all of these to the new subtype I1, with the exception of AB231909 which was identified as a I1/B4 recombinant. The next most similar strains for these distinct regions, were of subtypes C3 from Polynesia (only 3 C3 sequences available), subtype A3 from Gabon and genotype G from France respectively (Figure 27 and Table 18). On the C- and A-like sequences, strains of different subtypes (C1 from Uzbekistan and A4 from Gambia) were only 0.2% further away.



**Figure 26 : Phylogenetic comparison of genotype I complete genomes to a set of reference sequences**

Phylogenetic comparison of all complete genotype I genomes (n=15) obtained and compared to sequences of all known genotypes and subtypes. Non-genotype I genotypes identified in Lao PDR in

the present study are shown as full triangles. Numbers indicate bootstrap values of important nodes. Sequences AF241407 – AF241409 were published by Hannoun *et al.*



**Figure 27 : Bootscan analysis of genotype I**

Bootscan analysis of genotypes I, A, B, C and G compared to genotypes A-H. Data points correspond to the centre of sequence windows of 800 bp. For the analysis of the first 400 nt, the beginning of the genome was duplicated at the end of the sequence: nt 3200 to 4000 represent positions 0 to 800. Genotype A: black line and open squares; genotype B: grey line and crosses; genotype C: grey line and diamonds; genotype E: black line and triangles; genotype G: black lines and diamonds; genotype I: fat black line and circles. Genotypes D, F and H were included in the analysis but coincide with the baseline.

**Table 18 : Genotype, accession number and genetic distances and estimated years of evolution of the three sequence domains identified in genotype I.**

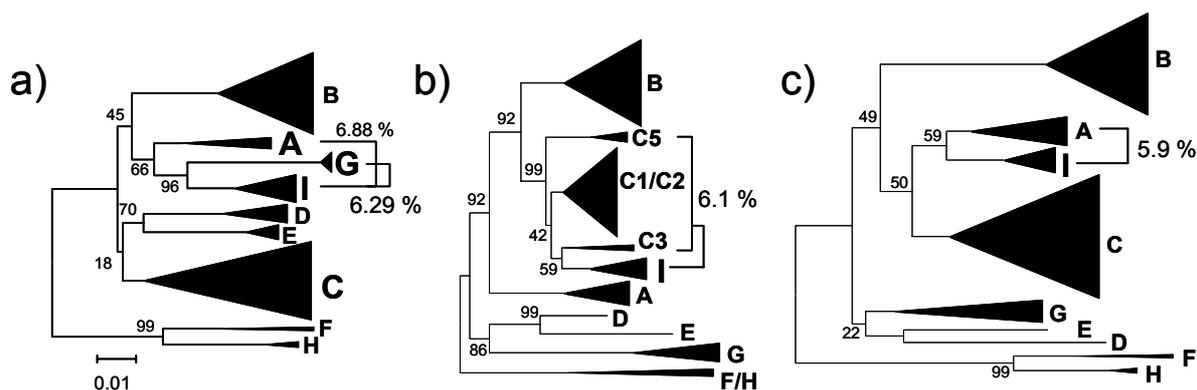
	nt	400-1400	1400-3000	3000-400
<b>most similar genotype</b>	G	6.29*	C	6.16
<b>2nd most similar genotype</b>	A	6.84	B	9.14
<b>most similar subtype</b>	NA		C3	5.42
<b>most similar strain</b>	G, AF160501	5.51 (650 yrs)	C3, X75656	4.1 (500 yrs)
<b>2nd most similar strain</b>	G, AB056513	5.62	C1, AB222715	4.3

\*Distances (in %) to genotypes or subtypes are average distances, *i.e.* the A-like sequence was in average most similar to subtype A4 but the minimal distance observed was to an A3 strain. NA: not applicable.

Surprisingly, when the stringency of bootscan analysis was reduced from a window size of 800 to 400 nt, similarities to genotype E appeared in nt positions 1200 to 1500 (data not shown). Also, the beginning of the region related to genotype C was shifted downstream by 100 nt. Non-human hepatitis B viruses showed much lower similarities to genotype I (minimum of 8.5% for the CHIMP genotype on positions 3000 to 400 versus 5.9% to for genotype A) and are not included in Figures 26 to 28.

With a substitution rate of  $4.2 \times 10^{-5}$  substitutions per site and per year (Fares and Holmes 2002), the 4.3% maximal diversity between all genotype I strains would have taken about 500 years to evolve from a common ancestor. The time of evolution of the G-like fragment of the current genotype I sequence from an ancestor sequence common with the closest genotype G strain (genetic distance of 5.51% between genotype G strain AF160501 and genotype I1 strain AF241408) would be about 650 years. Similarly, it would have taken about 500 years for the genotype C-like region to evolve (genetic distance of 4.10% between genotype C3 strain X7565 and genotype I2 strain M04-3223) and 530 years for the genotype A-like region (genetic distance of 4.50% between genotype A3 strain AM184126 and genotype I1 strain M04-0469) to develop (Figure 29a).

In order to exclude PCR artefacts in the recombination analysis of genotype I, the complete genome of one genotype I strain (M04-3665) was amplified using a single PCR reaction. The resulting fragment was cloned and its sequence was identical to the complete genome sequence assembled from the four semi-nested PCR fragments.



**Figure 28 : Phylogenetic comparison of genotype I with other genotypes on the distinct sequence fragments**

Phylogenetic comparison of positions 3000 to 400 (left), 0 to 1400 (middle) and 1400 to 3000 (right) of all genotype I strains (n=15) with all known

genotypes and subtypes. Percentages indicate average genetic distances between genotype I and G, C or A respectively.

#### 4.5. Amino acid sequence analysis

The comparison of all genotype I strains with a set of reference sequences of the known genotypes revealed several genotype I specific positions on the amino acid level. In the LHBS protein, 90V and 136I were found exclusively in all genotype I1 strains while in the core protein, 59V was often found in genotype I2 strains and not in others. In the X protein 48V was found exclusively in all genotype I strains while 40S was observed in all genotype I1 strains and rarely in others. On the polymerase gene, three subtype I1 specific (138Q, 318N and 823F), one subtype I2 specific (823C) and two genotype I specific (207S and 269K) positions were found. These substitutions did not allow establishing an evolutionary link between genotype I and the other genotypes. None of these specific mutations were located in positions known to affect viral protein expression or functions. All I1 strains were of serotype *adw*, while all I2 subtypes were of serotype *ayw*.

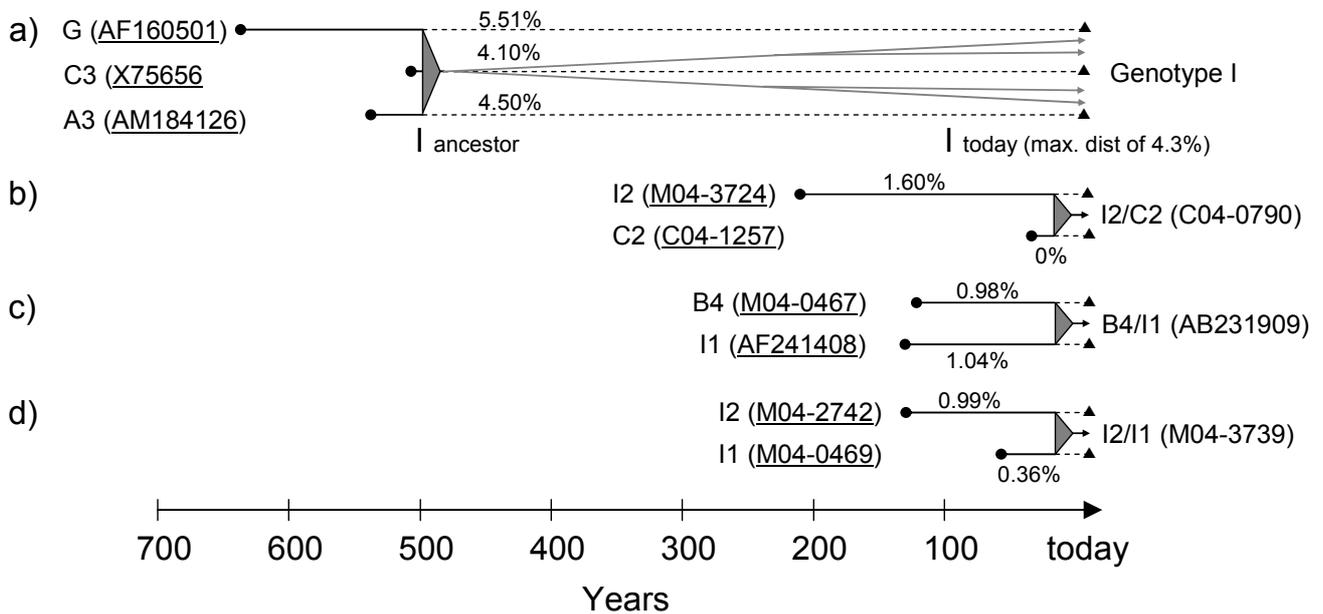
#### 4.6. Mixed infections and recombinations

In order to exclude artefacts due to mixed infections, 15 of the genotype I strains were cloned on various fragments. The 6 preS, 16 S, 11 X and 59 C clones analysed clustered with the same group than the uncloned sequence. Only in two cases (S fragment, M04-2769 and M05-0659) was a mixed infection by several genotypes found. The first (M05-0659) contained an I1 and I1/B4 recombinant sequence (B4 on the last 300 nt) while the second (M04-2769)

contained 5 different species: one B5/C2 recombinant (C2 on the last 400 nt), one C5/C2 recombinant (C2 on the last 400 nt), one genotype C2 sequence, four genotype C5 sequences and one C/I2 recombinant (C subtype unclear).

**Table 19 : Genotype, accession number and genetic distances (d, in %) and estimated years of evolution of three strains showing signs of recombination involving genotype I.**

	nt	most similar strain	d	years	nt	most similar strain	d	years
<b>C04-0790</b>	2500-1800	I2 M04-3724	1.6	200	1800-2500	C2 C04-1257	0	0
<b>AB231909</b>	1900-600	B4 M04-0467	0.98	115	600-1900	I1 AF241408	1.04	120
<b>M04-3739</b>	2500-1900	I2 M04-2742	0.99	120	1900-2500	I1 M04-0469	0.36	50



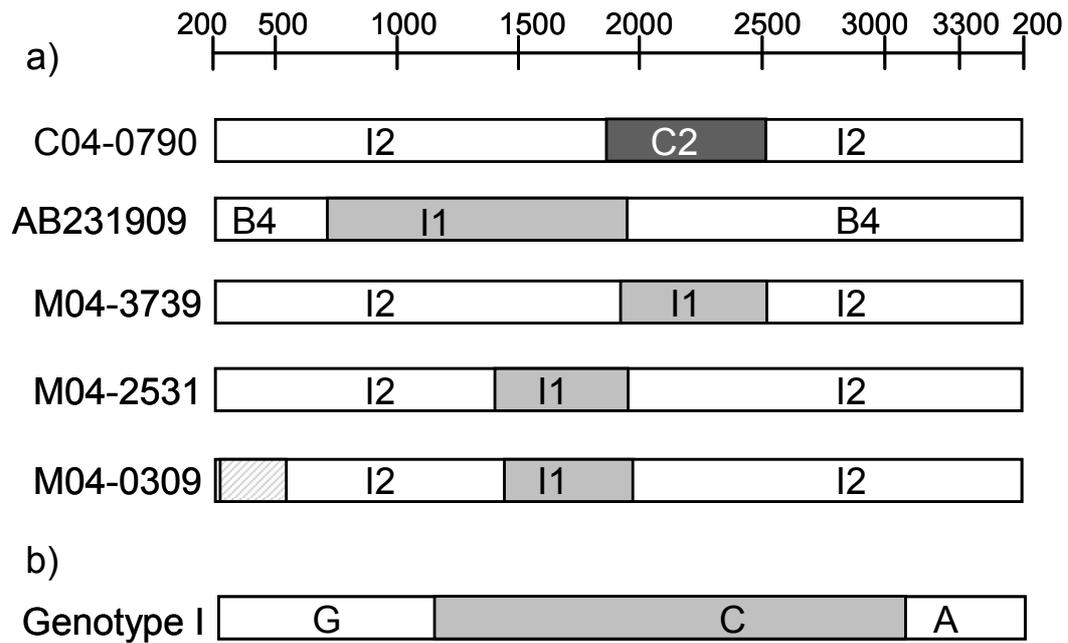
**Figure 29 : Evolutionary history of genotype I and three recombinant strains**

History of evolution of genotype I (a) as well as three (C04-0790, AB231309, M04-3739) recombinant strains (b, c and d). Numbers indicate genetic distances between the sequence fragments of interest (indicated with ▲) and the hypothetical

ancestor (indicated with ●) they share with their nearest known strain (underlined). The distance in brackets corresponds to the maximal distance observed between genotype I strains.

Four other sequences with no signs of mixed infections clustered inconclusively in the phylogenetic analysis. These were identified by cloning and bootscan analysis as recombinants between several genotypes and/or subtypes (Figure 30 and Table 19). In three recombinants, only subtypes I1 and I2 were involved (M04-3739, M04-2531 and M04-0309). The two latter had identical bootscan patterns while in the first the I2 sequence was shifted downstream by 200 nt. Strain C04-0790 showed a similar pattern than M04-3739 but with the I1 sequence replaced by a C2 sequence. Bootscan analysis of a previously reported strain (Tran and Abe, *unpublished*, AB231909) revealed a B4/I1 recombinant with the I1 sequence identified between positions 600 and 1864. Distance calculations (including BLAST searches) identified the most similar strains, for each of the distinct regions of the recombinants, to be almost exclusively circulating in Lao PDR (this study) (Figures 29b, c, d and Table 19). Genetic distances ranged from 0% to 1.6% and would correspond to a maximum of 200 years of evolution.

The recombinant strains C04-0790, AB231909 and M04-3739 were compared to sequences obtained in this study as well as by BLAST searches, to identify the closest relative for each recombinant sequence fragment. The distance between the sequence or sequence fragment of interest and the closest relative was considered to be twice the distance between each sequence and a common hypothetical ancestor which allowed to estimate how long ago the two sequences separated. The strains most similar to the I2/C2 recombinant were the genotype I2 strain M04-3724 (genetic distance: 1.60%) and the genotype C2 strain C04-1257 (100% identity). It would take an evolutionary time of 200 years for the I2 sequence to evolve from the common ancestor, while the two C2 sequences must have separated only recently (Figure 29b). Similarly, Figure 29c shows that the B4 and I1 fragments of the B4/I1 recombinant would take 115 and 120 years respectively and the two fragments of the I2/I1 recombinant would take 120 and 50 years (Figure 29d).



**Figure 30 : Diagrams of recombinants and genotype I**

Schematic bootscan diagrams of sequences showing evidence of recombination involving subtypes of genotype I and other subtypes (window

size 800). Hatched rectangles indicate missing sequence fragments. Sequence AB231909 was reported by Tran and Abe on GenBank.

## 5. Discussion

Here we describe a distinct set of 19 HBV strains found in the central provinces of Lao PDR, including the capital Vientiane. All strains clustered into two phylogenetically distinct groups and were more closely related to each other than to any strain of the known genotypes A to H. The 4.23% distance on the S gene and the 7.89% on the complete genome to the next closest genotype (genotype C), are well above the 4% proposed minimal distance on the S gene between genotypes (Okamoto *et al.* 1988) and the recently proposed 7.5% minimal distance between genotypes on a complete genome level (Kramvis and Kew 2007; Kramvis and Kew 2007). Formally, the definition of a 9<sup>th</sup> HBV genotype, tentatively designated genotype I, is therefore warranted. In addition, we propose the definition of two subtypes designated as I1 and I2. I1 strains were detected in donors living in Vientiane City and the surrounding provinces, while I2 strains were only found in donors living in Vientiane City. Strains reported as aberrant strains from Hanoi, Vietnam eight years ago (Hannoun *et al.* 2000) group with subtype I1, indicating that genotype I may be more widespread in Southeast Asia than suggested by the present study. Interestingly, one donor, infected with a subtype I1 strain, originated from Vietnam and a majority of genotype I infected donors originated from the Houaphan province which shares the border with Vietnam.

Genetic distance, phylogenetic and bootscan analysis indicated that this new genotype is a result of probably multiple recombination events involving mainly genotypes C and A, but surprisingly also genotype G and even genotype E. Most current genotypes of HBV seem to be the result of one or several recombination events (Bollyky *et al.* 1996; Simmonds and Midgley 2005; Szymaragd and Balloux 2007): In particular this is evident for the B/C recombinant which has spread in mainland Asia (Sugauchi *et al.* 2004) and has been defined as genotype B1 but also for genotypes B and C themselves which show similarities to genotype A (Figure 27). Genotype E is thought to be a recombinant between genotype D and another unknown or extinct genotype (Bowyer and Sim 2000). Genotype G is highly divergent from all currently known genotypes in most regions of the genome but shows similarities to genotype E in the end of the S gene (Figure 27). It is clear that the new genotype I described here has followed a similar series of recombination events. One of these events involved a genotype C-like fragment related to subtype C3, so far found only in the Pacific, except for a single incomplete strain in Lao PDR, and C1 found in Japan, Korea,

China and Uzbekistan (Kato *et al.* 2002; Norder *et al.* 2004; Kim *et al.* 2007). The genotype A-like fragment was most similar to subtypes A3 and A4, only recently found by us and confirmed by others in Sub Saharan Africa (Mulders *et al.* 2004; Kurbanov *et al.* 2005; Makuwa *et al.* 2006; Olinger *et al.* 2006). The apparent relatedness with the defective genotype G is even more surprising given the sporadic nature of this genotype found so far only in the United States, Japan, Germany and France (Stuyver *et al.* 2000; Kato *et al.* 2002; Vieth *et al.* 2002; Shibayama *et al.* 2005; Chudy *et al.* 2006). Although only a few sequences of the latter genotype have been reported it has clearly contributed to at least three other recombinations: one in San Francisco with genotype A (Kato *et al.* 2002), one in Thailand with genotype C (Suwannakarn *et al.* 2005) and one in Nigeria involving genotypes D and A (Olinger *et al.* 2006), all with distinct different breakpoints. Thus, none of the contributing genotypes or subtypes (C3, A3 and G) has ever been identified in South East Asia, questioning the Asian origin of genotype I.

When applying a mutation rate of  $4.2 \times 10^{-5}$  substitutions per site and per year (Fares and Holmes 2002), the time the different recombinant sequence fragments (nt 400 to 1400, 1400 to 3000 and 3000 to 400) would need to evolve from a hypothetical ancestor shared with the most similar strains would be 500 to 650 years. This is compatible with the 500 years it would have taken genotype I to develop the maximal diversity of 4.3% from a single ancestor virus.

Genotype I further recombined with regional strains in mixed infected patients. The time of evolution of the recombinant sequences from an ancestor shared with their most similar strains range from 0 to 200 years. Interestingly, all of the most likely ancestors were found in Lao PDR (this study). In fact, the sequence of one recombinant strain found in Vientiane was identical to a potential donor strain found in a different district of the capital. Thus the simplest possible evolutionary scenario would be the recombination of ancestral viruses to form genotype I some 500 years ago, the introduction into South East Asia and the recombination with local genotype during the past 200 years.

The identification and analysis of genotype I provide further evidence of the importance of recombination in the evolution of HBV, a complexity which may not be fully acknowledged by the current classification. According to this classification, viruses of genotype I fulfil the formal criteria of a new genotype and show a distinct geographic distribution with a consistent evolutionary history, although its origin remains unclear.

## Chapter 5: Conclusions and Perspectives

Currently it is estimated that a third of the world population has come in contact with HBV during their lifetime and that as a consequence, 400 million people are suffering from chronic hepatitis B infection. A majority will eventually die of liver complications or hepatocellular carcinoma. Although most of the infected people live in areas such as sub-Saharan Africa, Asia and the Pacific, in which HBV is endemic, the virus is circulating globally with highly variable prevalences.

Based on the accumulation of single nucleotide polymorphisms in the genome and phylogenetic reconstruction, the circulating strains have been classified into 8 genotypes (A – H) and multiple distinct subtypes. Although a majority of these genotypes and subtypes present distinct geographic distributions, the distribution of genotype E, one of the most prevalent genotypes and found only in West-Africa, is particularly noticeable. In 2004, a phylogenetic study covering hepatitis B strains from seven sub-Saharan countries, representing a third of the African continent, revealed that a majority of circulating strains were of genotype E, with one exception: in Cameroon the prevalence of genotype E strains was equal to the prevalence of strains of genotype A. This raised the question whether genotypes could co-infect the same host and whether such events could lead to recombinations between different genotypes. In addition, sequence analysis of the circulating genotype E strains found during this study, revealed a surprisingly low variability, less than half of what would normally be expected suggesting that genotype E was introduced relatively late in sub-Saharan Africa. The present study aimed to further characterize HBV genotype E and to advance our understanding of its widespread circulation but conspicuously low diversity. In addition, phylogenetic studies of HBV in Belarus and Lao PDR aimed at gaining further insights in the mechanisms of distribution and evolution of HBV genotypes and subtypes. During these studies, special interest was placed in detecting evidence of recombinations between genotypes and to evaluate their importance in the history of HBV evolution.

In a first approach, data was collected on the sensitivity and specificity of three currently available hepatitis B diagnostic kits by testing on sera from 200 Nigerian suspected HBV

positive donors. The objective was to evaluate the exceptionally high prevalences of hepatitis B found in sub-Saharan Africa and to evaluate a possible bias in regard to genotype E. HBsAg detection assays rely on the capture of the surface antigen by monoclonal antibodies and most of these antibodies are directed towards immunogenic epitopes, mainly in the a-determinant of genotype A surface antigen. Of all genotypes, the surface antigen of genotype E differs most from the one of genotype A. Results between the three HBsAg assays were mostly concordant positive or negative for the majority of samples, except for 36 for which results were discordant. All but one of the latter were confirmed negative by neutralization. 80% of discordant results were caused by a single assay which showed a particularly high false positivity rate. Nucleotide sequence analysis of DNA positive samples with no detectable HBsAg or reduced HBsAg detection signals revealed specific mutations mostly outside the a-determinant. Several of these mutations are found as wild type nucleotides normally in genotype A and only exceptionally in genotype E and failure to detect HBsAg antigen and differences in signal intensity was mainly associated with these mutations. These results indicate that the context, in which diagnostic kits are used, in this case Africa and genotype E, should be an important criterion for their selection and evaluation but also in their development. In fact, assays of different vendors show considerable discrepancies and some maybe suitable only for applications in restricted geographical areas. The existence of specific mutations that are able to affect an assay outcome, even with a positive bias, warrants further analysis of currently available commercial assays.

In light of the previous results, it is important to evaluate the extent of the genotype E prevalence in sub-Saharan Africa by characterizing HBV strains from additional countries. Of 196 serum samples, randomly collected from chronically HBV infected patients in the Central African Republic, we identified almost all as genotype E, thus showing that genotype E is circulating, as the most prevalent genotype, further to the east of Africa than previously known. The identification of only one genotype A1 strain and three genotype D strains revealed that North Africa, where mostly genotype D circulates, and South and East Africa, where mostly genotype A circulates, do not contribute to HBV in the Central African Republic.

Since previous results on genotype E diversity were mainly based on the preS/S region of the HBV genome, we obtained and analyzed 122 new HBV preC/C sequences and 3 complete genomes from three major countries in West Africa. The majority of these were of genotype

E, thus significantly increasing the number of genotype E sequences available. The only other genotype found was genotype A. Although for genotype E sequences the genetic diversity of the preC/C gene was about 2 – 3 times higher than the diversity of the preS/S gene, it was still considerably lower than for genotype A sequences. In fact, the genotype A strains from Cameroon, Mali and Nigeria could be phylogenetically divided into 3 subtypes, A3 and 2 new subtypes, tentatively designated A4 and A5, further showing the significant differences in the evolution and distribution between genotypes A and E. While genotype A presents multiple subtypes, each circulating in distinct locations in Africa, Europe and America, genotype E, while found in many different countries and only in Africa, shows a very low diversity and has not evolved into different subtypes. While a recent introduction would explain the low diversity, it would need to have been followed by a very fast spread throughout half of the African continent. Currently no route for such a spread is known and the possibility of unknown animal vectors, or even involuntary spread by humans, need to be considered. In fact, the latter is known to have played an important role in the spread of HCV.

During the study above, we also found that almost 20% of all patients analysed showed a co-infection of genotypes A and E and most originated from Cameroon, where both genotypes co-exist with equal prevalences. Evidence of recombination events between different genotypes, which prerequisites mixed infections, were found in two strains and one sample from Nigeria even showed evidence of a triple recombination of genotype E/D and A, separated by a genotype G specific insert. While recombination events are suspected to play a role in HBV evolution, evidences of their existence remain sparse, and their molecular mechanisms have not yet been elucidated. Thus, the recombinations reported here may help in further understanding the evolution of HBV and may give use indications in the past history of its spread. In fact, the presence of a genotype G specific sequence in an African patient, raises further questions about the origin of this genotype which, until now, had only been found sporadically in the USA, France and Germany.

Aiming to further understand the spread of HBV genotypes worldwide, we performed a phylogenetic study of 179 serum donors from Belarus. 69 were HBV DNA positive and infected with strains of genotypes A and D. Surprisingly, these genotype D strains could be identified to belong to four recently described subtypes, thus adding additional information to HBV nomenclature. This also indicated a genotype distribution in Belarus similar to Russia but with West-European influences. Since genotype D, the closest relative of genotype E,

shows multiple subtypes and a high diversity, the evolution and distribution of genotype E, which shows no subtypes and a low diversity must have taken a different course.

Adding to the recombination events identified during our studies, we also showed that 19 related strains identified in Lao PDR, are the result of recombination events between at least three genotypes. These strains, which did not cluster with any known HBV genotype, fulfilled the formal criteria for the definition of a new genotype (I) with two subgroups (I1 and I2). Further analysis indicated that this new genotype is a result of multiple recombination events involving genotypes G, C (C3), A (A3) and even E which probably occurred around 500 years ago. Interestingly none of the involved genotypes or subtypes except for C3 has ever been identified in Southeast Asia, blurring the geographic origin of genotype I. Compatible with a more recent introduction into Lao PDR are 6 strains showing recombination events involving I1/I2 and the most prevalent subtypes B4 and C2 in Lao PDR which seem to have occurred within the last 200 years only. The identification and analysis of the new genotype I provided further evidence of the importance of recombination in the evolution of HBV.

The results presented here indicate that genotype E must have followed a different course of evolution when compared to other HBV genotypes and warrant the investigation of alternative transmission routes such as animal vectors and human involvement. Also, the analysis of HBV in descendants of African origin living outside of Africa, could give insight in its evolutionary history. In addition our results show that recombination events are more important in the evolution of HBV than previously thought, further proven by our identification of a new genotype of HBV which clearly was the result of multiple recombination events. Finally, we also showed that the genotype can influence the diagnostics of HBV, which, combined with the fact that genotypes are known to influence the treatment and clinical outcome of HBV infections, warrants further studies especially in light of the new genotype identified.

## References

1. Abdulkarim, AS, NN Zein, JJ Germer, *et al.* (1998) Hepatitis C virus genotypes and hepatitis G virus in hemodialysis patients from Syria: identification of two novel hepatitis C virus subtypes. *Am J Trop Med Hyg* 59: 571-576.
2. Ahmed, SD, LE Cuevas, BJ Brabin, *et al.* (1998) Seroprevalence of hepatitis B and C and HIV in Malawian pregnant women. *J Infect* 37: 248-251.
3. Akuta, N and H Kumada (2005) Influence of hepatitis B virus genotypes on the response to antiviral therapies. *J Antimicrob Chemother* 55: 139-142.
4. Aldrich, CE, L Coates, TT Wu, *et al.* (1989) In vitro infection of woodchuck hepatocytes with woodchuck hepatitis virus and ground squirrel hepatitis virus. *Virology* 172: 247-252.
5. Alhababi, F, TA Sallam and CY Tong (2003) The significance of 'anti-HBc only' in the clinical virology laboratory. *J Clin Virol* 27: 162-169.
6. Allain, JP, D Candotti, K Soldan, *et al.* (2003) The risk of hepatitis B virus infection by transfusion in Kumasi, Ghana. *Blood* 101: 2419-2425.
7. Alter, HJ, RH Purcell, JL Gerin, *et al.* (1977) Transmission of hepatitis B to chimpanzees by hepatitis B surface antigen-positive saliva and semen. *Infect Immun* 16: 928-933.
8. Arauz-Ruiz, P, H Norder, KA Visona and LO Magnius (1997) Molecular epidemiology of hepatitis B virus in Central America reflected in the genetic variability of the small S gene. *J Infect Dis* 176: 851-858.
9. Arauz-Ruiz, P, H Norder, KA Visona and LO Magnius (1997) Genotype F prevails in HBV infected patients of hispanic origin in Central America and may carry the precore stop mutant. *J Med Virol* 51: 305-312.
10. Arauz-Ruiz, P, H Norder, BH Robertson and LO Magnius (2002) Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. *J Gen Virol* 83: 2059-2073.
11. Baptista, M, A Kramvis and MC Kew (1999) High prevalence of 1762(T) 1764(A) mutations in the basic core promoter of hepatitis B virus isolated from black Africans with hepatocellular carcinoma compared with asymptomatic carriers. *Hepatology* 29: 946-953.

## References

---

12. Barker, LF and R Murray (1972) Relationship of virus dose to incubation time of clinical hepatitis and time of appearance of hepatitis--associated antigen. *Am J Med Sci* 263: 27-33.
13. Bartenschlager, R and H Schaller (1988) The amino-terminal domain of the hepadnaviral P-gene encodes the terminal protein (genome-linked protein) believed to prime reverse transcription. *Embo J* 7: 4185-4192.
14. Bartholomeusz, A and S Schaefer (2004) Hepatitis B virus genotypes: comparison of genotyping methods. *Rev Med Virol* 14: 3-16.
15. Beasley, RP, LY Hwang, CC Lin and CS Chien (1981) Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22 707 men in Taiwan. *Lancet* 2: 1129-1133.
16. Beasley, RP, LY Hwang, CC Lin, *et al.* (1981) Hepatitis B immune globulin (HBIG) efficacy in the interruption of perinatal transmission of hepatitis B virus carrier state. Initial report of a randomised double-blind placebo-controlled trial. *Lancet* 2: 388-393.
17. Beck, J and M Nassal (1998) Formation of a functional hepatitis B virus replication initiation complex involves a major structural alteration in the RNA template. *Mol Cell Biol* 18: 6265-6272.
18. Beck, J, M Vogel and M Nassal (2002) dNTP versus NTP discrimination by phenylalanine 451 in duck hepatitis B virus P protein indicates a common structure of the dNTP-binding pocket with other reverse transcriptases. *Nucleic Acids Res* 30: 1679-1687.
19. Beck, J and M Nassal (2003) Efficient Hsp90-independent in vitro activation by Hsc70 and Hsp40 of duck hepatitis B virus reverse transcriptase, an assumed Hsp90 client protein. *J Biol Chem* 278: 36128-36138.
20. Berting, A, C Fischer, S Schaefer, *et al.* (2000) Hemifusion activity of a chimeric influenza virus hemagglutinin with a putative fusion peptide from hepatitis B virus. *Virus Res* 68: 35-49.
21. Blakely, TA, MN Bates, MG Baker and M Tobias (1999) Hepatitis B carriage explains the excess rate of hepatocellular carcinoma for Maori, Pacific Island and Asian people compared to Europeans in New Zealand. *Int J Epidemiol* 28: 204-210.
22. Blum, HE, L Stowring, A Figus, *et al.* (1983) Detection of hepatitis B virus DNA in hepatocytes, bile duct epithelium, and vascular elements by in situ hybridization. *Proc Natl Acad Sci U S A* 80: 6685-6688.
23. Blum, HE, ZS Zhang, E Galun, *et al.* (1992) Hepatitis B virus X protein is not central to the viral life cycle in vitro. *J Virol* 66: 1223-1227.
24. Blumberg, BS, BJ Gerstley, DA Hungerford, WT London and AI Sutnick (1967) A serum antigen (Australia antigen) in Down's syndrome, leukemia, and hepatitis. *Ann Intern Med* 66: 924-931.

## References

---

25. Bock, CT, P Schranz, CH Schroder and H Zentgraf (1994) Hepatitis B virus genome is organized into nucleosomes in the nucleus of the infected cell. *Virus Genes* 8: 215-229.
26. Bollyky, PL, A Rambaut, PH Harvey and EC Holmes (1996) Recombination between sequences of hepatitis B virus from different genotypes. *J Mol Evol* 42: 97-102.
27. Bortolotti, F, P Cadrobbi, C Crivellaro, *et al.* (1990) Long-term outcome of chronic type B hepatitis in patients who acquire hepatitis B virus infection in childhood. *Gastroenterology* 99: 805-810.
28. Bowyer, SM, L van Staden, MC Kew and JG Sim (1997) A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa. *J Gen Virol* 78 ( Pt 7): 1719-1729.
29. Bowyer, SM and JG Sim (2000) Relationships within and between genotypes of hepatitis B virus at points across the genome: footprints of recombination in certain isolates. *J Gen Virol* 81: 379-392.
30. Bozdayi, G, AR Turkyilmaz, R Idilman, *et al.* (2005) Complete genome sequence and phylogenetic analysis of hepatitis B virus isolated from Turkish patients with chronic HBV infection. *J Med Virol* 76: 476-481.
31. Brechot, C, F Jaffredo, D Lagorce, *et al.* (1998) Impact of HBV, HCV and GBV-C/HGV on hepatocellular carcinomas in Europe: results of a European concerted action. *J Hepatol* 29: 173-183.
32. Breiner, KM and H Schaller (2000) Cellular receptor traffic is essential for productive duck hepatitis B virus infection. *J Virol* 74: 2203-2209.
33. Breiner, KM, S Urban, B Glass and H Schaller (2001) Envelope protein-mediated down-regulation of hepatitis B virus receptor in infected hepatocytes. *J Virol* 75: 143-150.
34. Bruss, V and WH Gerlich (1988) Formation of transmembraneous hepatitis B e-antigen by cotranslational in vitro processing of the viral precore protein. *Virology* 163: 268-275.
35. Bruss, V and D Ganem (1991) The role of envelope proteins in hepatitis B virus assembly. *Proc Natl Acad Sci U S A* 88: 1059-1063.
36. Bukh, J, RH Miller and RH Purcell (1995) Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. *Semin Liver Dis* 15: 41-63.
37. Cha, TA, E Beall, B Irvine, *et al.* (1992) At least five related, but distinct, hepatitis C viral genotypes exist. *Proc Natl Acad Sci U S A* 89: 7144-7148.
38. Chamberlain, RW, N Adams, AA Saeed, P Simmonds and RM Elliott (1997) Complete nucleotide sequence of a type 4 hepatitis C virus variant, the predominant genotype in the Middle East. *J Gen Virol* 78 ( Pt 6): 1341-1347.

## References

---

39. Chang, SF, HJ Netter, M Bruns, *et al.* (1999) A new avian hepadnavirus infecting snow geese (*Anser caerulescens*) produces a significant fraction of virions containing single-stranded DNA. *Virology* 262: 39-54.
40. Chen, HS, MC Kew, WE Hornbuckle, *et al.* (1992) The precore gene of the woodchuck hepatitis virus genome is not essential for viral replication in the natural host. *J Virol* 66: 5682-5684.
41. Chen, HS, S Kaneko, R Girones, *et al.* (1993) The woodchuck hepatitis virus X gene is important for establishment of virus infection in woodchucks. *J Virol* 67: 1218-1226.
42. Chirillo, P, M Falco, PL Puri, *et al.* (1996) Hepatitis B virus pX activates NF-kappa B-dependent transcription through a Raf-independent pathway. *J Virol* 70: 641-646.
43. Chu, CJ, EB Keeffe, SH Han, *et al.* (2003) Hepatitis B virus genotypes in the United States: results of a nationwide study. *Gastroenterology* 125: 444-451.
44. Chu, CM, P Karayiannis, MJ Fowler, *et al.* (1985) Natural history of chronic hepatitis B virus infection in Taiwan: studies of hepatitis B virus DNA in serum. *Hepatology*: 431-434.
45. Chudy, M, M Schmidt, V Czudai, *et al.* (2006) Hepatitis B virus genotype G monoinfection and its transmission by blood components. *Hepatology* 44: 99-107.
46. Colin, JF, D Cazals-Hatem, MA Lioriot, *et al.* (1999) Influence of human immunodeficiency virus infection on chronic hepatitis B in homosexual men. *Hepatology* 29: 1306-1310.
47. Conjeevaram, HS and AS Lok (2003) Management of chronic hepatitis B. *J Hepatol* 38 Suppl 1: S90-103.
48. Cui, C, J Shi, L Hui, *et al.* (2002) The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. *J Gen Virol* 83: 2773-2777.
49. Dane, DS, CH Cameron and M Briggs (1970) Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *Lancet* 1: 695-698.
50. Davis, GL, JH Hoofnagle and JG Waggoner (1984) Spontaneous reactivation of chronic hepatitis B virus infection. *Gastroenterology* 86: 230-235.
51. Dejean, A, P Sonigo, S Wain-Hobson and P Tiollais (1984) Specific hepatitis B virus integration in hepatocellular carcinoma DNA through a viral 11-base-pair direct repeat. *Proc Natl Acad Sci U S A* 81: 5350-5354.
52. Di Martino, V, T Thevenot, JF Colin, *et al.* (2002) Influence of HIV infection on the response to interferon therapy and the long-term outcome of chronic hepatitis B. *Gastroenterology* 123: 1812-1822.
53. Dienstag, JL, AR Rhodes, AK Bhan, *et al.* (1978) Urticaria associated with acute viral hepatitis type B: studies of pathogenesis. *Ann Intern Med* 89: 34-40.

## References

---

54. Dikstein, R, O Faktor, R Ben-Levy and Y Shaul (1990) Functional organization of the hepatitis B virus enhancer. *Mol Cell Biol* 10: 3683-3689.
55. Doria, M, N Klein, R Lucito and RJ Schneider (1995) The hepatitis B virus HBx protein is a dual specificity cytoplasmic activator of Ras and nuclear activator of transcription factors. *Embo J* 14: 4747-4757.
56. Dzierzanowska-Fangrat, K, M Woynarowski, I Szczygielska, *et al.* (2006) Hepatitis B virus genotypes in children with chronic hepatitis B in Poland. *Eur J Gastroenterol Hepatol* 18: 655-658.
57. Echevarria, JM, P Leon, CJ Domingo, *et al.* (1991) Characterization of HBV2-like infections in Spain. *J Med Virol* 33: 240-247.
58. Echevarria, JM and P Leon (2004) Hepatitis B virus genotypes identified by a Line Probe Assay (LiPA) among chronic carriers from Spain. *Enferm Infecc Microbiol Clin* 22: 452-454.
59. Echevarria, JM, A Avellon and LO Magnus (2005) Molecular epidemiology of hepatitis B virus in Spain: Identification of viral genotypes and prediction of antigenic subtypes by limited sequencing. *J Med Virol* 76: 176-184.
60. Edmunds, WJ, GF Medley, DJ Nokes, *et al.* (1996) Epidemiological patterns of hepatitis B virus (HBV) in highly endemic areas. *Epidemiol Infect* 117: 313-325.
61. Fares, MA and EC Holmes (2002) A revised evolutionary history of hepatitis B virus (HBV). *J Mol Evol* 54: 807-814.
62. FitzSimons, DW and P Van Damme (1999) European Forum on Immunization: report of Second General Meeting, Munich, 2-3 June 1997. *Eur J Pediatr* 158: 359-361.
63. Flodgren, E, S Bengtsson, M Knutsson, *et al.* (2000) Recent high incidence of fulminant hepatitis in Samara, Russia: molecular analysis of prevailing hepatitis B and D virus strains. *J Clin Microbiol* 38: 3311-3316.
64. Frank, C, MK Mohamed, GT Strickland, *et al.* (2000) The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet* 355: 887-891.
65. Frosner, GG, H Schomerus, KH Wiedmann, *et al.* (1982) Diagnostic significance of quantitative determination of hepatitis B surface antigen in acute and chronic hepatitis B infection. *Eur J Clin Microbiol* 1: 52-58.
66. Fung, SK and AS Lok (2004) Hepatitis B virus genotypes: do they play a role in the outcome of HBV infection? *Hepatology* 40: 790-792.
67. Fung, SK and AS Lok (2004) Management of hepatitis B patients with antiviral resistance. *Antivir Ther* 9: 1013-1026.

## References

---

68. Funk, ML, DM Rosenberg and AS Lok (2002) World-wide epidemiology of HBeAg-negative chronic hepatitis B and associated precore and core promoter variants. *J Viral Hepat* 9: 52-61.
69. Galle, PR, J Hagelstein, B Kommerell, *et al.* (1994) In vitro experimental infection of primary human hepatocytes with hepatitis B virus. *Gastroenterology* 106: 664-673.
70. Ganne-Carrie, N, V Williams, H Kaddouri, *et al.* (2006) Significance of hepatitis B virus genotypes A to E in a cohort of patients with chronic hepatitis B in the Seine Saint Denis District of Paris (France). *J Med Virol* 78: 335-340.
71. Girones, R and RH Miller (1989) Mutation rate of the hepadnavirus genome. *Virology* 170: 595-597.
72. Gong, Y, E Yao and JE Tavis (2001) Evidence that the RNAseH activity of the duck hepatitis B virus is unable to act on exogenous substrates. *BMC Microbiol* 1: 12.
73. Gray, AH, JW Fang, GL Davis, *et al.* (1997) Variations of hepatitis B virus core gene sequence in Western patients with chronic hepatitis B virus infection. *J Viral Hepat* 4: 371-378.
74. Gripon, P, C Diot, N Theze, *et al.* (1988) Hepatitis B virus infection of adult human hepatocytes cultured in the presence of dimethyl sulfoxide. *J Virol* 62: 4136-4143.
75. Gripon, P, C Diot and C Guguen-Guillouzo (1993) Reproducible high level infection of cultured adult human hepatocytes by hepatitis B virus: effect of polyethylene glycol on adsorption and penetration. *Virology* 192: 534-540.
76. Gripon, P, S Rumin, S Urban, *et al.* (2002) Infection of a human hepatoma cell line by hepatitis B virus. *Proc Natl Acad Sci U S A* 99: 15655-15660.
77. Gunther, S, BC Li, S Miska, *et al.* (1995) A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. *J Virol* 69: 5437-5444.
78. Hannoun, C, P Horal and M Lindh (2000) Long-term mutation rates in the hepatitis B virus genome. *J Gen Virol* 81: 75-83.
79. Hannoun, C, H Norder and M Lindh (2000) An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam. *J Gen Virol* 81: 2267-2272.
80. Hannoun, C, K Krogsgaard, P Horal and M Lindh (2002) Genotype mixtures of hepatitis B virus in patients treated with interferon. *J Infect Dis* 186: 752-759.
81. Hannoun, C, A Soderstrom, G Norkrans and M Lindh (2005) Phylogeny of African complete genomes reveals a West African genotype A subtype of hepatitis B virus and relatedness between Somali and Asian A1 sequences. *J Gen Virol* 86: 2163-2167.
82. Hasegawa, I, Y Tanaka, A Kramvis, *et al.* (2004) Novel hepatitis B virus genotype a subtyping assay that distinguishes subtype Aa from Ae and its application in epidemiological studies. *J Virol* 78: 7575-7581.

## References

---

83. Haviv, I, D Vaizel and Y Shaul (1996) pX, the HBV-encoded coactivator, interacts with components of the transcription machinery and stimulates transcription in a TAF-independent manner. *Embo J* 15: 3413-3420.
84. Henkler, FF and R Koshy (1996) Hepatitis B virus transcriptional activators: mechanisms and possible role in oncogenesis. *J Viral Hepat* 3: 109-121.
85. Higgins, DG, JD Thompson and TJ Gibson (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* 266: 383-402.
86. Hoofnagle, JH and AM di Bisceglie (1997) The treatment of chronic viral hepatitis. *N Engl J Med* 336: 347-356.
87. Howard, C (1995). Hepadnaviridae. *Virus Taxonomy*. F. Murphy, C. Fauquet, D. Bishop and J. Smith. New York, Springer-Verlag.
88. Hraber, PT, RW Leach, LP Reilly, *et al.* (2007) Los Alamos hepatitis C virus sequence and human immunology databases: an expanding resource for antiviral research. *Antivir Chem Chemother* 18: 113-123.
89. Hsu, HY, MH Chang, SH Liaw, YH Ni and HL Chen (1999) Changes of hepatitis B surface antigen variants in carrier children before and after universal vaccination in Taiwan. *Hepatology* 30: 1312-1317.
90. Hu, J, DO Toft and C Seeger (1997) Hepadnavirus assembly and reverse transcription require a multi-component chaperone complex which is incorporated into nucleocapsids. *Embo J* 16: 59-68.
91. Hu, J and D Anselmo (2000) In vitro reconstitution of a functional duck hepatitis B virus reverse transcriptase: posttranslational activation by Hsp90. *J Virol* 74: 11447-11455.
92. Hu, J, D Toft, D Anselmo and X Wang (2002) In vitro reconstitution of functional hepadnavirus reverse transcriptase with cellular chaperone proteins. *J Virol* 76: 269-279.
93. Hu, KQ and A Siddiqui (1991) Regulation of the hepatitis B virus gene expression by the enhancer element I. *Virology* 181: 721-726.
94. Huang, J and TJ Liang (1993) A novel hepatitis B virus (HBV) genetic element with Rev response element-like properties that is essential for expression of HBV gene products. *Mol Cell Biol* 13: 7476-7486.
95. Huang, ZM and TS Yen (1994) Hepatitis B virus RNA element that facilitates accumulation of surface gene transcripts in the cytoplasm. *J Virol* 68: 3193-3199.
96. Huang, ZM and TS Yen (1995) Role of the hepatitis B virus posttranscriptional regulatory element in export of intronless transcripts. *Mol Cell Biol* 15: 3864-3869.

## References

---

97. Huy, TT, K Ishikawa, W Ampofo, *et al.* (2006) Characteristics of hepatitis B virus in Ghana: full length genome sequences indicate the endemicity of genotype E in West Africa. *J Med Virol* 78: 178-184.
98. Jardi, R, F Rodriguez, M Buti, *et al.* (2001) Quantitative detection of hepatitis B virus DNA in serum by a new rapid real-time fluorescence PCR assay. *J Viral Hepat* 8: 465-471.
99. Jeantet, D, I Chemin, B Mandrand, *et al.* (2004) Cloning and expression of surface antigens from occult chronic hepatitis B virus infections and their recognition by commercial detection assays. *J Med Virol* 73: 508-515.
100. Kane, M (1995) Global programme for control of hepatitis B infection. *Vaccine* 13 Suppl 1: S47-49.
101. Kann, M, B Sodeik, A Vlachou, WH Gerlich and A Helenius (1999) Phosphorylation-dependent binding of hepatitis B virus core particles to the nuclear pore complex. *J Cell Biol* 145: 45-55.
102. Kao, JH, PJ Chen, MY Lai and DS Chen (2002) Clinical and virological aspects of blood donors infected with hepatitis B virus genotypes B and C. *J Clin Microbiol* 40: 22-25.
103. Kaplan, PM, RL Greenman, JL Gerin, RH Purcell and WS Robinson (1973) DNA polymerase associated with human hepatitis B antigen. *J Virol* 12: 995-1005.
104. Kato, H, E Orito, RG Gish, *et al.* (2002) Characteristics of hepatitis B virus isolates of genotype G and their phylogenetic differences from the other six genotypes (A through F). *J Virol* 76: 6131-6137.
105. Kato, H, R Ruzibakiev, N Yuldasheva, *et al.* (2002) Hepatitis B virus genotypes in Uzbekistan and validity of two different systems for genotyping. *J Med Virol* 67: 477-483.
106. Kekule, AS, U Lauer, L Weiss, B Lubber and PH Hofschneider (1993) Hepatitis B virus transactivator HBx uses a tumour promoter signalling pathway. *Nature* 361: 742-745.
107. Kidd-Ljunggren, K, M Oberg and AH Kidd (1995) The hepatitis B virus X gene: analysis of functional domain variation and gene phylogeny using multiple sequences. *J Gen Virol* 76 ( Pt 9): 2119-2130.
108. Kim, CM, K Koike, I Saito, T Miyamura and G Jay (1991) HBx gene of hepatitis B virus induces liver cancer in transgenic mice. *Nature* 351: 317-320.
109. Kim, H, YM Jee, BC Song, *et al.* (2007) Molecular epidemiology of hepatitis B virus (HBV) genotypes and serotypes in patients with chronic HBV infection in Korea. *Intervirology* 50: 52-57.

## References

---

110. Kimbi, GC, A Kramvis and MC Kew (2004) Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa. *J Gen Virol* 85: 1211-1220.
111. Kock, J and HJ Schlicht (1993) Analysis of the earliest steps of hepadnavirus replication: genome repair after infectious entry into hepatocytes does not depend on viral polymerase activity. *J Virol* 67: 4867-4874.
112. Kock, J, EM Borst and HJ Schlicht (1996) Uptake of duck hepatitis B virus into hepatocytes occurs by endocytosis but does not require passage of the virus through an acidic intracellular compartment. *J Virol* 70: 5827-5831.
113. Kock, J, TF Baumert, WEt Delaney, HE Blum and F von Weizsacker (2003) Inhibitory effect of adefovir and lamivudine on the initiation of hepatitis B virus infection in primary tupaia hepatocytes. *Hepatology* 38: 1410-1418.
114. Kramvis, A, L Weitzmann, WK Owiredu and MC Kew (2002) Analysis of the complete genome of subgroup A' hepatitis B virus isolates from South Africa. *J Gen Virol* 83: 835-839.
115. Kramvis, A and MC Kew (2005) Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy. *J Viral Hepat* 12: 456-464.
116. Kramvis, A, K Restorp, H Norder, *et al.* (2005) Full genome analysis of hepatitis B virus genotype E strains from South-Western Africa and madagascar reveals low genetic variability. *J Med Virol* 77: 47-52.
117. Kramvis, A and MC Kew (2007) Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes. *Hepato Res* 37: S9-S19.
118. Kramvis, A and MC Kew (2007) Molecular characterization of subgenotype A1 (subgroup Aa) of hepatitis B virus. *Hepato Res* 37: S27-32.
119. Krekulova, L, V Rehak, HP da Silva Filho, M Zavoral and LW Riley (2003) Genotypic distribution of hepatitis B virus in the Czech Republic: a possible association with modes of transmission and clinical outcome. *Eur J Gastroenterol Hepatol* 15: 1183-1188.
120. Kumar, S, K Tamura and M Nei (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5: 150-163.
121. Kurbanov, F, Y Tanaka, K Fujiwara, *et al.* (2005) A new subtype (subgenotype) Ac (A3) of hepatitis B virus and recombination between genotypes A and E in Cameroon. *J Gen Virol* 86: 2047-2056.
122. Kuroki, K, F Eng, T Ishikawa, *et al.* (1995) gp180, a host cell glycoprotein that binds duck hepatitis B virus particles, is encoded by a member of the carboxypeptidase gene family. *J Biol Chem* 270: 15022-15028.

## References

---

123. Lai, CL, V Ratziu, MF Yuen and T Poynard (2003) Viral hepatitis B. *Lancet* 362: 2089-2094.
124. Landers, TA, HB Greenberg and WS Robinson (1977) Structure of hepatitis B Dane particle DNA and nature of the endogenous DNA polymerase reaction. *J Virol* 23: 368-376.
125. Laperche, S, A Girault, MJ Beaulieu, F Bouchardeau and AM Courouce (2001) Determination of hepatitis B virus subtypes by an enzyme immunoassay method using monoclonal antibodies to type-specific epitopes of HBsAg. *J Viral Hepat* 8: 447-453.
126. Lazarevic, I, M Cupic, D Delic, *et al.* (2007) Distribution of HBV genotypes, subgenotypes and HBsAg subtypes among chronically infected patients in Serbia. *Arch Virol* 152: 2017-2025.
127. Lee, WM (1997) Hepatitis B virus infection. *N Engl J Med* 337: 1733-1745.
128. Lemon, SM, TJ Layden, L Seeff, *et al.* (2000) The 20th United States-Japan Joint Hepatitis Panel Meeting. *Hepatology* 31: 800-806.
129. Liu, HF, E Sokal and P Goubau (2001) Wide variety of genotypes and geographic origins of hepatitis B virus in Belgian children. *J Pediatr Gastroenterol Nutr* 32: 274-277.
130. Lok, AS, RH Liang, EK Chiu, *et al.* (1991) Reactivation of hepatitis B virus replication in patients receiving cytotoxic therapy. Report of a prospective study. *Gastroenterology* 100: 182-188.
131. Lok, AS, EJ Heathcote and JH Hoofnagle (2001) Management of hepatitis B: 2000--summary of a workshop. *Gastroenterology* 120: 1828-1853.
132. Lok, AS (2002) Chronic hepatitis B. *N Engl J Med* 346: 1682-1683.
133. Lole, KS, RC Bollinger, RS Paranjape, *et al.* (1999) Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 73: 152-160.
134. Lopez-Cabrera, M, J Letovsky, KQ Hu and A Siddiqui (1991) Transcriptional factor C/EBP binds to and transactivates the enhancer element II of the hepatitis B virus. *Virology* 183: 825-829.
135. Madhava, V, C Burgess and E Drucker (2002) Epidemiology of chronic hepatitis C virus infection in sub-Saharan Africa. *Lancet Infect Dis* 2: 293-302.
136. Magnus, LO and H Norder (1995) Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. *Intervirology* 38: 24-34.
137. Mahoney, FJ (1999) Update on diagnosis, management, and prevention of hepatitis B virus infection. *Clin Microbiol Rev* 12: 351-366.

## References

---

138. Makuwa, M, S Souquiere, P Telfer, *et al.* (2006) Identification of hepatitis B virus subgenotype A3 in rural Gabon. *J Med Virol* 78: 1175-1184.
139. Makuwa, M, S Souquiere, O Bourry, *et al.* (2007) Complete-genome analysis of hepatitis B virus from wild-born chimpanzees in central Africa demonstrates a strain-specific geographical cluster. *J Gen Virol* 88: 2679-2685.
140. Mangia, A, YH Chung, JH Hoofnagle, *et al.* (1996) Pathogenesis of chronic liver disease in patients with chronic hepatitis B virus infection without serum HBeAg. *Dig Dis Sci* 41: 2447-2452.
141. Marion, PL, LS Oshiro, DC Regnery, GH Scullard and WS Robinson (1980) A virus in Beechey ground squirrels that is related to hepatitis B virus of humans. *Proc Natl Acad Sci U S A* 77: 2941-2945.
142. Marquardt, O, KH Heermann, M Seifer and WH Gerlich (1987) Cell type specific expression of pre S 1 antigen and secretion of hepatitis B virus surface antigen. *Brief Report. Arch Virol* 96: 249-256.
143. Mason, WS, G Seal and J Summers (1980) Virus of Pekin ducks with structural and biological relatedness to human hepatitis B virus. *J Virol* 36: 829-836.
144. Masuda, M, T Yuasa and H Yoshikura (1990) Effect of the preS1 RNA sequence on the efficiency of the hepatitis B virus preS2 and S protein translation. *Virology* 174: 320-324.
145. Mathej, C, E Wollants, J Verbeeck, *et al.* (2005) Molecular epidemiology of hepatitis C among drug users in Flanders, Belgium: association of genotype with clinical parameters and with sex- and drug-related risk behaviours. *Eur J Clin Microbiol Infect Dis* 24: 514-522.
146. Mathet, VL, ML Cuestas, V Ruiz, *et al.* (2006) Detection of hepatitis B virus (HBV) genotype E carried--even in the presence of high titers of anti-HBs antibodies--by an Argentinean patient of African descent who had received vaccination against HBV. *J Clin Microbiol* 44: 3435-3439.
147. Mathur, M, D Turbadkar and M Rele (2002) Prevalence of HIV infection in HBsAg positive cases. *Indian J Med Microbiol* 20: 225.
148. McMahon, BJ, P Holck, L Bulkow and M Snowball (2001) Serologic and clinical outcomes of 1536 Alaska Natives chronically infected with hepatitis B virus. *Ann Intern Med* 135: 759-768.
149. Michitaka, K, Y Tanaka, N Horiike, *et al.* (2006) Tracing the history of hepatitis B virus genotype D in western Japan. *J Med Virol* 78: 44-52.
150. Minami, M, K Poussin, M Kew, *et al.* (1996) Precore/core mutations of hepatitis B virus in hepatocellular carcinomas developed on noncirrhotic livers. *Gastroenterology* 111: 691-700.

## References

---

151. Mizuochi, T, Y Okada, K Umemori, S Mizusawa and K Yamaguchi (2006) Evaluation of 10 commercial diagnostic kits for in vitro expressed hepatitis B virus (HBV) surface antigens encoded by HBV of genotypes A to H. *J Virol Methods* 136: 254-256.
152. Morozov, V, M Pisareva and M Groudinin (2000) Homologous recombination between different genotypes of hepatitis B virus. *Gene* 260: 55-65.
153. Moyer, LA and EE Mast (1994) Hepatitis B: virology, epidemiology, disease, and prevention, and an overview of viral hepatitis. *Am J Prev Med* 10 Suppl: 45-55.
154. Mulders, MN, V Venard, M Njayou, *et al.* (2004) Low Genetic Diversity despite Hyperendemicity of Hepatitis B Virus Genotype E throughout West Africa. *J Infect Dis* 190: 400-408.
155. Nainan, OV, ML Khristova, K Byun, *et al.* (2002) Genetic variation of hepatitis B surface antigen coding region among infants with chronic hepatitis B virus infection. *J Med Virol* 68: 319-327.
156. Nassal, M (1992) The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. *J Virol* 66: 4107-4116.
157. Nassal, M and A Rieger (1993) An intramolecular disulfide bridge between Cys-7 and Cys61 determines the structure of the secretory core gene product (e antigen) of hepatitis B virus. *J Virol* 67: 4307-4315.
158. Natoli, G, ML Avantaggiati, P Chirillo, *et al.* (1995) Modulation of intracellular signal transduction pathways by the hepatitis B virus transactivator pX. *J Hepatol* 22: 14-20.
159. Neurath, AR, SB Kent, K Parker, *et al.* (1986) Antibodies to a synthetic peptide from the preS 120-145 region of the hepatitis B virus envelope are virus neutralizing. *Vaccine* 4: 35-37.
160. Neurath, AR, SB Kent, N Strick and K Parker (1986) Identification and chemical synthesis of a host cell receptor binding site on hepatitis B virus. *Cell* 46: 429-436.
161. Ni, YH, HY Hsu, MH Chang, DS Chen and CY Lee (1993) Absence or delayed appearance of hepatitis B core antibody in chronic hepatitis B surface antigen carrier children. *J Hepatol* 17: 150-154.
162. Ni, YH, LM Huang, MH Chang, *et al.* (2007) Two decades of universal hepatitis B vaccination in taiwan: impact and implication for future strategies. *Gastroenterology* 132: 1287-1293.
163. Nolte, FS (2001) Hepatitis C virus genotyping: clinical implications and methods. *Mol Diagn* 6: 265-277.
164. Norder, H, AM Courouce and LO Magnius (1992) Molecular basis of hepatitis B virus serotype variations within the four major subtypes. *J Gen Virol* 73 ( Pt 12): 3141-3145.

## References

---

165. Norder, H, B Hammas, SD Lee, *et al.* (1993) Genetic relatedness of hepatitis B viral strains of diverse geographical origin and natural variations in the primary structure of the surface antigen. *J Gen Virol* 74 ( Pt 7): 1341-1348.
166. Norder, H, AM Courouce and LO Magnius (1994) Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. *Virology* 198: 489-503.
167. Norder, H, AM Courouce, P Coursaget, *et al.* (2004) Genetic diversity of hepatitis B virus strains derived worldwide: genotypes, subgenotypes, and HBsAg subtypes. *Intervirology* 47: 289-309.
168. Odemuyiwa, SO, MN Mulders, OI Oyedele, *et al.* (2001) Phylogenetic analysis of new hepatitis B virus isolates from Nigeria supports endemicity of genotype E in West Africa. *J Med Virol* 65: 463-469.
169. Okamoto, H, M Imai, F Tsuda, *et al.* (1987) Point mutation in the S gene of hepatitis B virus for a d/y or w/r subtypic change in two blood donors carrying a surface antigen of compound subtype adyr or adwr. *J Virol* 61: 3030-3034.
170. Okamoto, H, F Tsuda, H Sakugawa, *et al.* (1988) Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *J Gen Virol* 69 ( Pt 10): 2575-2583.
171. Okamoto, H, F Tsuda, Y Akahane, *et al.* (1994) Hepatitis B virus with mutations in the core promoter for an e antigen-negative phenotype in carriers with antibody to e antigen. *J Virol* 68: 8102-8110.
172. Olinger, CM, V Venard, M Njayou, *et al.* (2006) Phylogenetic analysis of the precore/core gene of hepatitis B virus genotypes E and A in West Africa: new subtypes, mixed infections and recombinations. *J Gen Virol* 87: 1163-1173.
173. Olubuyide, IO, SO Ola, B Aliyu, *et al.* (1997) Prevalence and epidemiological characteristics of hepatitis B and C infections among doctors and dentists in Nigeria. *East Afr Med J* 74: 357-361.
174. Orito, E, M Mizokami, Y Ina, *et al.* (1989) Host-independent evolution and a genetic classification of the hepadnavirus family based on nucleotide sequences. *Proc Natl Acad Sci U S A* 86: 7059-7062.
175. Owiredu, WK, A Kramvis and MC Kew (2001) Molecular analysis of hepatitis B virus genomes isolated from black African patients with fulminant hepatitis B. *J Med Virol* 65: 485-492.
176. Pante, N and M Kann (2002) Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm. *Mol Biol Cell* 13: 425-434.
177. Park, SG and G Jung (2001) Human hepatitis B virus polymerase interacts with the molecular chaperonin Hsp60. *J Virol* 75: 6962-6968.

## References

---

178. Parkin, DM, F Bray, J Ferlay and P Pisani (2001) Estimating the world cancer burden: Globocan 2000. *Int J Cancer* 94: 153-156.
179. Pawlotsky, JM, L Tsakiris, F Roudot-Thoraval, *et al.* (1995) Relationship between hepatitis C virus genotypes and sources of infection in patients with chronic hepatitis C. *J Infect Dis* 171: 1607-1610.
180. Pontisso, P, MG Ruvoletto, WH Gerlich, *et al.* (1989) Identification of an attachment site for human liver plasma membranes on hepatitis B virus particles. *Virology* 173: 522-530.
181. Posada, D and KA Crandall (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
182. Pourquier, P, AD Jensen, SS Gong, Y Pommier and CE Rogler (1999) Human DNA topoisomerase I-mediated cleavage and recombination of duck hepatitis B virus DNA in vitro. *Nucleic Acids Res* 27: 1919-1925.
183. Power, JP, E Lawlor, F Davidson, *et al.* (1995) Molecular epidemiology of an outbreak of infection with hepatitis C virus in recipients of anti-D immunoglobulin. *Lancet* 345: 1211-1213.
184. Prassolov, A, H Hohenberg, T Kalinina, *et al.* (2003) New hepatitis B virus of cranes that has an unexpected broad host range. *J Virol* 77: 1964-1976.
185. Pult, I, HJ Netter, M Bruns, *et al.* (2001) Identification and analysis of a new hepadnavirus in white storks. *Virology* 289: 114-128.
186. Puoti, M, M Airoidi, R Bruno, *et al.* (2002) Hepatitis B virus co-infection in human immunodeficiency virus-infected subjects. *AIDS Rev* 4: 27-35.
187. Qadri, I, HF Maguire and A Siddiqui (1995) Hepatitis B virus transactivator protein X interacts with the TATA-binding protein. *Proc Natl Acad Sci U S A* 92: 1003-1007.
188. Rabe, B, A Vlachou, N Pante, A Helenius and M Kann (2003) Nuclear import of hepatitis B virus capsids and release of the viral genome. *Proc Natl Acad Sci U S A* 100: 9849-9854.
189. Ramos, B, M Nunez, L Martin-Carbonero, *et al.* (2007) Hepatitis B virus genotypes and lamivudine resistance mutations in HIV/hepatitis B virus-coinfected patients. *J Acquir Immune Defic Syndr* 44: 557-561.
190. Ramos, B, M Nunez, C Toro, *et al.* (2007) Changes in the distribution of hepatitis C virus (HCV) genotypes over time in Spain according to HIV serostatus: implications for HCV therapy in HCV/HIV-coinfected patients. *J Infect* 54: 173-179.
191. Robertson, BH and HS Margolis (2002) Primate hepatitis B viruses - genetic diversity, geography and evolution. *Rev Med Virol* 12: 133-141.
192. Robinson, WS, DA Clayton and RL Greenman (1974) DNA of a human hepatitis B virus candidate. *J Virol* 14: 384-391.

## References

---

193. Robinson, WS and RL Greenman (1974) DNA polymerase in the core of the human hepatitis B virus candidate. *J Virol* 13: 1231-1236.
194. Rossner, MT (1992) Review: hepatitis B virus X-gene product: a promiscuous transcriptional activator. *J Med Virol* 36: 101-117.
195. Schaefer, S and WH Gerlich (1995) In vitro transformation by hepatitis B virus DNA. *Intervirology* 38: 143-154.
196. Schaefer, S (2005) Hepatitis B virus: significance of genotypes. *J Viral Hepat* 12: 111-124.
197. Schaller, H and M Fischer (1991) Transcriptional control of hepadnavirus gene expression. *Curr Top Microbiol Immunol* 168: 21-39.
198. Schlicht, H, R Bartenschlager and H Schakker (1991). Biosynthesis and Enzymatic Functions of the Hepadnaviral Reverse Transcriptase. Molecular Biology of Hepatitis B viruses. A. McLachlan. Boca Raton, CRC Press.
199. Schlicht, HJ and H Schaller (1989) The secretory core protein of human hepatitis B virus is expressed on the cell surface. *J Virol* 63: 5399-5404.
200. Schreier, E, M Roggendorf, G Driesel, M Hohne and S Viazov (1996) Genotypes of hepatitis C virus isolates from different parts of the world. *Arch Virol Suppl* 11: 185-193.
201. Schuster, R, E Hildt, SF Chang, *et al.* (2002) Conserved transactivating and pro-apoptotic functions of hepadnaviral X protein in ortho- and avihepadnaviruses. *Oncogene* 21: 6606-6613.
202. Scott, RM, R Snitbhan, WH Bancroft, HJ Alter and M Tingpalapong (1980) Experimental transmission of hepatitis B virus by semen and saliva. *J Infect Dis* 142: 67-71.
203. Seeger, C, D Ganem and HE Varmus (1986) Biochemical and genetic evidence for the hepatitis B virus replication strategy. *Science* 232: 477-484.
204. Sells, MA, ML Chen and G Acs (1987) Production of hepatitis B virus particles in Hep G2 cells transfected with cloned hepatitis B virus DNA. *Proc Natl Acad Sci U S A* 84: 1005-1009.
205. Shibayama, T, G Masuda, A Ajisawa, *et al.* (2005) Characterization of seven genotypes (A to E, G and H) of hepatitis B virus recovered from Japanese patients infected with human immunodeficiency virus type 1. *J Med Virol* 76: 24-32.
206. Shimoda, T, T Shikata, T Karasawa, *et al.* (1981) Light microscopic localization of hepatitis B virus antigens in the human pancreas. Possibility of multiplication of hepatitis B virus in the human pancreas. *Gastroenterology* 81: 998-1005.

## References

---

207. Simmonds, P, F McOmish, PL Yap, *et al.* (1993) Sequence variability in the 5' non-coding region of hepatitis C virus: identification of a new virus type and restrictions on sequence diversity. *J Gen Virol* 74 ( Pt 4): 661-668.
208. Simmonds, P and S Midgley (2005) Recombination in the genesis and evolution of hepatitis B virus genotypes. *J Virol* 79: 15467-15476.
209. Sprengel, R, EF Kaleta and H Will (1988) Isolation and characterization of a hepatitis B virus endemic in herons. *J Virol* 62: 3832-3839.
210. Standring, DN, JH Ou, FR Masiarz and WJ Rutter (1988) A signal peptide encoded within the precore region of hepatitis B virus directs the secretion of a heterogeneous population of e antigens in *Xenopus* oocytes. *Proc Natl Acad Sci U S A* 85: 8405-8409.
211. Stevens, CE, RA Neurath, RP Beasley and W Szmunes (1979) HBeAg and anti-HBe detection by radioimmunoassay: correlation with vertical transmission of hepatitis B virus in Taiwan. *J Med Virol* 3: 237-241.
212. Stuyver, L, S De Gendt, C Van Geyt, *et al.* (2000) A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *J Gen Virol* 81: 67-74.
213. Su, F and RJ Schneider (1996) Hepatitis B virus HBx protein activates transcription factor NF-kappaB by acting on multiple cytoplasmic inhibitors of rel-related proteins. *J Virol* 70: 4558-4566.
214. Sugauchi, F, E Orito, T Ichida, *et al.* (2002) Hepatitis B virus of genotype B with or without recombination with genotype C over the precore region plus the core gene. *J Virol* 76: 5985-5992.
215. Sugauchi, F, E Orito, H Kato, *et al.* (2003) Genotype, serotype, and phylogenetic characterization of the complete genome sequence of hepatitis B virus isolates from Malawian chronic carriers of the virus. *J Med Virol* 69: 33-40.
216. Sugauchi, F, H Kumada, H Sakugawa, *et al.* (2004) Two subtypes of genotype B (Ba and Bj) of hepatitis B virus in Japan. *Clin Infect Dis* 38: 1222-1228.
217. Summers, J, A O'Connell and I Millman (1975) Genome of hepatitis B virus: restriction enzyme cleavage and structure of DNA extracted from Dane particles. *Proc Natl Acad Sci U S A* 72: 4597-4601.
218. Summers, J, JM Smolec and R Snyder (1978) A virus similar to human hepatitis B virus associated with hepatitis and hepatoma in woodchucks. *Proc Natl Acad Sci U S A* 75: 4533-4537.
219. Summers, J and WS Mason (1982) Replication of the genome of a hepatitis B--like virus by reverse transcription of an RNA intermediate. *Cell* 29: 403-415.
220. Sun, Z, P Lu, MH Gail, *et al.* (1999) Increased risk of hepatocellular carcinoma in male hepatitis B surface antigen carriers with chronic hepatitis who have detectable urinary aflatoxin metabolite M1. *Hepatology* 30: 379-383.

## References

---

221. Sunbul, M and H Leblebicioglu (2005) Distribution of hepatitis B virus genotypes in patients with chronic hepatitis B in Turkey. *World J Gastroenterol* 11: 1976-1980.
222. Sureau, C, JL Romet-Lemonne, JI Mullins and M Essex (1986) Production of hepatitis B virus by a differentiated human hepatoma cell line after transfection with cloned circular HBV DNA. *Cell* 47: 37-47.
223. Suwannakarn, K, P Tangkijvanich, A Theamboonlers, K Abe and Y Poovorawan (2005) A novel recombinant of Hepatitis B virus genotypes G and C isolated from a Thai patient with hepatocellular carcinoma. *J Gen Virol* 86: 3027-3030.
224. Suzuki, S, F Sugauchi, E Orito, *et al.* (2003) Distribution of hepatitis B virus (HBV) genotypes among HBV carriers in the Cote d'Ivoire: complete genome sequence and phylogenetic relatedness of HBV genotype E. *J Med Virol* 69: 459-465.
225. Swofford, DL (2003). *PAUP\* – Phylogenetic Analysis Using Parsimony (\*and other methods)*. MA, Sinauer Associates.
226. Szmargd, C and F Balloux (2007) The population genomics of hepatitis B virus. *Mol Ecol* 16: 4747-4758.
227. Szomor, KN, A Dencs, G Toth, *et al.* (2007) Variability of the PreS1/PreS2/S regions of hepatitis B virus in Hungary. *Arch Virol* 152: 697-704.
228. Tai, PC, D Banik, GI Lin, *et al.* (1997) Novel and frequent mutations of hepatitis B virus coincide with a major histocompatibility complex class I-restricted T-cell epitope of the surface antigen. *J Virol* 71: 4852-4856.
229. Takada, N, S Takase, A Takada and T Date (1993) Differences in the hepatitis C virus genotypes in different countries. *J Hepatol* 17: 277-283.
230. Takahashi, K, B Brotman, S Usuda, S Mishiro and AM Prince (2000) Full-genome sequence analyses of hepatitis B virus (HBV) strains recovered from chimpanzees infected in the wild: implications for an origin of HBV. *Virology* 267: 58-64.
231. Tallo, T, H Norder, V Tefanova, *et al.* (2004) Hepatitis B virus genotype D strains from Estonia share sequence similarity with strains from Siberia and may specify ayw4. *J Med Virol* 74: 221-227.
232. Tassopoulos, NC, GJ Papaevangelou, MH Sjogren, *et al.* (1987) Natural history of acute hepatitis B surface antigen-positive hepatitis in Greek adults. *Gastroenterology* 92: 1844-1850.
233. Testut, P, CA Renard, O Terradillos, *et al.* (1996) A new hepadnavirus endemic in arctic ground squirrels in Alaska. *J Virol* 70: 4210-4219.
234. Theamboonlers, A, P Jantaradsamee, N Kaew-In, *et al.* (1999) The predominant genotypes of hepatitis B virus in Thailand. *Ann Trop Med Parasitol* 93: 737-743.
235. Thomas, H, S Lemon and Z AJ., Eds. (2005). *Viral Hepatitis*. Oxford, Blackwell Publishing.

## References

---

236. Thomas, HC (1982) Immunological mechanisms in chronic hepatitis B virus infection. *Hepatology*: 116S-119S.
237. Toh, H, H Hayashida and T Miyata (1983) Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. *Nature* 305: 827-829.
238. Tong, S, J Li and JR Wands (1995) Interaction between duck hepatitis B virus and a 170-kilodalton cellular protein is mediated through a neutralizing epitope of the pre-S region and occurs during viral infection. *J Virol* 69: 7106-7112.
239. Torre, F and NV Naoumov (1998) Clinical implications of mutations in the hepatitis B virus genome. *Eur J Clin Invest* 28: 604-614.
240. Torresi, J, L Earnest-Silveira, G Civitico, *et al.* (2002) Restoration of replication phenotype of lamivudine-resistant hepatitis B virus mutants by compensatory changes in the "fingers" subdomain of the viral polymerase selected as a consequence of mutations in the overlapping S gene. *Virology* 299: 88-99.
241. Tsubota, A, H Kumada, K Takaki, *et al.* (1998) Deletions in the hepatitis B virus core gene may influence the clinical outcome in hepatitis B e antigen-positive asymptomatic healthy carriers. *J Med Virol* 56: 287-293.
242. Tur-Kaspa, R, RD Burk, Y Shaul and DA Shafritz (1986) Hepatitis B virus DNA contains a glucocorticoid-responsive element. *Proc Natl Acad Sci U S A* 83: 1627-1631.
243. Tur-Kaspa, R, Y Shaul, DD Moore, *et al.* (1988) The glucocorticoid receptor recognizes a specific nucleotide sequence in hepatitis B virus DNA causing increased activity of the HBV enhancer. *Virology* 167: 630-633.
244. Tuttleman, JS, JC Pugh and JW Summers (1986) In vitro experimental infection of primary duck hepatocyte cultures with duck hepatitis B virus. *J Virol* 58: 17-25.
245. Valliammai, T, JM Echevarria, P Leon, C Tusets and TJ Harrison (1995) Amplification and sequence analysis of the precore and core region of the HBV genome from sera of Spanish patients with HBV2-like infections. *J Med Virol* 46: 375-379.
246. van Steenberg, JE, HG Niesters, EL Op de Coul, *et al.* (2002) Molecular epidemiology of hepatitis B virus in Amsterdam 1992-1997. *J Med Virol* 66: 159-165.
247. Vento, S, G di Perri, R Luzzati, *et al.* (1989) Clinical reactivation of hepatitis B in anti-HBs-positive patients with AIDS. *Lancet* 1: 332-333.
248. Viazov, S, S Kuzin, N Paladi, *et al.* (1997) Hepatitis C virus genotypes in different regions of the former Soviet Union (Russia, Belarus, Moldova, and Uzbekistan). *J Med Virol* 53: 36-40.

## References

---

249. Vieth, S, C Manegold, C Drosten, T Nippraschk and S Gunther (2002) Sequence and phylogenetic analysis of hepatitis B virus genotype G isolated in Germany. *Virus Genes* 24: 153-156.
250. Wang, J, AS Lee and JH Ou (1991) Proteolytic conversion of hepatitis B virus e antigen precursor to end product occurs in a postendoplasmic reticulum compartment. *J Virol* 65: 5080-5083.
251. Wasenauer, G, J Kock and HJ Schlicht (1992) A cysteine and a hydrophobic sequence in the noncleaved portion of the pre-C leader peptide determine the biophysical properties of the secretory core protein (HBe protein) of human hepatitis B virus. *J Virol* 66: 5338-5346.
252. Weber, B, T Dengler, A Berger, HW Doerr and H Rabenau (2003) Evaluation of two new automated assays for hepatitis B virus surface antigen (HBsAg) detection: IMMULITE HBsAg and IMMULITE 2000 HBsAg. *J Clin Microbiol* 41: 135-143.
253. Weber, B (2005) Recent developments in the diagnosis and monitoring of HBV infection and role of the genetic variability of the S gene. *Expert Rev Mol Diagn* 5: 75-91.
254. Weber, B, N Van der Taelen-Brule, A Berger, *et al.* (2006) Evaluation of a new automated assay for hepatitis B surface antigen (HBsAg) detection VIDAS HBsAg Ultra. *J Virol Methods* 135: 109-117.
255. Weber, M, V Bronsema, H Bartos, *et al.* (1994) Hepadnavirus P protein utilizes a tyrosine residue in the TP domain to prime reverse transcription. *J Virol* 68: 2994-2999.
256. Weinberger, KM, T Bauer, S Bohm and W Jilg (2000) High genetic variability of the group-specific a-determinant of hepatitis B virus surface antigen (HBsAg) and the corresponding fragment of the viral polymerase in chronic virus carriers lacking detectable HBsAg in serum. *J Gen Virol* 81: 1165-1174.
257. WHO (2000). World Health Organization. World Health Organization Fact Sheet 2004 (Revised October 2000).
258. Will, H, W Reiser, T Weimer, *et al.* (1987) Replication strategy of human hepatitis B virus. *J Virol* 61: 904-911.
259. Wu, HL, PJ Chen, MH Lin and DS Chen (1991) Temporal aspects of major viral transcript expression in Hep G2 cells transfected with cloned hepatitis B virus DNA: with emphasis on the X transcript. *Virology* 185: 644-651.
260. Xu, Z, TS Yen, L Wu, *et al.* (2002) Enhancement of hepatitis B virus replication by its X protein in transgenic mice. *J Virol* 76: 2579-2584.
261. Xu, ZY, CB Liu, DP Francis, *et al.* (1985) Prevention of perinatal acquisition of hepatitis B virus carriage using vaccine: preliminary report of a randomized, double-blind placebo-controlled and comparative trial. *Pediatrics* 76: 713-718.

## References

---

262. Yang, SQ, M Walter and DN Standring (1992) Hepatitis B virus p25 precore protein accumulates in *Xenopus* oocytes as an untranslocated phosphoprotein with an uncleaved signal peptide. *J Virol* 66: 37-45.
263. Yeh, CT, YF Liaw and JH Ou (1990) The arginine-rich domain of hepatitis B virus precore and core proteins contains a signal for nuclear transport. *J Virol* 64: 6141-6147.
264. Yokosuka, O, M Omata, YZ Zhou, F Imazeki and K Okuda (1985) Duck hepatitis B virus DNA in liver and serum of Chinese ducks: integration of viral DNA in a hepatocellular carcinoma. *Proc Natl Acad Sci U S A* 82: 5180-5184.
265. Yoo, TW, S Donfield, A Lail, HS Lynn and ES Daar (2005) Effect of hepatitis C virus (HCV) genotype on HCV and HIV-1 disease. *J Infect Dis* 191: 4-10.
266. Yu, MC, JM Yuan, S Govindarajan and RK Ross (2000) Epidemiology of hepatocellular carcinoma. *Can J Gastroenterol* 14: 703-709.
267. Yuen, MF, CC Cheng, IJ Lauder, *et al.* (2000) Early detection of hepatocellular carcinoma increases the chance of treatment: Hong Kong experience. *Hepatology* 31: 330-335.
268. Yuen, MF, WK Seto, DH Chow, *et al.* (2007) Long-term lamivudine therapy reduces the risk of long-term complications of chronic hepatitis B infection even in patients without advanced disease. *Antivir Ther* 12: 1295-1303.
269. Zein, NN and DH Persing (1996) Hepatitis C genotypes: current trends and future implications. *Mayo Clin Proc* 71: 458-462.
270. Zein, NN, J Rakela, EL Krawitt, *et al.* (1996) Hepatitis C virus genotypes in the United States: epidemiology, pathogenicity, and response to interferon therapy. Collaborative Study Group. *Ann Intern Med* 125: 634-639.
271. Zoulim, F, J Saputelli and C Seeger (1994) Woodchuck hepatitis virus X protein is required for viral infection in vivo. *J Virol* 68: 2026-2030.

# Annexes

## 1. Conference participations

### 1.1. Talks

2004	• SarLorLux meeting, Homburg, Germany	13.09.
2005	• SarLorLux meeting, Luxembourg, Luxembourg	21.09.
	• Société Luxembourgeoise de Biologie Clinique (SLBC), Luxembourg, Luxembourg	23.09 - 24.09.
2006	• SarLorLux meeting, Nancy, France	13.09.
	• Gesellschaft für Virologie, Munich, Germany	15.03. - 18.03.
2007	• Molecular Diagnosis of Infectious Diseases Minsk, Belarus	17.05. - 18.05.
	• Annual Meeting of Reference labs for AIV and ND, Stralsund, Germany	23.05. - 25.05.
	• SarLorLux meeting, Homburg, Germany	10.09

### 1.2. Posters

2005	• European Society of Clinical Virology (ESCV), Geneva, Switzerland	27.04. - 30.04.
	• International Meeting on the Molecular Biology of Hepatitis B viruses, Heidelberg, Germany	18.09. - 21.09.

## 2. Publications

1. **Olinger CM**, Jutavijittum P, Hübschen J, Yousukh A, Samounry B, Thammavong T, Toriyama K and Muller CP. (2008) Identification and molecular characteristics of a new Hepatitis B genotype I in Lao People's Democratic Republic. Submitted
2. **Olinger CM**, Lazouskaya NV, Eremin VF and Muller CP. (2008) Multiple genotypes and subtypes of hepatitis B and C viruses in Belarus: similarities with Russia and West European influences. Clinical Microbiology and Immunology. In press.
3. **Olinger CM**, Weber B, Otegbayo JA, Ammerlaan W, van der Taelem-Brulé N, Muller CP. (2007) Hepatitis B virus genotype E surface antigen detection with different immunoassays and diagnostic impact of mutations in the preS/S gene. Medical Microbiology and Immunology. 196:247-252.
4. Bekondi C, **Olinger CM**, Boua N, Talarmin A, Muller CP, Le Faou A, Venard V. (2007) Central African Republic is part of the West-African hepatitis B virus genotype E crescent. Journal of Clinical Virology. 40:31-37.
5. Billing AM, Fack F, Renaut J, **Olinger CM**, Schote AB, Turner JD, Muller CP. (2007) Proteomic analysis of the cortisol-mediated stress response in THP-1 monocytes using DIGE technology. Journal of Mass Spectrometry. 42:1433-1444.
6. **Olinger CM**, Ducatez MF, Owoade AA, Tarnagda Z, Tahita MC, Sow A, De Landtsheer S, Ammerlaan W, Ouedraogo JB, Osterhaus AD, Fouchier RA, Muller CP. (2007) Molecular and antigenic evolution and geographical spread of H5N1 highly pathogenic avian influenza viruses in western Africa. Journal of General Virology. 88:2297-22306.
7. Nébié KY, **Olinger CM**, Kafando E, Dahourou H, Diallo S, Kientega Y, Domo Y, Kienou K, Ouattara S, Sawadogo I, Kya L, Muller CP. (2007) Lack of knowledge of blood donors in Burkina Faso (West Africa): Potential obstacle to transfusion security. Transfusion Clinique et Biologique, In press.

8. **Olinger CM**, Venard V, Njayou M, Oyefolu AO, Maïga I, Kemp AJ, Omilabu SA, le Faou A, Muller CP. (2006) Phylogenetic analysis of the precore/core gene of hepatitis B virus genotypes E and A in West Africa: new subtypes, mixed infections and recombinations. Journal of General Virology. 87:1163-1173.
9. Ducatez MF, **Olinger CM**, Owoade AA, De Landtsheer S, Ammerlaan W, Niesters HG, Osterhaus AD, Fouchier RA, Muller CP. (2006) Avian flu: multiple introductions of H5N1 in Nigeria. Nature. 442:37.
10. Ducatez MF, **Olinger CM**, Owoade AA, De Landtsheer S, Ammerlaan W, Niesters HG, Osterhaus AD, Fouchier RA, Muller CP. (2006) L'Etude de l'épidémiologie moléculaire de H5N1 hautement pathogène au Nigéria indique que plusieurs virus distincts y ont été introduits. Epidémiologie et Santé Animale. 50.
11. Kremer JR, Fack F, **Olinger CM**, Mulders MN, Muller CP. (2004) Measles virus genotyping by nucleotide-specific multiplex PCR. Journal of Clinical Microbiology.42:3017-3022.

## Acknowledgements

I would like to thank my promoter Prof. Dr. Claude P. Muller for offering me the opportunity of doing this work as a member of his team in the Institute of Immunology of the Laboratoire National de Santé in Luxembourg and for his valuable scientific guidance and continuous confidence during the last years.

I would like to thank the numerous collaborators in Nigeria, Cameroon, Burkina Faso, Mali, Benin, Togo, the Central African Republic, the Democratic Republic of Congo, Lao PDR, Thailand, Belarus, France, Germany and Luxembourg, without whom none of this work would have been possible. I would like to especially mention Patrick Mbah, Joseph Forbi, Jesse Otegbayo, Prapan Jutavijitum, Bola Oyefolu, Natallia Lazouskaya, Ademola Owoade, Zekiba Tarnagda and Nebie Yacouba, for giving me the possibility to share all my scientific knowledge with them and to show them Luxembourg but foremost, for expanding my field of view and sharing with me their cultures and views of this world.

I would like to thank Prof. Dr. Bernard Weber from the Laboratoires Réunion, Junglinster for his extensive support during the quantification and serology experiments and his confidence in my scientific expertise.

I want to thank all the present and former colleagues of the Institute, especially Wim for being the best companion during the time in Nigeria and Mick, for convincing me to start a career in science after all. Also, Jacques, Mariette, Judith, Tom, Carole, Ulla, Fred, Andrea, Sebastien Stefan, Fabienne, Corinne, Sophie, Stéphanie, Jon, Nathalie, Anna, Aurélie, Chantal, Denis, Dominique, Emilie, Iris, Joana, Julia, Latitia, Lynn, Alain, Lei, Mario, Nancy, Patrick, Simone and Vitor deserve my sincerest gratitude.

Financial support from the Fonds National de Recherche, the Centre de Recherche Public-Santé, the Ministère de la Culture, de l'Enseignement Supérieur et de la Recherche, the Ministère des Affaires étrangères et de l'Immigration is also gratefully acknowledged.

I would like to thank Olivier, Marc, Michèle, Mike, Manuel, Phillippe, Tom and Alain for their invaluable friendship and support during all these years.

I want to thank my girlfriend Anja for her love and much more than this ...

I wish to thank my parents and my brother and sister for having been with me all my life and always giving me their unconditional support.

# Curriculum

## Personal details

Surname / First name: **Olinger, Christophe M.**

Address: 17, rue d'Ernster, L-6955 Rodenbourg, Luxembourg

Telephone: +352 26787043 (Home), +352 661 789 288 (Mobile)

E-mail: [olingerc@binarylooks.com](mailto:olingerc@binarylooks.com)

Nationality: Luxemburgish

Date of birth: 13.05.1980

## Education and training

October 2004 – February 2008 **PhD**, molecular and cellular biology, virology, immunology, National Public Health Laboratory (LNS), Institute of Immunology, Luxembourg, Luxembourg  
Universität des Saarlandes, medical faculty, Homburg, Germany

September 2003 – July 2004 **Master of Science (DEA)**, molecular and cellular biology, microbiology  
DKFZ, Heidelberg, Germany  
Université Louis Pasteur, Strasbourg, France

September 2001 – July 2003 **Bachelor of Science (Maîtrise)**, molecular and cellular biology, microbiology  
Université Louis Pasteur, Strasbourg, France

September 1999 – July 2001 **DEUG**, biology and chemistry  
Centre Universitaire, Luxembourg, Luxembourg