

https://dspace.org/

Molecular characterization of hepatitis B virus (HBV) from mono-infected and HBV/human immunodeficiency virus (HIV) co-infected individuals in Sudan

Yousif, Mukhlid

2014-09-09
http://hdl.handle.net/10539/15438

Downloaded from a DSpace repository. https://dspace.org/

MOLECULAR CHARACTERIZATION OF HEPATITIS B VIRUS (HBV) FROM MONO-INFECTED AND HBV/HUMAN IMMUNODEFICIENCY VIRUS (HIV) CO-INFECTED INDIVIDUALS IN SUDAN

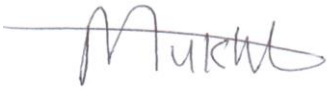
Mukhlid Yousif

A thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, in fulfilment of the requirements for the degree of Doctor of Philosophy

Johannesburg, June 2014

Declaration

I, Mukhlid Yousif declare that this thesis is my own work. It is being submitted for the degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.



Signed: .....

09 day of June 2014

Structure of the thesis

- Chapter 1: Introduction and Literature Review.
- Chapter 2: **Paper I:** *Genotype D of hepatitis B virus and its subgenotypes: An update*.
- Chapter 3: **Paper II:** *Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan.*
- Chapter 4: **Paper III:** *Genotyping and virological characteristics of hepatitis B virus in human immunodeficiency virus infected individuals in Sudan.*
- Chapter 5: **Paper IV:** *Analysis of ultra-deep pyrosequencing and cloning based sequencing of the basic core promoter/precore/core region of hepatitis B virus using newly developed bioinformatics tools.*
- Chapter 6: General Discussion and Conclusion
- Appendix A: **Paper V***: Overt and occult hepatitis B virus infection in adult Sudanese HIV patients*.
- Appendices B to F.

References for literature cited in the published papers appear within the reference section of the papers themselves. References for all other citations appear in the References section at the end of the thesis. Some references may appear in the reference lists of the papers and of the thesis.

Abstract

Hepatitis B virus (HBV), the prototype member of the family *Hepadnaviridae,* is hepatotropic and replicates by reverse transcription. HBV is responsible for the chronic infection of more than 240 million people worldwide, of which 65 million reside in Africa. The nine HBV genotypes (A to I) identified to date, are geographically distributed and exhibit different clinical manifestations and treatment responses. The term occult HBV infection (OBI) refers to a HBV infection in which HBV surface antigen (HBsAg) cannot be detected by conventional serological assays as has been defined by the Taormina expert panel. . HBV and human immune deficiency virus (HIV) are both endemic in many parts of the world and share common transmission routes. Worldwide, 10% of those infected with HIV are also chronically infected with HBV. HIV co-infection has been shown to be a risk factor for the development of OBI in individuals infected with HBV.

The aim of this study was to characterize, at the molecular level, HBV from mono-infected and HBV/HIV co-infected individuals in Sudan

The objectives of this study were the systematic and comparative analysis of HBV genotype D sequences, available in the public databases; the molecular characterization of HBV from mono-infected Sudanese liver disease patients and from HBV/HIV co-infected Sudanese patients; and the development and testing of bioinformatics tools to explore HBV sequence data generated using ultradeep pyrosequencing (UDPS) and comparison of UDPS results with those obtained from cloning based sequencing (CBS).

All available complete genomes of genotype D of HBV from the GenBank database were analyzed. The intra-group divergence of the subgenotypes ranged from 0.8% $\pm$ 0.5 for subgenotype D6 to 3.0% $\pm$ 0.3 for subgenotype D8. Phylogenetic analysis of genotype D

showed separation into six distinct clusters (subgenotypes D1, D2, D3/D6, D4, D5 and D7/D8), with good bootstrap support. The mean intergroup divergence between subgenotype D3 and subgenotype D6 was 2.6%, falling below the accepted threshold of 4% required to define a subgenotype. This suggests that subgenotypes D3 and D6 are the same subgenotype because they also share signature amino acids. Furthermore, subgenotype D8 is a genotype D/E recombinant, which clusters with subgenotype D7. This analysis provided an update on the classification of the subgenotypes of genotype D of HBV.

Although HBsAg seroprevalence in Sudan, a central-African country, is greater than 8%, the only sequencing data for HBV, available prior to our study, was from asymptomatic blood donors, where genotype E predominates, followed by genotype D and subgenotype A2. Ninety-nine HBV-positive liver disease patients were enrolled in our study, including: 15 with hepatocellular carcinoma (HCC), 42 with cirrhosis, 30 asymptomatic carriers, 7 with acute hepatitis and 5 with chronic hepatitis. The surface and basic core promoter/precore (BCP/PC) regions, and the complete genome of HBV were sequenced. Eighty-two percent of the samples from HBV mono-infected liver disease patients were genotyped. Fifty-nine percent were infected with genotype D (74% D1, 10% D2, 3% D3 and 13% D6), 30% with genotype E, 8.5% with genotype A and 2.5% with a genotype D/E recombinant. Patients infected with genotype E had a higher frequency of HBeAg-positivity (29.2%) and higher viral loads compared to patients infected with genotype D. BCP/PC region mutations, including the G1896A mutation, seen in 37% of the HBeAg-negative individuals, could account for the HBeAg-negativity.

A total of 358 Sudanese HIV-positive patients were enrolled. HBsAg was detected in 11.7% of the participants, indicating chronic HBV infection. HBV DNA was detected in 26.8% of the participants: 11.7% were HBsAg positive (overt infection) and the remaining 15.1% were

HBsAg-negative (OBI). Fifty serum samples from the HBV/HIV DNA-positive co-infected participants were selected for genomic analysis of HBV. Of these, the HBV genotype of 37 was determined. The genotype distribution of HBV isolates from the HBV/HIV co-infected participants did not differ significantly from those from the HBV mono-infected patients: genotype D (46%), E (21.6%), A (18.9%) and a D/E recombinant (13.5%). Compared to the HBV isolates from mono-infected liver disease patients, the frequency of the D/E recombinant and genotype A was higher in HBV/HIV co-infected patients, as was the intra-group divergence of genotype E. No difference in BCP/PC mutations affecting HBeAg expression at the transcriptional and translational levels between genotype D and E was observed. The following mutations could account for the HBsAg-negativity: sM133T, sE164G, sV168G and sS174N. No primary drug resistance mutations were found.

Two online bioinformatics tools, the “Deep Threshold Tool (DDT)” and the “Rosetta Tool”, were built to analyze data generated from UDPS and CBS of the BCP/PC region of four Sudanese serum samples, infected with either genotype D or E of HBV, from HBeAg-positive and HBeAg negative patients. A total of 10952 reads were generated by UDPS on the 454 GS Junior platform. The Threshold was calculated using DDT based on probability of error of 0.5%. In total, 39 unique mutations were identified by UDPS, of which 25 were non-synonymous. The ratio of nucleotide substitutions between isolates from HBeAg-negative and HBeAg-positive patients was 3.5:1. From the sequences analyzed, compared to genotype E isolates, genotype D isolates showed greater variation in the X, BCP/PC/C regions. Only 18 of the 39 positions identified by UDPS were detected by CBS.

Using the specific criteria, that have been suggested previously, to define genotypes/subgenotypes of HBV, we determined that genotype D has six and not eight subgenotypes. The importance of HBV genotypes in clinical consequences of infection and

response to antiviral treatment has led us to characterize HBV genotypes circulating in Sudan. HBV mono-infected patients and HBV/HIV co-infected individuals, were mainly infected with genotype D or E. HBV mono-infected patients, infected with genotype E, had higher HBeAg-positivity and higher viral loads than those infected with genotype D. The ratio of genotype A to non- A, as well as the genotype E intra-group divergence were higher in HBV/HIV co-infected individuals compared to HBV mono-infected individuals. OBI was found in 15.1% HBV/HIV co-infected patients and its clinical relevance remains to be determined. In order to overcome the limitations of Sanger sequencing, which include its high cost and inability to detect minor populations in quasispecies, next generation sequencing techniques have been developed. It was demonstrated that correct analysis of UDPS data required appropriate curation of read data, in order to clean the data and eliminate artefacts and that the appropriate consensus (reference) sequence should be used in order to identify variants correctly. CBS detected fewer than 50% of the substitutions detected by UDPS. This new technology may allow the detection of minor variants between the different genotypes of HBV and provide biomarkers for the prediction of clinical manifestation of HBV and response to antiviral therapy.

Research outputs

1.1. Publications

Publications included as chapters in this thesis are marked with an asterisk*

- ***YOUSIF, M**. & KRAMVIS, A. 2013. Genotype D of hepatitis B virus and its subgenotypes: An update. Hepatology Research, 43, 355-364. *http://onlinelibrary.wiley.com/doi/10.1111/j.1872-034X.2012.01090.x/full* **[Chapter 2]**
- ***YOUSIF, M**., MUDAWI, H., BAKHIET, S., GLEBE, D. & KRAMVIS, A. 2013. Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan. BMC Infectious Diseases, 13, 328. **[Chapter 3]**
- ***YOUSIF, M**., MUDAWI, H., HUSSEIN, W., MUKHTAR, M., NEMERI, O., GLEBE, D. & KRAMVIS, A. Genotyping and Virological Characteristics of Hepatitis B Virus in Human Immunodeficiency Virus Infected Individuals in Sudan . *Revision submitted* to International Journal of Infectious Disease, 2014. **[Chapter 4]**
- MUDAWI, H., HUSSEIN, W., MUKHTAR, M., **YOUSIF, M**., NEMERI, O., GLEBE, D. & KRAMVIS, A. Overt and occult hepatitis B virus infection in adult Sudanese HIV patients. *Revision submitted* to International Journal of Infectious Disease, 2014. **[Appendix A]**
- ***YOUSIF, M**, Bell, T G, MUDAWI, H., GLEBE, D. & KRAMVIS, A. Analysis of Ultra-deep Pyrosequencing and Cloning Based Sequencing of the Basic Core Promoter/Precore/Core Region of Hepatitis B Virus Using Newly Developed Bioinformatics Tools. PlosOne, 2014. **[Chapter 5]**

2.1. Conferences outputs

- **YOUSIF, M**., MUDAWI, H., BAKHIET, S., GLEBE, D., KRAMVIS, A. 2014. *Occult hepatitis B virus infection and reactivation in Africans in the context of the human immunodeficiency virus pandemic.* German Research Foundation (DFG) Meeting: German-African Cooperative Projects in Infectology. Oral presentation. Dar es Salam, Tanzania (January 2014).

- **YOUSIF, M.** & KRAMVIS, A. 2012. *Genotype D of hepatitis B virus and its subgenotypes: An update*. Research Day 2012 Poster Presentation. Faculty of Health Sciences, University of the Witwatersrand, South Africa (September 2012).

- **YOUSIF, M**., MUDAWI, H., BAKHIET, S., GLEBE, D., KRAMVIS, A. 2012. *Occult hepatitis B virus infection and reactivation in Africans in the context of the human immunodeficiency virus pandemic.* German Research Foundation (DFG) Meeting: German-African Cooperative Projects in Infectology. Oral and poster presentation. Bonn, Germany (June 2012).

- **YOUSIF, M**., MUDAWI, H., BAKHIET, S., GLEBE, D., KRAMVIS, A. 2011. *Molecular characterization of hepatitis B virus (HBV) from liver diseases patients and asymptomatic carriers of the virus in Sudan.* Oral presentation at "Virology Africa" 2011. University of Cape Town, Cape Town, South Africa (December 2011).

- **YOUSIF, M**., MUDAWI, H., BAKHIET, S., GLEBE, D., KRAMVIS, A. 2011. *Molecular characterization of hepatitis B virus (HBV) from HBV mono-infected and from HBV/human immunodeficiency virus (HIV) co-infected individuals in Sudan.* Hepatitis Virus Diversity Research Programme (HVDRP) Symposium 2011. Hepatitis B Vi-

rus Genotypes: *Global diversity and Evolution.* Oral presentation. University of the Witwatersrand, Johannesburg, South Africa (September 2011).

- **YOUSIF, M**., MUDAWI, H., BAKHIET, S., GLEBE, D., KRAMVIS, A. 2011. *Molecular characterization of hepatitis B virus (HBV) from HBV mono-infected and from HBV/human immunodeficiency virus (HIV) co-infected individuals in Sudan.* German Research Foundation (DFG) Meeting 2011. German-African Cooperative Projects in Infectology. Poster presentation. Accra, Ghana (March 2011).

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Anna Kramvis, for giving me the opportunity to undertake my PhD research in the Hepatitis Virus Diversity Research Programme. I feel privileged to have worked under her supervision and to have been a part of her research group. I am very thankful to Professor Kramvis for the continuous support she gave me during my PhD studies, for her patience, motivation, enthusiasm, encouragement, and immense knowledge. Her guidance helped me throughout my research and during the writing of this thesis. I could not imagine having a better supervisor and mentor for my PhD.

My thanks are extended to all members of the Hepatitis Virus Diversity Research Programme, for their support and encouragement. A special thanks to Dr. Nimisha Bhoola for her help with cloning and to Dr. Trevor Bell for his collaboration on the Pyrosequencing project and for proof-reading parts of this thesis.

I would like to acknowledge and thank Dr. Hatim Mudawi for his support and guidance during the early phase of this study, while I was collecting serum samples, and for referring patients who were enrolled in the study. Thanks also to Dr. Dieter Glebe for his collaboration in our project.

I would also like to thank the staff at IbnSina Hospital and Khartoum Hospital for assisting in collecting serum samples from HBV-infected patients, and Mr Waleed Abdallateif, who collected serum samples from HBV/HIV co-infected patients.

Funding for both bursaries and running expenses for this work was received from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). I thank them for making this funding available, as this work would not have been possible without their financial support.

To my parents, Imam and Alawia, for their unconditional love and support, and for believing in me; to my brother Madi and my sister Mayada, for their constant support along the way, and to Reem and Noora, my ladies, I will always love you.

Table of Contents

Declaration i

Structure of the thesis ii

Abstract iii

Research outputs vii

1.1. Publications vii

2.1. Conferences outputs viii

Acknowledgements x

List of Figures xv

List of Tables xvi

List of Abbreviations xvii

Dedication xx

CHAPTER 1: Introduction 1

1.1 The Hepatitis B Virus 1

1.1.1 Classification 1

1.1.2 HBV virion structure 2

1.1.3 HBV Genome Structure 4

1.1.4 HBV Replication 5

1.1.5 HBV genotypes 15

1.1.6 Clinical relevance of the HBV genotypes 22

1.1.7 HBV recombination 24

1.1.8 HBV mutation 25

1.2 Epidemiology, Transmission and Natural History of HBV infection 29

1.2.1 Epidemiology 29

1.2.2 Transmission 30

1.2.3 Natural History of HBV infection 31

1.3 HBV/ HIV co-infection 37

1.3.1 HBV, HIV and HBV/HIV co-infection in Sudan 38

1.4 Nucleic acid sequencing techniques 41

1.4.1 First generation sequencing technologies: 42

1.4.2 Second (next) generation sequencing technologies (NGS): 43

1.5 Aims and Objectives 48

Chapter 2: Paper I 49

Genotype D of hepatitis B virus and its subgenotypes: An update 49

Chapter 3: Paper II 65

Molecular Characterization of Hepatitis B Virus in Liver Disease Patients and Asymptomatic Carriers of the Virus in Sudan 65

Chapter 4: Paper III 77

Genotyping and Virological Characteristics of Hepatitis B Virus in Human Immunodeficiency Virus Infected Individuals in Sudan 77

Chapter 5: Paper IV 112

Analysis of Ultra-deep Pyrosequencing and Cloning Based Sequencing of the Basic Core

Promoter/Precore/Core Region of Hepatitis B Virus Using Newly Developed Bioinformatics Tools....112

Chapter 6: General Discussion and Conclusion....127

Appendix A: Paper v....137

Overt and occult hepatitis B virus infection in adult Sudanese HIV patients.....137

Appendix B....163

Appendix C....167

Appendix D....170

Appendix E....174

Appendix F....176

Appendix G....181

References....185

List of Figures

Figure 1.1: The Structure of the Hepatitis B Virion 3

Figure 1.2: Structure and organization of hepatitis B virus genome..... 5

Figure 1.3: Hepatitis B virus replication cycle……..…..... 7

Figure 1.4: A schematic representation of the hepatitis B virus polymerase shows the length of each polymerase domain in amino acids for the different genotypes..... 12

Figure 1.5: The structure of the encapsidation signal of *Hepadnaviridae*..... 14

Figure 1.6: HBV genotypes. Phylogenetic tree shows the 9 HBV genotypes..... 16

Figure 1.7: Geographic distribution of chronic Hepatitis B virus infection..... 30

Figure 1.8: Natural history of Hepatitis B virus infection……..... 32

Figure 1.9: Overview of the 454 sequencing system..... 46

List of Tables

Table 1.1: Virological characteristics of HBV genotypes..... 16

Table 1.2: Clinical characteristics associated with HBV genotype..... 24

Table 1.3: Prevalence of HBV, HIV and HBV/HIV co-infection in Sudan.....40

List of Abbreviations

AH	Acute hepatitis
ALT	Alanine aminotransferase
Anti-HBc	Antibodies to HBV core antigen
Anti-HBe	Antibodies to HBV e antigen
Anti-HBs	Antibodies to HBV surface antigen
ASC	Asymptomatic carriers
ASHV	Arctic squirrel hepatitis virus
BCP	Basal core promoter
BQW	Best quality water
CBS	Cloning based sequencing
cccDNA	Covalently closed DNA
CCD	Cooled Charge-Coupled Device camera
CH	Chronic hepatitis
ChHBV	Chimpanzee hepatitis B
CR	Cirrhosis
CURS	Core upstream regulatory sequences
DHBV	Duck hepatitis B
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
ELISA	Enzyme linked immunosorbent assay
Enh1	Enhancer I
Enh2	Enhancer II
GiHBV	Gibbon hepatitis virus
GoHBV	Gorilla hepatitis virus

GSHV	Ground squirrel hepatitis virus
HBcAg	Hepatitis B core antigen
HBeAg	Hepatitis B e antigen
HBsAg	Hepatitis B surface antigen
HBV	Hepatitis B virus
HBx	Hepatitis B X protein
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HHBV	Heron hepatitis virus
HIV	Human immunodeficiency virus
HNF	Hepatocyte nuclear factors
ICT	Immunochromatography test
IQR	Interquartile range
LHBs	Hepatitis B virus large surface protein
MHBs	Hepatitis B virus medium surface protein
mRNA	messenger RNA
NGS	Next generation sequencing
NTCP	Sodium taurocholate cotransporting polypeptide
OBI	Occult hepatitis B infection
ORF	Open reading frame
OuH	Orang-utan hepatitis
P	Polymerase gene
PCR	Polymerase chain reaction
PgRNA	Pregenomic RNA
PreC/C	Precore/Core gene

PreS/S	PreS/Surface gene
rcDNA	Relaxed Circular DNA
RFLP	Restriction fragment length polymorphism
RGHBV	Ross’s goose HBV
RH	RNase H domain of the polymerase
RNA	Ribonucleic acid
RT	Reverse transcriptase domain of the polymerase
SGHBV	Snow goose HBV
SHBs	Hepatitis B virus small surface protein
SNAP	Sudan National AIDS control Programme
SP	Spacer domain of the polymerase
TP	Terminal protein domain of the polymerase
UPDS	Ultra deep pyrosequencing
URR	Upstream regulatory region
USA	United States of America
WHO	World Health Oranization
WHV	woodchuck hepatitis
WMHV	Woolly monkey hepatitis B virus
ε	Encapsidation signal

Dedication

I dedicate this thesis to my beloved family

CHAPTER 1: INTRODUCTION

Hepatitis B virus (HBV) infects the liver and causes acute and chronic liver disease. The majority of acute HBV infections are self-limiting, whereas chronic HBV infection usually has a lifelong course that may lead to the development of cirrhosis and hepatocellular carcinoma (HCC; liver cancer) (Donato *et al.,* 1998). Despite the presence of an effective prophylactic vaccine it is estimated that there are 2 billion people who have been infected with HBV during their lifetime and over 240 million are chronic carriers of the virus (WHO, 2012), with 65 million chronic carriers residing in Africa (Kramvis and Kew, 2007).

1.1 The Hepatitis B Virus

1.1.1 Classification

HBV, with a genome of 3.1-3.2 kb is the smallest DNA virus infecting man and is the prototype member of the family *Hepadnaviridae*. This family consists of 2 genera: the *Orthrohepadnaviridae,* which infect mammals and the *Avihepadnaviridae,* which infect birds. The members of the genus *Orthohepadnavirus* include: HBV, the woodchuck hepatitis virus (WHV) (Hodgson and Michalak, 2001), the ground squirrel hepatitis virus (GSHV) (Marion *et al.*, 1980), the arctic squirrel hepatitis virus (ASHV) (Testut *et al.*, 1996), the chimpanzee HBV (ChHBV) (Takahashi *et al.*, 2000), the gorilla HBV (GoHBV) (Makuwa *et al.*, 2003), the gibbon HBV (GiHBV) (Norder *et al.*, 1996), the orang-utan hepatitis (OuH) (Takahashi *et al.*, 2000), the woolly monkey hepatitis B virus (WMHV) (Lanford *et al.*, 1998), the Barbay macaques hepatitis virus (Gheit *et al.*, 2002), the tree shrew hepatitis virus (Cao *et al.*, 2003) and recently, Dickens *et al*., 2013 provided evidence for occult HBV infection (OBI) in chacma baboons (Dickens *et al.*, 2013). The genus *Avihepadnavirus* consists of the duck hepatitis B virus (DHBV) (Mason *et al.*, 1980), the heron hepatitis B virus (HHBV)

(Sprengel *et al.*, 1988), the snow goose HBV (SGHBV) (Chang *et al.*, 1999), the Ross's goose HBV (RGHBV) (Pult *et al.*, 2001), the white stork HBV (STHBV) (Pult *et al.*, 2001) and grey crowned crane HBV (CHBV) (Prassolov *et al.*, 2003).

1.1.2 HBV virion structure

HBV is an enveloped DNA virus with a diameter between 42 - 47 nm (Tiollais *et al.*, 1985). The viral DNA and DNA polymerase are contained in the icosahedral nucelocapsid composed of core protein (HBcAg), enclosed by the outer lipid membrane (envelope) in which large (LHBs), middle (MHBs) and small (SHBs) surface proteins are embedded (Kann, 2002). This complete infectious virion is known as the Dane particle (Dane *et al.*, 1970, Seeger and Mason, 2000, Locarnini *et al.*, 2003). In addition, non-infectious filamentous and spherical particles, 22 nm in diameter, which lack the core protein and viral nucleic acid, can circulate in the serum. These are composed only of the lipid membrane containing the surface proteins (Seeger and Mason, 2000, Locarnini *et al.*, 2003). The HBV virion structure is illustrated in Figure 1.1.

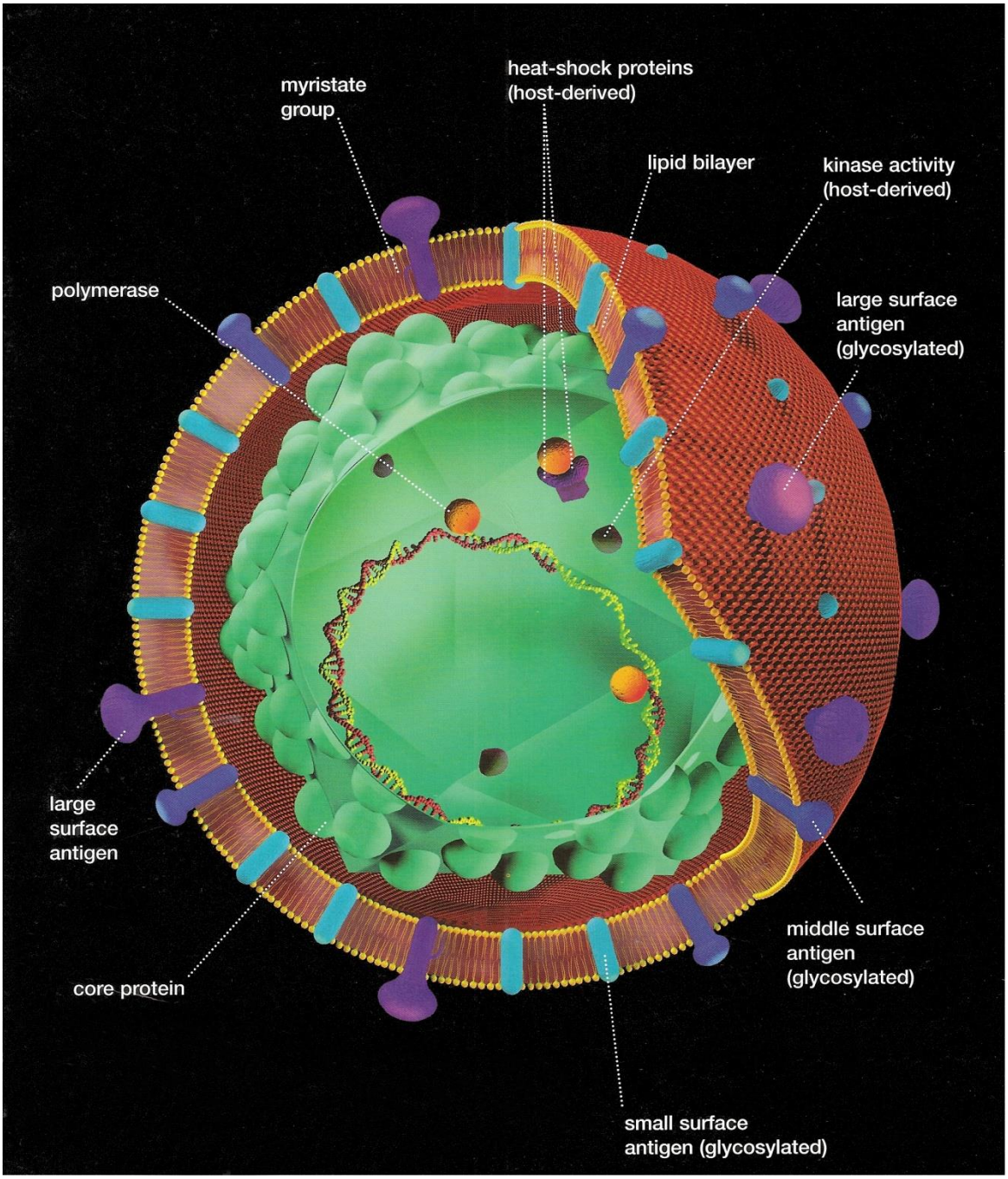


Figure 1.1: The Structure of the Hepatitis B Virion

Reproduced with permission from [Hepatitis B Virus, Human Virus Guide, Second Edition, (Lai and Locarnini, 2002). © International Medical Press].

1.1.3 HBV Genome Structure

As shown in Figure 1.2 the HBV genome is composed of circular, partially double-stranded DNA, ranging in size between 3182 to 3248 nucleotides, depending on the genotype (Feitelson, 1994). The partially double-stranded DNA is composed of a full-length negative strand (3020 to 3320 nucleotides) and an incomplete positive strand (1600 to 2800 nucleotides) (Beck and Nassal, 2007, Seeger and Mason, 2000). This compact genome consists of four partially overlapping open reading frames (ORFs), which include:

(i) Precore/Core gene (PreC/C), which encodes the hepatitis B e antigen (HBeAg) and the core protein (HBcAg).

(ii) Polymerase gene (P), which encodes the viral polymerase (reverse transcriptase).

(iii) PreS/Surface gene (PreS/S), which encodes: LHBs, MHBs, SHBs envelope proteins

(iv) X gene, which encodes the small regulatory hepatitis B X protein (HBx) (Tiollais *et al.*, 1985, Ganem and Varmus, 1987).

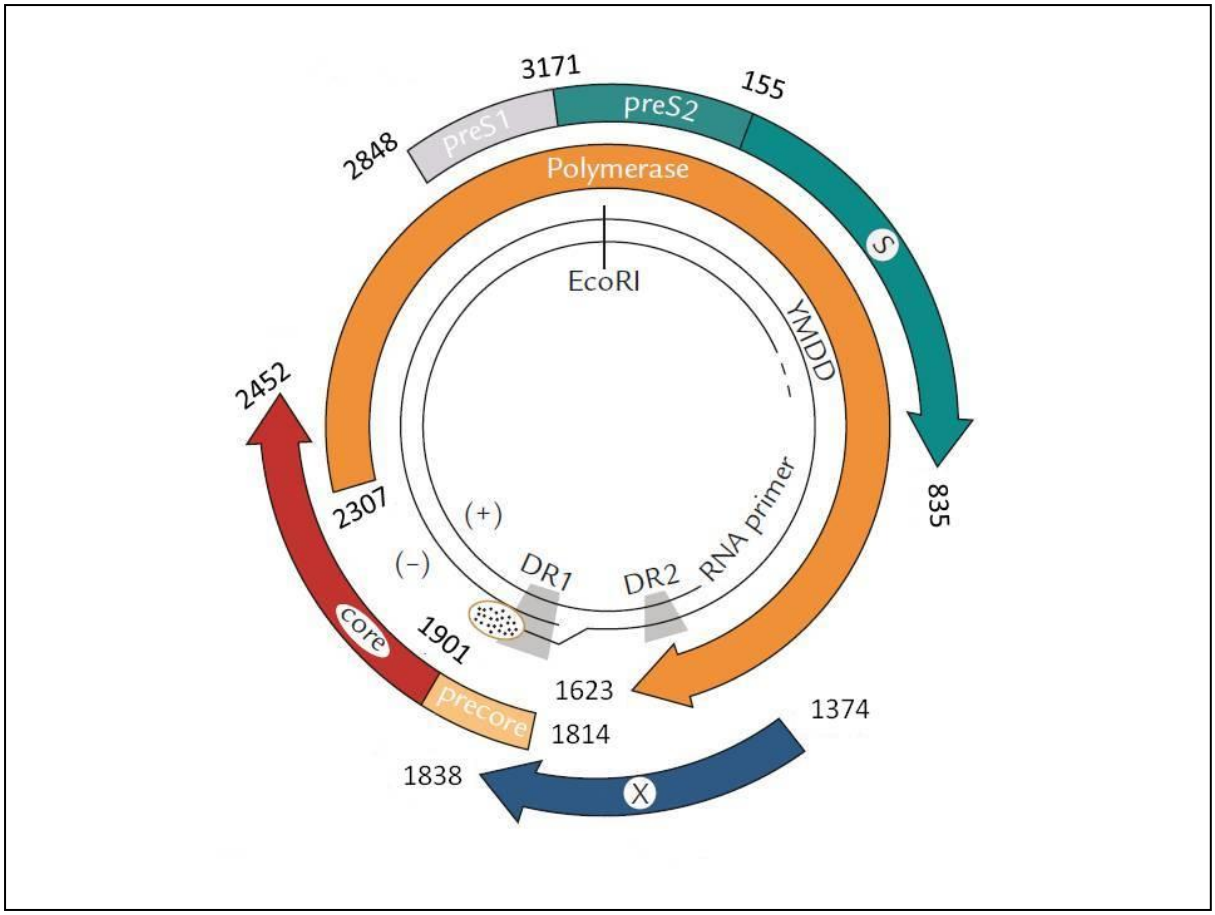


Figure 1.2: Structure and organization of hepatitis B virus genome. Coordinates are relative to genotype D (GU456684). Modified from (Lin and Kao, 2010), with permission from Elsevier.

1.1.4 HBV Replication

The exact mechanism by which HBV replicates is not fully understood. The replication cycle, as illustrated in figure 1.3, begins when the infectious virion attaches to a hepatocyte-specific preS1 receptor (Yan *et al.*, 2012). Evidence, from a recent study, revealed that the receptor-binding region of pre-S1 interacts specifically with sodium taurocholate cotransporting polypeptide (NTCP), which is a multiple trans-membrane transporter, expressed mainly in the liver (Yan *et al.*, 2012). The virus enters the cell in one of two ways: either by fusion and penetration at the plasma membrane or by endocytosis of the viral nuclocapsid followed by fusion to the cell (Pawlotsky, 2005, Xie *et al.*, 2010, Urban *et al.*, 2010). Entry of the virus is

followed by the release of the viral nucleocapsid into the cytoplasm (Urban *et al.*, 2010). The nucleocapsid is transported to the nucleus, where the relaxed circular DNA (rcDNA) is released into the nucloplasm (Kann *et al.*, 2007). Here the rcDNA is repaired by the viral polymerase, which completes the synthesis of the positive strand (Nassal, 2008). Thereafter the ligation of both positive and negative DNA strands occurs and the covalently closed circular DNA (cccDNA) is formed (Nassal, 2008), which serves as the transcriptional template for RNA polymerase II to produce six genomic and subgenomic RNAs:

(i) 3.5 kb pregenomic RNA (pgRNA) translated into the HBcAg and polymerase.

(ii) 3.5 kb preC/C mRNA translated into the HBeAg.

(iii) 2.4 kb preS1 mRNA translated into LHBs (preS1/preS2/S).

(iv) 2.1 kb preS2 mRNA translated into MHBs (preS2/S).

(v) 2.1 kb S mRNA translated into the SHBs (S).

(vi) 0.7 kb X mRNA translated into the X protein ((Summers and Mason, 1982, Moolla *et al.*, 2002, Feitelson, 1992, Ganem and Varmus, 1987, Beck and Nassal, 2007).

This transcription is regulated by both viral and host factors such as viral proteins (core, the regulatory X protein) and CCAAT/enhancer-binding protein (C/EBP)/hepatocyte nuclear factors (HNF), respectively (Levrero *et al.*, 2009).

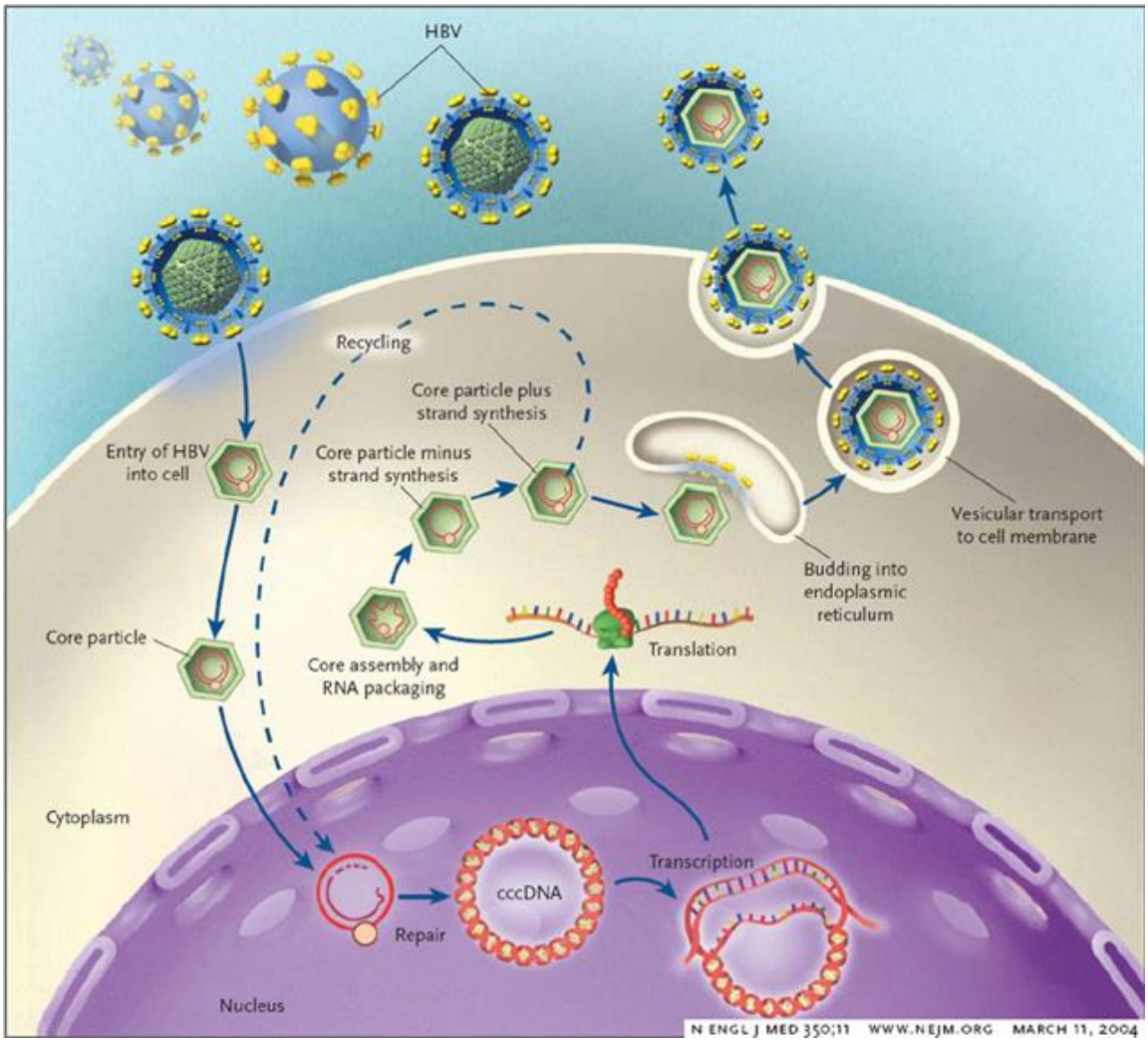


Figure 1.3: Hepatitis B virus replication cycle. Reproduced with permission from (Ganem and Prince, 2004), Copyright Massachusetts Medical Society.

1.1.4.1 *Cis*-acting elements

The synthesis of the six RNA transcripts is controlled by *cis*-acting elements including four promoters and two enhancers that regulate the promoters (Kramvis and Kew, 1999). These promoters are:

(i) The core promoter (CP), which is 232 bp long (1591-1822 from *Eco*R1 site) directs the transcription of the 3.5 kb preC/C mRNA and the 3.5 kb pgRNA. It consists of the basic core promoter (BCP), whose function is to initiate transcription of both pre/C

mRNA and pgRNA. The second part of the CP is the upstream regulatory region (URR), which is made up of core upstream regulatory sequences (CURS) that are known to activate the BCP (Yuh *et al.*, 1992, Kramvis and Kew, 1999, Moolla *et al.*, 2002). In addition, the CP modulates core protein synthesis at post-transcriptional or translational level without affecting polymerase expression (Baumert *et al.*, 1996, Baumert *et al.*, 1998, Kramvis and Kew, 1999).

(ii) The PreS1 promoter, located between bp 2710–2800 from the *Eco*R1 site, regulates the transcription of preS1 mRNA (2.4 kb transcript).

(iii) The PreS2 promoter, at 2960 – 3180 from the *Eco*R1 site, regulates the transcription of preS2 mRNA (2.1 kb transcript).

(iv) The X promoter regulates the transcription of 0.7 kb mRNA (Lu *et al.*, 1995).

The two enhancers that control the promoters are:

(i) Enhancer I, located between bp 1074-1234 from the *Eco*R1 site, which functions to activate the CP (Guo *et al.*, 1991).

(ii) Enhancer II, located 600 nucleotides downstream of enhancer I, between bp 1636 - 1741 from the *Eco*R1 site, which functions to stimulate the transcriptional activity of the two surface promoters and the X promoter (Yuh and Ting, 1993, Yuh and Ting, 1990, Moolla *et al.*, 2002).

All transcripts are exported from the nucleus to the cytoplasm, where seven proteins are translated (Urban *et al.*, 2010, Kann, 2002, Ganem and Prince, 2004).

1.1.4.2 Viral gene products

1.1.4.2.1 Surface proteins

Three distinct surface proteins, LHBs, MHBs and SHBs are embedded in the lipid membrane and serve to protect the nucleocapsid from lysis. These three proteins are translated in the endoplasmic reticulum (ER) of the host cell from three separate start sites and all terminate the same stop codon. The surface proteins exist in two forms, a glycosylated and an unglycosylated form, and are covalently linked to the HBcAg by disulphide bridges formed by cysteine residues (Kann, 2002, Seeger and Mason, 2000).

(i) SHBs: The 226 amino acid S protein (HBsAg) (unglycosylated 24 kD; glycosylated 27 kD) (Gerlich *et al.*, 1992) is the most abundant protein of all three surface proteins and is encoded by the SHBs domain. The loop located between amino acids 124 - 147, forms the major antigenic determinant, known as the 'a' determinant, which is the initial target of the anti-HBs immune response during acute hepatitis (Carman *et al.*, 1990). Amino acid substitutions within this determinant, can lead to conformational changes affecting the binding of neutralising antibodies (Lada *et al.*, 2006, Chongsrisawat *et al.*, 2006, Locarnini *et al.*, 2003, Seeger and Mason, 2000).

(ii) MHBs: The 281 amino acid MHBs protein (preS2/S), encoded by the MHBs domain, consists of the SHBs and a 55 amino acid preS2 (Ganem and Schneider, 2001). MHBs contains overlapping B and T cell epitopes and a binding site for neutralizing antibodies, located between amino acids 120 to 145 (Chisari and Ferrari, 1995) and the transactivation domain (Hildt *et al.*, 2002).

(iii) LHBs: The LHBs is the largest surface protein, encoded by LHBs domain 389 - 400 amino acids in length, which depending on the genotype, can be between 108 and 119 amino acids long. The LHBs is more prevalent than MHBs in virions and filaments

but less prevalent in HBs spheres. The preS1 has two main functions: the binding of the Dane particle to the hepatocyte during the infection and providing a ligand for the core particle during assembly of the viral envelope (Seeger and Mason, 2000).

1.1.4.2.2 HBcAg and HBeAg

The core protein or the HBcAg (21.5 kD) is translated from the pgRNA and composed of 183 to 195 amino acids, depending on the genotype. Dimers of HBcAg assemble spontaneously into the icosahedral viral capsid (nucleocapsid) in the cytoplasm of hepatocytes (Kann, 2002). The major HBc epitope appears to be a conformational epitope, which requires HBc aggregation because when the aggregates are denatured, HBc antigenicity disappears (Yokosuka and Arai, 2006).

The HBeAg (17 kD) is translated from the preC/C mRNA start codon and differs from the core protein in that it contains an extra 29 amino acids at the amino end (Jean-Jean *et al.*, 1989). The HBeAg is post-translationally modified by cleavage of 10 amino acids from the N-terminus, and at variable sites on the carboxyl end (Kramvis *et al.*, 1997) and is a soluble, non-particulate protein that is secreted in the serum (Kramvis *et al.*, 1997, Jean-Jean *et al.*, 1989) or expressed on the surface of the hepatocyte. Conventionally, HBeAg expression is an indicator of active HBV infection and ongoing viral replication (Locarnini *et al.*, 2003, Kramvis *et al.*, 1997); however, the presence of mutations affecting HBeAg expression, results in the absence of HBeAg expression, when active viral replication is occurring.

1.1.4.2.3 Polymerase

The polymerase protein (95 kD) is encoded by the polymerase ORF, the longest ORF, spanning almost 80% of the genome and overlapping the other ORFs (Kann, 2002, Seeger

and Mason, 2000). Opinion on the mechanism by which the polymerase is translated remains controversial (Kann, 2002). Some studies showed that translation of the polymerase protein is initiated at the first start codon of the polymerase ORF (Roychoudhury and Shih, 1990, Lin and Lo, 1992); while others showed that polymerase synthesis may involve leaky scanning that does not result in core protein translation (Kozak, 1989, Lin and Lo, 1992).

The polymerase protein is between 833 - 846 amino acids long, depending on the genotype, and has four domains (Beck and Nassal, 2007), (Figure 1.4).

(i) The terminal protein (TP) spans amino acids at the N-terminus of the protein. It remains covalently linked to the minus strand DNA during replication, eventually preventing phosphorylation of the minus strand DNA (Bosch *et al.*, 1988). TP has no sequence similarity to any other protein in the public databases (Ferrer-Orta *et al.*, 2006).

(ii) The spacer domain (SP) joins the (TP) and the reverse transcriptase (RT) domain. It is a highly variable region and has an important function in maintaining the heterogeneity of the polymerase and providing flexibility for conformational changes and response to immune selective pressure (Chen *et al.*, 2013).

(iii) The RT contains the YMDD motif, which is the catalytic domain essential for reverse transcriptase activity.

(iv) RNase H domain (RH) encompasses amino acids in the C-terminus and serves to degrade pgRNA (Bosch *et al.*, 1988). The RT and RNase H domains together are required as structural components for pgRNA encapsidation.

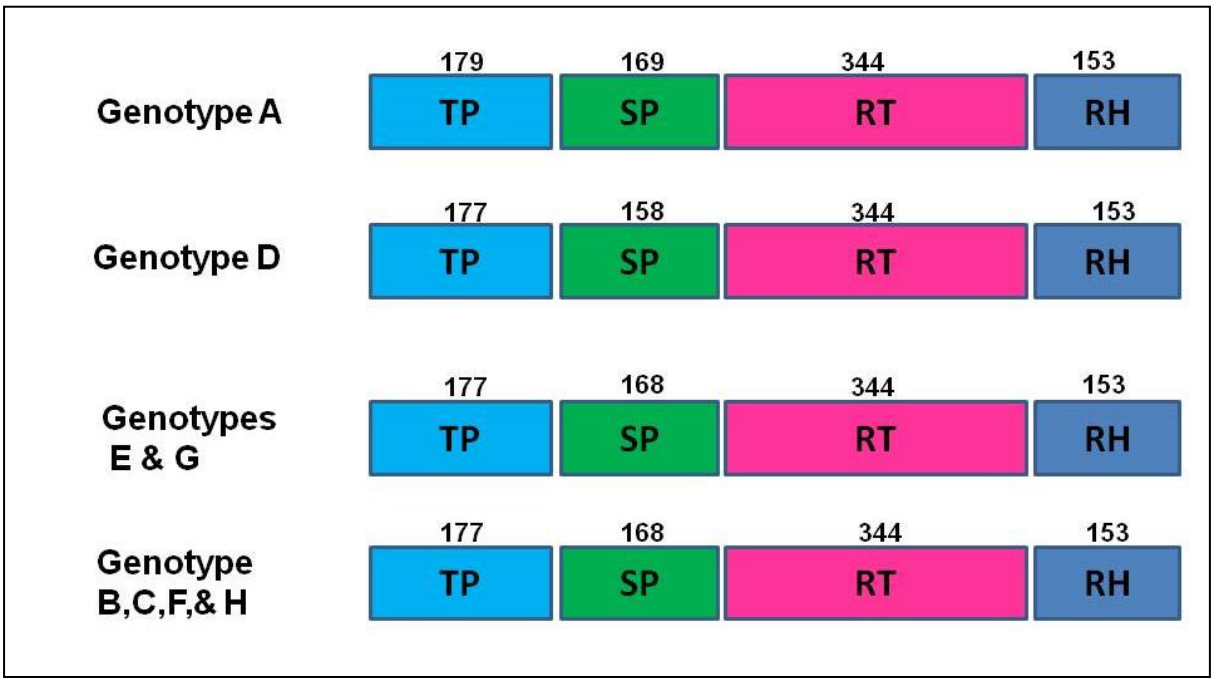


Figure 1.4: A schematic representation of the hepatitis B virus polymerase shows the different lengths of each polymerase domain in amino acids for the different genotypes. TP: Terminal protein, SP: Spacer, RT: Reverse transcriptase and RH: RNase H

1.1.4.2.4 HBx protein

The X protein, 154 amino acids in length, is encoded by the X ORF and translated from the 0.7 kb X mRNA (Moolla *et al.*, 2002, Locarnini *et al.*, 2003). Although this protein is essential for *in vivo* HBV infection, its exact function is poorly understood (Seeger and Mason, 2000). Among other functions, the X protein is known to activate various signal pathways *in vitro* and to bind to p53 (Locarnini *et al.*, 2003). Carboxyl terminally truncated forms of X ORF have been known to integrate into the host genome of chronically infected individuals and these integrants are believed to play a role in the development of HCC (Yokosuka and Arai, 2006).

1.1.4.3 Encapsidation

The penultimate step in the HBV replication cycle is encapsidation, which is the packaging of the pgRNA, together with the reverse transcriptase, into the capsid to form the core particle (Hu *et al.*, 2004). Reverse transcription occurs in this particle and starts by the binding of the polymerase to the bulge of the encapsidation signal (ε) of the pgRNA (Seeger and Mason, 2000). Later the polymerase/ε interaction was found to constitute the first step in the initiation of reverse transcription as well (Wang and Seeger, 1993, Nassal and Rieger, 1996, Beck and Nassal, 2007) (Figure 1.5). ε is a stem-loop structure initially defined as the sequence at the 5' end of HBV pgRNA that mediates polymerase-dependent encapsidation of pgRNA, and also of heterologous transcripts to which it is fused (Junker-Niepmann *et al.*, 1990). Reverse transcription of the pgRNA leads to the formation of the negative strand DNA and then the pgRNA is gradually degraded by the RNaseH activity of the polymerase, leaving 18 nucleotides (Seeger and Mason, 2000). These 18 nucleotides are used as the template for the synthesis of the partial length positive-strand DNA, which is prematurely terminated because of the depletion of nucleotides within the capsid. Encapsidation and reverse transcription within the capsid result in the maturation of the core particle (Seeger and Mason, 2000). The mature core particle can be either enveloped by host-derived lipids and viral envelope proteins in the ER and secreted/budded out of the hepatocyte; or recycled to the nucleus, where it replenishes the cccDNA pool. This cccDNA pool constitutes a risk factor for chronic infection, and serves as a reservoir for viral reactivation, when antiviral therapy is interrupted (Beck and Nassal, 2007, Urban *et al.*, 2010).

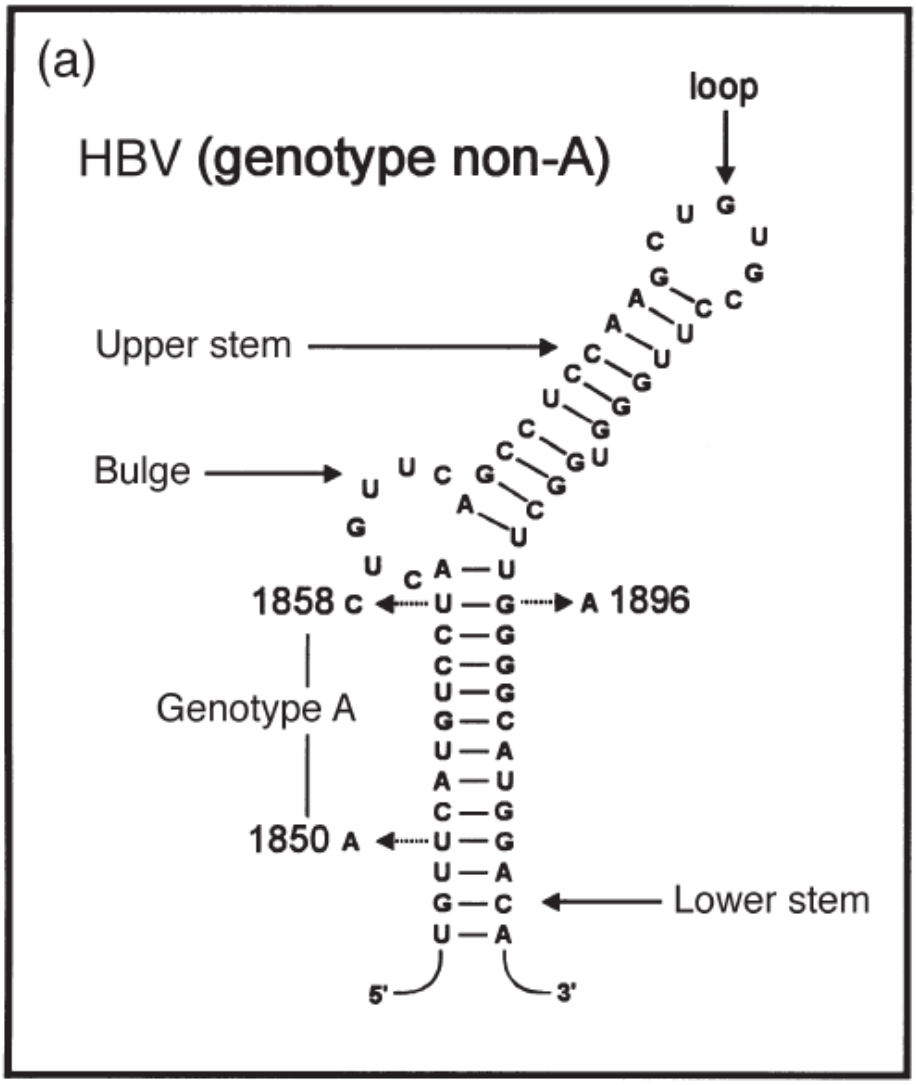


Figure 1.5: The structure of the encapsidation signal of *Hepadnaviridae*. Reproduced from (Kramvis and Kew, 1998), with permission from John Wiley and Sons.

As a result of the lack of the viral polymerase proofreading activity, HBV exhibits considerable sequence heterogeneity, hence its classification into genotypes.

1.1.5 HBV genotypes

The genotypes of HBV are defined by inter-genotypic differences of more than 7.5% in the complete nucleotide sequence of the genome (Okamoto *et al.*, 1988, Norder *et al.*, 1992a, Kramvis *et al.*, 2005a) and more than 4% in the S gene (Norder *et al.*, 1992b, Kramvis *et al.*, 2005a, Kramvis *et al.*, 2008). Four genotypes, A to D, were recognized initially (Okamoto *et al.*, 1988), with genotypes, E and F (Naumann *et al.*, 1993, Norder *et al.*, 1994, Norder *et al.*, 1992b), G (Stuyver *et al.*, 2000), H (Arauz-Ruiz *et al.*, 2002) and I (Yu *et al.*, 2010), being recognized subsequently. A tenth genotype, J, isolated from a single individual, has been proposed (Tatematsu *et al.*, 2009) (Figure 1.6).

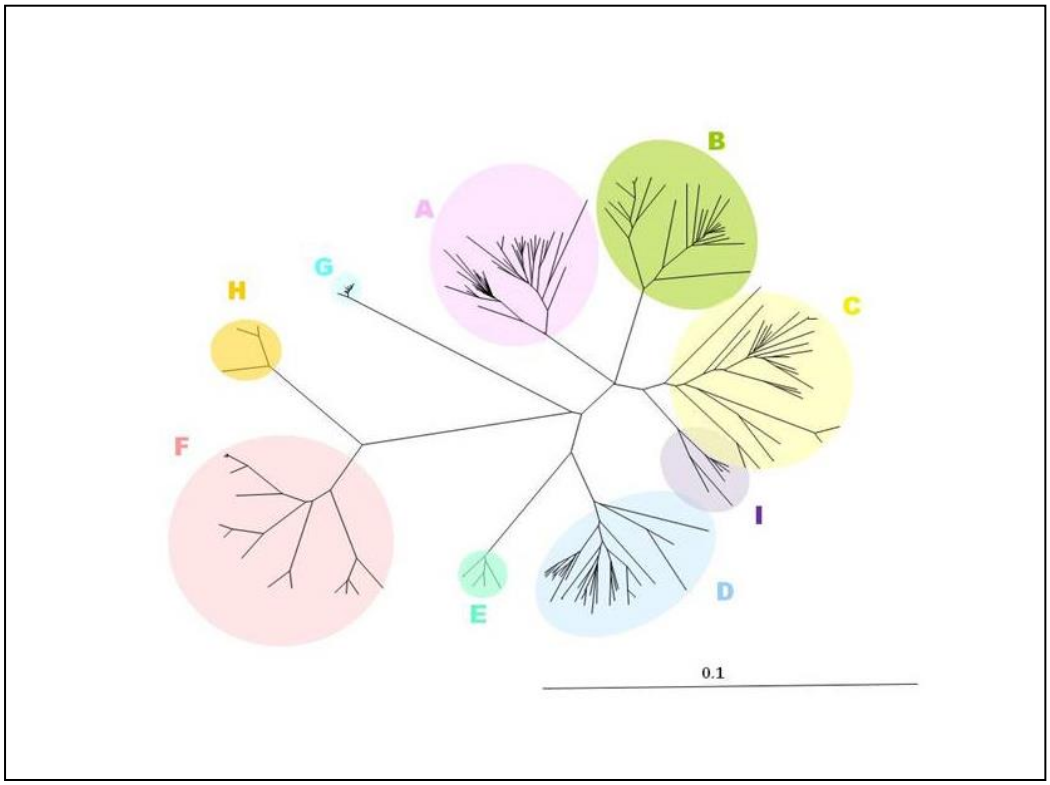


Figure 1.6: HBV genotypes. Phylogenetic tree showing the 9 genotypes of HBV.

Subgenotypes are defined by the inter-group difference of more than 4% and less than 7.5% in the complete nucleotide sequence and significant bootstrap support following phylogenetic analysis (Kramvis *et al.*, 2008). Subgenotypes, recognized in genotypes A, B, C, D, F and I,

are named numerically, with the letter of the genotype followed by a number (Kramvis *et al.*, 2008).

Table 1.1: Virological characteristics of HBV genotypes

Genotype		A	B	C	D	E	F	G	H
Length		3221[¶]	3215	3215	3182[€]	3212[£]	3215	3248[¥]	3215
HBV DNA level		NS[a]	↓[b]	↑[c]	NS	NS	NS	NS	NS
Frequency of precore G1896A mutation		↓	↑	↓	↑	NS	NS	NS	NS
Frequency of basic core promoter T1762A/A1764G mutation		↑	↓	↑	↓	NS	NS	NS	NS
Frequency of pre-S deletion mutation		NS	↓	↑	NS	NS	NS	NS	NS
Serotype	*ayw2*	•		•	•				
	ayw3			•	•		•		
	ayw4				•	•	•		
	ayr			•					
	adw2	•	•	•				•	
	adw3		•		•				
	adw4	•					•		•
	adrq-		•	•					
	adrq+			•					

Adapted from Kramvis et al 2005, Purdy et al 2007 and Schaefer, 2007a

[¶] Six nucleotides insertion in the core region, [€] 33 nucleotides deletion at the start of preS1 region, [£] Three nucleotides deletion at the start of the preS1 region, [¥] Three nucleotides deletion at the start of the preS1 region and 36 nucleotides insertion at the core region, [a]NS: Not studied, [b]↓ Low, [c]↑ High.

Before the identification of the HBV genotypes, HBV was classified according to serological subtypes. Ten serological subtypes were identified: *ayw1*, *ayw2*, *ayw3*, *ayw4*, *ayr*, *adw2*, *adw3 adw4*, *adrq−* and *adrq+* (Okamoto *et al.*, 1988, Norder *et al.*, 1992a). These serological

subtypes are classified according to the amino acid substitutions on the HBsAg at positions 122, 127, 134 and 160 (Norder *et al.*, 1992a). There is a limited correlation between the HBV genotypes and the serological subtypes because several serological subtypes are encoded by more than one genotype (Norder *et al.*, 1994, Norder *et al.*, 2004, Kramvis *et al.*, 2008). Table 1.1 summarizes the virological characteristics and the consequences of the infection of different HBV genotypes, which are discussed subsequently.

1.1.5.1 Genotype A

Genotype A (3221 bp), (eg: AY233275), has a characteristic 6 nucleotide insertion at the 3' end of the core gene, which results in an additional two amino acids at positions 153 and 154 at the core gene. These are not found in the other genotypes (Kramvis *et al.*, 2005a). The first subgenotype of A, A1 was discovered using analysis of preS/S sequences of South African isolates (Bowyer *et al.*, 1997). This was later confirmed using whole genome sequencing (Kramvis *et al.*, 2002, Kimbi *et al.*, 2004). Subgenotype A1 has subsequently been found in Malawi (Sugauchi *et al.*, 2003), Somalia (Sánchez *et al.*, 2002), Brazil (Moraes *et al.*, 1996), Philippines (Sugauchi *et al.*, 2004), India and Nepal {Gopalakrishnan, 2013 #5545;Sugauchi, 2004 #987}, Yemen (Sallam and William Tong, 2004), Haiti (Andernach *et al.*, 2009b), Zimbabwe (Gulube *et al.*, 2011), Kenya (Ochwoto *et al.*), Martinique (Brichler *et al.*, 2013), Uganda (Hannoun *et al.*, 2005) and Tanzania (Hasegawa *et al.*, 2006). Subgenotype A2 has been reported to be widely distributed outside Africa, in European countries and the USA (Sugauchi *et al.*, 2004). In the BCP/PC region, subgenotype A1 can be differentiated from subgenotype A2 by TCAT, instead of GCAC at the Kozak sequence 1809-1812, and 1888A instead of 1888G (Kramvis *et al.*, 2008). Subgenotype A3 has been identified in the Cameroon and Gabon (Kurbanov *et al.*, 2005). Subgenotypes A4 and A5 circulate in Mali

and Nigeria, respectively (Olinger *et al.*, 2006). Subgenotype A6 was isolated from three African-Belgium patients (Pourkarim *et al.*, 2010b) and subgenotype A7 from Cameroonians (Hübschen *et al.*, 2011). However, the existence of subgenotypes A4, A5 and A7 has been questioned. Based on extensive phylogenetic and divergence analyses of complete genomes sequences, Pourkarim and colleagues found that there is misclassification of the subgenotypes of A and proposed that genotype A should be classified into subgenotypes A1, A2, quasispecies A3 (A3, A4, A5 and A7) and A4 (previously A6) (Pourkarim *et al.*, 2011a).

1.1.5.2 Genotype B

Initially genotype B (3215 bp) was classified into 2 subgenotypes, B1 (Bj) mainly distributed in Japan, and B2 (Ba), which is a B/C recombinant at preC/C genes, found predominatly in the rest of Asia (Sugauchi *et al.*, 2002, Norder *et al.*, 2004). Subsequently 7 additional subgenotypes of B have been identified (B3 to B9) (Norder *et al.*, 2004, Sakamoto *et al.*, 2006, Sakamoto *et al.*, 2007, Nurainy *et al.*, 2008, Mulyanto *et al.*, 2009, Thedja *et al.*, 2011). However, recent analyses questioned the validity of this classification. Based on phylogenetic analyses and divergence calculations of the entire genomes, genotype B has been reclassified into 5 subgenotypes: B1, B2, B4, B6 and quasi-subgenotype B3 of Southeast Asian and Chinese origin, which includes the previously described: B3, B5, B7, B8 and B9. Furthermore, previously described B6 from the Arctic was reclassified as B5 (Shi *et al.*, 2012a).

1.1.5.3 Genotype C

To date genotype C (3215 bp) has been classified into 16 subgenotypes distributed geographically in Southern Asia, East of Asia, Australia, Philippines, Indonesia and Papua province of Indonesia (Depamede *et al.*, 2010, Mulyanto *et al.*, 2012, Huy *et al.*, 2004,

Norder *et al.*, 2004, Kramvis *et al.*, 2008, Sugauchi *et al.*, 2001, Sakamoto *et al.*, 2006, Utsumi *et al.*, 2009, Cavinta *et al.*, 2009a, Cavinta *et al.*, 2009b, Mulyanto *et al.*, 2011). However, whether the number of sample sequences used to define these subgenotypes can be considered as sufficient, remains controversial.

1.1.5.4 Genotype D

Compared to the other genotypes, D has the shortest genome length of 3182 nucleotides and is characterized by a 33-nucleotide deletion at the beginning of the preS1 region (Kramvis *et al.*, 2005a, Norder *et al.*, 2004). Genotype D is found worldwide, with its subgenotypes, showing distinct geographical distribution. This genotype predominates in the Mediterranean region and is also found in many parts of Africa, northern and central Americas, India and Australia (Kramvis *et al.*, 2008, Meldal *et al.*, 2009, Abdou Chekaraou *et al.*, 2010, Norder *et al.*, 2004, Utsumi *et al.*, 2009, Banerjee *et al.*, 2006, Ghosh *et al.*, 2012).

Until 2008, genotype D was classified into four subgenotypes (D1-D4) (Norder *et al.*, 2004, Kramvis *et al.*, 2008). Since then an additional five subgenotypes of D have been identified, D5 isolated from India in 2006 (Banerjee *et al.*, 2006), D6 from the Papua province of Indonesia in 2009 (Utsumi *et al.*, 2009), D7 from Tunisia in 2009 (Meldal *et al.*, 2009) and D8 from Niger in 2010 (Abdou Chekaraou *et al.*, 2010). However, an extensive updated analyses presented in Chapter 2, revealed that there are 6 subgenotypes of D, D1 to D6. Subsequently, subgenotype D9 was described from India in 2012 (Ghosh *et al.*, 2012).

1.1.5.5 Genotype E

Genotype E (3212 bp) is characterized by a three nucleotide deletion in preS1 (Norder *et al.*, 1994, Kramvis *et al.*, 2005b). The amino acids of the preS1, preS2 and S genes are well conserved, with signature motifs specific for genotype E: Leu, Ser, Trp, Thr, Val, Pro, Leu, Glu and Trp at positions 3 to 11 of the preS1 (Kramvis *et al.*, 2005b). In addition, signature amino acids are also found in the preS1: Thr, Arg, His, Thr, Met, Lys and Thr at positions 18, 38, 44, 52, 83, 85 and 108, respectively. This genotype has no subgenotypes and encodes for serological subtype *ayw*4 (Kramvis *et al.*, 2005b, Norder *et al.*, 2004). Genotype E predominates in Western and Central Africa reviewed in (Kramvis *et al.*, 2005b). Studies have shown that genotype E has low genetic diversity in a large geographical area and this has led to this suggestion that this genotype emerged recently (Andernach *et al.*, 2009a, Odemuyiwa *et al.*, 2001, Suzuki *et al.*, 2003, Mulders *et al.*, 2004, Kramvis *et al.*, 2005b).

1.1.5.6 Genotype F

Genotype F (3215 bp) has the highest intergroup divergence value (14%) and may represent the first split from the human hepadnaviral ancestor (Norder *et al.*, 1994, Kramvis *et al.*, 2005a, Naumann *et al.*, 1993). It has been classified into four subgenotypes. Subgenotype F1 is found mainly in Central America, and subgenotype F2 mainly in South America (y Leone *et al.*, 2003, Devesa *et al.*, 2004, Mbayed *et al.*, 2001, Norder *et al.*, 1993, Arauz-Ruiz *et al.*, 1997). Subgenotype F3 and F4 are found in Bolivia and Argentina, respectively (Devesa *et al.*, 2004, Huy *et al.*, 2006). Subgenotype F1 has characteristic of T1858 at the PC gene and sT45 at the S region, while subgenotype F2 has characteristics of C1858 and sL45 (Norder *et al.*, 2003, Kramvis *et al.*, 2005a).

1.1.5.7 Genotype G

Genotype G bp (3248) is characterized by a 36 nucleotide insertion in the core gene and a three nucleotide deletion in the preS1 region (Norder *et al.*, 1994, Kramvis *et al.*, 2005b, Kramvis *et al.*, 2005a, Stuyver *et al.*, 2000). This genotype circulates in Mexico, USA and Europe (Stuyver *et al.*, 2000) and has no subgenotypes. A distinctive characteristic of this genotype is the absence of HBeAg expression caused by the presence of two translational stop codons at positions 2 and 28 of the precore region (Kato *et al.*, 2002).

1.1.5.8 Genotype H

Genotype H (3215 bp) has been isolated from Nicaragua, Mexico and California (Arauz-Ruiz *et al.*, 2002). It was classified previously as genotype F (Arauz-Ruiz *et al.*, 1997), but upon sequencing of the complete genome, this genotype was found to have unique amino acids substitutions separating it from genotype F (Arauz-Ruiz *et al.*, 2002).

1.1.5.9 Genotype I

Initially genotype I was described as a recombinant of genotypes A and C in Vietnamese HBV strains (Hannoun *et al.*, 2000). Later a single HBV strain, a recombinant of genotypes A, C and G was isolated in Vietnam and found to be phylogentically closely related to the strains isolated by Hannoun *et al*., (2000). This strain was assigned to the ninth genotype (I) (Huy *et al.*, 2008). Kurbanov and colleagues (2008), however, questioned the existence of this genotype, because of the limitations of the methods used in the analysis (Kurbanov *et al.*, 2008). Recently, a similar recombinant strain was found in northwest China and analyzed using more advanced methods, including TreeOrderScan, GroupScan, unique amino acids

signature and *in vitro* experiments. This study provided conclusive evidence for the existence of the new genotype I (Yu *et al.*, 2010). Genotype I is divided into two subgenotypes I1 and I2 (Yu *et al.*, 2010).

1.1.5.10 Genotype J

An isolate from a single Japanese HCC patient, has been proposed to belong to the 10th genotype, J (Tatematsu *et al.*, 2009). The designation for genotype J was based on its sequence divergence and unique phylogenetic position relative to human and ape genotypes and the absence of recombination (Tatematsu *et al.*, 2009). Additional isolates would be required to confirm the presence of this 10th genotype.

1.1.6 Clinical relevance of the HBV genotypes

There are several factors that can affect the clinical manifestation of HBV infection. Host factors such as age and immunological status and viral factors such as the genomic variability of the virus (Kao, 2002b). Global heterogeneity of the HBV genotypes may play a significant role in the clinical outcome of infection (Kramvis and Kew, 2007). HBV genotypes have demonstrated significant impact on the severity of liver disease and response to antiviral therapies and vaccination (Kramvis and Kew, 2005). The co-distribution of HBV genotypes has made it possible to reliably compare genotypes A and D in Western Europe, the USA and southern Africa and genotypes B and C in south-eastern Asia in terms of their association with clinical characteristics (see Table 1.2).

Comparative studies in Asia have shown that patients infected with genotype C generally have more severe liver disease, including cirrhosis and HCC, than patients infected with genotype B (Kao, 2002b, Yuen *et al.*, 2004). Patients infected with genotype C experience delayed HBeAg seroconversion to anti-HBe compared to patients infected with genotype B.

This usually increases the duration of viral replication, which may ultimate result in fibrosis, cirrhosis and HCC (Lin and Kao, 2011). In Taiwan, genotype B is associated with the early onset of HCC, while HCC develops later in those infected with genotype C (Kao, 2002b, Lin and Kao, 2011). This, however, is not the case in Japan where genotype B is associated with the later development of HCC compared to genotype C (Orito *et al.*, 2005). This may be as a result of the different prevalence of the subgenotypes of B, with subgenotype B2 circulating in many parts of Asia and subgenotype B1 confined to Japan (Sugauchi *et al.*, 2002).

In sub-Saharan Africa, subgenotype A1 is the predominant subgenotype (Bowyer *et al.*, 1997, Kramvis *et al.*, 2002, Kimbi *et al.*, 2004) and has it been associated with high rates of HCC (Kew *et al.*, 2005). Patients infected with subgenotype A1 have a higher rate of HBeAg negativity, low HBV DNA levels (Tanaka *et al.*, 2004) and develop HCC at a younger age compared to patients infected with other genotypes (Kew *et al.*, 2005). A South African case control study showed a 4.5 times higher relative risk of HCC in carriers infected with genotype A compared to those who infected with other genotypes (Kew *et al.*, 2005). Furthermore, the course of infection with subgenotype A2, which is found in Europe and USA, is different to that of subgenotype A1, with HBeAg seroconversion occurring at a later stage of infection and HCC found to occur in older patients (McMahon, 2009).

Genotype D has a worldwide distribution. Most patients infected with genotype D generally seroconvert to anti-HBe positivity at a more advanced age; which means they may undergo multiple hepatitis flares, which accelerates the progression of chronic hepatitis to cirrhosis (McMahon, 2009, Lin and Kao, 2011). This however was found not to be the case in India, where genotype D correlates with severe liver disease and the development of HCC in younger patients (Thakur *et al.*, 2002). In a study carried out in the USA, genotype D was

found to be an independent risk factor for fulminate hepatitis (Wai *et al.*, 2005). These differences may be the result of different subgenotypes of D circulating in the different geographical regions of the world.

Genotype E predominates in Africa, particularly west and central Africa. However, very few studies have looked at the clinical outcome of genotype E infection.

Table 1.2: Clinical characteristics associated with HBV genotypes

	Genotype	A	B	C	D	E	F	G	H
Mode of transmission	Horizontal	•			•	•	•	•	•
	Perinatal		•	•					
Tendency of chronicity	High	•		•					
	Low		•		•				
	Not studied					•	•	•	•
Positivity of HBeAg	High	•		•					
	Low		•		•				
	Not studied					•	•	•	•
HBeAg seroconversion	Early	•	•						
	Late			•	•				
	Not studied					•	•	•	•
HBsAg seroconversion	More	•	•						
	Less			•	•				
	Not studied					•	•	•	•

Adapted from (Lin and Kao, 2010).

1.1.7 HBV recombination

When two or more genotypes co-circulate in one geographical region, co-infection with more than one genotype may occur and lead to the recombination events between these genotypes and provide a variation within individuals and in the population in general (Kramvis *et al.*, 2005a). Recombination may also help the virus evade the host's immune response by changing the antigenetic configurations (Owiredu *et al.*, 2001, Shi *et al.*, 2012b). Several papers report recombination events between genotypes: A and D in Africa (Owiredu

et al., 2001), B and C in Asia (Sugauchi *et al.*, 2002), C and D in Tibet (Cui *et al.*, 2002), A, C and G in Vietnam (Tran *et al.*, 2008) and D and E in Niger and Sudan (Abdou Chekaraou *et al.*, 2010, Mahgoub *et al.*, 2011). Recombination events increase genetic diversity in the HBV population and lead to the designation of an increased number of genotypes/subgenotypes. Although a number of researchers have suggested that recombinant strains should not be classified as novel subgenotypes, but be reported separately instead, this suggestion has not been followed widely (Shi *et al.*, 2013, Kramvis *et al.*, 2008, Pourkarim *et al.*, 2010a).

1.1.8 HBV mutation

A mutation can be defined as a random error occurring in the genome, which may lead to genetic variation in the organism (Bonhoeffer and Sniegowski, 2002). Because of the lack of proof reading activity by the HBV reverse transcriptase, the mutation rate of HBV is relatively high with the nucleotide substitution rate per site per year estimated at 1.5×10^{-6} to 3.0×10^{-6}(Paraskevis *et al.*, 2013). Errors in genomic replication play a critical role in viral evolution and survival because may they allow the virus to escape host immune responses (immune/vaccine escape mutants) and to withstand of antiviral treatment (drug resistance mutants) (Bonhoeffer and Sniegowski, 2002). There are a number of mutations that can affect the virus by affecting HBeAg expression, its infectivity and its sensitivity to antiviral drugs.

1.1.8.1 Mutations that can affect expression of HBeAg

There are two types of mutations that affect HBeAg expression, the precore and BCP mutations. The G to A mutation at 1896 (G1896A), located in the ε structure, creates a stop codon (TGG to TAG) in the second last codon of the precore gene. This mutation truncates

the HBeAg precursor at 28 amino acids abolish HBeAg expression (Carman *et al.*, 1989). In order to maintain the stability of the ε signal, the G1896A mutation requires a T at position 1858, to form a Watson-Crick base-pair. This mutation is genotype specific because T1858 occurs in genotypes B, D, E G and some strains of C, while C1858 only occurs in genotypes A, F, H and some strains of genotype C. This is the reason for the absence of G1896A in genotypes A, F, H and some strains of genotype C (Tong *et al.*, 2013, Locarnini *et al.*, 2003, Kramvis *et al.*, 2005a, Li *et al.*, 1993, Lok *et al.*, 1994). Mutations that can abolish the translation of HBeAg include the precore start codon mutations at 1814, 1815 and 1816 (Günther *et al.*, 1999). Mutations at 1817, 1874 and 1897 cause HBeAg truncation (Günther *et al.*, 1999). The silent mutation G1899A frequently occurs together with the precore stop codon mutation G1896A (Günther *et al.*, 1999).

G1862T, which occurs in the bulge of the ε signal, may affect HBeAg expression (Kramvis *et al.*, 1997, Kramvis *et al.*, 1998). This mutation results in a valine to phenylalanine mutation in the precore region of HBV, interfering with post-translational modification of the HBeAg precursor, causing intracellular retention and impaired secretion of HBeAg (Chen *et al.*, 2008). G1862T occurs more frequently in genotype A and particularly in subgenotype A1 (Kramvis *et al.*, 1997).

The BCP mutations were discovered by Okamoto and colleagues (Okamoto *et al.*, 1994). The most common mutation is the double mutation A1762T/G1764A, which results in the down regulation of transcription of pre-core mRNA, thereby decreasing HBeAg production up to 70%, while increasing pre-genomic RNA packaging and viral progeny production (Buckwold *et al.*, 1996, Takahashi *et al.*, 1995, Hunt *et al.*, 2000, Laskus *et al.*, 1995). This mutation is seen more often in HBV isolated from HCC patients compared to asymptomatic carriers (Baptista *et al.*, 1999, Kramvis and Kew, 1999). The A1762T/G1764A mutation coupled with

the T1753 mutation may be a marker for advanced liver disease (Kramvis and Kew, 1999). Other mutations at C1653, T1753, T1766, and A1768 have been reported to reduce HBeAg production and when occurring together with the A1762T/G1764A may increase the risk of HCC and fulminant hepatitis (Kramvis and Kew, 1999, Tsai *et al.*, 2009, Cao, 2009).

The region (1809-1812), which precedes the pre-core start codon, is termed the "Kozak" sequence and is critical in the initiation of translation (Ahn *et al.*, 2003, Baptista *et al.*, 1999). Mutations in the Kozak sequence, which occur mainly in subgenotype A1, affect HBeAg expression at the translational level and reduce HBeAg synthesis (Ahn *et al.*, 2003).

1.1.8.2 Mutations at the complete S region

The preS1 and preS2 regions have the highest heterogeneity in the HBV genome. Mutations and deletions at the preS regions have been identified from HBV isolated from chronic HBV patients (Günther *et al.*, 1999). Deletions and mutations at the preS2 start codon prevent preS2 protein synthesis, indicating that the expression of preS2 protein is not essential for viral replication, virion morphogenesis, and secretion (Yokosuka and Arai, 2006, Santantonio *et al.*, 1992).

The surface gene has a vital role in the life cycle of HBV and infectivity of the virus. HBsAg contains the 'a' determinant, which is the initial target of the anti-HBs immune response (Carman *et al.*, 1990). Mutations of the surface gene at the 'a' determinant region lead to variations in the antigenicity and immunogenicity of the HBsAg, which may result in the impaired detection of the virus by the immune system (Torresi, 2002). The most frequently described mutations are the immune/vaccine escape mutants, which can occur following vaccination, and the treatment of liver transplantation recipients with HBV immunoglobulin (Carman *et al.*, 1990). The most common mutation is G to R at 145 in the S gene (sG145R),

which minimizes the binding of antibodies raised against S protein, to virions and subviral particles, and thus results in breakthrough infections (Cooreman *et al.*, 1999, Tong *et al.*, 2013). Another mutation at the same position sG145A, leads to diagnostic failure (Cooreman *et al.*, 1999, Tong *et al.*, 2013). Mutations sP120L, sT123N, sT126N, sQ129R, sM133L, sM133T and sD144A also have been reported to lead to vaccine escape (Carman *et al.*, 1990, Yokosuka and Arai, 2006). Therefore, there is a potential risk of false negative HBsAg results because commercially available monoclonal or polyclonal antibody may not detect the mutant HBV strains (Yokosuka and Arai, 2006).

1.1.8.3 Drug resistance mutations

Immunomodulatory agents and nucleoside/nucleotide analogue (NA) antiviral agents have been used for the treatment of chronic HBV infection (Ching-Lung Lai, 2002). Long-term use of NAs is associated with the evolution of viral resistance mutations in the HBV polymerase gene, resulting in treatment failure (Ching-Lung Lai, 2002). The most common mutation, which confers lamivudine resistance and occurs in the RT domain, is the rtM204V/I at the YMDD motif. This mutation can occur together with mutations: rtV173L or rtL180M. Mutations associated with adefovir resistance are rtA181V/T and rtN236T (Tong *et al.*, 2013). Entecavir has a higher genetic barrier to resistance because three mutations are needed to generate resistance (Tong *et al.*, 2013, Gish *et al.*, 2012). After 5 years of therapy with the NAs, resistance to lamivudine is 80%, 29% to adefovir and only 1.2% to entecavir (Zoulim and Locarnini, 2009). Moreover, because the polymerase ORF overlaps the ORF coding for HBsAg, amino acid changes in the HBsAg selected by NAs can alter its antigenicity (Torresi, 2002). The HBsAg silent mutations overlapping polymerase region mutations selected by NAs include the mutants sE164D and sM198I (Torresi, 2002).

1.2 Epidemiology, Transmission and Natural History of HBV infection

1.2.1 Epidemiology

According to the most recent data from the World Health Organization, two billion people have evidence of past or present HBV infection and there are over 240 million chronic carriers of HBV globally (WHO, 2012). At least 600 000 people die annually as a result of the clinical outcomes of HBV infection. The prevalence of HBV infection, which is measured by HBsAg-positivity, varies geographically. Areas of high endemicity (>8% HBV carriers) include South-east Asia, China, sub-Saharan and Southern Africa and the Amazon Basin (Hou *et al.*, 2005, Burnett *et al.*, 2005, Kramvis and Kew, 2007). Approximately 45% of the world's population lives in these regions (Kew, 2010). Areas of intermediate endemicity (2-8%) include North Africa, southern and eastern Europe, Middle East and part of South America (Kramvis and Kew, 2007, Burnett *et al.*, 2005, Hou *et al.*, 2005). Low endemicity areas (<2%) encompass the developed countries: North America, western Europe and Australia. Recently Japan has been classified as a low endemicity region following the implementation of a successful vaccination program (Figure 1.7) (Andre, 2000).

In Africa, 65 million people are infected with the virus and approximately 250 000 die every year because of HBV infection (Kramvis and Kew, 2007). Most of the African countries fall within the high endemicity regions and 70-95% of the population have been exposed to the virus at some stage of their lives. Exposure to HBV, measured by the prevalence of anti-HBc, ranges from 1.8% to 98%. This varies between the different regions in Africa, with western Africa having a higher exposure of 85% and eastern parts of Africa showing 65-85% (Kramvis and Kew, 2007).

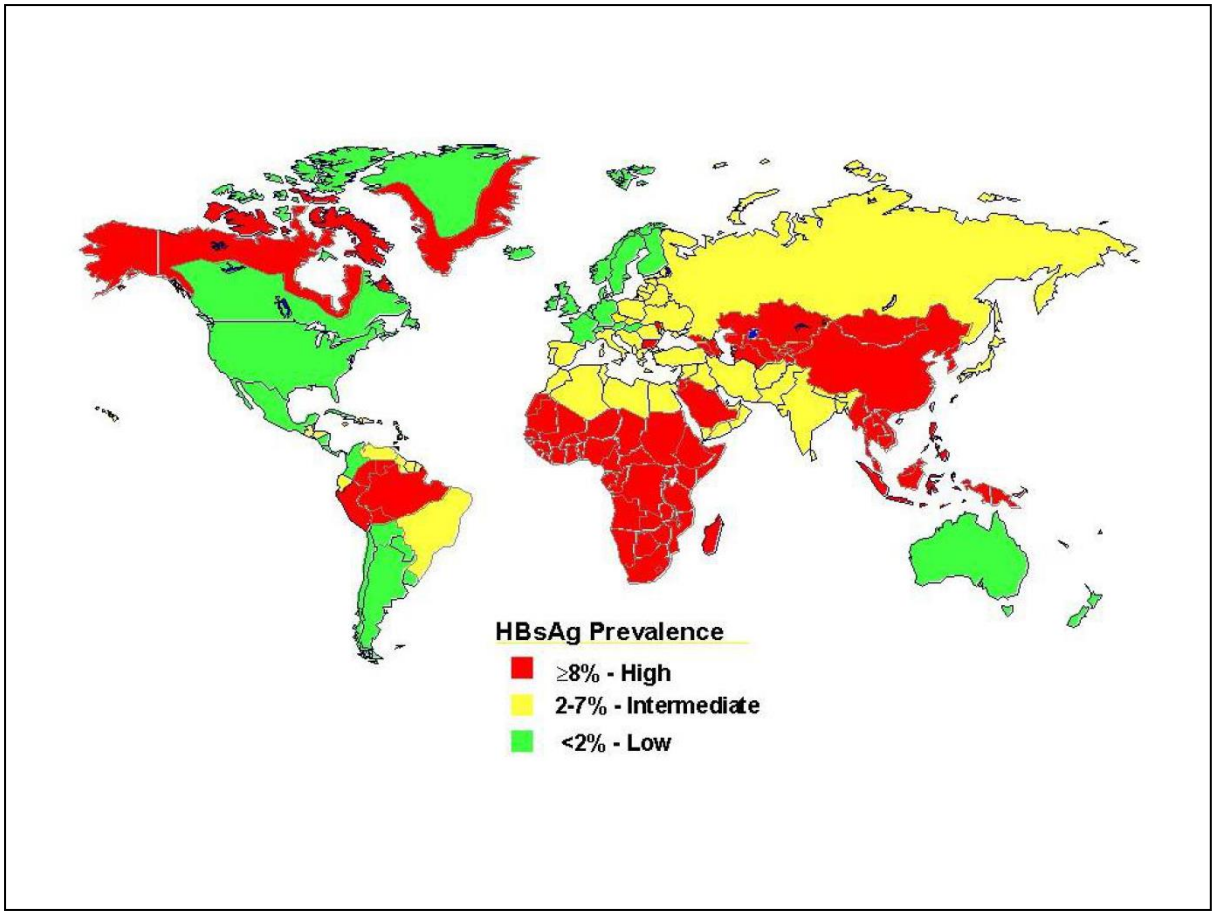


Figure 1.7: Geographic distribution of chronic Hepatitis B virus infection. Source: http://virology-online.com/viruses/HepatitisB.htm. Accessed on July 19, 2013

1.2.2 Transmission

Blood is the main carrier for HBV infection, although other body fluids, like semen and saliva, may also play a role in transmission. There is no evidence that HBV can be transmitted by airborne means or through the oral faecal route (Bancroft *et al.*, 1977, Scott *et al.*, 1980, Hou *et al.*, 2005). HBV transmission routes can be divided into two types:

(i) Perinatal transmission: when the infection passes from the mother to the child, through percutaneous and per mucosal exposure to the mother's blood from 28 weeks of gestation up to 7 days after birth (Beasley and Hwang, 1983, Jonas, 2009, Burnett *et al.*, 2012). This type of transmission is also called vertical infection.

Usually it occurs in HBeAg positive mothers and leads to chronic infection in the infant. This type of transmission has been observed mainly in China and southeast Asia (Kramvis and Kew, 2007, Hou *et al.*, 2005).

(ii) Horizontal transmission occurs when HBV is passed from person to person or from person to surface to person. This route can be divided into four routes:

a) Sexual, and especially homosexual activity, can transmit HBV, and this mode of transmission occurs in all areas of the world but is very common in regions of low endemicity (Alter, 2003, Wasmuth, 2009).

b) Transmission between toddlers has been found to be the major route of infection in Africa, where there is a high prevalence of HBV. Infection during childhood can lead to chronic HBV infection and once these individuals become sexually active, sexual transmission becomes the dominant route of infection (Burnett *et al.*, 2012).

c) Worldwide percutaneous/parenteral transmission is an important mode of transmission. Infection, following unsafe injection drug use occurs mainly in developed countries. In developing countries, health care professionals and their patients are at risk of getting HBV infection when good clinical practice for blood transfusions and organ transplants is not followed. Cultural customs like tattooing and scarification can also lead to the transmission of HBV.

d) Transmission via HBV contaminated blood products in patients.

1.2.3 Natural History of HBV infection

The incubation period, which is the time from exposure to HBV to the onset of jaundice, ranges from 60 to 150 days, with an average of 90 days (Shepard *et al.*, 2006). There is a

wide spectrum of clinical presentation of HBV infection, ranging from subclinical or asymptomatic carriage to acute self-limiting hepatitis or fulminant hepatitis to chronic hepatitis, which can lead to cirrhosis and/or HCC. Viral factors and the host immune response have been implicated in the pathogenesis and the clinical outcome of the infection. The age of the patient at the time of infection is also critical factor in determining the clinical outcome of HBV infection. The majority of perinatal or childhood infections are asymptomatic and usually progress to chronicity. While infections in adults are usually acute and self resolving (Figure 1.8) (McMahon *et al.*, 1985, de Franchis *et al.*, 2003).

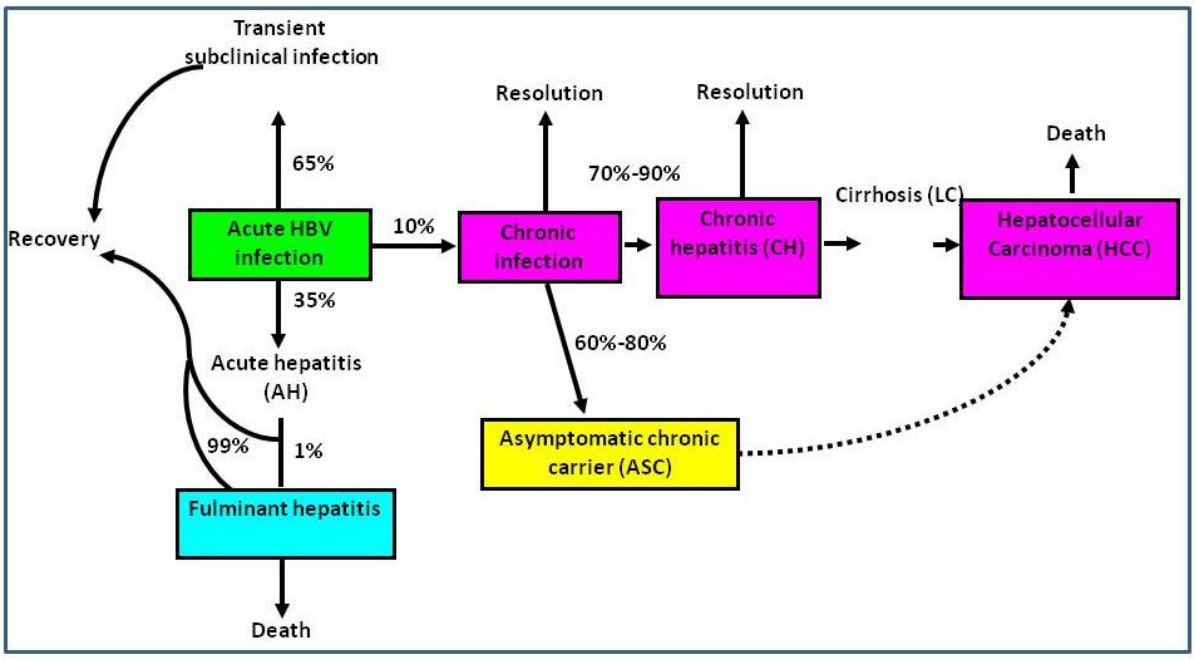


Figure 1.8: Natural history of Hepatitis B virus infection. Adapted from (Feitelson, 1992)

1.2.3.1 Acute hepatitis

An acute HBV infection appears within 6 months after exposure to the virus. The clinical symptoms include nausea, vomiting, abdominal pain, fever, dark urine, change in the colour of the stool and hepatomegaly (Shepard *et al.*, 2006). HBsAg and anti-HBc-IgM are the first markers to appear in acute infections and alanine transaminase (ALT) levels are usually high,

while HBV DNA can be detected in one to two weeks after exposure. Fulminant hepatitis with liver injury is a complication of acute infection. Progression to chronic infection or the resolution of the acute infection is dependent on the host's immune response to the virus (Shepard *et al.*, 2006).

1.2.3.2 Chronic hepatitis

Chronic HBV infection is the persistence of HBsAg for more than 6 months in the serum of an infected person, with no anti-HBc-IgM (Shepard *et al.*, 2006). The progression of chronic infection from acute or subclinical infection is dependent on host factors, mainly the age of the patient. Ninety percent of infants infected develop chronic infection, while only 5% of adults will develop chronicity (Yim and Lok, 2006, de Franchis *et al.*, 2003, Huang *et al.*, 2006). Patients chronically infected with HBV and lacking the anti-HBs, develop HBeAg, which is a marker of HBV replication and correlates with greater infectivity. Seroconversion to anti-HBe is usually associated with resolution of the infection and/or the development of BCP/PC mutations in the virus (Shepard *et al.*, 2006, McMahon *et al.*, 2001).

Chronic HBV infection can be divided into four phases:

(i) The immune tolerance phase, during which infected individuals have high HBV DNA levels and HBeAg positivity (Villeneuve, 2005), normal ALT levels and no liver inflammation. This phase usually occurs in infants who were infected at birth or early childhood, and it can last for up to four decades (Yim and Lok, 2006, de Franchis *et al.*, 2003).

(ii) The immune clearance phase, in which there is seroconversion from HBeAg-positivity to anti-HBe, followed by inflammation of the hepatocytes and increased

ALT levels (Yim and Lok, 2006, de Franchis *et al.*, 2003, Villeneuve, 2005, Yim *et al.*, 2006).

(iii) The inactive or asymptomatic carrier state, where ALT levels become normal, HBV DNA is undetectable, there is no liver inflammation and anti-HBe persists. This phase can last a lifetime.

(iv) The reactivation phase, which is characterized by elevated ALT levels and the presence of HBV DNA, with the possibility of further liver damage, leading to fibrosis and cirrhosis (Yim *et al.*, 2006).

1.2.3.3 Cirrhosis

Patients chronically infected with HBV can develop cirrhosis. The aetiology of cirrhosis established alcohol consumption or viral hepatitis as culprits, with chronic HBV infection being the most common cause (Schuppan and Afdhal, 2008). Cirrhosis can be defined as an advanced stage of liver fibrosis that is accompanied by distortion of the hepatic vasculature leading to portal hypertension, liver function impairment and HCC (Wanless *et al.*, 2000, Desmet and Roskams, 2004, Schuppan and Afdhal, 2008). During fibrosis, collagen replaces the infected, injured liver tissue (Sherlock and Dooley, 2008, Schuppan and Afdhal, 2008). Although, the prevalence of cirrhosis worldwide is not known, in the USA alone, more than 25,000 deaths are as a result of cirrhosis. This number is probably more where there is a high prevalence of HBV infection (in areas such as Asia and Africa) (Schuppan and Afdhal, 2008). Unless clinical symptoms appear, cirrhosis remains undetected without a liver biopsy for histological examination and/or bioimaging by ultrasonography (Epstein *et al.*, 1979, Schuppan and Afdhal, 2008). Therefore, patients chronically infected with HBV should be checked regularly (Epstein *et al.*, 1979, Schuppan and Afdhal, 2008).

1.2.3.4 Hepatocellular Carcinoma (HCC)

HCC is one of the main malignant diseases, and one of the main causes of death worldwide (Parkin *et al.*, 2008). HBV-associated cirrhosis is one of main risk factors for the development of HCC, although it has been reported that HCC can occur with or without cirrhosis (Bréchot *et al.*, 2000). The global distribution of chronic HBV infection and HCC incidence correspond. The incidence of HCC is particularly high in East and South East Asia and sub-Saharan Africa, where HBV carrier rates are between 10-15%. The incidence of HCC is intermediate in areas such as western Asia, Central America, eastern and southern Europe with carrier rates between 2-7%. The remaining countries have low carrier rates (<2%) and low prevalence of HCC (Parkin *et al.*, 1999, Parkin, 2001). Current American Association for the Study of Liver Diseases (AASLD) and European Association for the Study of Liver diseases (EASL) guidelines, recommend at least one annual screening for HCC using imaging techniques, in patients with cirrhosis (Schuppan and Afdhal, 2008).

1.2.3.5 **Occult Hepatitis B infection (OBI)**

According to the Taormina expert meeting in 2008, OBI is defined as 'the presence of HBV DNA in the liver (with detectable or undetectable HBV DNA in the serum) of individuals testing HBsAg negative by currently available assays. When detectable, the amount of HBV DNA in the serum is usually very low (<200IU/ml)' (Raimondo *et al.*, 2008). It has been suggested that the molecular basis of OBI is related to the long term persistence of viral cccDNA in the nuclei of hepatocytes (Raimondo *et al.*, 2007). Individuals, who are HBsAg-negative, with serum HBV DNA levels < 200 IU/ml, are considered to have "true" OBI. Samples in which serum HBV DNA levels are similar to HBsAg positive overt infection, but are HBsAg-negative, have been termed "false" OBI (Raimondo *et al.*, 2008), or HBsAg-

covert infection (Bell *et al.*, 2012).

OBI can be classified into: seronegative (negative for all serological markers), which accounts for approximately 20% of all OBI cases, and seropositive (anti-HBs and anti-HBc-positive), which accounts for 80% of OBI cases (Hollinger and Sood, 2010, Torbenson and Thomas, 2002, Raimondo *et al.*, 2008). Detection of OBI involves amplifying HBV DNA from liver and/or blood samples, using real time PCR, with a detection limit of < 10 copies of HBV DNA per reaction or nested PCR for three different, highly conserved regions of the HBV genome. Samples which yield amplicons from at least two different genomic regions, are considered to be representative of occult infections (Raimondo *et al.*, 2007).

The occurrence of OBI is worldwide and the distribution of OBI may reflect the general prevalence of HBV (Raimondo *et al.*, 2007). OBI is found in HCV-infected individuals, cryptogenetic liver disease, dialysis patients, HIV-infected patients, blood donors and the general population (Gutiérrez-García *et al.*, 2011).

Essentially, OBI is found in four clinical contexts: (i) recovery from acute infection leading to seroconversion from HBsAg to anti-HBs, (ii) chronic HBV infection with mutant strains that have a mutation in the S region and this could result in diagnostic failure of HBsAg by routinely used assays, (iii) chronic infection without any marker except HBV DNA, (iv) chronic infection with HBsAg levels too low to be detected by serological assays the most common type occurring in endemic areas (Allain, 2004), (v) OBI in HIV-infected individuals. OBI has many impacts on different clinical aspects, including the possible transmission of infection, risk of reactivation and enhancing liver disease progression that can lead to HCC (Raimondo *et al.*, 2007). The clinical importance of OBI remains controversial, which is the main reason for the growing interest in this topic (Raimondo *et al.*, 2013). OBI

has been reported more commonly in HIV-positive individuals (Mphahlele *et al.*, 2006, Bell *et al.*, 2012). In resource-limited countries, the diagnosis of co-infection may be missed, because it is difficult to measure the HBV DNA viral loads, or use nucleic acid tests, as a result of the high cost, lack of human resources and infrastructure. If OBI is not treated, this infection may lead to a higher risk of cirrhosis and of hepatic-related deaths in HIV patients (Mphahlele *et al.*, 2006, Firnhaber and Ive, 2009).

1.3 HBV/ HIV co-infection

HIV is a retrovirus, a member of family *Retroviridae*. It is an enveloped RNA virus and replicates by reverse transcription (Vogt, 1997). $CD4^+$ T cells and macrophages of the immune system of infected individuals are specifically targeted and killed by HIV, resulting in host immunodeficiency (Carmichael *et al.*, 1993). HBV and HIV share common routes of blood borne transmission, but they differ in efficiency by which certain types of exposures transmit them and in their prevalence by geographic region (Alter, 2006). Worldwide, it has been estimated that 10% of the 34 million people infected with HIV, have chronic HBV infection (Alter, 2006). It has been reported that HIV can negatively affect the natural history, physiology, diagnosis and therapeutic responses to HBV (Kottilil *et al.*, 2005). As a result of liver related diseases, patients co-infected with HIV and HBV have a high mortality rate, up to 19-fold higher than those infected with HBV alone, and are up to 6-fold more likely to develop chronic hepatitis B than are HIV-negative individuals (Thio, 2009). In a study conducted in the USA, it was found that patients infected with both HBV and HIV are more likely to develop chronic HBV infection compared to those patients infected with HBV only (Bodsworth *et al.*, 1991). In addition, HIV decreases the rate of seroconversion of HBeAg to anti-HBe and increases the level of HBV DNA replication (Krogsgaard *et al.*, 1987, Colin *et al.*, 1999, Thio, 2009). Cirrhosis is one of the major complications of

HBV/HIV co-infection despite the decreased levels of ALT resulting from the HIV infection (di Martino *et al.*, 2002, Thio, 2009). Moreover, anti-HBs-positive patients infected with HIV are at risk of losing this antibody, leading to the reactivation of the HBV infection (Thio, 2009, Biggar *et al.*, 1987, Laukamm-Josten *et al.*, 1988). HIV infection can complicate the diagnosis of HBV because of reverse seroconversion of anti-HBs to HBsAg, especially if the $CD4^{+}$T cell count is < 200 cells/mm^3, which also leads to re-emergence of HBV replication. Therefore, if liver disease emerges in the presence of anti-HBs in HIV- infected individuals, it is recommended that tests for HBV serological markers and HBV DNA be repeated (Thio, 2009).

1.3.1 HBV, HIV and HBV/HIV co-infection in Sudan

Sudan, or officially the Republic of the Sudan is a country located in the north to east-central Africa. Before the southern Sudanese declared their independence and established South Sudan, Sudan was the biggest country in Africa. The collection of serum samples from patients for the present study was done prior to this separation, in July 2011.

The majority of previous studies carried out in Sudan determined HBV infection using HBsAg as the marker by ELISA, although some of researchers used Immunochromatography tests (ICTs), followed by ELISA for confirmation. For HIV infection, most of the studies followed the WHO recommendations for the diagnosis of HIV (WHO, 2006). All the studies carried out between 1989 and 2012, to determine the prevalence of HBV, HIV and co-infection in Sudan, are summarized in table 3.

The prevalence of HBV infection ranges from 3.9% in Khartoum to 26% in patients seen in out-patient clinics in Juba in southern Sudan. The median HBV prevalence for all the studies reported in table 3 is 5.6%.

Detection of HBV DNA by PCR is not routinely done in Sudan because of financial and human resource constraints and only two studies describe the detection of HBV DNA. In the first study, when 404 HBsAg-negative serum samples from the Central Blood Bank were tested for anti-HBc, 36% (145/404) were found to be positive. When these 145 were tested for HBsAg, 11% (16/145) were found to be overt infection i.e. HBsAg-positive, HBV DNA-positive, and 10% (15/145) were HBsAg-negative, HBV DNA-positive i.e. OBI (Mahgoub *et al.*, 2011). Because only anti-HBc-positive samples were tested, this may not be a true reflection of the prevalence of either overt or occult infection in Sudanese blood donors. In addition to determining the prevalence of HBV infection in Sudanese asymptomatic blood donors, Maghoub *et* al., (2011) also determined the HBV genotypes: 57.5% of blood donors were infected with genotype E, 40.5% with genotype D and 2% with subgenotype A2 (Mahgoub *et al.*, 2011). In the second study, when 100 serum samples from HBsAg-negative blood donors were tested, 42 were anti-HBc-positive. Amplification for HBV DNA was only carried out for the 42 anti-HBc-positive sera and 38 of these were HBV DNA-positive (Mahmoud *et al.*, 2013). This translates to an OBI prevalence of 90% in the samples tested, which is rather high. Therefore these results should be treated with caution for a number of reasons. Firstly, the samples that were anti-HBc-negative were not tested. Secondly, amplification of S, C and X genes was carried out using SYBR green real-time PCR.

Thirdly and paradoxically, an attempt to quantify the HBV DNA using the Taqman probe technique (Artus-Biotech, Germany), proved unsuccessful.

Table 1.3 Prevalence of HBV, HIV and HBV/HIV co-infection in Sudan

	Year	Geographical area	Participants	Number of participants	HBsAg %	Anti-HBc %	Anti-HBs %	HIV %	HIV/HBV %	References
HBV studies	1989	Gezira- central Sudan	Villagers	851	18.7	16.2	63.9	NA[¶]	NA	(Hyams *et al.*, 1989)
	1994	Juba- south Sudan	Out patients	651	26	67	NM[€]	NA	NA	(McCarthy *et al.*, 1994)
	2007	Omduman, central Sudan	Pregnant women	728	5.6	NM	NM	NA	NA	(Elsheikh *et al.*, 2007)
	2008	Shendi, north Sudan	Blood donors	78	5.1	NM	NM	NA	NA	(Nail *et al.*, 2008)
	2009	Darfur, western Sudan	Blood donors	400	6.5	NM	NM	NA	NA	(Abou *et al.*, 2009)
	2010	Gezira- central Sudan	Out patients	404	6.9	47.5	47.5	NA	NA	(Mudawi *et al.*, 2007)
	2011	Khartoum	hospital staff	245	4.9	NM	NM	NA	NA	(Elduma and Saeed, 2011)
	2011	Khartoum	Blood donors	404	3.9	36	50.4	NA	NA	(Mahgoub *et al.*, 2011)
	2012	Khartoum	Haemodialysis patients	353	4.5	NM	NM	NA	NA	(Gasim *et al.*, 2011)
HIV studies	1995	Juba- south Sudan	Out patients	401	NA	NA	NA	6.2	NA	(McCarthy *et al.*, 1995)
	1997	Khartoum	Children	1118	NA	NA	NA	1.2	NA	(Hashim *et al.*, 1997)
	2006	Different sites	TB patients	1797	NA	NA	NA	4.0	NA	(El-Sony, 2006)
	2006	Khartoum maternity hospital	Pregnant women	305	NA	NA	NA	1	NA	(Gassmelseed *et al.*, 2006)
	2008	NA	NA	NA	NA	NA	NA	1.4	NA	{UNAIDS., 2008 #1267}
	2010	Khartoum	Children	106	NA	NA	NA	5.7	NA	(Abbas *et al.*, 2010)
	2010	Khartoum maternity hospital	Female sex workers	321	NA	NA	NA	0.9	NA	(Abdelrahim, 2010)
	2011	Eastern Sudan	pregnant women	430	NA	NA	NA	0.23	NA	(Mohammed *et al.*, 2011)
HBV/HIV co-infection studies	1989	North east	Sexually active heterosexuals	536	14	63	49	0	0	(McCarthy *et al.*, 1989a)
	1989	North east, east and south	Soldiers	773	25	52	52	1.7	0.8	(McCarthy *et al.*, 1989b)

¶ NA: Not applicable, [€]ND: Not measured

It is very difficult to surmise, from the data available and summarized in table 1.3, the true prevalence of HIV in Sudan because of the paucity of data and the wide time span over which these studies were performed. HIV prevalence ranges from 0.23% in pregnant women from eastern Sudan to 6.2% in out-patient clinics in Juba in southern Sudan. The median prevalence for HIV infection is 1.3%. In 2008, the estimated HIV prevalence in adults in Sudan was 1.4% {UNAIDS., 2008 #1267} and according to Sudan National AIDS control Programme (SNAP), this figure dropped to 052% in 2011 (SNAP, 2012).

The only two studies, carried out on HBV/HIV co-infection in Sudan, were carried out 24 years ago, using serological techniques. The first study showed no HIV infection among sexually active heterosexuals in northeast Sudan and 14% had HBV infection (McCarthy *et al.*, 1989a). The second study, among soldiers in different region of eastern Sudan, showed 1.7% HIV infection and 25% HBsAg positivity, with only 0.8% being co-infected with HBV and HIV (McCarthy *et al.*, 1989b).

1.4 Nucleic acid sequencing techniques

DNA is the molecule which encodes the genetic information used in the development and functioning of all known living organisms and many viruses. After the discovery of the genetic code by Watson and Crick in 1953 (Watson and Crick, 1953), it became increasingly apparent that the order in which nucleotides appeared within a sequence was critical. This ordering appeared to determine both the structure and function of proteins and other cellular processes. Moreover, the DNA was shown to code for regulatory elements that can control transcription and replication of DNA. Thus determining the order of nucleotides – termed DNA sequencing – became a vital biological research method. Sequencing of viral genomes, derived from clinical samples, may reveal the extent of viral genetic variation, response to

antiviral therapies and enhance the understanding of the course of infection (Ahmadian *et al.*, 2006). Below follows a review on the evolution of sequencing technologies, which comprises two categories, termed first generation and second (next) generation sequencing (Zhou *et al.*, 2010).

1.4.1 First generation sequencing technologies:

1.4.1.1 Maxam and Gilbert method:

In 1977, Maxam and Gilbert invented a sequencing technique based on sequencing by chemical cleavage. DNA fragments are generated by digestion of the sequencing template using restriction enzymes, with the ends of the fragments being radioactively labelled (Maxam and Gilbert, 1977). This method allows breaking an end-labelled DNA strand at specific bases using chemical reagents. The labelled fragments are separated using electrophoresis in a polyacrylamide gel and are then detected by autoradiography. This process was not used extensively because it was technically complex and required hazardous chemicals, including radioactivity.

1.4.1.2 Sanger technology

In 1977, Frederick Sanger and colleagues developed a sequencing technique based on sequencing by chain-termination. Briefly, this method involves a single strand DNA template, DNA polymerase, DNA primer, normal deoxynucleotide triphosphates (dNTPs) and modified dideoxynucleotide triphosphates (ddNTPs) that terminate DNA elongation (Sanger *et al.*, 1977). The incorporation of a ddNTP terminates the extension of the new strand, because the modified nucleotide lacks a hydroxyl group, which is necessary for the formation of a phosphodiester bond. The ddNTPs, which are labelled with radioactive markers, are

visualized by autoradiography on a polyacrylamide gel after electrophoresis. Later, radioactive markers were replaced by colour fluorescent dyes (Sanger *et al.*, 1977). Sanger sequencing is the most frequently used method and is semi-automated, requiring machines such as the. ABI 3130XL (Ahmadian *et al.*, 2006).

Despite the fact that the Sanger sequencing method produces long sequence fragments, it is expensive, both in terms of monetary cost and time required per nucleotide. In addition Sanger sequencing is capable of detecting variation of only ≥20% of the entire population (Solmone *et al.*, 2009). An alternative technique used to detect the quasispecies in the viral genome is cloning, which is a process that amplifies identical copies of DNA fragments. Previously, cloning followed by direct Sanger sequencing was the only method used to study the genetic diversity of a viral population (Ramírez *et al.*, 2013) and as a reference technique to study viral quasispecies. This technique, however, has a number of limitations, it is expensive time consuming and labour intensive, especially when a large number of clones is required for a study (Ramírez *et al.*, 2013). The continued improvement of sequencing technology lead to the development of next generation technologies, which are capable of sequencing many thousands of nucleotides faster and at a lower cost per nucleotide.

1.4.2 Second (next) generation sequencing technologies (NGS):

Next generation sequencing (NGS) technologies comprise a number of advanced methods, which have been invented recently. These technologies are able to produce a very high throughput of data and allow for the detection of mutations, even when present at very low levels in the quasispecies population (Hoffmann *et al.*, 2007, Ijaz *et al.*, 2008). Furthermore, the moderate cost per base, and the rapid turnaround time, has made this technology popular. Three phases of sample preparation are required when using NGS technologies:

(i) DNA library preparation, which involves DNA fragmentation, addition of adapters to each end of the fragments, denaturation and the immobilization of the resulting single stranded DNA on a solid surface, such as supporting beads; (ii) DNA clonal amplification by emulsion PCR or bridge PCR, and (iii) sequence reads by optical events produced from a cyclic chain extension using an enzyme, such as polymerase. Optical events are recorded by a Cooled Charge-Coupled Device camera (CCD) camera and analyzed using a computer algorithm (Metzker, 2010, Zhou *et al.*, 2010). The three most commonly used NGS platforms are:

1.4.2.1 GS FLX and GS Junior platforms:

The GS FLX and GS Junior platforms use pyrosequencing technology (Figure 1.9). Pyrosequencing is a "sequence-by-synthesis approach", in which beads covered with DNA primers are mixed with single-stranded templates from the DNA library. A relative number of beads and DNA single stranded fragments are mixed to ensure that each bead carries only one single-strand DNA molecule. This solution is mixed with oil to form an emulsion for the emulsion PCR. Thereafter, the beads are deposited onto a picotiter plate. When a new nucleotide is incorporated into the sequence, a pyrophosphate molecule is released, which is converted to ATP by ATP sulfurylase. The ATP drives the conversion of luciferin to oxyluciferin, releasing light energy, which is captured by a camera (Ronaghi, 2001, Metzker, 2010, Zhou *et al.*, 2010). The GS Junior machine and GS FLX machine from 454 Roche Diagnostics can generate reads of up to 400 bases in length, to a maximum depth of 0.035 gigabases and up to 1000 bases in length at a maximum depth of 0.7 gigabases. (Capobianchi *et al.*, 2013, Zhou *et al.*, 2010).

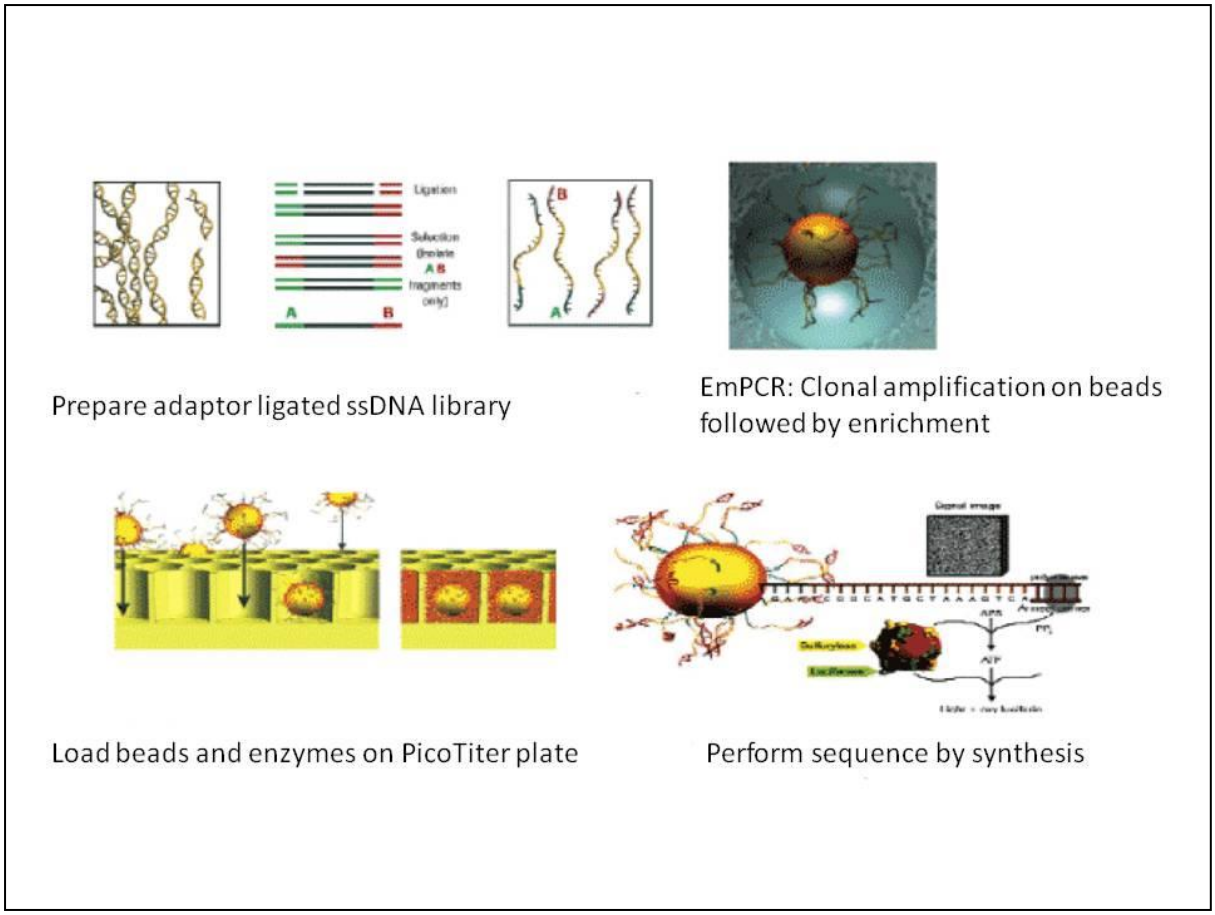


Figure 1.9: Overview of the 454 sequencing system. Source: www.roche-applied-science.com. **Accessed on August 17, 2013**

1.4.2.2 Hi Seq 2000 (Illumina):

The second platform is the Hi Seq 2000 from Illumina. This platform is based on bridge amplification. The reaction is carried out on the surface of a flow cell covered with oligonucleotides corresponding to the sequences of the adapters ligated during the sample preparation step. The single-stranded DNA fragments from the DNA library are ligated with adapters and deposited on the surface of flow cell. Then, PCR of the DNA fragments is carried out, using bridge amplification. After several cycles of PCR, DNA copies accumulate on the surface. Fluorescently labelled nucleotides and polymerase are added, a fluorescence image is captured and the base unique fluorescence reveals the identity of newly incorporated

nucleotides, which are detected by the camera. The Illumina genome analyzer produces sequences of approximately 100 bases in length and a maximum depth of 600 gigabases of data per run (Capobianchi *et al.*, 2013, Zhou *et al.*, 2010, Illumina, 2013).

1.4.2.3 SOLiD platform:

This platform from Applied Biosystems is similar to the 454 Roche systems, in that it uses emulsion PCR for DNA amplification with beads. The difference, however, is that the SOLiD platform uses a ligase enzyme instead of polymerase, and the method is called "sequencing by ligation". This technology produces reads of only 75 bases in length. However, up to 100 gigabases of data per run is possible (Capobianchi *et al.*, 2013, Mardis, 2008, Technologies, 2013).

Although NGS technology has been used extensively in HIV studies, the ability of NGS technology to detect mutations at very low percentages is only now beginning to be implemented and applied for HBV research.

In 2004, Lindstrom *et al.,* used pyrosequencing to detect lamivudine resistance in clinical samples from HBV-infected individuals (Lindström *et al.*, 2004). The authors showed that pyrosequencing has a number of advantages compared to the Sanger sequencing method. Firstly, pyrosequencing is faster compared to Sanger sequencing. Ten years ago it was shown that ABI3100 machine, which utilizes Sanger sequencing could process 16 samples per hour, whereas pyrosequencing can process 96 samples in 10 minutes (Lindström *et al.*, 2004). Secondly, the sample preparation step in pyrosequencing was less complicated and quicker when compared to the preparation needed in Sanger sequencing method and lastly the dispensation order can be designed according to the variation expected in the genome (Lindström *et al.*, 2004). The limitation of the technology at the time meant that reads were

only 40 bases in length. A major improvement in NGS occurred when the ultra deep pyrosequencing (UDPS) methods were developed. The advantage of using UDPS is that it can produce about 200 base pair reads compared to only 40 base pairs by earlier methods (Solmone *et al.*, 2009). UDPS have been used extensively in HIV research and specifically to study HIV drug resistance. Very few studies have used UDPS to characterize the HBV drug resistance (Solmone *et al.*, 2009). A study conducted in 2009, used UDPS for the first time to analyze the HBV drug resistance mutations in the RT domain of the polymerase region. The results showed that HBV strains from patients on NAs treatment have higher variability in the RT region, compared to strains from drug-naïve patients (Solmone *et al.*, 2009). Both previously described and novel drug resistance mutations were identified in this study, which were also found in HBV isolated from drug-naïve patients, thus confirming the findings of previous studies (Solmone *et al.*, 2009). More recently, Homs *et al.*, confirmed that pyrosequencing is able to detect minor populations of the lamivudine resistance rtM240V mutation at the YMDD motif in drug-naïve HBV patients, even when occurring at significantly low levels. Other YMDD variants were also detected (Homs *et al.*, 2011). In the same study, the authors found that mutations in the polymerase region that, lead to drug resistance, occur together with the precore mutations in the same genome, in anti-HBe patients. This finding suggested that mutations in the precore region and YMDD polymerase motif are common in HBV quasispecies in chronic HBV patients (Homs *et al.*, 2011). Furthermore, these authors have shown that two isolates from HBeAg-positive cases have a high percentage of stop codon mutations in the precore region, while isolates from HBeAg negative carriers also have a low percentage of strains wild-type at codon 28 (Homs *et al.*, 2011). In 2013, Rodriguez *et al.*, used UDPS to detect variants with amino acid substitutions at rtA181and rtN236 as a minor population in treatment-naive patients, who later developed adefovir resistance (Rodriguez *et al.*, 2013). Recently, it has been shown that pyrosequencing

can detect quasispecies in the main core immunodominant motifs of the core gene of HBV (Homs *et al.*, 2012).

1.5 Aims and Objectives

The objectives of this study were:

(i) to carry out an in-depth analysis of the relationship of the subgenotypes of genotype D using complete genome sequence data, nucleotide divergence calculations, phylogenetic analysis and bioinformatics to detect amino acid signature motifs for each subgenotype and to redefine their geographical distribution.

(ii) To molecularly characterize HBV in HBsAg-positive Sudanese liver disease patients and asymptomatic carriers of the virus, by genotyping as well as sub-genomic and full genomic sequence analysis of HBV isolates.

(iii) To molecularly characterize HBV infection in HBV/HIV co-infected Sudanese individuals with either overt or OBI.

(iv) To develop and test bioinformatics tools to explore UDPS HBV sequence data and to compare UDPS results with those obtained from cloning based sequencing (CBS).

Chapter 2: Paper I

Genotype D of hepatitis B virus and its subgenotypes: An update

YOUSIF, M. & KRAMVIS, A. 2013

Hepatology Research, 43, 355-364.

Journal: Hepatology *Research* is a peer-reviewed, ISI-accredited journal, with an impact factor of 2.07

Hepatology Research

Hepatology Research 2013; **43**: 355–364 doi: 10.1111/j.1872-034X.2012.01090.x

Original Article

Genotype D of hepatitis B virus and its subgenotypes: An update

Mukhlid Yousif and Anna Kramvis

Hepatitis Virus Diversity Research Programme, Department of Internal Medicine, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

Aim: The aim of the present study was to systematically and comparatively analyze the subgenotypes of genotype D of hepatitis B virus.

Methods: In total, 304 complete genomes of all genotype D subgenotypes were downloaded from the public databases. The sequences were analyzed using nucleotide divergence calculations, phylogenetic analysis and bioinformatics to detect amino acids signature motifs for each subgenotype and to define their geographical distribution.

Results: Intragroup divergence ranged from 0.8 ± 0.5 (% ± standard deviation) for subgenotype D6 to 3.0 ± 0.3 for D8. Inter-subgenotype divergence mostly ranged 4–7.5%. Phylogenetic analysis of genotype D showed separation into six distinct clusters (subgenotypes D1, D2, D3/D6, D4, D5 and D7/D8) with good bootstrap support. The mean intergroup divergence between D3 and D6 was the lowest and fell below the threshold of 4%, which is required to define a subgenotype, suggesting that subgenotypes D3 and D6 belong to one subgenotype. "D8" is a genotype D/E recombinant, which clusters with D7. A number of signature amino acids were found in all four open reading frames that could differentiate the subgenotypes, which also showed distinct geographical distribution.

Conclusion: There are six and not eight subgenotypes of D, D1–D6, which can be differentiated by distinct clustering with high bootstrap support and signature amino acids. Subgenotypes D3 and "D6" should be reclassified as a single subgenotype D3 and it would be more correct to classify "D8" as a genotype D/E recombinant rather than a subgenotype.

Key words: bioinformatics, genotype, phylogenetic analysis, serotype, signature amino acids, subgenotype

INTRODUCTION

HEPATITIS B VIRUS (HBV), the prototype member of the family *Hepadnaviridae*, is hepatotropic and replicates by reverse transcription. HBV is responsible for chronic infection of approximately 350 million people worldwide[1] and causes both acute and chronic liver disease. The majority of acute HBV infections are self-limiting, whereas chronic HBV infection usually has a lifelong course that may lead to the development of liver cirrhosis and hepatocellular carcinoma (HCC; liver cancer).[2,3]

Hepatitis B virus has a partially double-stranded genome composed of approximately 3.2 kb, with four overlapping open reading frames (ORF: preS/S, prec/core, X and polymerase).[4–6] As a result of its unusual mechanism of replication by reverse transcription and the lack of proofreading ability of its reverse transcriptase,[7] HBV displays sequence heterogeneity. To date, nine genotypes, A–I, defined by a intergroup divergence of at least 7.5% over the entire genome and approximately 39 subgenotypes, with intergroup divergence of 4–7.5%, have been identified.[8–10] A 10th genotype J has been proposed.[11] Most genotypes and some subgenotypes have a distinct geographical distribution.[6,9] HBV genotypes and, in some cases, subgenotypes have been shown to play an important role in the clinical consequences of the infection as well as in the response to antiviral treatment.[12,13]

Compared to the other genotypes, genotype D has the shortest genome of 3182 nucleotides in length and is characterized by a 33 nucleotide deletion at the beginning of the pre-S1 region.[6] Genotype D has been classified into eight subgenotypes, D1–D8, showing distinct geographical distributions.[8,14–18]

Correspondence: Professor Anna Kramvis, Hepatitis Virus Diversity Research Programme, University of the Witwatersrand, 7 York Road, Parktown, Johannesburg, 2193, South Africa.
Email: Anna.Kramvis@wits.ac.za
Received 12 July 2012; revision 16 August 2012; accepted 19 August 2012.

© 2012 The Japan Society of Hepatology

The aim of this study was to carry out an in-depth analysis of the relationship of the subgenotypes of genotype D using complete genomes, nucleotide divergence calculations, phylogenetic analysis and bioinformatics to detect amino acids signature motifs for each subgenotype and to redefine their geographical distribution.

METHODS

HBV nucleotide sequence selection

IN TOTAL, 301 complete genome sequences of 3182 nucleotides in length of subgenotypes D1–D7 of HBV and three sequences of subgenotype D8 (3212 nucleotides in length) were downloaded from the public databases (GeneBank/The European Molecular Biology Laboratory [EMBL]/DNA Databank of Japan [DDBJ]) as follows: subgenotypes D1 (153), D2 (64), D3 (30), D4 (six), D5 (12), D6 (six), D7 (30) and D8 (three), up to 15 March 2012.

Phylogenetic analysis

The complete genome sequences were aligned using Gendoc,[19] and neighbor-joining phylogenetic trees with a bootstrap of 1000 replicates were constructed using Molecular Evolutionary Genetic Analysis MEGA ver. 5 (http:/megasoftware.net) software program based on Kimura two-parameter distance estimation.

Nucleotide divergence calculations

Nucleotide divergence calculations were carried out using DAMBE,[20] and the means and standard deviation (SD) were calculated to classify the subgenotypes.

Analysis for intergenotypic recombination

SimPlot ver. 3.5.1, an interactive 32-bit software program that plots distances (or similarity) versus position,[21] was used to show recombination in isolates FN594769–FN594771, previously classified as subgenotype D8.[15] The window size of 200 bp and step sizes of 20 bp were used. The putative recombinants were compared to representative sequences belonging to genotypes A–I, with outgroup genotype F.

The SimPlot ver. 3.5.1 program was used to find informative sites.[22,23] Informative sites are sites at which two sequences have a nucleotide in common and the other two sequences share another one. Each informative site supports one of three possible phylogenetic relationships among the four taxa. Bootscanning[24] and cluster analysis maximizing the χ^2 were used to identify the breakpoints in the intergenotypic recombinants.[21,25] *P*-values for the subsequent division of the sequence into genotypes were calculated using Fisher's exact test.

Identification of amino acids signatures

All nucleotide sequences were translated using GenDoc[19] and individual subgenotypes were searched for signature amino acids in the different ORFs of the HBV genome.

RESULTS AND DISCUSSION

UNTIL 2008, GENOTYPE D of HBV was classified into four subgenotypes (D1–D4).[8,16] However, since then, an additional four subgenotypes of D have been identified, D5 in 2006,[18] D6 in 2009,[17] D7 in 2009[14] and D8 in 2010.[15] These subgenotypes were proposed using an intergroup divergence of 4–7.5% and significant bootstrap support in the phylogenetic tree.[8,16]

Divergence calculations and phylogenetic analysis

Table 1 summarizes the intragroup and intergroup nucleotide divergences (% ± SD) of 304 complete HBV genome sequences belonging to subgenotypes D1–D8. The intragroup divergences ranged from 0.8 ± 0.5 (% ± SD) for subgenotype D6 to 3.0 ± 0.3 (% ± SD) for subgenotype D8. The intergroup divergence between the subgenotypes ranged between 4% and 7.5% with some exceptions, which are shown in bold in Table 1. The mean intergroup nucleotide divergence was the lowest between subgenotype D3 and subgenotype D6 at 2.6 ± 0.6%.

The phylogenetic tree comparing the 304 complete HBV genomes, belonging to genotype D, showed separation into six clusters, subgenotypes D1, D2, D3/D6, D4, D5 and D7/D8, supported by high bootstrap values (Fig. 1). Using the algorithm described by Purdy in 2007,[26] the serological subtypes or serotypes were determined. The subgenotypes of D show different frequencies of serological subtypes: 94% of subgenotype D1 strains were classified as *ayw*2; 91% of subgenotype D2 as *ayw*3; 57% of subgenotype D3 as *ayw*2, 40% *ayw*3 and 3% *ayw*4; and 92% of subgenotype D5 was classified as *ayw*3. All the strains belonging to subgenotypes D4, D6, D7 and D8 were classified as *ayw*2 (Fig. 1). We confirmed the significant association of *ayw*2 with D1 and *ayw*3 with D2, as previously shown.[8] In the present study, we demonstrate that *ayw2* has a statistically significant association with D4, D6, D7 and D8, while *ayw*3 is associated with D5 ($P < 0.05$) (Fig. 1).

© 2012 The Japan Society of Hepatology

Table 1 Mean intragroup and intergroup divergence of the subgenotypes of genotype D determined using 304 sequences downloaded from public databases

	D1	D2	D3	D4	D5	D6	D7	D8
No. of sequences	153	64	30	6	12	6	30	3
D1	2.6 ± 0.5†							
D2	**3.1 ± 0.5**	2.1 ± 0.8						
D3	**3.8 ± 0.6**	**3.9 ± 0.6**	2.2 ± 0.7					
D4	4.6 ± 0.6	4.5 ± 0.7	4.5 ± 0.7	2.2 ± 1.0				
D5	5.0 ± 0.6	5.0 ± 0.5	4.9 ± 0.6	5.0 ± 0.6	1.7 ± 0.7			
D6	**3.7 ± 0.5**	**3.6 ± 0.5**	**2.6 ± 0.6**	4.4 ± 0.6	5.0 ± 0.4	0.8 ± 0.5		
D7	4.7 ± 0.4	4.7 ± 0.4	4.7 ± 0.6	4.0 ± 0.5	5.4 ± 0.4	4.8 ± 0.4	2.8 ± 0.6	
D8	5.2 ± 0.5	5.2 ± 0.4	5.2 ± 0.5	4.5 ± 0.6	5.8 ± 0.4	5.3 ± 0.4	4.8 ± 1.6	3.0 ± 0.3
D8–E‡	5.0 ± 0.4	5.2 ± 0.4	5.4 ± 0.5	4.2 ± 0.4	5.8 ± 0.2	5.6 ± 0.3	**3.4 ± 0.3**	2.9 ± 0.5

†The intragroup divergence shaded in grey and the intergroup divergence of less than 4% are shown in bold.
‡Mean intergroup and intragroup divergence determined after removing sequences at positions 1640–2430 and 2830–3128 from the *EcoR*1 site, which represent the region of recombination between genotype D and E in subgenotype D8.

The separation of subgenotype D6, from Papua, from subgenotype D3, was not statistically supported whereas its clustering with subgenotype D3 was well supported (bootstrap = 99%), consistent with the fact that the mean intergroup divergence between subgenotypes D3 and D6 was the lowest (Table 1). Using six new sequences, subgenotype D6 was identified by Utsumi and colleagues as a separate subgenotype to D3, with 3.3–4.7% intergroup divergence and 88% bootstrap support in the phylogenetic tree.[17] However, by using 30 sequences of subgenotype D3 and the six sequences of D6, we found a lower intergroup divergence of 2.6 ± 0.6 (% ± SD), with lower bootstrap support of 56% for separate clustering of D3 and D6 and a significantly high bootstrap support of 99% for the co-clustering of D3 and D6. This new analysis suggests that D6 is not a separate subgenotype but rather a clade of subgenotype D3.

Although the previous analysis showed a 100% bootstrap support for the separation of subgenotypes D7 and D8[15], our present analysis does not support this separation but supports the co-clustering of D7 with D8 at 75%. Moreover, using SimPlot analysis we corroborated the previous findings that the three strains FN594769–FN594771, proposed to belong to subgenotype D8,[15] are a recombinant of genotype D and E. The fourth isolate, FN594768, classified as subgenotype D8,[15] was excluded from the present analyses because it contained an insertion. As demonstrated for FN594769 in Figure 2, by mapping the informative sites we estimated the positions of transition between genotypes D and E. The recombination profile determined for the three strains did not differ significantly from that determined in the earlier analysis of Abdou Chekaraou *et al*., who showed that FN594769–FN594771 clustered with genotype E within the regions ~1640–2430, and ~2830–3128 from the *EcoR*1 site[15] After removing these genotype E regions from all sequences compared, the remaining region showed a decreased mean intergroup divergence of subgenotype D8 with D7 being 3.4 ± 0.3 (% ± SD) instead of 4.8 ± 1.6 (% ± SD) when the complete sequence was compared (Table 1). Furthermore, the length of the subgenotype D8 of 3212 nucleotides is longer than the characteristic length for genotype D of 3182 nucleotides[6,16] and it also has the three nucleotide deletion in the pre-S1, which is characteristic of genotype E. Therefore, we propose that subgenotype D8 should be considered as a genotype D/E recombinant rather than a subgenotype of D as has been done for the C/D recombinant circulating in Tibet.[27]

Even though, in agreement with others,[9,28] we found the intergroup divergence between subgenotypes D1 and D2 to be 3.1 ± 0.5 (% ± SD) (Table 1), below the threshold of 4%, which is used to separate subgenotypes, the separation into subgenotypes is justified. First, the separate clustering of D1 and D2 is well supported with significant bootstraps of 87% and 94% for D1 and D2, respectively; second, they express different serological subtypes, D1 *ayw*2 and D2 *ayw*3; and third, they show distinct geographical distribution, with D1 found mainly in the Middle East and D2 in Europe (Fig. 1).

© 2012 The Japan Society of Hepatology

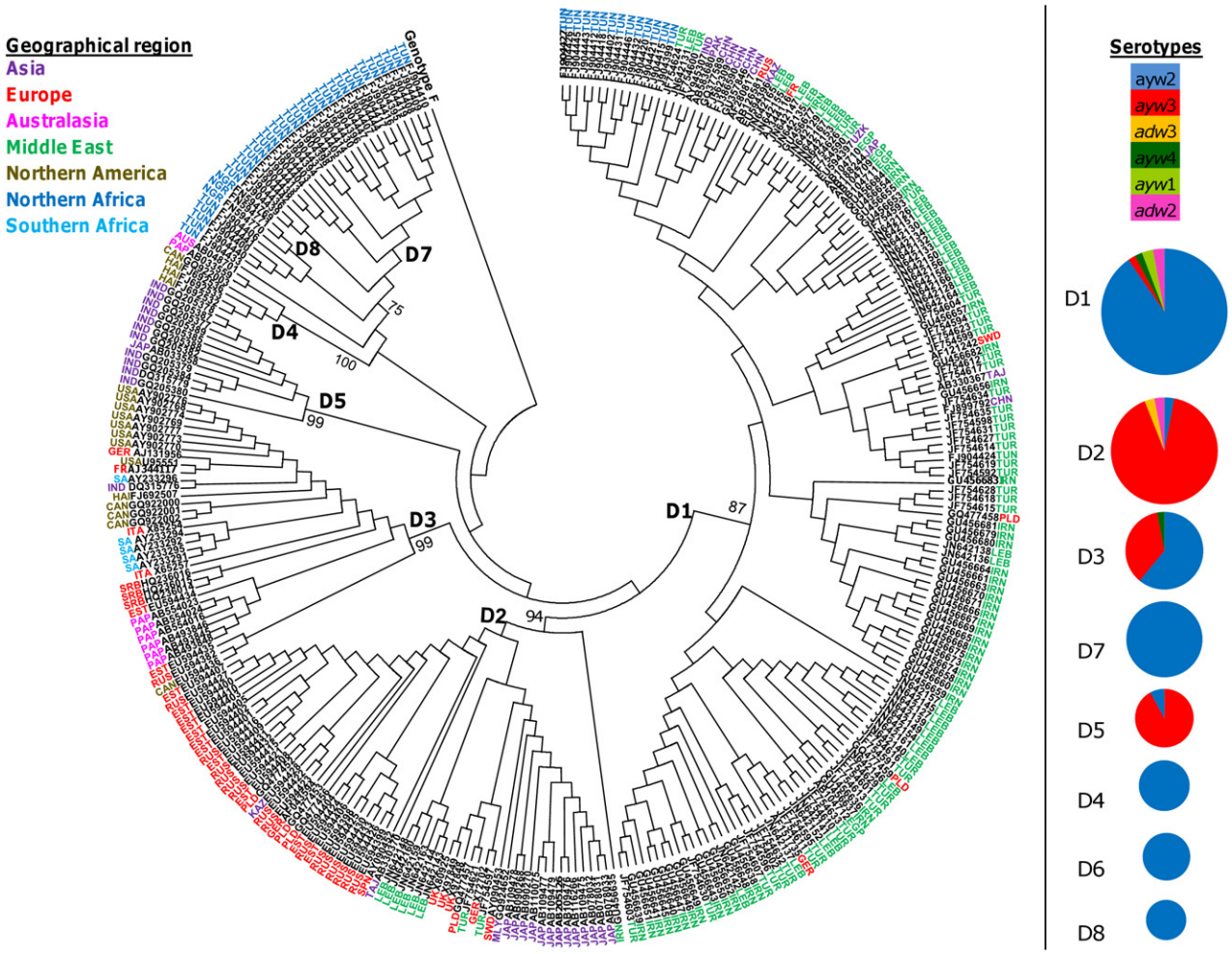


Figure 1 A rooted phylogenetic tree of 304 complete genomes sequences of hepatitis B virus of the "eight" subgenotypes of genotype D using neighbor-joining analysis. Bootstrap statistical analysis was performed using 1000 datasets, indicated as percentages on the nodes. The sequences are labeled by their accession numbers and their country of origin: AUS, Australia; CAN, Canada; CHN, China; EGP, Egypt; EST, Estonia; FR, France; GER, Germany; HAI, Haiti; IND, India; IRN, Iran; ITA, Italia; JAP, Japan; KAZ, Kazakhstan; LEB, Lebanon; MLY, Malaysia; NIG, Niger; PAK, Pakistan; PAP, Papua New Guinea/Indonesia; PLD, Poland; RUS, Russia; SRB, Serbia; SA, South Africa; SPN, Spain; SDW, Sweden; TAJ, Tajikistan; TUN, Tunisia; TUR, Turkey; UZK, Uzbekistan. The size of the pie charts indicate the relative sample size of the sequences analyzed. The letter D followed by the numbers represent the subgenotypes of D.

Comparison of amino acids sequences of the subgenotypes of D

Table 2 summarizes the comparison of amino acids of the 304 HBV complete genomes belonging to the subgenotypes of genotype D. The amino acids that were found predominantly in the individual subgenotypes are shaded in grey, whereas those that were specific to an individual subgenotype and found only in that subgenotype are encircled.

Although signature amino acids in the S region of subgenotypes D1–D4 were previously compared,[16] this is the first study to compare amino acids signatures in all the ORF of the subgenotypes D1 to "D8". Some subgenotypes had specific signature amino acids that can differentiate them from the other subgenotypes of D. D5 had ps1:28R in the pre-S1, sp:88S in the spacer of the polymerase region, while D4 had tp:161V in the terminal protein and rh:107P in RNAse H in the polymerase region. Moreover, there are a number of amino acids

© 2012 The Japan Society of Hepatology

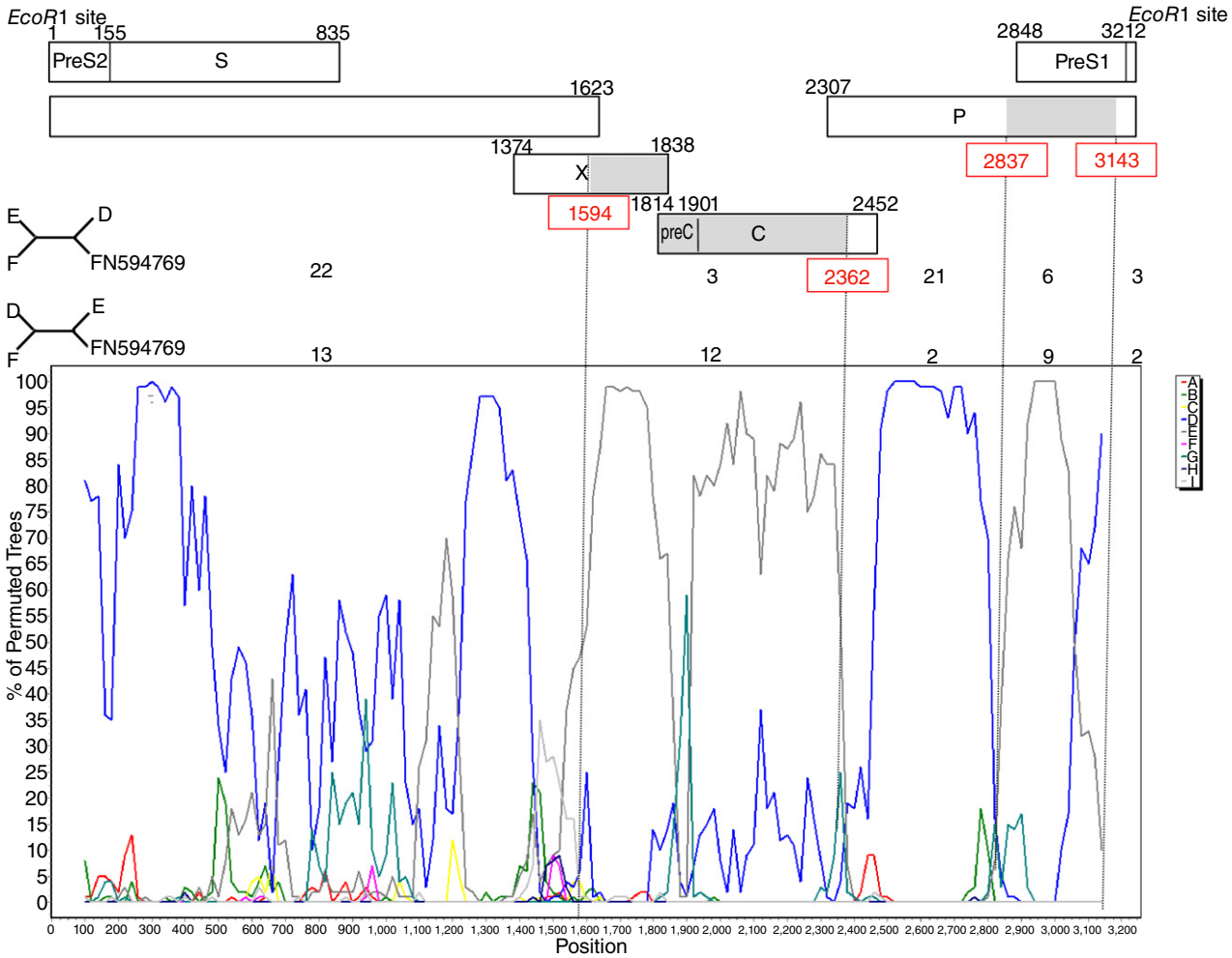


Figure 2 Resolution of the chimeric structure of hepatitis B virus (HBV) genome using a bootscanning graph, which shows the bootscanning values of query sequence of isolate FN594769 to representative sequences of genotypes A–I as indicated on the figure key with outgroup, genotype F. The graph was generated using Simplot ver. 3.5.1 with window size 200 bp, step size 20 bp, gap-strip off, 100 bootstrap replicates, Kimura transition/transversion ratio:2 and neighbor-joining. Dashed vertical lines indicate the position of the breakpoints determined using maximization of χ^2 as described in Methods. The *P*-values were calculated using Fisher's exact test and were less than 0.05. The numbers of informative sites are shown on the right hand side of the four-member trees. The informative sites shared by the isolate with genotype D are shown above and with genotype E, below. The genomic regions belonging to genotype E are shaded in grey on the top of the figure, which shows the genomic arrangement of the HBV open reading frames. Numbers above the open reading frames denote nucleotide positions from the start of *EcoR*1 restriction site.

that are predominately found in certain subgenotypes. Amino acids ps1:85A, ps2:43L and ps2:54P in pre-S2 x:36S; x:102T and x:110Q in X region; and tp:45T, sp:100T, sp:115D, sp:122S, sp:139N and rt:336M in the reverse transcriptase were found predominantly in subgenotype D5. Subgenotype D4 could be differentiated by tp:90N, sp: 95T and rt:122I. Subgenotypes D3/"D6" had the same signature amino acids in all regions, ps2:42L, sp:59F, sp:80F, sp:153L, rt:21S, rt:122L and rt:130P, further supporting our proposal that these are not separate subgenotypes of D. Subgenotype "D7" could be distinguished by rh:138V and subgenotype "D8" by sp:153S, which is also found in genotype E. Subgenotype "D7" and "D8" share the signature amino acid at rt:237T. Subgenotype D1 could be differentiated by ps2:139V and sp:100S, while D2 was differentiated by sp:128G and rt:126R. No signature amino acids were found in the hepatitis B surface antigen and preC/C

© 2012 The Japan Society of Hepatology

Table 2 A comparison of the amino acid residues encoded by the different subgenomic regions of HBV belonging to the subgenotypes of D

Region	Amino acids position	Subgenotype of D (304)†								Genotype E (12)
		D1 (153)	D2 (64)	D3 (30)	D6 (6)	D4 (6)	D5 (12)	D7 (30)	D8 (3)	
Pre-S1 (ps1)	28	A/S,T,K‡	A/T	A/T	A	N	(R)	N/T	R	R
	85	S/A,T	S/P	S	S	S	A	S	S	S
Pre-S2 (ps2)	39	V/A	A/V	A	A	A	A/V	A/V	A	A/V
	42	I/T,A,L	I	L/I	L	I/T	I	I	I	I
	43	S/L	S/L	S	S	S/L	L/S	S/L	S	S
	54	L/M,Q,T,R	L	L/Q,P	L/Q	L	P/L	L/P,V	L	P/L
X (x)	36	T/S,A,P,D,N	T/A,P,D,G	T/S	T	A/T,P	S	D/G	D	D/G
	102	V/G,A,M	V/A	A/T	A	A	T/A	A/T	A	A/V
	110	A	A	A	A	A	Q/P,A	A	A	A
Terminal Protein (tp)	45	N/D,T	N/T	N	N	N	T/N	N/D	N	L
	90	K/T,N,Q,E	K/N	K	K	N/Y,K	K	K/N	K	N
	161	E/D	E	E	E	(V)	E	E	E	E
Spacer (sp)	59	I/T,N,V	I/V,T,F,L	F/V,S	F	V	V/A	V	V	V
	80	L/H,S,P,F,R	L/P,R,F	F/C	F	L/F	L/F	L/F	L	L
	88	L/F,I,R,H,Q	I/L,F,T,V	L/I	L	L/F	(S)	L/F	L/I	L
	95	K/T,Q,E	K	K/Q	K	T	K	K/T	K	K
	100	S/A,T	A/T,V	A/T,V,S	A	A	T/S	A/T,P	A	N
	115	E/D	E	E	E	E	D/E	E	E	E
	122	N/S,H,T	N/K	N/H	N	N	S/N	N/K	N	S
	128	S/G	G/S	S	S	S	S	S	S	S
	139	Q	Q	Q	Q/R	Q	N/Q	Q	Q/H	Q
	153	H/R,L,Y	H	L/H	L	H	H	H	S	S
Reverse transcriptase (rt)	7	A/T,D,V	A/T,V	A	A	A	T/A	A/V,T,G	A	T
	21	A/T,G,N,S,V	A	S/A	S	A	A	A/T	A	A
	122	F/V,L,I,D	F/V,L,I,C	L/F	L	I	F	F/V	F	I/T
	126	H	R/H	H	H	H	H	H	H	Y/H
	130	Q/P	Q/P	P/Q	P	Q	Q/P	Q	Q	P
	237	P	P	P	P	P	P	T	T	P
	269	I/L	I/L	I	I	I	L	I	I	I
	336	L/M	L/M	L/M	L	L	M	L/M	L	M/L
RNAse H (rh)	107	L/I	L/I,V	L	L	(P)	L	I/V	I	V/I,G
	138	D/V	D/V,G	D	D	D	D	V/D	D	V

The amino acids that were found predominantly in the individual subgenotypes of D are shaded in grey, whereas those that were specific to an individual subgenotype and found only in that subgenotype are encircled.

†Number of the sequences included in the analysis.

‡The relative size of the font is an indication of the frequency of the amino acid at that position for each subgenotype.

© 2012 The Japan Society of Hepatology

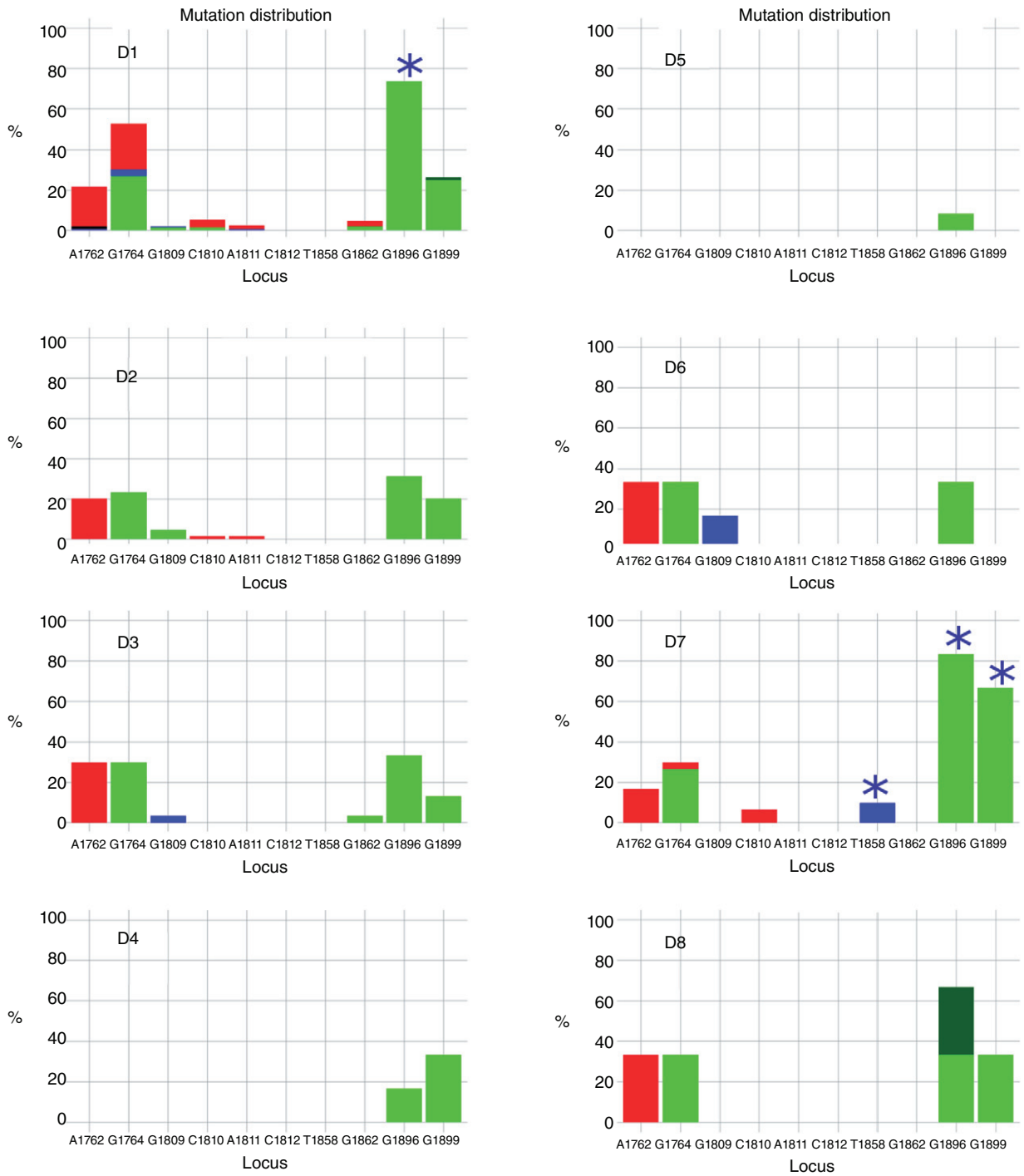


Figure 3 Mutation distribution graphs, for the basic core promoter/precore region (1742–1901 from the *EcoR*1 site), showing the percentage of mutant residues relative to the reference motif found at the 10 loci of interest (1762, 1764, 1809–1812, 1858, 1862, 1896 and 1899). Eight datasets, one for each subgenotype of D, were submitted to the Mutation Reporter Tool to produce the graphs (153 sequences for D1, 64 for D2, 30 for D3, six for D4, 12 for D5, six for D6, 30 for D7 and three for D8). The reference motif used was AGGCACTGGG. This is also shown by the letter preceding each locus on the *X*-axis. To facilitate direct comparisons between the graphs, conserved loci were not suppressed and the *Y*-axis was scaled to 100% by selecting the appropriate controls on the input page of the Mutation Reporter Tool (Bell TG and Kramvis A, unpublished data). Residue A (■), C (■), G (■), R (■), T (■).

© 2012 The Japan Society of Hepatology

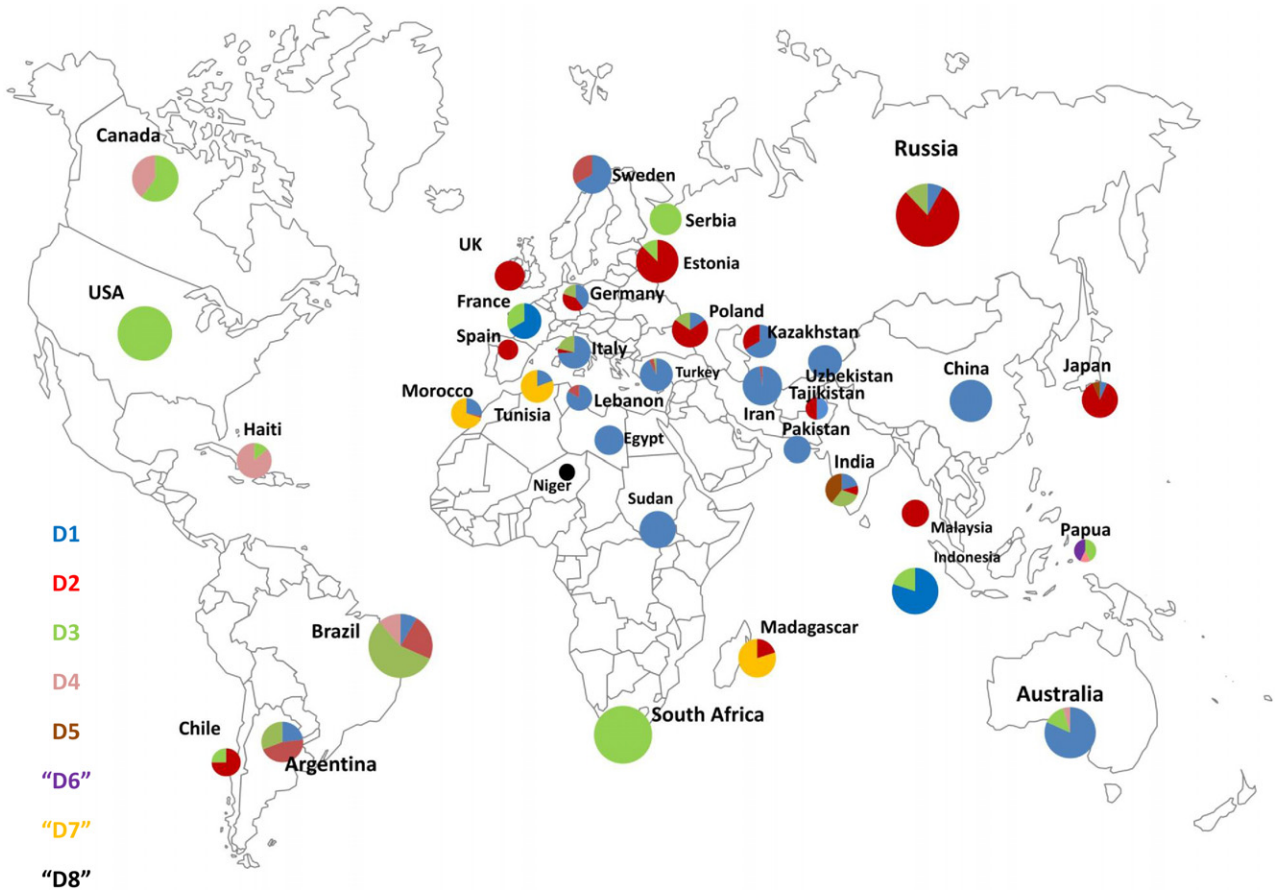


Figure 4 Global distribution of the "eight" subgenotypes of genotype D. The size of the pie charts indicates the relative sample size of the number of sequences analyzed. The geographical distribution of the subgenotypes of D was updated using 304 complete genomes (Fig. 1) and 488 subgenomic sequences from which the subgenotype could be determined (accession numbers available upon request). According to the present analysis "D6" is reclassified as D3, "D7" becomes D6 and "D8" is more correctly classified as a D/E recombinant.

hepatitis B e-antigen (HBeAg) precursor proteins of the different subgenotypes of D.

Mutations in the basal core promoter (BCP) and precore (PC) regions

Using our team's Mutation Reporter Tool (Bell TG and Kramvis A, unpublished data) for the analysis of 10 loci in the BCP and PC (1742–1899 from the *EcoR*1 site), described previously,[8] certain correlations were found between the frequency of the mutations in these loci and the subgenotypes of D. The frequency of the mutations in the different subgenotypes are illustrated in Figure 3. The BCP A1762T/G1764A occurred in all subgenotypes, except for subgenotypes D4 and D5, but this may be as a result of the relatively low numbers of sequences available for these subgenotypes and this possibility can only be precluded once additional sequences become available. The mutation G1896A, which abolishes the translation of HBeAg,[29] was found in significantly higher frequency in subgenotypes D1 and "D7", at 74% and 83%, respectively ($P < 0.05$) (Fig. 3). G1899A, which often occurs together with G1896A, occurred in 67% of subgenotype "D7", at a significantly higher frequency than in the other subgenotypes ($P < 0.05$). 1858C occurred in 10% of subgenotype "D7" strains, also at a statistically higher frequency than in the other subgenotypes ($P < 0.05$). The 1858C is positively associated with genotype A whereas it is negatively associated with genotype D.[8] In order to exclude the possibility that the subgenotype "D7" sequences with 1858C were genotype A/D recombinants in the PC region, we carried out Simplot analysis,[21] which did not

© 2012 The Japan Society of Hepatology

reveal any recombination between genotype A and subgenotype "D7" (data not shown).

Drug resistance and immune escape mutations

Long-term antiviral treatment directed against HBV can lead to the development of drug resistance as well as immune and vaccine escape mutations.[30,31] Although a number of clinically relevant mutations were found to occur (Tables S1 and S2), there was no correlation between these mutations and the subgenotypes of D.

Geographical distribution of subgenotypes of genotype D

Figure 4 updates the geographical distribution of the subgenotypes of D using the 304 complete genomes sequences and 488 partial subgenomic sequences. Subgenotype D1 was found mainly in the Middle East, subgenotype D2 in Europe and Asia, subgenotype D3 in South Africa, Northern America and Europe, subgenotype D4 in Northern and Central Americas and Australia, subgenotype D5 only in India, subgenotype "D6" reclassified as D3 from Papua New Guinea /Indonesia, subgenotype "D7", which becomes subgenotype D6 after the present analyses, in the Maghreb (Morocco, Algeria and Tunisia) and subgenotype "D8", the genotype D/E recombinant from Niger.

To conclude, these comprehensive analyses of the subgenotypes of genotype D clearly demonstrate that there are six rather than eight subgenotypes of D, D1–D6, which can be differentiated by distinct clustering with high bootstrap support and signature amino acids. Subgenotypes D3 and "D6" should be reclassified as a single subgenotype, D3, and it would be more correct to classify the previously named subgenotype "D8" as a genotype D/E recombinant rather than a separate subgenotype of genotype D.

ACKNOWLEDGMENTS

WE ACKNOWLEDGE THE financial support from the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) and the National Research Foundation of South Africa (NRF; GUN#65530).

REFERENCES

1 WHO. Hepatitis B virus Fact sheet N°204 2008.
2 Kao JH, Chen DS. Global control of hepatitis B virus infection. *Lancet Infect Dis* 2002; **2:** 395–403.
3 Seeger C, Mason WS. Hepatitis B virus biology. *Microbiol Mol Biol Rev* 2000; **64:** 51–68.
4 Tiollais P, Pourcel C, Dejean A. The hepatitis B virus. *Nature* 1985; **317:** 489–95.
5 Ganem D, Varmus HE. The molecular biology of the hepatitis B viruses. *Annu Rev Biochem* 1987; **56:** 651–93.
6 Kramvis A, Kew M, Francois G. Hepatitis B virus genotypes. *Vaccine* 2005; **23:** 2409–23.
7 Steinhauer DA, Holland JJ. Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. *J Virol* 1986; **57:** 219–28.
8 Kramvis A, Arakawa K, Yu MC, Nogueira R, Stram DO, Kew MC. Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. *J Med Virol* 2008; **80:** 27–46.
9 Kurbanov F, Tanaka Y, Mizokami M. Geographical and genetic diversity of the human hepatitis B virus. *Hepatol Res* 2010; **40:** 14–30.
10 Mulyanto, Pancawardani P, Depamede SN *et al*. Identification of four novel subgenotypes (C13–C16) and two intergenotypic recombinants (C12/G and C13/B3) of hepatitis B virus in Papua Province, Indonesia. *Virus Res* 2012; **163:** 129–40.
11 Tatematsu K, Tanaka Y, Kurbanov F *et al*. A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. *J Virol* 2009; **83:** 10538–47.
12 Kramvis A, Kew MC. Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy. *J Viral Hepat* 2005; **12:** 456–64.
13 Kao JH. Hepatitis B viral genotypes: clinical relevance and molecular characteristics. *J Gastroenterol Hepatol* 2002; **17:** 643–50.
14 Meldal BHM, Moula NM, Barnes IHA, Boukef K, Allain J. A novel hepatitis B virus subgenotype, D7, in Tunisian blood donors. *J Gen Virol* 2009; **90:** 1622–8.
15 Abdou Chekaraou M, Brichler S, Mansour W *et al*. A novel hepatitis B virus (HBV) subgenotype D (D8) strain, resulting from recombination between genotypes D and E, is circulating in Niger along with HBV/E strains. *J Gen Virol* 2010; **91:** 1609–20.
16 Norder H, Courouce AM, Coursaget P *et al*. Genetic diversity of hepatitis B virus strains derived worldwide: genotypes, subgenotypes, and HBsAg subtypes. *Intervirology* 2004; **47:** 289–309.
17 Utsumi T, Lusida MI, Yano Y *et al*. Complete genome sequence and phylogenetic relatedness of hepatitis B virus isolates in Papua, Indonesia. *J Clin Microbiol* 2009; **47:** 1842–7.
18 Banerjee A, Kurbanov F, Datta S *et al*. Phylogenetic relatedness and genetic diversity of hepatitis B virus isolates in Eastern India. *J Med Virol* 2006; **78:** 1164–74.

© 2012 The Japan Society of Hepatology

19 Nicholas KBN, Nicholas HB, Deerfield DW. GeneDoc: analysis and visualization of genetic variation. *EMBnewnews* 1997; **4:** 14.
20 Xia X, Xie Z. DAMBE: software package for data analysis in molecular biology and evolution. *J Hered* 2001; **92:** 371–3.
21 Lole KS, Bollinger RC, Paranjape RS *et al*. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 1999; **73:** 152–60.
22 Smith JM. Analyzing the mosaic structure of genes. *J Mol Evol* 1992; **34:** 126–9.
23 Robertson DL, Hahn BH, Sharp PM. Recombination in AIDS viruses. *J Mol Evol* 1995; **40:** 249–59.
24 Salminen MO, Carr JK, Burke DS, McCutchan FE. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* 1995; **11:** 1423–5.
25 Robertson DL, Sharp PM, McCutchan FE, Hahn BH. Recombination in HIV-1. *Nature* 1995; **374:** 124–6.
26 Purdy MA, Talekar G, Swenson P, Araujo A, Fields H. A new algorithm for deduction of hepatitis B surface antigen subtype determinants from the amino acid sequence. *Intervirology* 2007; **50:** 45–51.
27 Zhou B, Xiao L, Wang Z, Chang ET, Chen J, Hou J. Geographical and ethnic distribution of the HBV C/D recombinant on the Qinghai-Tibet Plateau. *PLoS ONE* 2011; **6:** e18708.
28 Ghosh S, Banerjee P, RoyChoudhury A *et al*. Unique hepatitis B virus subgenotype in a primitive tribal community in Eastern India. *J Clin Microbiol* 2010; **48:** 4063–71.
29 Carman W, Hadziyannis S, McGarvey M *et al*. Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. *Lancet* 1989; **334:** 588–91.
30 Sloan RD, Ijaz S, Moore PL, Harrison TJ, Teo CG, Tedder RS. Antiviral resistance mutations potentiate hepatitis B virus immune evasion through disruption of its surface antigen a determinant. *Antivir Ther* 2008; **13:** 439–47.
31 Sheldon J, Ramos B, Garcia-Samaniego J *et al*. Selection of hepatitis B virus (HBV) vaccine escape mutants in HBV-infected and HBV/HIV-coinfected patients failing antiretroviral drugs with anti-HBV activity. *J Acquir Immune Defic Syndr* 2007; **46:** 279–82.

SUPPORTING INFORMATION

ADDITIONAL SUPPORTING INFORMATION may be found in the online version of this article at the publisher's website:

Table S1 Genotype D strains with mutations in the reverse transcriptase domain that may lead to drug resistance.
Table S2 Genotype D strains with mutations in the HBsAg that may lead to vaccine, detection and/or immune escape.

© 2012 The Japan Society of Hepatology

Table S1: Genotype D strains with mutations in the reverse transcriptase domain that may lead to drug resistance

			HBV reverse transcriptase domain												
Codon			173	180	181	184	191	194	202	204	207	213	214	215	233	236
			V	L	A	T	V	A	S	M	V	S	V	Q /P/H	I	N
Subgenotype	Accession numbers	Country														
D1	GU456682	Iran	.	.	.	.	.	V	.	.	.	.	.	S	.	.
	GU456646		.	.	.	.	.	.	.	.	.	.	.	.	V	.
	GU456638		M	.	.	.	.	.	.	.	.	.	.	.	.	.
	GU456647		.	.	.	.	.	.	.	.	.	.	.	S	.	.
	GU456645		.	.	.	.	.	.	.	.	.	.	.	S	.	.
	GU456639		.	.	.	.	.	.	.	.	M	.	E	.	.	.
	GU456637		.	.	.	.	.	.	.	.	.	T	.	.	.	.
	JF754633	Turkey	.	.	.	.	.	.	.	.	.	T	.	.	.	.
	JF754631		.	.	.	.	.	.	.	.	.	.	.	S	.	.
	JF754629		.	.	.	.	.	.	.	.	.	.	A	S	.	.

JF754623		.	.	.	.	.	.	.	.	.	T	.		.		.	.	
JF754619		.	.	.	.	.	.	.	.	.	.	.		S		.	.	
JF754599		.	.	.	.	.	.	.	.	.	T	.		.		.	.	
JF754614		.	.	.	.	.	.	.	.	.	T	.		.		.	.	
JF754598		.	.	.	.	.	.	.	.	.	T	.		.		.	.	
FJ904431	Tunisia	.	.	.	.	.	.	.	.	.	.	.		S		.	.	
FJ904446		.	.	.	.	.	.	.	.	.	T	.		.		.	.	
JN642155		.	.	.	.	I	.	.	.	.	.	.		.		.	.	
JN642153		.	.	.	.	.	.	.	.	.	.	.		S		.	.	
JN642149		.	.	.	.	.	.	.	.	.	T	.		.		.	.	
JN642145		.	.	.	.	.	.	.	.	.	.	.		S		.	.	
JN642137	Lebanon	.	.	.	.	.	.	.	.	.	.	E		.		.	.	
JN642129		.	.	.	.	.	.	.	.	.	.	E		.		.	.	
JN642158		.	.	.	.	.	.	.	.	M	.	.		.		.	.	
JN642154		.	.	.	.	.	.	.	.	.	T	.		.		.	.	

	JN642150		.	.	.	.	.	.	.	.	.	.	.	S	.	.
	JN642140		.	.	.	.	.	.	.	.	.	.	G	.	.	.
	AF121242	Sweden	.	.	.	.	.	.	.	.	.	T	.	.	.	.
	GQ477453	Poland	M	.	.	.	.	.	.	.	.	.	.	.	.	.
	EU594398	Russia	.	.	.	.	.	.	.	.	.	.	.	S	.	.
	JN642163	Lebanon	.	.	.	.	.	T	.	.	.	T	.	.	.	.
D2	AF043594	Germany	.	.	.	.	.	.	.	.	.	.	.	S	.	.
	AB109478	Japan	.	.	.	.	.	.	.	.	.	T	.	.	.	.
	X80925	UK	.	.	.	.	.	.	T	.	.	.	.	.	.	.
D4	GQ922003	Canada	.	.	.	.	.	.	.	.	.	.	.	.	V	.
	FJ904441		.	.	.	.	.	.	.	.	.	.	E	.	.	.
	FJ904433		.	.	.	.	.	.	.	.	.	.	I	.	.	.
D7	FJ904395	Tunisia	.	.	.	.	.	.	.	.	.	.	A	.	.	.
	FJ904430		.	.	.	.	.	.	.	.	.	.	E	.	.	.
	FJ904414		.	.	.	.	.	.	.	.	.	.	.	S	.	.

	FJ904404		.	.	.	.	.	.	.	.	.	T	D	.		.	.
D6	AB493848	Papua	.	.	.	.	.	.	.	.	.	.	.	S		.	.
	AB493845		.	.	.	.	.	.	.	.	.	.	.	S		.	.
	AB554024		.	.	.	A	.	.	.	.	.	.	.	.		.	.
D8	FN594769	Niger	.	.	.	.	.	T	.	.	.	.	.	.		.	.

Dots indicate amino acids identity and the mutations are highlighted in grey. Only sequences, which had mutations, are shown. All other sequences were wild-type.

Table S2: Genotype D strains with mutations in the HBsAg that may lead to vaccine, detection and/or immune escape

Subgenotype	Accession numbers	Country	HBsAg					
			120	133	144	145	153	158
			P	M	D	G	P	F
D1	GU456682	Iran	.	.	.	.	.	L
	FJ904427	Tunisia	.	.	E	R	.	.
	JN642133	Lebanon	.	S	.	.	.	.
	JN642166		S	.	.	.	.	.
	JN642134		T	.	.	.	.	.
	AF151735	Germany	.	.	E	.	.	.
	AY945307	India	.	.	.	.	.	L
D2	EU594433	Russia	T	.	.	.	.	.
D5	GQ205389	India	.	.	.	E	.	.
	GQ205385		.	.	.	E	.	.
D8	FN594770	Niger	.	T	.	.	.	.

Dots indicate amino acids identity and the mutations are highlighted in grey. Only sequences, which had mutations, are shown. All other sequences were wild-type.

Chapter 3: Paper II

Molecular Characterization of Hepatitis B Virus in Liver Disease Patients and Asymptomatic Carriers of the Virus in Sudan

YOUSIF, M., MUDAWI, H., BAKHIET, S., GLEBE, D. & KRAMVIS, A (2013)

BMC Infectious Diseases, 13**,** 328.

Journal: *BMC Infectious Diseases,* is a peer-reviewed, ISI-accredited, open-access journal with an impact factor of 3.03

Yousif et al. BMC Infectious Diseases 2013, 13:328
http://www.biomedcentral.com/1471-2334/13/328

BMC Infectious Diseases

RESEARCH ARTICLE Open Access

Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan

Mukhlid Yousif[1], Hatim Mudawi[2], Sahar Bakhiet[3], Dieter Glebe[4] and Anna Kramvis[1*]

Abstract

Background: Hepatitis B virus is hyperendemic in Sudan. Our aim was to molecularly characterize hepatitis B virus from Sudanese individuals, with and without liver disease, because genotypes play an important role in clinical manifestation and treatment management.

Methods: Ninety-nine patients - 30 asymptomatic, 42 cirrhotic, 15 with hepatocellular carcinoma, 7 with acute hepatitis and 5 with chronic hepatitis- were enrolled. Sequencing of surface and basic core promoter/precore regions and complete genome were performed.

Results: The mean ± standard deviation, age was 45.7±14.8 years and the male to female ratio 77:22. The median (interquartile range) of hepatitis B virus DNA and alanine aminotransferase levels were 2.8 (2.2-4.2) log IU/ml and 30 (19–49) IU/L, respectively. Using three genotyping methods, 81/99 (82%) could be genotyped. Forty eight percent of the 99 patients were infected with genotype D and 24% with genotype E, 2% with putative D/E recombinants and 7% with genotype A. Patients infected with genotype E had higher frequency of hepatitis B e antigen-positivity and higher viral loads compared to patients infected with genotype D. Basic core promoter/precore region mutations, including the G1896A in 37% of HBeAg-negative individuals, could account for hepatitis B e antigen-negativity. Pre-S deletion mutants were found in genotypes D and E. Three isolates had the vaccine escape mutant sM133T.

Conclusion: Sudanese hepatitis B virus carriers were mainly infected with genotypes D or E, with patients infected with genotype E having higher HBeAg-positivity and higher viral loads. This is the first study to molecularly characterize hepatitis B virus from liver disease patients in Sudan.

Keywords: Bioinformatics, Genotype, Serotype, Sudan, Subgenotype, Africa, Phylogenetics

Background

Hepatitis B virus (HBV), the prototype member of the family *Hepadnaviridae* is responsible for chronic infection of more than 240 million people worldwide [1], of which 65 million reside in Africa [2].

Sudan is an African country with high HBV seroprevalence of greater than 8% HBsAg-positivity, ranging from 6.8% in central Sudan to 26% in southern Sudan [3-5]. HBV infection occurs in early childhood in southern Sudan, with the infection increasing with age in northern Sudan [3,6]. HBV was shown to cause 22% of fulminant hepatitis cases in Sudan [7] and 18.5% of Sudanese blood donors were exposed to the virus and 4% were infected at the time of donation [8]. Nine genotypes of HBV, A–I, with a distinct geographic distribution have been recognized [9-11]. A tenth genotype, J, has been proposed but was found only in one person [12]. Genotype A, D and E circulate in Africa [2]. Genotype A prevails in southern, eastern and central Africa. Genotype D is the dominant genotype in northern Africa, whereas in western Africa genotype E predominates. Subgenotypes have also been identified within genotypes A and D [9,10].

Considering Sudan's unique position and the flux of people across its borders, it is important that the HBV genotypes prevailing in this country are determined. In a single study, in Sudanese asymptomatic blood donors, it

* Correspondence: Anna.Kramvis@wits.ac.za
[1]Hepatitis Virus Diversity Research Programme, Department of Internal Medicine, Faculty of Health Sciences, University of Witwatersrand, 7 York Road, Parktown, Johannesburg 2193, South Africa
Full list of author information is available at the end of the article

BioMed Central

© 2013 Yousif et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

was found that 57.5% were infected with genotype E, 40.5% with genotype D and 2% with subgenotype A2 [8]. Moreover, a diversity of genotypes are distributed in neighbouring countries [2]: genotype D in Egypt to the north [13], genotype E in the Democratic Republic of Congo to the west [14] and genotype A in Kenya [15,16] and Uganda [17] to the south. Knowledge of the genotypes prevailing in Sudanese, with and without liver disease, is important in treatment management, as well as disease prognosis because genotypes play a role in both of these aspects [18,19].

The Regional committee for the World Health Organization (WHO) Eastern Mediterranean Region (EMR), to which Sudan belongs, urged member states to:

> *"Improve the epidemiological surveillance systems, develop a hepatitis registry and implement serosurveys in order to produced reliable data to guide prevention and control measures and monitor impact of preventive strategies"* [20].

Thus our objective was to molecularly characterize HBV from HBsAg-positive persons with known clinical status.

Methods

Serum samples

A cross-sectional, laboratory based study was conducted. Ninety-nine sera were collected from HBsAg-positive patients referred to the hepatology and general medical clinics at IbnSina Hospital, Soba University Hospital and Khartoum Teaching Hospital in Khartoum State between August 2008 and March 2009. The clinical report forms (CRFs) were completed by qualified practitioners. Informed consent was obtained from each patient included in the study and the study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki as reflected in *a priori* approval by the Human Ethics Committees of the University of the Witwatersrand and the University of Khartoum and the Ministry of Health of Sudan. Alanine amino transferase (ALT) levels were determined (reference range 5–40 IU/L) [21] and samples stored at -20°C. Chronic carriers were infected for longer than six months and were classified as asymptomatic carriers (ASCs) if they had normal ALT or chronic hepatitis (CH) patients if they had abnormal ALT. Acute hepatitis (AH) cases were diagnosed based on clinical presentation (symptoms and clinical presence of jaundice) plus high ALT and the presence of hepatitis B core IgM antibody (HBcAbIgM). Clinical and ultrasonographic evidence were used to diagnose cirrhotic (CR) and hepatocellular carcinoma (HCC) cases.

HBV serology

HBsAg was assayed using Monolisa™ HBsAg ULTRA and HBeAg using Monolisa™ HBeAg-Ab PLUS kit (Bio-rad, Hercules, CA). Anti-HBe and anti-HBc antibodies were determined using HBeAg/Anti-HBe ELISA and Anti-HBc Total ELISA, respectively (DIAsource ImmunoAssays S.A, Nivelles, Belgium). Anti-HBc IgM was determined using the ARCHITECT® kit (Abbott Diagnostics, Wiesbaden, Germany).

DNA extraction

DNA was extracted from 200 μL of serum using QIAamp DNA mini blood kit (QIAGEN GmbH, Germany), according to manufacturer's instructions, and eluted in 100 μL of buffer.

Real-time PCR quantification of HBV DNA

PCR primers, HBV-Taq1 and HBV-Taq2 covering a region of the S gene (321 to 401 from the *Eco*RI site) with a FAM/TAMRA labelled TaqMan BS-1 probe [22] were used to quantify HBV DNA in an ABI 7500 Real Time PCR System (Applied Biosystems, Foster City, Ca, USA). A serial dilution of cloned plasmid DNA containing a single genome of HBV DNA, with concentrations ranging from 2×10^1 to 2×10^{11} IU/ ml, was used as template to generate the standard curve. The second WHO International Standard for HBV Nucleic Acid Amplification Techniques (product code 97/750 National Institute for Biological Standards and Controls (NIBSC); Hertfordshire, UK), which has a final concentration of 10^6 IU/ml was used as the internal standard. The standard curve, blank, positive and negative controls, and samples were all tested in duplicate. The measured IU/ml for each reaction was calculated using the Ct (cycle threshold) value of each PCR interpolated against the linear regression of the standard curve. The lower detection limit of our assay is ~20 IU/ ml. The conversion formula of IU =copies/4.7 was used [23-25].

Polymerase chain reaction (PCR) and restriction fragment polymorphism assay (RFLP)

The basic core promoter/precore (BCP/PC) region and complete S open reading frame (ORF) were amplified in a MyCycler™ thermocycler (Bio-Rad, Hercules, Ca, USA) using Promega Taq DNA polymerase (Promega, Madison, WI). The BCP/PC PCR was amplified using a slight modification of the method described by Takahashi and colleagues [26] primers 1606 (+) (1606–1625 from *Eco*RI site) and 1974 (–) (1974–1955 from *Eco*RI site) were used for the first round (denaturation 94°C for 1 min, annealing 55°C for 1 min, extension 72°C for 2 min, 40 cycles) and 1653(+) (1653–1672 from *Eco*RI site) and 1959(–) (1959–1940 from *Eco*RI site) for the second round, with the identical cycling conditions as the first-round PCR [25,27]. A

nested PCR was carried out to amplify the complete S ORF : primers 2410(+) (2410–2439 from *Eco*RI site) and 1314(–) (1314–1291 from *Eco*RI site) were used for the first round (denaturation 94°C for 1 min, annealing 66°C for 1 min, extension 72°C for 3 min, 40 cycles) and 2451 (+) (2451–2482 from *Eco*RI site) and 1280 (–) (1280–1254 from *Eco*RI site) for the second round (denaturation 94°C for 1 min, annealing 65°C for 1 min, extension 72°C for 3 min, 40 cycles) [27]. Another nested PCR reaction was carried out to amplify a short region of the S (459–710 from *Eco*RI site) to determine the genotype: primers 255 (+) (253–275 from *Eco*RI site) and 759 (–) (759–739 from *Eco*RI site) were used for the first round (denaturation 94°C for 30 sec, annealing 56°C for 40 sec, extension 72°C for 1 min, 40 cycles) and [459(+) (460–480 from *Eco*RI site)] and [710(–) (711–691 from *Eco*RI site)] for the second round (denaturation 94°C for 30 sec, annealing 56°C for 40 sec, extension 72°C for 50 sec, 40 cycles) [28]. When the complete S ORF and the short S could not be amplified, a RFLP assay was used to determine the genotype of the HBV isolates. Primers P7 (256–278 from *Eco*RI site) and P8 (796–776 from *Eco*RI site) were used to amplify nucleotides 256–796 of the S region. The amplicon was then cleaved using restriction enzymes *Hin*fI and *Tsp*509I, in separate reactions, to give the characteristic RFLP patterns for the different genotypes [29]. The complete genome was amplified using a single amplification method with primers P1 (1821–1841 from *Eco*RI site) and P2 (1825–1806 from *Eco*RI site), with modifications in the cycling conditions [30]. Initial denaturation at 98°C for 1 min, 80°C for 30 sec followed by adding the polymerase mix, 98°C for 1 min, then 35 cycles of: 98°C for 10 sec, 57°C for 30 sec and 72°C for 1 min.

Sequencing

The BigDye Terminator v3.0 Cycle Sequencing Ready Reaction Kit (Applied Biosystems., Foster City, USA) was used and sequencing performed using the ABI 3130XL Genetic analyzer (Applied Biosystems, Foster City, CA). In addition to sequencing primers: 2497F (2497–2516), 3188F (3188–3206) and 591F (591–611) used previously [27,31], four new primers: 1069F (1069–1088) (5′- TGT ATT CAA TCT AAG CAG GC-3′), *Post-Bgl* III (2000–2017) (5′-CCG ATA CAG AGC TGA GGC-3′), 2022F (2022–2041) (5′-CCT TAG AGT CTC CTG AGC AT-3 ′), 2480R (2461–2480) (5′-CAC CTT ATG AGT CCA AGG AA-3′), were used to sequence the complete genome. All positions are numbered from the *Eco*RI site of genotype A (AY233286). The three overlapping fragments of the complete S, and the 7 fragments of the complete genome thus obtained, were assembled using the Fragment Merger Tool [32]. Sequences were deposited in GenBank, accession numbers KF170739-KF170812.

Phylogenetic analysis

Neighbour-joining using MEGA5 [33] or PHYLIP (Phylogeny inference package version 3.69) or DNADIST consecutively with NEIGHBOR were used to generate dendrograms [34].

Analysis for intergenotypic recombination

SimPlot 3.5.1©, an interactive 32-bit software program, that plots distances (or similarity) versus position [35] was used to show recombination in isolate SDAC031 [33].

Results

HBV serology and viral loads

Of the 99 patients, 77 were males and 22 were females. Mean age ± standard deviation (SD) was [45.7±14.8] years. All sera were HBsAg- and anti-HBc-positive; 12 were HBeAg-positive/anti-HBe-negative, 75 were HBeAg-negative/anti-HBe-positive, and 12 had neither HBeAg nor anti-HBe. All patients were HBV DNA-positive, with a median (interquartile range) (IQR) viral load of [2.8 (2.2–4.2)] log IU/ml and ALT median (IQR) level was [30 (19–49)] IU/L (reference range 5–40 IU/L) [21].

Clinical and demographic characteristics

The 99 cases were classified into five clinical groups: HCC (n=15), CR (n=42), ASCs (n=30), AH (n=7) and CH (n=5). The HCC patients were significantly older than patients in the other groups ($p < 0.05$) and the AH patients were significantly younger ($p < 0.05$). The ALT levels were the highest in the AH ($p < 0.05$) (Table 1). The median viral load of HBeAg-positive patients [6.99 (5.63–7.57) log IU/ml] was significantly higher than in HBeAg-negative patients [2.55 (2.09–3.42) log IU/ml] ($p<0.05$). The viral loads and frequency of HBeAg-positivity did not differ between clinical groups.

HBV genotyping and phylogenetic analysis

In order to maximize the number of isolates genotyped, three methods were used sequencially (Figure 1).

81 HBV isolates were thus successfully genotyped

Firstly, 53 of 99 were genotyped using amplification and sequencing of the complete S (2848–835 from *Eco*RI), followed by phylogenetic analysis (Figure 2), secondly,17 of 46 were genotyped by amplification and sequencing of a short S fragment (520–704 from *Eco*RI), followed by phylogenetic analysis (trees not shown) and thirdly,11 of 29 by RFLP analysis.

The relatively longer amplicon of the S region of 2.1 kb in length, compared to the shorter regions used for the short S and RFLP amplifications, meant that not all samples could be amplified in the longer region successfully. 18.2% (18/99) could not be genotyped using any of the three methods. No sequences could be obtained for 10 of these isolates and for the remaining 8 only the

Table 1 Demographic and clinical characteristics of 99 patients

	Clinical status				
	HCC (15)	ASC (30)	CR (42)	AH (7)	CH (5)
Male: Female	13:02	22:08	37:05:00	02:05	03:02
Age mean ±SD	57.3±13.8[a]	42.8±15.1	46.5±12.9	32.1±12.4[a]	37.6±41.0
ALT median IU/L (IQR)	40 (24–50)	22 (18–33)	27 (18–45)	461 (312–834)[a]	38 (29–47)
HBeAg positive (%)	2 (13.3)	2 (6.7)	7 (16.7)	1 (14.3)	0
Log viral loads	3.8	2.5	2.8	3	2.1
IU/ml median (IQR)	(2.4-4.7)	(2.2–3.3)	(2.2-5.3)	(2.2-3.8)	(2.0-3.6)

HCC Hepatocellular carcinoma, *ASC* Asymptomatic carriers, *CR* Cirrhosis, *AH* Acute hepatitis, *CH* Chronic hepatitis.
[a]Significant difference with other groups.

BCP/PC region was sequenced, which is not sufficient to differentiate between genotypes D and E.

The genotype distribution in the 81 samples was 59% genotype D: 30% genotype E: 8.5% genotype A: 2.5% putative D/E recombinant. If we include the 18 samples that could not be genotyped, 48.5% (48/99) were infected with genotype D, 24.2% (24/99) with genotype E, 7.1% (7/99) with genotype A and 2% (2/99) were infected with a putative D/E recombinant.

Phylogenetic analysis of the complete S region allowed for further classification into subgenotypes (Figure 2). Based on the recently suggested classification system [33], 23 of 31 genotype D isolates belonged to subgenotype D1 (74%), 3 to D2 (10%), 1 to D3 (3%) and 4 to D6 (13%). One genotype A isolate belonged to subgenotype A1 (Figure 2).

The complete genome of four isolates was amplified, two (SDAC024 and SDAC031) belonged to genotype D and two (SDAC047 and SDAC125) to genotype E.

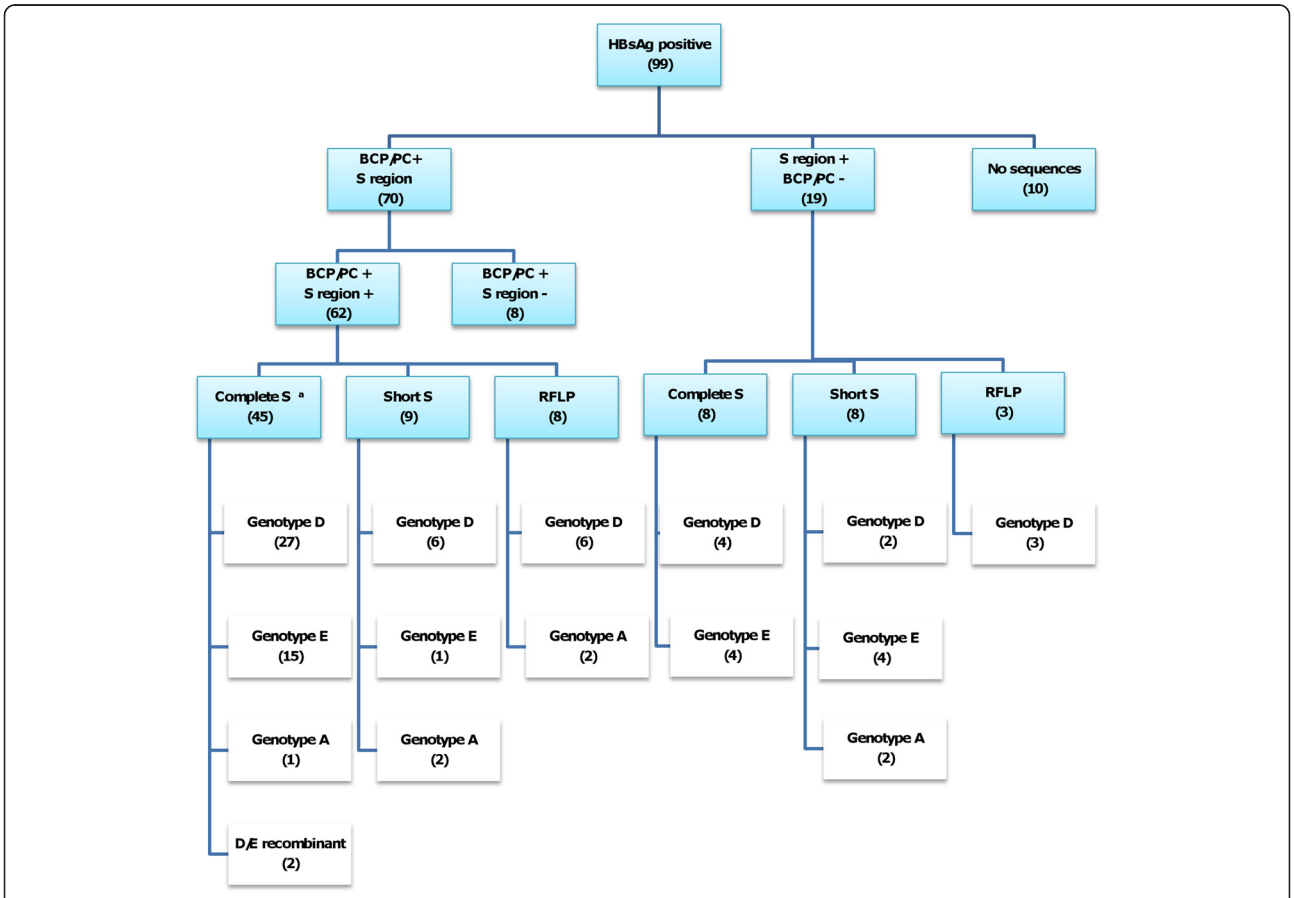


Figure 1 Algorithm summarizing protocol and results of the molecular characterization of BCP/PC region and genotyping of HBV.
[a] Three isolates showed genotype D based on S region and genotype A based on BCP/PC region.

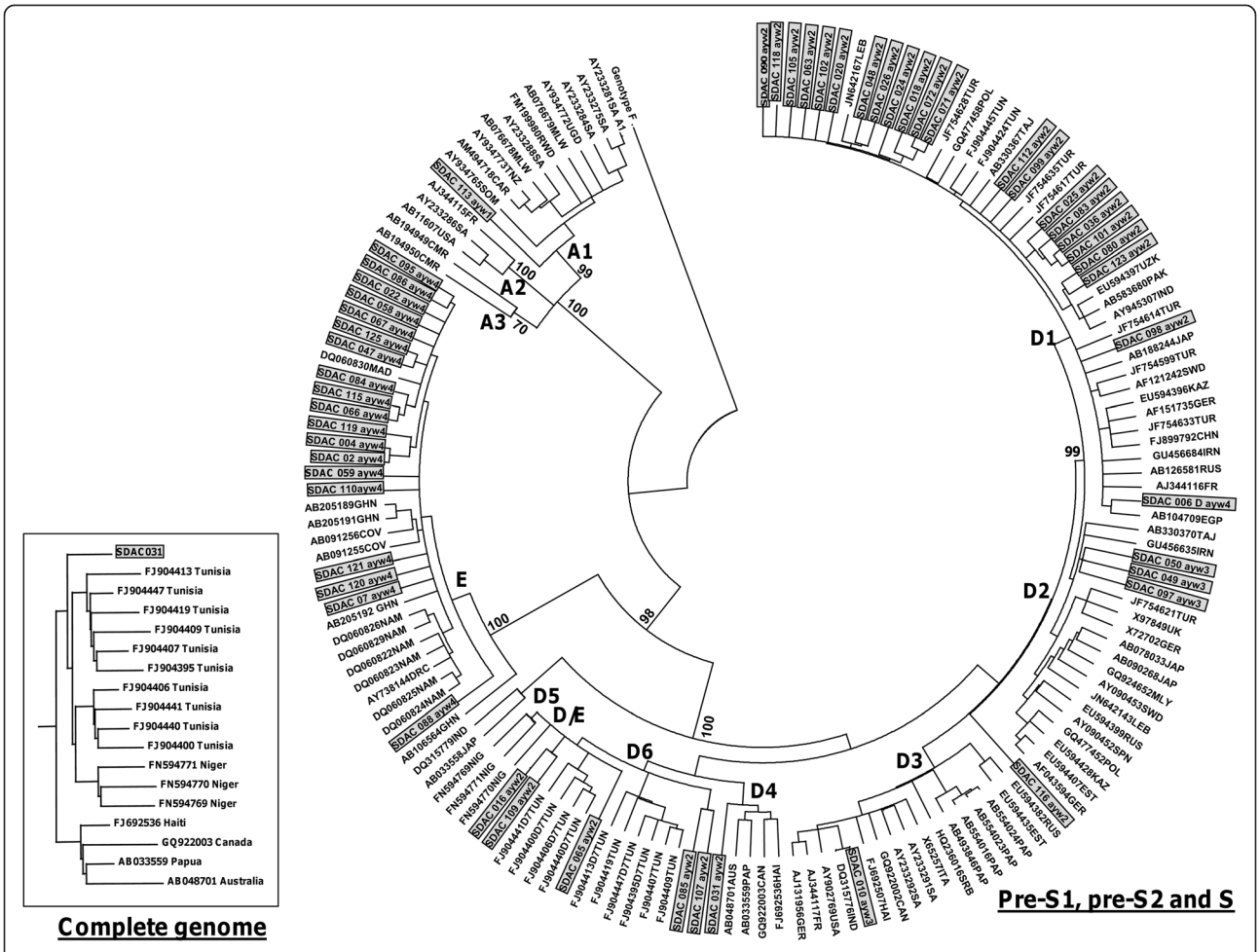


Figure 2 A rooted phylogenetic tree of 53 complete S sequences of HBV obtained from Sudanese liver disease patients (shaded in box) with 104 reference HBV sequences, using neighbour-joining. Bootstrap statistical analysis was performed using 1000 datasets, indicated as percentages on the nodes. The sequences are labeled by their accession numbers and country. [AUS, Australia; CMR, Cameron; CAN, Canada; CAR, Central African Republic; CHN, China; COV, Côte d'Ivoire; DRC, Democratic Republic of Congo; EGP, Egypt; EST, Estonia; FR, France; GHN, Ghana; GER, Germany; HAI, Haiti; IND, India; IRN, Iran; ITA, Italia; JAP, Japan; KAZ, Kazakhstan; LEB, Lebanon; MAD, Madagascar; MLY, Malaysia; MLW, Malawi; NAM, Namibia; NIG, Niger; PAK, Pakistan; PAP, Papua New Guinea/Indonesia; POL, Poland; RUS, Russia; RWD, Rwanda; SRB, Serbia; SOM, Somalia; SA, South Africa; SPN, Spain; SDW, Sweden; TNZ, Tanzania; TAJ, Tajikistan; TUN, Tunisia; TUR, Turkey; UGN, Uganda; UK, United Kingdom; USA, United States; UZK, Uzbekistan]. The letters, A, D and E, represent the genotypes and the numbers the subgenotypes. A rooted phylogenetic tree of the complete genome of SDAC 031 (shaded) relative to 17 reference HBV sequences, using neighbour-joining is shown in the left hand box. When the complete genome was analyzed SDAC031 clustered with D6 whereas when the S region alone was analyzed it clustered with D4.

Although the genotypes determined using the complete genome agreed with that determined using the S ORF, there was a discrepancy in the subgenotype classification of SDAC031. Using phylogenetic analysis of the complete S, it clustered as an outlier of subgenotype D4, whereas following complete genome phylogenetic analysis it clustered with D6 (Figure 2). Simplot analysis showed SDAC031 to be a recombinant of D6 and D4. By mapping the informative sites, we estimated the transition positions between D6 and D4 (Figure 3). The genotype E isolates (SDAC047 and SDAC125) clustered together when the complete S was compared, whereas they separated into different clades in the complete genome analysis (*data not shown*).

Although the majority of subgenotype D isolates showed the respective distinct subgenotype amino acids signatures [33], there were exceptions. SDAC031 (D6) showed preS2I42T, spV59F and rtT237P. SDAC085 (D6) had spV59S. SDAC107 (D6) had preS2I42T and rtT237P. All 17 genotype E isolates had genotype E signature amino acid motifs.

Comparison of patients infected with genotypes D and E

Because the majority were infected with either genotype D or E, further analyses compared the patients' clinical and virological features infected with these genotypes. No significant difference was observed in gender, age, ALT and

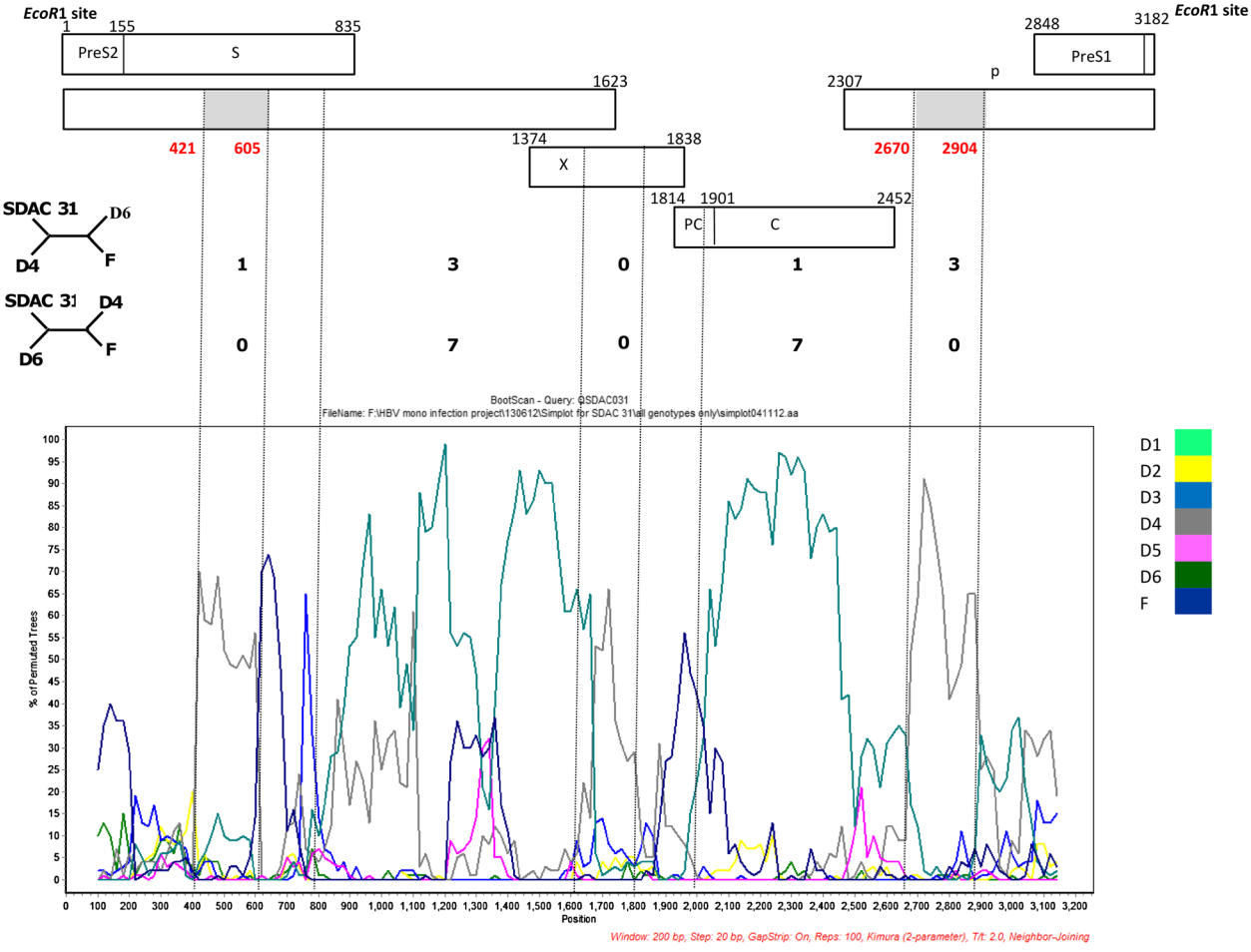


Figure 3 Graphs show bootscanning values of query sequence SDAC031 to representative sequences of subgenotypes D1 to D6 and outgroup, genotype F. Graphs were generated using Simplot version 3.5.1© with window size: 200 bp, step size 20 bp, gap-strip off, 100 bootstrap replicates, Kimura transition/transversion ratio:2 and NEIGHBOR. Dashed vertical lines indicate breakpoints. The numbers of informative sites, which are shared by the isolate with subgenotype D4 are shown above and with subgenotype D6, below, on the right hand side of the four-member trees representing these sites. The genomic regions belonging to subgenotype D4 are shaded in *grey*.

clinical groups of patients infected with either genotype. Patients infected with genotype E, showed a significantly higher frequency of HBeAg-positivity compared to patients infected with genotype D ($p < 0.05$) and HBV DNA levels in patients infected with genotype E were significantly higher ($p < 0.05$) (Table 2).

Molecular characterization of HBV isolates

Analysis of basic core promoter/precore (BCP/PC) region

Of the 70 isolates amplified and sequenced in the BCP/PC region, 62 amplfied in the S region and thus had genotype assignments. 57 isolates belonged to either genotype D or E, and 5 to genotype A. Seven of 8 genotype D or E isolates

Table 2 Clinical and demographic characteristics of 81 patients, for which HBV was genotyped

	Clinical status					Demographic characteristics		Biochemical and virological characteristics		
	HCC (13)	ASC (24)	CR (33)	AH (7)	CH (4)	Male: Female	Age mean ±SD	ALT median IU/L (IQR)	HBeAg positive (%)	Viral loads log median IU/ml (IQR)
Genotype D (%)	4(30.7)	14(58.3)	23(69.7)	4(57.1)	3(75)	38:10	44.6 ±15.1	29 (19–57)	4 (8.3)	2.8 (2.0-3.9)
Genotype E (%)	6(46.2)	6(25)	8(24.3)	3(42.9)	1(25)	17:07	49.3 ±14.3	29 (18–65)	7 (29.2)[a]	4 (2.6-5.7)[a]
Genotype D/E (%)	1(7.7)	0	1(3)	0	0	2:0	56.5 ±7.0	NA	0	2.4 (2.3-2.41)
Genotype A (%)	2(15.4)	4(16.7)	1(3)	0	0	5:2	38.14 ±14.51	33 (28–54)	1 (14.3)	2.6 (2.3-3.9)

HCC, Hepatocellular carcinoma; *ASC*, Asymptomatic carriers; *CR*, Cirrhosis; *AH*, Acute hepatitis, *CH*, Chronic hepatitis.
[a]Significant difference between patients infected with genotype D or E.

from HBeAg-positive patients had wild-type BCP/PC, and one (SDAC118) had G1764T. The mutations found in genotype D or E isolates from HBeAg-negative individuals are shown in Figure 4: 15/49 genotype D or E isolates from HBeAg-negative sera, had wild-type BCP/PC region. In 22/49 HBeAg-negative sera, the absence of HBeAg was as a result of missense mutations affecting the translation of the HBeAg precursor: 18 had the classical stop codon mutation G1896A ($p < 0.05$, when comparing isolates from HBeAg-negative and –positive sera) and 4 had initiation codon mutations, 2 A1814T and 2 T1815C. A1762T/G1764A was found in 11 G1896A mutants and in one T1815C mutant. Six isolates from HBeAg-negative sera had A1762T/G1764A alone. The Kozak sequence was GCAC at 1809–1812 in 45 isolates, GTAC in isolate SDAC 121 and TCAT with 1858C and 1888A in isolates SDAC018, SDAC090 and SDAC091. SDAC050 had 1858C with 1888A alone.

Although generally the BCP/PC region cannot be used to determine specific genotypes, genotype A and its subgenotypes can be differentiated from genotypes D or E because of specific sequence characteristics. One genotype A isolate from a HBeAg-positive patient, (SDAC108) had A1762T/G1764A. From the BCP/PC sequences it was deduced that this isolate belonged to subgenotype A2, because it had GCAC at 1809–1812, 1858C and 1888G. The other four genotype A isolates were from HBeAg-negative individuals. These isolates, SDAC062, SDAC073 SDAC100 and SDAC113, belonged to subgenotype A1 because they had TCAT at 1809–1812, 1858C and 1888A. One isolate had wild-type BCP/PC region, two had A1762T/G1764A, one with G1862T and the fourth isolate had G1862T alone. No significant association was found between the BCP/PC mutations and genotypes or clinical groups.

Analysis of the pre-S1, pre-S2, Surface and Polymerase regions

The complete S region was sequenced for 53 isolates and pre-S mutations were detected in 14 (26%). Four mutational patterns were identified: pre-S1 deletions (2/14; 14%), pre-S2 start codon mutation (4/14; 29%), pre-S2 start codon mutation with pre-S2 deletion (1/14; 7%) and pre-S2 deletion alone (7/14; 50%) (Table 3). Transcription regulatory box domain substitutions [CCAAT→CTAAT] (803–807 from *Eco*RI site), were found in two isolates (SDAC095 genotype E and SDAC098 genotype D). Three *cis*-acting elements mutations were found: C1092T in the Enhancer I in one isolate, C1155T in the Enhancer I in 6 isolates and T3109G S2 promoter in 2 isolates. PreS2F22L was found in 6 genotype E and three genotype D isolates. The vaccine escape mutation, sM133T, was observed in genotype E isolates, SDAC022, SDAC067 and SDAC088. In the reverse transcriptase domain of the polymerase, the following mutations were detected: rtA194T in SDAC049 (genotype D), rtV207M in SDAC020 (genotype D), rtS213T in SDAC063 and SDAC098 (genotype D), SDAC002, SDAC121 and SDAC004 (genotype E), rtV214A in SDAC059 (genotype E) and rtS215Q in SDAC049 and SDAC090 (genotype D).

Discussion

In Sudan, where HBV is hyperendemic, little is known about the genotypes and mutants of HBV in liver disease patients. Thus HBV, from 99 patients, belonging to five clinical groups, were studied. Of the 81 samples that were successfully genotyped, close to 60% belonged to genotype D whereas the opposite was true in Sudanese blood donors, where genotype E prevailed (57%) [8]. HCC patients in this study, who were infected with either genotype D or E, were significantly older than ASC, CH, AH, CR patients, with the AH patients being significantly younger (Table 1).

Pattern		1762 1764	1809	1812	1814 1815	1858	1888	1896	Frequency	%
1	AGGCACATTG	A-G--//--GCAC-	AT--//--T--//--G--//--G						15	30.9
2	TAGCACATTA	T-A--//--GCAC-	AT--//--T--//--G--//--A						10	20.4
3	TAGCACATTG	T-A--//--GCAC-	AT--//--T--//--G--//--G						6	12.2
4	ATGCACATTA	A-T--//--GCAC-	AT--//--T--//--G--//--A						5	10.2
5	AGTCATATCG	A-G--//--TCAT-	AT--//--C--//--A--//--G						3	6.1
6	AGGCACTTTG	A-G--//--GCAC-	TT--//--T--//--G--//--G						2	4.1
7	AGGCACATTA	A-G--//--GCAC-	AT--//--T--//--G--//--A						2	4.1
8	TAGTACATTA	T-A--//--GTAC-	AT--//--T--//--G--//--A						1	2
9	ATGCACACTG	A-T--//--GCAC-	AC--//--T--//--G--//--G						1	2
10	TAGCACACTG	T-A--//--GCAC-	AC--//--T--//--G--//--G						1	2
11	ATGCACATTG	A-T--//--GCAC-	AT--//--T--//--G--//--G						1	2
12	TGGCACATTG	T-G--//--GCAC-	AT--//--T--//--G--//--G						1	2
13	AGGCACATCG	A-G--//--GCAC-	AT--//--C--//--A--//--G						1	2

Figure 4 Mutation distribution in the basic core promoter/precore region, at 11 loci of interest (1762, 1764, 1809–1812, 1814, 1815, 1858, 1888 and 1896), of 49 genotype D or E HBV isolates from HBeAg-negative patients. Boxed motif represents the wild- type of genotypes D or E. Data were obtained using the Mutation Reporter Tool [36]. Mutations are shaded in *grey*. The frequency of the mutational patterns did not differ between genotype D and E.

Table 3 Pre-S region mutations and deletions

	Clinical group	Genotype	Pre-S2 S tart codon	Deletions					
				Region	Nucleotides		Amino acids		Region and functions affected
ID					Size	Position	Size	position	
SDAC024	CR	D	ATG	Pre-S2	21	35-55[a]	7	16-22[a]	
SDAC048	CR	D	ATG	Pre-S1	72	2998-3069	24	51-74	S promoter, Transactivator domain
SDAC083	CR	D	ATG	Pre-S2	21	35-55	7	16-22	
SDAC097	CR	D	*GTG*[b]	Pre-S 2	9	45- 53	3	19-21	
SDAC098	CH	D	ATG	Pre-S1	18	3129-3146	6	94-99	Hsc70 binding site, CCAAT binding factor (CBF) binding site, Cytosolic anchorage determinant (CAD)
SDAC112	ASC	D	ATG	Pre-S2	6	49-54	2	20-21	
SDAC118	CR	D	ATG	Pre-S2	54	4-57	18	5 to 22	pHSA binding site, Viral secretion
SDAC107	CR	D	*ATA*	-	-	-	-	-	-
SDAC090	CR	D	*ATA*	-	-	-	-	-	-
SDAC058	CR	E	ATG	Pre-S2	9	58-66[c]	3	20-22[c]	Transactivator domain
SDAC059	HCC	E	*GTG*	Pre-S2	12	52-63	4	18-21	Transactivator domain
SDAC088	HCC	E	ATG	Pre-S2	12	52-63	4	18-21	Transactivator domain
SDAC095	HCC	E	ATG	Pre-S2	15	52-66	5	18-22	Transactivator domain
SDAC067	ASC	E	*GTG*	-	-	-	-	-	-
SDAC113	ASC	A	*ACA*	-	-	-	-	-	-

[a] nucleotide or amino acid position relative to genotype D (GU456684).
[b] mutant pre-S2 start codon *italicized*.
[c] nucleotide or amino acid position relative to genotype E (AB205191).

This differs from HCC patients, infected with genotype A, who develop cancer at a significantly younger age [28]. In agreement with others [37,38], the HCC group had the highest median viral load. High viral loads have been implicated as a risk factor for HCC development [39].

In agreement with the study in Sudanese blood donors [8], the predominant subgenotype of D was D1. Furthermore, following phylogenetic analysis of the complete S region, two isolates, which lacked the 33 nucleotide deletion, characteristic of genotype D, were genotype D/E recombinants [8,33,40]. Following phylogenetic analysis of the complete S, one isolate belonged to subgenotype A1, and from the BCP/PC sequence, a further four were deduced to belong to subgenotype A1 [41]. This is the first time that subgenotype A1 has been identified in Sudan whereas a previously sequenced genotype A isolate from Sudan belonged to A2 [8]. Subgenotype A1 circulates in southern and eastern Africa and Southern Asia and A2 is found in Northern and Central Europe and North America [42].

Discrepant results were obtained in the subgenotype assignment of SDAC031 when the complete S and complete genome were analyzed. The complete S clustered as an outlier of D4 whereas the complete genome clustered with D6. D4 and D6 are phylogenetically closely related (Figure 2). Further analysis using Simplot, showed that it was a D6/D4 recombinant (Figure 3) and the majority of signature amino acids were of D6, with the exception of a number of amino acids, which were of D4 [33]. Geographically, D6 is distributed in theMaghreb and Madagascar, whereas D4 in the Americas and Australia. It is possible that D4 originated in Africa but has subsequently been replaced by other subgenotypes of D and the recombinant is a remnant of the original strain/s.

The HBV genotypes show a distinct geographical distribution in Africa, with genotype D predominating in the North, genotype E in the West and genotype A in the South-East [9]. This is the first study to describe the co-circulation of genotypes D and E in liver disease patients and to allow the comparison of patients infected with these two genotypes. In agreement with a study carried out in Sudanese blood donors [8], viral loads were significantly higher in genotype E-infected patients compared to genotype D-infected, with patients infected with genotype E, showed a significantly higher frequency of HBeAg-positivity. The small size of HBeAg allows it to traverse the placenta and elicit HBe/HBcAg-specific T helper cell tolerance *in utero* [43]. Thus babies born to HBeAg-positive mothers have high chronicity rates than those born to HBeAg-negative mothers [43]. The high frequency of HBeAg-positivity in mothers infected with genotype E, would lead to its vertical transmission and

explain the high prevalence and geographical restriction of this genotype in Africa and to African emigrants to other regions [14]. The higher HBeAg-positivity seen in individuals infected with genotype E, could confer tolerance and less serious clinical manifestations than genotype D, where HBeAg-positivity was lower. This could explain why genotype E prevails in the Sudanese blood donors [8], whereas genotype D prevailed in the liver disease patients in the present study.

The majority of the HBeAg-negativity was as a result of the classical G1896A, which abolishes HBeAg expression [44] and occurs in genotype D or E but not A because the encapsidation signal secondary structure precludes this mutation in genotype A [45,46]. Other mutations including transcriptional A1762T/G1764A and translation initiation mutations were responsible of HBeAg-negativity in a number of patients. Three isolates had TCAT instead of GCAC in the Kozak sequence preceding the precore initation codon and can affect HBeAg expression at the translational level [47]. This Kozak mutation occurred together with 1858C and 1888A (pattern 5, Figure 4), which are characteristics of subgenotype A1 [41]. One isolate has 1858C and 1888A alone. However, following phylogenetic analysis of the complete S region, these four isolates were found to belong to genotype D. It is possible that these patients were co-infected with genotypes D and A or with D/A recombinants. These possibilities can only be discriminated by complete genome cloning and sequencing.

Four different pre-S mutational patterns were identified (Table 3). In genotype E, pre-S2 deletions were found mainly in HCC patients, whereas in genotype D, the deletion mutants were from non-HCC patients. However, the numbers were small to reach any firm conclusions. Pre-S deletion mutants were found in genotype E isolates from ASCs from Guinea [48] and Sudan [8]. Pre-S deletion/mutations affect the progression to serious liver disease in patients infected with either genotype B or C [49].

Six genotype E and three genotype D isolates had preS2F22L, which is a risk factor for HCC [50,51]. Interestingly, three genotype E isolates had the sM133T mutation, which could possibly compromise antibody neutralization and may represent potential vaccine escape mutants [52]. However, because these individuals were not vaccinated for HBV, this mutation may have emerged as a result of host immune pressure. The reverse transcriptase mutations rtA194T, rtV207M, rtS213T, rtV214A and rtS215Q detected in the present study are neither primary resistance mutations, nor have they ever been seen in overt resistance during therapy [53]. When mutations rtA194T, rtV207M rtS213T rtV214A and rtQ215S were tested in our sensitive and reliable *in vitro* resistance test system, the mutants showed no resistance to lamivudine (LMV), entecavir (ETV), adenofovir (ADF) and tenofovir (TDF). (Glebe *et al.*, unpublished observations).

Conclusion

This is the first study to molecularly characterize HBV from Sudanese liver disease patients, who were predominantly infected with genotypes D and E, allowing comparison of the effect of these genotypes on clinical manifestation in the same ethnic group.

Abbreviations

AH: Acute hepatitis; ALT: Alanine amino transferase; ASC: Asymptomatic carriers; BCP/PC: Basic core promoter/precore; CH: Chronic hepatitis; CRF: clinical report form; DFG: Deutsche Forschungsgemeinschaft (German Research Foundation); DNA: Deoxyribonuclease; HBcAb: Hepatitis B core antibody; HBeAb: Hepatitis B e antibody; HBeAg: Hepatitis B e antigen; HBsAb: hepatitis B surface antibody; HBsAg: Hepatitis B surface antigen; HBV: hepatitis B virus; HCC: Hepatocellular carcinoma; IQR: Interquartile range; CR: cirrhosis; ORF: open reading frame; PCR: Polymerase chain reaction; RFLP: Restriction fragment polymorphism; rt: Reverse transcriptase; s: Surface; sp: Spacer; WHO: World Health Organization.

Competing interests

The authors declare that there are no competing interests.

Authors' contributions

Conceived the study: AK, DG, HM; Designed the experiments: MY, DG, AK; Performed the experiments:MY; Analyzed the data: MY, AK; Contributed reagents/materials: AK, SB, HM; Wrote the paper: MY, AK Read, contributed to and approved the paper: MY, HM, SB, DG, AK. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) of the special African-German collaborative programme, "Africa Initiative", grant GL595/3-1 to AK, HM and DG) and the National Research Foundation of South Africa (NRF; GUN#65530 to AK). We thank Wolfram Gerlich for critical reading, Waleed Mohammed for technical assistance and clinicians at IbnSina, Soba University and Khartoum Hospitals.

Author details

[1]Hepatitis Virus Diversity Research Programme, Department of Internal Medicine, Faculty of Health Sciences, University of Witwatersrand, 7 York Road, Parktown, Johannesburg 2193, South Africa. [2]Department of Medicine, Faculty of Medicine, University of Khartoum, Khartoum, Sudan. [3]Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan. [4]Institute of Medical Virology, National Reference Centre of Hepatitis B and D, Justus Liebig-University of Giessen, Giessen, Germany.

Received: 5 February 2013 Accepted: 16 July 2013
Published: 18 July 2013

References

1. WHO: *Hepatitis B virus Fact sheet N° 204.* 2012. http://www.who.int/mediacentre/factsheets/fs204/en/. Accessed January 2013.
2. Kramvis A, Kew MC: **Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes.** *Hepatol Res* 2007, **37**(s1):S9–S19.
3. Mudawi HMY: **Epidemiology of viral hepatitis in Sudan.** *Clin Exp Gastroenterol* 2008, **1**:9–13.
4. Mudawi HM, Smith HM, Rahoud SA, Fletcher IA, Saeed OK, Fedail SS: **Prevalence of hepatitis B virus infection in the Gezira state of central Sudan.** *Saudi J Gastroenterol* 2007, **13**(2):81–83.
5. McCarthy MC, El-Tigani A, Khalid IO, Hyams KC: **Hepatitis B and C in Juba, southern Sudan: results of a serosurvey.** *Trans R Soc Trop Med Hyg* 1994, **88**(5):534–536.

6. McCarthy MC, Hyams KC, el-Tigani el-Hag A, el-Dabi MA, el-Sadig el-Tayeb M, Khalid IO, George JF, Constantine NT, Woody JN: **HIV-1 and hepatitis B transmission in Sudan.** *AIDS* 1989, **3**(11):725–729.
7. Mudawi HMY, Yousif BA: **Fulminant hepatic failure in an African setting: Etiology, clinical course, and predictors of mortality.** *Dig Dis Sci* 2007, **52:**3266–3269.
8. Mahgoub S, Candotti D, El-Ekiaby M, Allain JP: **Hepatitis B virus (HBV) infection and recombination between HBV genotypes D and E in asymptomatic blood donors from Khartoum Sudan.** *J Clin Microbiol* 2011, **49**(1):298–306.
9. Kramvis A, Kew M, Francois G: **Hepatitis B virus genotypes.** *Vaccine* 2005, **23**(19):2409–2423.
10. Norder H, Courouce AM, Coursaget P, Echevarria JM, Lee SD, Mushahwar IK, Robertson BH, Locarnini S, Magnius LO: **Genetic diversity of hepatitis B virus strains derived worldwide: genotypes, subgenotypes, and HBsAg subtypes.** *Intervirology* 2004, **47**(6):289–309.
11. Yu H, Yuan Q, Ge SX, Wang HY, Zhang YL, Chen QR, Zhang J, Chen PJ, Xia NS: **Molecular and phylogenetic analyses suggest an additional hepatitis B virus genotype "I".** *PLoS One* 2010, **5**(2):e9297.
12. Tatematsu K, Tanaka Y, Kurbanov F, Sugauchi F, Mano S, Maeshiro T, Nakayoshi T, Wakuta M, Miyakawa Y, Mizokami M: **A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J.** *J Virol* 2009, **83**(20):10538–10547.
13. Saudy N, Sugauchi F, Tanaka Y, Suzuki S, Aal AA, Zaid MA, Agha S, Mizokami M: **Genotypes and phylogenetic characterization of hepatitis B and delta viruses in Egypt.** *J Med Virol* 2003, **70**(4):529–536.
14. Mulders MN, Venard V, Njayou M, Edorh AP, Bola Oyefolu AO, Kehinde MO, Muyembe Tamfum JJ, Nebie YK, Maiga I, Ammerlaan W, *et al*: **Low genetic diversity despite hyperendemicity of hepatitis B virus genotype E throughout West Africa.** *J Infect Dis* 2004, **190**(2):400–408.
15. Mwangi J, Nganga Z, Songok E, Kinyua J, Lagat N, Muriuki J, Lihana R, Khamadi S, Osman S, Lwembe R, *et al*: **Molecular genetic diversity of hepatitis B virus in Kenya.** *Intervirology* 2008, **51**(6):417–421.
16. Usuda S, Okamoto H, Tanaka T, Kidd-Ljunggren K, Holland PV, Miyakawa Y, Mayumi M: **Differentiation of hepatitis B virus genotypes D and E by ELISA using monoclonal antibodies to epitopes on the preS2-region product.** *J Virol Methods* 2000, **87**(1–2):81–89.
17. Hannoun C, Soderstrom A, Norkrans G, Lindh M: **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* 2005, **86**(Pt 8):2163–2167.
18. Kramvis A, Kew MC: **Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy.** *J Viral Hepat* 2005, **12**(5):456–464.
19. Kao JH, Chen PJ, Lai MY, Chen DS: **Genotypes and clinical phenotypes of hepatitis B virus in patients with chronic hepatitis B virus infection.** *J Clin Microbiol* 2002, **40**(4):1207–1209.
20. Mohsni E: **HBV and HCV Contorl in the WHO Eastern Mediterranean Region (WHO EMR).** *Viral Hepatitis* 2010, **18**(2):26–27.
21. Papatheodoridis GV, Manolakopoulos S, Liaw YF, Lok A: **Follow-up and indications for liver biopsy in HBeAg-negative chronic hepatitis B virus infection with persistently normal ALT: a systematic review.** *J Hepatol* 2012, **57**(1):196–202.
22. Weinberger KM, Wiedenmann E, Bohm S, Jilg W: **Sensitive and accurate quantitation of hepatitis B virus DNA using a kinetic fluorescence detection system (TaqMan PCR).** *J Virol Methods* 2000, **85**(1–2):75–82.
23. Firnhaber C, Viana R, Reyneke A, Schultze D, Malope B, Maskew M, Di-Bisceglie A, MacPhail P, Sanne I, Kew M: **Occult hepatitis B virus infection in patients with isolated core antibody and HIV co-infection in an urban clinic in Johannesburg South Africa.** *Int J Infect Dis* 2009, **13**(4):488–492.
24. Firnhaber C, Chen CY, Evans D, Maskew M, Schulz D, Reyneke A, Kramvis A: **Prevalence of hepatitis B virus (HBV) co-infection in HBV serologically-negative South African HIV patients and retrospective evaluation of the clinical course of mono- and co-infection.** *Int J Infect Dis* 2012, **16**(4):e268–e272.
25. Bell TG, Makondo E, Martinson NA, Kramvis A: **Hepatitis B virus infection in human immunodeficiency virus infected southern African adults: occult or overt–that is the question.** *PLoS One* 2012, **7**(10):e45750.
26. Takahashi K, Aoyama K, Ohno N, Iwata K, Akahane Y, Baba K, Yoshizawa H, Mishiro S: **The precore/core promoter mutant (T1762A1764) of hepatitis B virus: clinical significance and an easy method for detection.** *J Gen Virol* 1995, **76**(Pt 12):3159–3164.
27. Vermeulen M, Dickens C, Lelie N, Walker E, Coleman C, Keyter M, Reddy R, Crookes R, Kramvis A: **Hepatitis B virus transmission by blood transfusion during 4 years of individual-donation nucleic acid testing in South Africa: estimated and observed window period risk.** *Transfusion* 2012, **52:**880–892.
28. Kew MC, Kramvis A, Yu MC, Arakawa K, Hodkinson J: **Increased hepatocarcinogenic potential of hepatitis B virus genotype A in Bantu-speaking sub-saharan Africans.** *J Med Virol* 2005, **75**(4):513–521.
29. Lindh M, Andersson AS, Gusdal A: **Genotypes, nt 1858 variants, and geographic origin of hepatitis B virus–large-scale analysis using a new genotyping method.** *J Infect Dis* 1997, **175**(6):1285–1293.
30. Gunther S, Li BC, Miska S, Kruger DH, Meisel H, Will H: **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* 1995, **69**(9):5437–5444.
31. Makondo E, Bell TG, Kramvis A: **Genotyping and Molecular Characterization of Hepatitis B Virus from Human Immunodeficiency Virus-Infected Individuals in Southern Africa.** *PLoS One* 2012, **7**(9):e46345. doi:46310.41371/journal.pone.0046345.
32. Bell TG, Kramvis A: **Fragment merger: an online tool to merge overlapping long sequence fragments.** *Viruses* 2013, **5**(3):824–833.
33. Yousif M, Kramvis A: **Genotype D of hepatitis B virus and its subgenotypes: An update.** *Hepatol Res* 2013, **43**(4):355–364.
34. Gulube Z, Chirara M, Kew M, Tanaka Y, Mizokami M, Kramvis A: **Molecular characterization of hepatitis B virus isolates from Zimbabwean blood donors.** *J Med Virol* 2011, **83**(2):235–244.
35. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC: **Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination.** *J Virol* 1999, **73**(1):152–160.
36. Bell TG, Kramvis A: **Mutation Reporter Tool: An online tool to interrogate loci of interest, with its utility demonstrated using hepatitis B virus.** *Virol J* 2013, **10**(1):62. doi:10.1186/1743-422X-10-62.
37. Liaw Y-F, Chien R-N, Yeh C-T, Tsai S-L, Chu C-M: **Acute exacerbation and hepatitis B virus clearance after emergence of YMDD motif mutation during lamivudine therapy.** *Hepatology* 1999, **30**(2):567–572.
38. Rehermann B, Nascimbeni M: **Immunology of hepatitis B virus and hepatitis C virus infection.** *Nat Rev Immunol* 2005, **5**(3):215–229.
39. Liu TT, Fang Y, Xiong H, Chen TY, Ni ZP, Luo JF, Zhao NQ, Shen XZ: **A case–control study of the relationship between hepatitis B virus DNA level and risk of hepatocellular carcinoma in Qidong China.** *World J Gastroenterol* 2008, **14**(19):3059–3063.
40. Abdou Chekaraou M, Brichler S, Mansour W, Le-Gal F, Garba A, Deny P, Gordien E: **A novel hepatitis B virus (HBV) subgenotype D (D8) strain, resulting from recombination between genotypes D and E, is circulating in Niger along with HBV/E strains.** *J Gen Virol* 2010, **91**(Pt 6):1609–1620.
41. Kramvis A, Arakawa K, Yu MC, Nogueira R, Stram DO, Kew MC: **Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus.** *J Med Virol* 2008, **80**(1):27–46.
42. Kramvis A, Paraskevis D: **Subgenotype A1 of Hepatitis B Virus - Tracing Human Migrations in and out of Africa.** *Antivir Ther* 2013, **18**(3 Pt B):513–521.
43. Milich D, Liang TJ: **Exploring the biological basis of hepatitis B e antigen in hepatitis B virus infection.** *Hepatology* 2003, **38**(5):1075–1086.
44. Carman WF, Jacyna MR, Hadziyannis S, Karayiannis P, McGarvey MJ, Makris A, Thomas HC: **Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection.** *Lancet* 1989, **2**(8663):588–591.
45. Li JS, Tong SP, Wen YM, Vitvitski L, Zhang Q, Trepo C: **Hepatitis B virus genotype A rarely circulates as an HBe-minus mutant: possible contribution of a single nucleotide in the precore region.** *J Virol* 1993, **67**(9):5402–5410.
46. Lok AS, Akarca U, Greene S: **Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal.** *Proc Natl Acad Sci USA* 1994, **91**(9):4077–4081.
47. Ahn SH, Kramvis A, Kawai S, Spangenberg HC, Li J, Kimbi G, Kew M, Wands J, Tong S: **Sequence variation upstream of precore translation initiation codon reduces hepatitis B virus e antigen production.** *Gastroenterology* 2003, **125**(5):1370–1378.
48. Garmiri P, Loua A, Haba N, Candotti D, Allain JP: **Deletions and recombinations in the core region of hepatitis B virus genotype E strains**

from asymptomatic blood donors in Guinea, west Africa. *J Gen Virol* 2009, **90**(Pt 10):2442–2451.
49. Chen BF, Liu CJ, Jow GM, Chen PJ, Kao JH, Chen DS: **High Prevalence and Mapping of Pre-S Deletion in Hepatitis B Virus Carriers With Progressive Liver Diseases.** *Gastroenterology* 2006, **130**(4):1153–1168.
50. Mun H-S, Lee S-A, Kim H, Hwang E-S, Kook Y-H, Kim B-J: **Novel F141L Pre-S2 Mutation in Hepatitis B Virus Increases the Risk of Hepatocellular Carcinoma in Patients with Chronic Genotype C Infections.** *J Virol* 2011, **85**(1):123–132.
51. Biswas A, Panigrahi R, Banerjee A, Pal M, De BK, Chakrabarti S, Chakravarty R: **Differential pattern of pre-S mutations/deletions and its association with hepatitis B virus genotypes in Eastern India.** *Infect Genet Evol* 2012, **12**(2):384–391.
52. Carman WF: **The clinical significance of surface antigen variants of hepatitis B virus.** *J Viral Hepat* 1997, **4**(Suppl 1):11–20.
53. Dupouey J, Gerolami R, Solas C, Colson P: **Hepatitis B virus variant with the a194t substitution within reverse transcriptase before and under adefovir and tenofovir therapy.** *Clinics and Research in Hepatology and Gastroenterology* 2012, **36**(2):e26–e28.

doi:10.1186/1471-2334-13-328
Cite this article as: Yousif *et al.*: Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan. *BMC Infectious Diseases* 2013 **13**:328.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

Chapter 4: Paper III

Genotyping and Virological Characteristics of Hepatitis B Virus in Human Immunodeficiency Virus Infected Individuals in Sudan

YOUSIF, M., MUDAWI, H., HUSSEIN, W., MUKHTAR, M., NEMERI, O., GLEBE, D. & KRAMVIS, A (2014)

International Journal of Infectious Diseases

Journal: International Journal of Infectious Diseases, is a peer-reviewed, ISI-accredited, open-access journal, with an impact factor of 3.03

This is a follow-up study of the work carried out by our collaborative team. The sister manuscript entitled: *Hepatitis B or C virus co-infection and occult hepatitis B virus infection in adult Sudanese HIV patients* is presented in Appendix A

This paper is a follow up of the paper submitted for publication together with our Sudanese collaborators, which has been included in Appendix A (Mudawi et al). In order to facilitate the comprehension of this section, a summary of the major findings of the Mudawi et al paper is provided by way of introduction.

Human Immunodeficiency (HIV) is a serious health problem in sub-Saharan Africa. HIV infection is complicated by co-infection with other pathogens including hepatitis B and C viruses (HBV and HCV) that share similar routes of infection. Co-infection with HBV or HCV and HIV is characterized by more rapid progression of liver disease including accelerated fibrosis, cirrhosis and hepatocellular carcinoma. No clear knowledge about HBV in the setting of HIV infection is available in Sudan, and it was important to determine this in order to select appropriate antiretroviral therapy. Moreover, no studies have been carried out to determine the presence of HBV DNA, in the absence of HBsAg, in HIV-positive Sudanese. Thus the aims of the study were to determine the prevalence of HBV and HCV infection and the prevalence of occult hepatitis B infection (OBI) among adult Sudanese HIV individuals.

A total of 358 HIV- infected patients were enrolled in the study with a mean age $\pm$ standard deviation (SD) of 35.2 $\pm$ 9.3 years, with male: female ratio of 1.3:1.0. Mean alanine aminotransferase $\pm$ SD (ALT) levels of the study population was 10.9$\pm$ 18.0 U/L. All participants tested positive to HIV antibodies using indirect ELISA.

Of the 358 patients, 225 (62.8%) were anti- HBc IgG-positive indicating previous exposure to HBV infection. The mean $\pm$ SD, age and ALT levels of this subpopulation were 35.3 $\pm$ 9.4 years and 11.0 + 20.9 U/L, respectively, and these values did not differ from those of the population as a whole. Forty two patients (11.7%) were HBsAg-positive. Mean $\pm$ SD, age and ALT levels of the group was 34.7$\pm$ 8.7 years and 9.8 $\pm$ 7.2 U/L respectively and these

for both HBeAg and anti-HBe. HBV DNA was detected by PCR in 96 patients (26.8%); all were anti-HBc IgG–positive. Forty two patients (11.7%) were HBsAg-positive and 54 patients (15.1%) were HBsAg -negative indicating occult infection. Overall median [interquartile range (IQR)] viral load was 2.9[1.7-4.1] $\log_{10}$ IU/ml, with HBsAg-positive individuals having significantly higher viral loads compared to HBsAg-negative patients. Eight HBsAg-positive, HBeAg-positive, anti-HBe-negative patients had viral loads of 5.6 $\log_{10}$ IU/ml, 24 HBsAg-positive, HBeAg-negative, anti-HBeAg-positive patients had viral loads of 3.8 $\log_{10}$ IU/ml and 10 HBsAg-positive, HBeAg/anti-HBeAg-negative had viral loads of 3.9 $\log_{10}$ IU/ml. Thirty five (9.8%) HBsAg-negative patients had a viral load < 2.3 $\log_{10}$ IU/ml (< 200 IU/ml) (true OBI) and nineteen (5.3%) had viral loads >2.3 $\log_{10}$ IU/ml (>200 IU/ml) (false OBI or HBsAg-covert). One patient was reactive in both anti-HBc IgG and anti-HBc IgM, the patient was HBsAg-negative with normal ALT and had detectable HBV DNA on PCR indicating a recent HBV infection or a HBV reactivation in a patient with OBI.

In conclusion in Sudan, HBsAg-negative HBV infection was detected in 15.1% of patients. Prevention of HBV and HIV through education and vaccination against HBV is essential in our setting. All HIV infected patients should be carefully screened for HBV infection, with HBsAg and anti-HBc Ig G antibodies prior to starting ART. Although PCR testing would detect OBI, which would be missed by serological testing alone, it is too expensive to include as a routine screening test in Sudan

Genotyping and Virological Characteristics of Hepatitis B Virus in Human Immunodeficiency Virus Infected Individuals in Sudan

Mukhlid Yousif[1], Hatim Mudawi[2], Waleed Hussein[3], Maowia Mukhtar[3], Omer Nemeri[4], Dieter Glebe[5] and Anna Kramvis[1*]

[1]Hepatitis Virus Diversity Research Programme, Department of Internal Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

[2]Department of Medicine, Faculty of Medicine, University of Khartoum, Sudan

[3]Institute of Endemic Diseases, University of Khartoum, Sudan

[4]College of Medicine, Bahri University

[5]Institute of Medical Virology, National Reference Centre of Hepatitis B and D, Justus Liebig-University of Giessen, Giessen, Germany

[*]**Corresponding author:** Professor A. Kramvis,
Hepatitis Virus Diversity Research Programme,
University of the Witwatersrand,
7 York Road, Parktown, Johannesburg, 2193,
South Africa

E-mail address: Anna.Kramvis@wits.ac.za

Telephone: +27(0) 11 488 3100

Fax: +27(0) 86 529 6806

Objectives: Hepatitis B virus (HBV) and human immunodeficiency virus (HIV) share common routes of blood borne transmission. In HBV mono-infected Sudanese individuals, genotypes D, E and A circulate. The objective of this study was to molecularly characterize HBV from HBV/HIV co-infected individuals. *Methods*: From 32 HBsAg-positive and 18 HBsAg-negative sera, the polymerase overlapping the S region and the basic core promoter (BCP/PC) of HBV were amplified and sequenced. *Results*: HBV from 37 samples was successfully genotyped and the genotype distribution was 46% D, 21.6% E, 18.9% A and 13.5% D/E recombinant. Compared to mono-infected individuals, the frequency of the D/E recombinant and genotype A was higher in HBV/HIV co-infected patients, as was the intragroup divergence of genotype E. BCP/PC mutations affecting HBeAg expression at the transcriptional and translational levels were detected. Two HBsAg-positive individuals had preS deletion mutants. In the S region, the following mutations could account for the HBsAg-negativity: sM133T, sE164G, sV168G and sS174N. No primary drug resistance mutations were found. *Conclusion*: In HBV/HIV co-infected Sudanese, the ratio of genotype A to non-A was higher than that in mono-infected patients. The genotype E intragroup divergence in HBV/HIV co-infected individuals was significantly higher than that in HBV mono-infected patients.

Key words: occult HBV infection, genetic diversity, deletion mutants, subgenotypes

INTRODUCTION

Hepatitis B virus (HBV), the smallest DNA virus that infects humans, is a member of family *Hepadnaviridae* and it causes acute and chronic liver disease. Human immunodeficiency virus (HIV) is an enveloped RNA virus that replicates by reverse transcription [1]. HBV and HIV share common routes of blood borne transmission, but they differ in efficiency of transmission and in their geographic distribution [2]. Worldwide, it has been estimated that 10% of the 34 million people infected with HIV, have chronic hepatitis B infection [2].

In Sudan, the prevalence of HBV infection ranges from 3.9% in blood donors in Khartoum [3] to 26% in patients seen in out-patient clinics in Juba in southern Sudan [4]. It has been estimated that HIV prevalence in adults in Sudan was 1.4% [5] and according to Sudan National AIDS control Programme (SNAP), this figure has decreased to 0.52% in 2011 [6]. It is very difficult to surmise, from the data available, the true prevalence of HBV/HIV co-infection in Sudan because of the paucity of data and the wide time span between studies. Using serological techniques, two studies carried out 24 years ago determined the prevalence of HBV/HIV co-infection in Sudan. The first study showed 14% HBV infection among HIV-negative individuals who were sexually active heterosexuals in northeast Sudan [7]. The second study, among soldiers in eastern Sudan, showed 1.7% HIV infection and 25% HBsAg positivity, with only 0.8% being co-infected with HBV and HIV [8]. In a recently completed study, our group has shown of the 358 HIV-infected individuals from Khartoum, 225 (62.8%) had at least one HBV serological marker, with 129 (36%) being HBV DNA negative and 96 (26.8%) being HBV DNA-positive by real-time polymerase chain reaction (Rt-PCR) [9] HBV DNA was detected together with HBsAg in 42 (11.7%) and without HBsAg in 54 (15.1%) of the participants [9]. Of these, 35 (9.8%) had a viral load < 200 IU/ml [< 2.3 log10 IU/ml] and

therefore true occult hepatitis B infection (OBI) [10] and 19 (5.3%) had viral loads >200 IU/ml [> 2.3 log10 IU/ml] and were therefore HBsAg-covert [11].

As a result of the lack of proofreading activity of the viral polymerase, HBV displays considerable sequence heterogeneity, leading to its classification into at least 9 genotypes [12-20], which are defined by the inter-genotypic differences of more than 7.5% in the complete HBV genome [12, 21, 22]. It is considered that genotypes play an important role in treatment management and disease prognosis [23-25]. In Sudan, close to 60% of liver disease patients were infected with genotype D (predominantly subgenotype D1), 30% infected with genotype E, 8.5% infected with genotype A and 2.5% infected with D/E recombinant [26]. On the other hand, in Sudanese blood donors, genotype E prevails (55%), followed by genotype D (39%), D/E recombinant (4%) and genotype A (2%) [3]. No information exists about the distribution of HBV genotypes in HIV-infected individuals in Sudan. The objectives of this study were to genotype and molecularly characterize HBV and to determine possible virological characteristics that could differentiate isolates from HBsAg-positive and HBsAg- negative HIV-infected Sudanese.

PATIENTS AND METHODS

Ethical considerations:

The study is the continuation of our previous study on HBV/HIV co-infection among adult Sudanese HIV patients[9]. The study was originally approved by the ethics committee of the Faculty of Medicine, University of Khartoum and the National Ethics Committee at the Federal Ministry of Health. All patients signed informed consent and the clinical report forms were completed by qualified practitioners. The study was approved by the Human Ethics Committee of the University of the Witwatersrand.

Serum samples:

Serum samples collected from 358 HIV-positive treatment naïve patients and 96 were positive HBV DNA using realtime PCR[9]. Fifty of these 96 were randomly selected for downstream genotyping and molecular analysis of the HBV.

DNA Extraction

DNA was extracted from 200 µL of serum using QIAamp DNA mini blood kit (QIAGEN GmbH, Germany), according to manufacturer's instructions, and eluted in 100 µL of elution buffer.

Polymerase Chain Reaction (PCR)

Two regions of the HBV genome were amplified: the basic core promoter/ precore (BCP/PC) region (1742-1900 from *Eco*R1) was amplified using a minor modifications of the method described by Takahashi and colleagues [27], primers 1606 (+) (1606–1625 from *Eco*R1 site) and 1974 (−) (1974–1955 from *Eco*R1 site) were used for the first round and 1653(+) (1653–1672 from *Eco*R1 site) and 1959(−) (1959–1940 from *Eco*R1 site) for the second round. The region 2624 – 1240 from *Eco*R1, covering part of the polymerase region and overlapping complete S (2848 to 835 from the *Eco*R1 site) (P/S) region was amplified using primers 2410(+) (2410–2439 from *Eco*R1 site) and 1314(−) (1314–1291 from *Eco*R1 site) for the first round and 2451 (+) (2451–2482 from *Eco*R1site) and 1280 (−) (1280–1254 from *Eco*R1 site) for the second round [28].

Sequencing

The BigDye Terminator v3.0 Cycle Sequencing Ready Reaction Kit (Applied Biosystems., Foster City, USA) was used and sequencing was performed using the ABI 3130XL Genetic analyzer (Applied Biosystems, Foster City, CA). The three overlapping fragments of the

region 2624 – 1240 from *Eco*R1, obtained were assembled using the Fragment Merger Tool [29]. Sequences were deposited in GenBank, accession numbers XXXXXXXX – XXXXXXXX..

Phylogenetic analysis

Neighbor-joining phylogenetic tree with a bootstrap of 1000 replicates was constructed using MEGA5 [30] or PHYLIP (Phylogeny inference package version 3.69). DNADIST consecutively with NEIGHBOR (neighbor joining) were used to generate dendrograms and viewed in TreeView [31, 32].

RESULTS

HBV serology, amplification and quantification of HBV DNA

Of the 50 HBV DNA-positive patients randomly selected, 28 were males. Mean age ± standard deviation (SD) was 35 ± 8.1 years. ALT median [interquartile range (IQR)] was [6.15 (3.3-11.8)] IU/L (reference range 5–40 IU/L) [33] and the median of HBV DNA viral load (IQR) was [4.2 (3.56-6.67)] log IU/ml. Thirty-two (64%) were HBsAg-positive and 18 (36%) were HBsAg-negative. Only 6 were HBeAg-positive. No difference in the age, gender, HBsAg prevalence and ALT levels was observed between 50 samples (randomly selected for further analyses), and the original 96 HBV DNA-positive samples [9]. Of the 18 HBsAg-negative patients, 10 had viral loads < 200 IU/ml [< 2.3 log10 IU/ml], and thus according to the Taormina definition, were true OBI [10] . The remaining 8 had viral loads > 200 IU/ml [> 2.3 log10 IU/ml] and therefore HBsAg-*covert* infections [11], (Table 1). No difference in the age, gender and ALT levels was observed between the true OBI and HBsAg- *covert*.

HBV genotyping and phylogenetic analysis

From the 50 samples, sequences were obtained from 46: 27 samples in both the BCP/PC and the P/S region, 12 in the BCP/PC alone and 7 in the P/S region only. For three samples, however, the sequences of of the S region were not of good quality and therefore could not be used for genotyping.

In total, 37 HBV isolates were successfully genotyped using one of two methods:

1. 31 out of 37 were genotyped using phylogenetic analysis of the P/S region (Figure 1).
2. The BCP/PC region sequences could be used to differentiate between genotype A and non-A. Six out of 37 had 1858C characteristic of genotype A and all belonged to subgenotype A1 because they had 1809-1812 TCAT and 1888A. Six samples had 1858T characteristic of non-A genotypes but this region cannot differentiate between genotype D and E and thus these samples could not be genotyped.

The genotype distribution of the 37 isolates was: 17 (46%) genotype D, 8 (21.6%) genotype E, 7 (18.9%) genotype A and 5 (13.5%) putative D/E recombinant. If we include the 13 (26%) samples that could not be genotyped, the distribution in the 50 samples was 34% (17/50) infected with genotype D, 16% (8/50) with genotype E, 14% (7/50) with genotype A and 10% (5/50) were infected with a putative D/E recombinant. The genotype distribution did not differ between HBsAg-positive and HBsAg-negative sera nor between HBsAg-*covert* and OBI patients. Comparing the genotype distribution of HBV isolates in the Sudanese HBV/HIV co-infected individuals to mono-infected liver disease patients [26], the frequency of genotype A and the D/E recombinant was higher in HIV co-infected patients ($p<0.05$) (Figure 2).

Using the following criteria: phylogenetic analysis of the P/S region, inter-group divergence, signature amino acids and geographical distribution, the HBV genotypes were further

classified into subgenotypes [20]. Ten of 17 genotype D isolates belonged to subgenotype D1 (58.8%), 2 to D2 (11.7%), and 5 to D6 (29.5%) and the one genotype A isolate belonged to subgenotype A1 (Figure 1). SDH042-OBI was an outlier of the cluster containing subgenotypes of D1- D3 and D5. However, when the analysis was repeated using only the complete S region (2848 to 835 from the *Eco*R1 site) SDH042-OBI clustered with subgenotype D1 (Figure 1A).

Genetic diversity of Sudanese strains

The intra-group divergence of the HBV isolated from the HBV/HIV-co-infected individuals in the present study were compared to those isolated from HBV mono-infected liver disease patients [26] (Figure 3). The intra-group divergence of genotype E from HBV mono-infected liver disease patients was lower than the intra-group divergence of genotype D from HBV mono- infected liver disease patients ($p<0.05$), whereas in HBV/HIV co-infected individuals this difference was not observed (Figure 3). Moreover, only the intra-group divergence of genotype E from HBV/HIV co-infected individuals was higher than intra-group divergence of genotype E from HBV mono-infected liver disease patients ($p<0.05$). The intra-group divergence of the HBV/HIV co-infected individuals was 83% higher than that in HBV mono-infected liver disease patients for genotype E, whereas it was only 3.7% higher for genotype D.

Analysis of BCP/PC region

BCP/PC mutations can down regulate or abolish the expression of the HBeAg at the transcriptional, translational or post-translational levels [34]. Of the 6 isolates from HBeAg-positive sera, two had mutations at the BCP: [SDH052 (HBsAg-positive- genotype D] had A1762G and [SDH084 (HBsAg-positive- genotype E)] had A1762T/G1764A. The remaining 4 isolates, from HBeAg-positive patients, had wild type BCP/PC. Thirty of the 44 isolates

from HBeAg-negative individuals (17 HBsAg-positive: 13 HBsAg-negative) were successfully amplified and sequenced in the BCP/PC region. Twenty-five isolates had genotype assignments: 11 belonged to genotype D, 5 to genotype E, 6 to genotype A and 3 were D/E recombinants. The mutations found in the isolates from HBeAg-negative patients are shown in Table 2. Three isolates had A1762T alone and 3 isolates had start codon mutations at 1814-1816. Ten of the 26 isolates had G1896A. Even though 1858T is found in both genotypes D and E and allows for G1896A to occur, G1896A was found more frequently in genotype E and the D/E recombinant compared to genotype D ($p<0.05$). Moreover, G1896A occurred more frequently in isolates from HBsAg-positive compared to HBsAg-negative individuals ($p< 0.05$). G1896A occurred together with A1762T/G1764A in two genotype E isolates. The mutation C1766T was found in 5 genotype A isolates only ($p<0.05$).

Analysis of the pre-S1, pre-S2, Surface and polymerase regions

The sequence of the P/S region (2624- 1240 from *Eco*R1 site) was obtained for 31 HBV isolates from HBV/HIV co-infected individuals. Of the 31 isolates, 20 from HBsAg-positive and 11 from HBsAg-negative. PreS deletions were found in two genotype E isolates: SDH084-HBsAg-positive had a 93 nucleotide deletion in the preS1 and preS2 regions, from 3178 to 58 from *Eco*R1 site, which lead to an 8 amino acid deletion from position 111 to 118 in the preS1 and 23 amino acids, were deleted from position 1 to 23, in the preS2. SDH120-HBsAg-positive had a 36 nucleotide deletion in the preS2 starting from nucleotide 5 to 40 from *Eco*R1 site, which results in a 12 amino acid deletion from position 6 to 17. Table 3 shows the various mutations in the polymerase and overlapping S regions.

DISCUSSION

In our recently completed study of 358 HIV-positive participants, 225 (62.8%) had serological evidence of past or current HBV infection. HBV DNA was detected in 96 (26.8%): 42 (11.7%) with overt HBsAg-positive infection and 54 (15.1%) with HBsAg-negative and anti-HBc as a marker of past HBV infection [9]. Fifty of the 96 HBV DNA-positively samples were randomly selected for molecular characterization of the HBV.

The distribution ratio of HBV genotypes D and E, found in HBV/HIV co-infected individuals (Figure 2), did not differ from that found in HBV mono-infected liver disease patients and there was no difference in their clustering in the phylogenetic tree (Figure 1) [26]. However, genotype A and the putative D/E recombinant were found in a higher frequency in HBV/HIV co-infected individuals compared to HBV mono-infected liver disease patients (genotype A: 18.4% *versus* 8.6%, and D/E: 13.2% *versus* 2.5%) ($p < 0.05$). Genotypes D, E, A and D/E recombinant were found in Sudanese blood donors, with genotype E predominating [3].

With regard to the subgenotype distribution, in agreement with the previous studies of HBV in Sudan [3, 26], subgenotype D1 was found to predominate in the HBV/HIV co-infected individuals (Figure 1). Thus the predominant subgenotype of D found in Sudan is D1. This is the subgenotype of D that prevails in the Mediterranean, Arabian subcontinent and in China [30]. Subgenotype A1 was found circulating in in HBV/HIV-co-infected individuals and the liver disease patients [26], whereas subgenotype A2 was found in blood donors [3]. Subgenotype A1 predominates in eastern and southern Africa and subgenotype A2 is the subgenotype of A predominating outside Africa [35] .

Following the calculation of the intra-group divergence of the genotype D and E isolates from HBV mono-infected liver disease patients and HBV/HIV co-infected individuals, it was

found that genotype E from HBV mono-infected liver disease patients had a low diversity of 1.8% compared to 2.7% for genotype D from HBV mono-infected liver disease patients ($p<0.05$) (Figure 4). Genotype E has been characterized to have a low diversity compared to other genotypes and subgenotypes of HBV. Using more than 400 sequences of HBV isolated from west African countries, the genetic diversity of genotype E was found to be 0.73% for the S gene and 1.7% for the complete genome [36]. Complete genomes of genotype E isolates from Namibia, Angola and Madagascar differed from each other by only 1.2% [37]. Genotype D has been shown to be the most divergent genotype after genotype C [38]. The significant difference in the intragroup divergence of genotypes D and E seen in mono-infected individuals, however, was not seen in HBV/HIV co-infected individuals and the intra-group divergence of genotype E in HBV/HIV co-infected individuals was significantly higher than that in HBV mono-infected liver disease patients (Figure 4). It appears that in HBV mono-infected liver disease patients, who are immune competent, the genetic diversity of genotype E is constrained whereas in HBV/HIV co-infected individuals there is an expansion of the quasispecies. HIV infection and its associated immune suppression might alter the heterogeneity of the virus and increase the genetic variability of HBV [39], which may be as a result of the reactivation of occult HBV that can survive in the presence of immune selection.

The 81% prevalence of HBeAg-negativity in HBV/HIV co-infected patients [9] was lower than in HIV- negative patients, where 88 % were HBeAg –negative [26]. When comparing the mutation distribution in BCP/PC, at 10 loci that can affect HBeAg-expression (1762, 1764, 1809 –1812, 1814–1816 and 1896), we found that mutations at 1762 and 1764 occurred less frequently in isolates from HBV/HIV co-infected individuals compared to those from mono-infected liver disease patients ($p<0.05$). Mutations at 1762 1764 can reduce the transcription of precore mRNA, resulting in lower expression of HBeAg [40]

Two genotype E HBV isolates from HBsAg-positive HBV/HIV co-infected individuals had deletions in the preS1 and preS2 regions. No deletions were detected in genotype D isolates. The 12 amino acid deletion mutant was found in the pre-S2 whereas the 31 amino acid deletion covered the 3' pre-S1 and 5'pre-S2. Similar deletion mutants were found in genotype D and E isolates from Sudanese HBV mono-infected liver disease patients [26] but only in genotype E isolates from blood donors in Sudan [3] and Guinea [41]. The deletion mutants in HBV genotype E isolates from HBV mono-infected liver disease patients were mainly in hepatoceullar carcinoma patients whereas this was not the case for genotype D isolates [26]. Deletion mutants in the pre-S have been implicated in the progression of liver diseases in genotypes B and C [42] and the development of hepatoceullar carcinoma in genotype A [43]. The presence of deletion mutants in HBV from HBV/HIV co-infected individuals has been described in genotype A, both subgenotype A1 [44] and subgenotype A2 [45]. PreS2F22L was found in four isolates from HBsAg-positive HBV/HIV co-infected individuals and was found to be significantly associated with G1896A. PreS2F22L has been significantly associated with cirrhosis [46] and has been shown to be a s risk factor for development of HCC in Chinese patients infected with genotype C [47] and in Indian patients infected with genotype D [46, 48]. HIV infection has been intimated to increase the hepatocarcinogenic potential of HBV [49].

In agreement with other studies in HIV-infected individuals, HBsAg-negativity was frequent [44, 50]. The HBsAg-negativety resulted in both OBI (viral loads < 200 IU/ml[< 2.3 log10 IU/ml]) [10] and HBsAg-covert infection (viral loads > 200 IU/ml [> 2.3 log10 IU/ml]) [11]. In the case of OBI, the inability to detect HBsAg could be as a result of its presence at very low concentrations, below the detection limits of the assay. In the case of the HBsAg-covert infection, the HBsAg-negativity may be the result of the presence of detection escape mutations in the 'a' determinant. Although we found no mutational pattern that could

differentiate the HBV isolated from HBsAg-negative individuals from those isolated from HBsAg-positive ones, there were a number of mutations that could account for the HBsAg-negativity (Table 3). These mutations include ps2R48T, sM133T, sE164G, sV168G and sS174N and have also been found in strains circulating in HBsAg-negative HIV infected individuals infected with subgenotype A1 of HBV [44]. Of particular interest is S174N that has only been detected in HBV from HBsAg-negative HBV/HIV co-infected individuals [44, 51] and has been described as a reactivation marker in a single patient infected with subgenotype A2 [51]. No primary drug resistance mutations were detected in the present study. This is not unexpected because all patients recruited in this study were treatment naïve.

The distribution of genotypes D and E of HBV found in HBV/HIV co-infected individuals did not differ from those found in HBV mono-infected liver disease patients. However, the ratio of genotype A to non-A in HIV co-infected individuals was found to be higher than that found in mono-infected patients. A predominance of genotype A of HBV in HIV co-infected individuals has also been reported by others [44, 45, 52]. Subgenotype A1 is the predominant strain circulating in countries neighbouring Sudan, including Kenya [53] and Uganda [54]. In a previous study, it was shown that the only factor that differentiated HBV-positive HIV-infected individuals from those who were HBV-negative was the number of lifetime sexual contacts [11], thus implicating sexual transmission of this subgenotype [44] Furthermore, the intragroup divergence of genotype E in HBV/HIV co-infected individuals was significantly higher than that in HBV mono-infected patients. We detected preS deletions, mutations associated with severe liver disease, detection-escape mutations as well as reactivation markers in the HBV isolated from HIV co-infected individuals. These alterations can have important implications in exacerbating the clinical manifestations of HBV infection, its detection and clinical management.

ACKNOWLEDGMENTS

We acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) of the special African-German collaborative programme, "Africa Initiative", Grant GL595/3-1). The funding agency had no role in the study. We thank Trevor Bell for some statistical calculations. The authors have no conflict of interest to declare. The study was approved by the Ethics committee of the Faculty of Medicine, University of Khartoum, the National Ethics committee, Federal Ministry of Health, Sudan and by the Human Ethics Committee of the University of the Witwatersrand, South Africa. Written consent was obtained from all participants before their enrollment.

REFERENCES

1.Vogt V. Retroviral virions and genomes. 1997.

2.Alter MJ. Epidemiology of viral hepatitis and HIV co-infection. Journal of Hepatology 2006;44:S6.

3.Mahgoub S, Candotti D, El Ekiaby M, Allain JP. Hepatitis B virus (HBV) infection and recombination between HBV genotypes D and E in asymptomatic blood donors from Khartoum, Sudan. Journal of Clinical Microbiology 2011;49:298-306.

4.McCarthy MC, el-Tigani A, Khalid IO, Hyams KC. Hepatitis B and C in Juba, southern Sudan: results of a serosurvey. Trans R Soc Trop Med Hyg 1994;88:534-6.

5.UNAIDS. 2008 Report on the global AIDS epidemic. World Health Organization; 2008.

6.SNAP SNACP. Global AIDS response progress reporting 2010-2011. 2012.

7.McCarthy MC, Burans JP, Constantine NT, EL-Hag AA, El-Tayeb M. Hepatitis B and HIV in Sudan: a serosurvey for hepatitis B and human immunodeficiency virus antibodies among sexually active heterosexuals. DTIC Document; 1989.

8.McCarthy MC, Hyams KC, El-Hag AE-T, El-Dabi MA, El-Tayeb ME-S, Khalid IO, George JF, Constantine NT, Woody JN. HIV-1 and hepatitis B transmission in Sudan. AIDS 1989;3:725-30.

9.Mudawi H, Hussein, W., Mukhtar, M, Yousif, M, Nemeri, O., Glebe, D, Kramvis, A Overt and occult hepatitis B virus infection in adult Sudanese HIV patients. International Journal of Infectious Diseases submitted, 2013.

10.Raimondo G, Allain JP, Brunetto MR, Buendia MA, Chen DS, Colombo M, Craxi A, Donato F, Ferrari C, Gaeta GB, Gerlich WH, Levrero M, Locarnini S, Michalak T, Mondelli MU, Pawlotsky JM, Pollicino T, Prati D, Puoti M, Samuel D, Shouval D, Smedile A, Squadrito G, Trepo C, Villa E, Will H, Zanetti AR, Zoulim F. Statements

from the Taormina expert meeting on occult hepatitis B virus infection. J Hepatol 2008;49:652-7.

11.Bell TG, Makondo E, Martinson NA, Kramvis A. Hepatitis B Virus Infection in Human Immunodeficiency Virus Infected Southern African Adults: Occult or Overt–That Is the Question. Plos One 2012;7:e45750.

12.Okamoto H, Tsuda F, Sakugawa H, Sastrosoewignjo RI, Imai M, Miyakawa Y, Mayumi M. Typing Hepatitis B Virus by Homology in Nucleotide Sequence: Comparison of Surface Antigen Subtypes. Journal of General Virology 1988;69:2575-83.

13.Naumann H, Schaefer S, Yoshida CF, Gaspar AM, Repp R, Gerlich WH. Identification of a new hepatitis B virus (HBV) genotype from Brazil that expresses HBV surface antigen subtype adw4. J Gen Virol 1993;74 (Pt 8):1627-32.

14.Norder H, Courouce AM, Magnius LO. Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. Virology 1994;198:489-503.

15.Norder H, Hammas B, Lofdahl S, Courouce AM, Magnius LO. Comparison of the amino acid sequences of nine different serotypes of hepatitis B surface antigen and genomic classification of the corresponding hepatitis B virus strains. J Gen Virol 1992;73 (Pt 5):1201-8.

16.Stuyver L, De Gendt S, Van Geyt C, Zoulim F, Fried M, Schinazi RF, Rossau R. A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. Journal of General Virology 2000;81:67-74.

17.Arauz-Ruiz P, Norder H, Robertson BH, Magnius LO. Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. Journal of General Virology 2002;83:2059-73.

18.Yu H, Yuan Q, Ge S-X, Wang H-Y, Zhang Y-L, Chen Q-R, Zhang J, Chen P-J, Xia N-S. Molecular and Phylogenetic Analyses Suggest an Additional Hepatitis B Virus Genotype "I". Plos One 2010;5:e9297.

19.Tatematsu K, Tanaka Y, Kurbanov F, Sugauchi F, Mano S, Maeshiro T, Nakayoshi T, Wakuta M, Miyakawa Y, Mizokami M. A Genetic Variant of Hepatitis B Virus Divergent from Known Human and Ape Genotypes Isolated from a Japanese Patient and Provisionally Assigned to New Genotype J. Journal of Virology 2009;83:10538-47.

20.Kramvis A, Arakawa K, Yu MC, Nogueira R, Stram DO, Kew MC. Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. J Med Virol 2008;80:27-46.

21.Norder H, Courouce AM, Magnius LO. Molecular basis of hepatitis B virus serotype variations within the four major subtypes. The Journal of general virology 1992;73 (Pt 12):3141-5.

22.Kramvis A, Kew M, Francois G. Hepatitis B virus genotypes. Vaccine 2005;23:2409-23.

23.Kramvis A, Kew MC. Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy. J Viral Hepat 2005;12:456-64.

24.Kao JH, Chen PJ, Lai MY, Chen DS. Genotypes and clinical phenotypes of hepatitis B virus in patients with chronic hepatitis B virus infection. J Clin Microbiol 2002;40:1207-9.

25.McMahon BJ. The influence of hepatitis B virus genotype and subgenotype on the natural history of chronic hepatitis B. Hepatol Int 2009;3:334-42.

26.Yousif M, Mudawi H, Bakhiet S, Glebe D, Kramvis A. Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan. BMC Infectious Diseases 2013;13:328.

27.Takahashi K, Aoyama K, Ohno N, Iwata K, Akahane Y, Baba K, Yoshizawa H, Mishiro S. The precore/core promoter mutant (T1762A1764) of hepatitis B virus: clinical significance and an easy method for detection. J Gen Virol 1995;76 (Pt 12):3159-64.

28.Vermeulen M, Dickens C, Lelie N, Walker E, Coleman C, Keyter M, Reddy R, Crookes R, Kramvis A. Hepatitis B virus transmission by blood transfusion during 4 years of individual-donation nucleic acid testing in South Africa: estimated and observed window period risk. Transfusion 2012;52:880-92.

29.Bell TG, Kramvis A. Fragment Merger: An Online Tool to Merge Overlapping Long Sequence Fragments. Viruses 2013;5:824-33.

30.Yousif M, Kramvis A. Genotype D of hepatitis B virus and its subgenotypes: An update. Hepatology Research 2013;43:355-64.

31.Gulube Z, Chirara M, Kew M, Tanaka Y, Mizokami M, Kramvis A. Molecular characterization of hepatitis B virus isolates from Zimbabwean blood donors. Journal of Medical Virology 2011;83:235-44.

32.Page RDM. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences 1996;12:357-8.

33.Papatheodoridis GV, Manolakopoulos S, Liaw YF, Lok A. Follow-up and indications for liver biopsy in HBeAg-negative chronic hepatitis B virus infection with persistently normal ALT: a systematic review. J Hepatol 2012;57:196-202.

34.Kramvis A, Kew MC. Molecular characterization of subgenotype A1 (subgroup Aa) of hepatitis B virus. Hepatology Research 2007;37:S27-S32.

35.Kramvis A, Paraskevis D. Subgenotype A1 of HBV-tracing human migrations in and out of Africa. Antiviral therapy 2012;18:513-21.

36.Hübschen JM, Mugabo J, Peltier CA, Karasi J-C, Sausy A, Kirpach P, Arendt V, Muller CP. Exceptional genetic variability of hepatitis B virus indicates that Rwanda is east

of an emerging African genotype E/A1 divide. Journal of Medical Virology 2009;81:435-40.

37.Kramvis A, Restorp K, Norder H, Botha JF, Magnius LO, Kew MC. Full genome analysis of hepatitis B virus genotype E strains from South-Western Africa and madagascar reveals low genetic variability. Journal of Medical Virology 2005;77:47-52.

38.Norder H, Courouce AM, Coursaget P, Echevarria JM, Lee SD, Mushahwar IK, Robertson BH, Locarnini S, Magnius LO. Genetic diversity of hepatitis B virus strains derived worldwide: Genotypes, subgenotypes, and HBsAg subtypes. Intervirology 2004;47:289-309.

39.Pal A, Panigrahi R, Biswas A, Datta S, Sarkar N, Guha SK, Saha B, Banerjee A, Chakrabarti S, Chakravarty R. Influence of HIV-associated degree of immune suppression on molecular heterogeneity of hepatitis B virus among HIV co-infected patients. Virology 2013;436:134-42.

40.Buckwold VE, Xu Z, Chen M, Yen TS, Ou JH. Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication. J Virol 1996;70:5845-51.

41.Garmiri P, Loua A, Haba N, Candotti D, Allain JP. Deletions and recombinations in the core region of hepatitis B virus genotype E strains from asymptomatic blood donors in Guinea, west Africa. J Gen Virol 2009;90:2442-51.

42.Chen BF, Liu CJ, Jow GM, Chen PJ, Kao JH, Chen DS. High Prevalence and Mapping of Pre-S Deletion in Hepatitis B Virus Carriers With Progressive Liver Diseases. Gastroenterology 2006;130:1153-68.

43.Kew MC, Kramvis A, Yu MC, Arakawa K, Hodkinson J. Increased hepatocarcinogenic potential of hepatitis B virus genotype A in Bantu-speaking sub-saharan Africans. Journal of Medical Virology 2005;75:513-21.

44.Makondo E, Bell TG, Kramvis A. Genotyping and Molecular Characterization of Hepatitis B Virus from Human Immunodeficiency Virus-Infected Individuals in Southern Africa. Plos One 2012;7:e46345.

45.Audsley J, Littlejohn M, Yuen L, Sasadeusz J, Ayres A, Desmond C, Spelman T, Lau G, Matthews GV, Avihingsanon A, Seaberg E, Philp F, Saulynas M, Ruxrungtham K, Dore GJ, Locarnini SA, Thio CL, Lewin SR, Revill PA. HBV mutations in untreated HIV-HBV co-infection using genomic length sequencing. Virology 2010;405:539-47.

46.Biswas A, Panigrahi R, Banerjee A, Pal M, De BK, Chakrabarti S, Chakravarty R. Differential pattern of pre-S mutations/deletions and its association with hepatitis B virus genotypes in Eastern India. Infection, Genetics and Evolution 2012;12:384-91.

47.Mun H-S, Lee S-A, Kim H, Hwang E-S, Kook Y-H, Kim B-J. Novel F141L Pre-S2 Mutation in Hepatitis B Virus Increases the Risk of Hepatocellular Carcinoma in Patients with Chronic Genotype C Infections. Journal of Virology 2011;85:123-32.

48.Gopalakrishnan D, Keyter, M., Shenoy, K.T., Balakumaran, L.K., Barathi, T., Thomas, V., Vinayakumar, K.R., Panackel, T., Korah, A.T., Nair, R., Kramvis. Predominance of subgenotype A1 of hepatitis b virus in liver disease patients from the state of Kerala, Southern India. World J Gastroenterol, In press 2013.

49.Kew MC, Smuts H, Stewart A. Does HIV infection enhance the hepatocarcinogenic potential of chronic hepatitis B virus infection? J Acquir Immune Defic Syndr 2010;53:413-4.

50.Mphahlele MJ, Lukhwareni A, Burnett RJ, Moropeng LM, Ngobeni JM. High risk of occult hepatitis B virus infection in HIV-positive patients from South Africa. Journal of Clinical Virology 2006;35:14-20.

51.Henke-Gendo C, Amini-Bavil-Olyaee S, Challapalli D, Trautwein C, Deppe H, Schulz TF, Heim A, Tacke F. Symptomatic Hepatitis B Virus (HBV) Reactivation despite

Reduced Viral Fitness Is Associated with HBV Test and Immune Escape Mutations in an HIV-Coinfected Patient. Journal of Infectious Diseases 2008;198:1620-4.

52.Quarleri J, Moretti F, Bouzas M, Laufer N, Carrillo MG, Giuliano SF, Perez H, Cahn P, Salomon H. Hepatitis B virus genotype distribution and its lamivudine-resistant mutants in HIV-coinfected patients with chronic and occult hepatitis B. AIDS research and human retroviruses 2007;23:525-31.

53.Ochwoto M, Chauhan R, Goplakrishnan D, Chen C-Y, Ng'ang'a Z, Okoth F, Kioko H, Kimotho J, Kaiguri P, Kramvis A. Genotyping and molecular characterization of hepatitis B virus in liver disease patients in Kenya. Infection, Genetics and Evolution 2013.

54.Forbi JC, Ben-Ayed Y, Xia G-l, Vaughan G, Drobeniuc J, Switzer WM, Khudyakov YE. Disparate distribution of hepatitis B virus genotypes in four sub-Saharan African countries. Journal of Clinical Virology 2013.

Table 1: Characteristics of HBV isolated from HIV-positive individuals

		Patient characteristics			Viral characteristics			HBV Infection
HBsAg status	**isolate**	**Gender**	**Age**	**ALT IU/L**	**Genotype**	**Subgenotype**	**Viral load IU/ml log**	
Positive	SDH001	Male	50	24	D	D1	6.82	Overt Infection
	SDH020	Female	26	19.9	E	NA	4.25	Overt Infection
	SDH025	Female	23	7.6	D	D2	6.86	Overt Infection
	SDH028	Male	31	10.6	-	-	4.74	Overt Infection
	SDH043	Female	34	4.6	-	-	3.46	Overt Infection
	SDH049	Male	45	11.8	D	D1	5.65	Overt Infection
	SDH050	Male	48	0.4	D	D1	9.16	Overt Infection
	SDH052	Male	38	16.1	D	D6	7.28	Overt Infection
	SDH065	Male	32	18.2	D	D1	8.19	Overt Infection
	SDH071	Male	35	1.2	D/E	NA	4.56	Overt Infection
	SDH075	Female	32	5.5	D	D1	3.73	Overt Infection
	SDH078	Male	25	7.6	-	-	4.08	Overt Infection
	SDH079	Male	25	11.7	-	-	3.34	Overt Infection
	SDH084	Male	25	8.9	E	NA	6.23	Overt Infection
	SDH097	Female	23	6.4	D/E	NA	4.53	Overt Infection
	SDH111	Female	36	7.6	-	-	3.67	Overt Infection
	SDH113	Female	42	27.5	D/E	NA	3.37	Overt Infection
	SDH116	Male	23	5.5	E	NA	4.19	Overt Infection
	SDH120	Male	27	17.8	E	NA	7.83	Overt Infection

Positive	SDH122	Female	38	8.4	-	-	4.00	Overt Infection
	SDH123	Female	35	15.2	D	D1	7.20	Overt Infection
	SDH146	Female	33	8	-	-	4.17	Overt Infection
	SDH153	Female	28	5.9	D	D1	4.16	Overt Infection
	SDH184	Female	30	5.9	A	A1	7.00	Overt Infection
	SDH188	Male	28	4.2	E	NA	4.95	Overt Infection
	SDH189	Female	30	0.8	-	-	3.42	Overt Infection
	SDH193	Male	45	5.9	E	-	8.16	Overt Infection
	SDH194	Female	37	0.8	D	D6	3.14	Overt Infection
	SDH195	Male	38	3.3	A	A1	3.59	Overt Infection
	SDH196	Male	38	13.1	-	-	3.13	Overt Infection
	SDH199	Female	35	3.8	A	A1	3.08	Overt Infection
	SDH200	Male	45	5.6	A	A1	7.12	Overt Infection
Negative	SDH005	Female	45	12.3	D	D1	<2.3	OBI
	SDH007	Female	30	14	D	D2	3.55	HBsAg- *covert*
	SDH022	Male	27	7.2	D	D1	<2.3	OBI
	SDH036	Female	30	5.9	-	-	<2.3	OBI
	SDH042	Male	41	0.8	D	D1	<2.3	OBI
	SDH058	Male	42	8.4	D	D6	2.33	HBsAg- *covert*
	SDH061	Male	50	1.4	D	D6	<2.3	OBI
	SDH087	Male	22	11.8	D	D2	<2.3	OBI
	SDH119	Female	35	11.8	D/E	NA	<2.3	OBI
	SDH126	Male	45	2.9	A	A1	<2.3	OBI

Negative	SDH130	Male	37	2.2	-	-	3.92	HBsAg- *covert*
	SDH147	Female	23	4.6	-	-	5.45	HBsAg- *covert*
	SDH152	Female	38	2.1	E	NA	4.62	HBsAg- *covert*
	SDH160	Male	37	1.2	-	-	4.01	HBsAg- *covert*
	SDH172	Female	35	3.8	A	A1	<2.3	OBI
	SDH174	Male	40	3.3	E	NA	4.38	HBsAg- *covert*
	SDH192	Male	54	7.9	D/E	NA	2.78	HBsAg- *covert*
	SDH197	Male	40	2.5	A	A1	<2.3	OBI

(-): No genotype assignment, NA: not applicable, OBI: occult hepatitis B infection, HBsAg-covert: HBsAg-negative with viral loads >200 IU/ml [> 2.3 log10 IU/ml]

Table 2: Molecular characteristics of the BCP/PC region of HBV isolated from HBeAg-negative HIV co-infected individuals

HBsAg	Isolate	Genotype	Subgenotype	viral load IU/ ml log	1753	1762-1764	1766	1768	1809-1812	1814-1816	1817	1850	1858	1862	1888	1896	1899
Positive	SDH025	D	D2	6.86	T	AG	C	T	GCAC	ATG	C	T	T	G	G	A	G
	SDH049	D	D1	5.65	T	AG	C	T	GCAC	ATT	C	T	T	G	G	G	A
	SDH065	D	D1	8.19	T	AG	G	T	GCAC	ATG	C	T	T	G	G	A	G
	SDH071	D/E	NA	4.56	T	AG	C	T	GCAC	ATG	C	T	T	G	G	A	G
	SDH075	D	D1	3.73	T	AG	C	T	GCAC	ATG	C	T	T	G	G	G	C
	SDH097	D/E	NA	4.53	T	AG	C	T	GCAC	ATG	C	T	T	G	G	A	A
	SDH116	E	NA	4.19	T	AG	C	T	GCAC	ATG	C	T	T	G	G	A	A
	SDH120	E	NA	7.83	T	AG	G	T	GCAC	ATG	C	T	T	G	G	A	G
	SDH123	D	D1	7.2	G	TG	T	A	GCAC	ATG	C	T	T	G	G	G	G
	SDH153	D	D1	4.16	T	AG	C	T	GCAC	ATG	C	T	T	G	G	A	G
	SDH188	E	NA	4.95	T	TA	C	T	GCAC	ATG	C	T	T	G	G	A	G
	SDH194	D	D6	3.14	G	AG	C	T	GCAC	ATG	C	T	T	G	G	G	G
	SDH195	A	A1	3.59	T	AG	T	T	TCAT	ATG	C	A	C	G	A	G	G
	SDH199	A	A1	3.08	T	AG	C	T	TCAT	ATG	C	A	C	G	A	G	G
	SDH200	A	A1	7.12	T	GT	T	T	TCAT	ATG	C	A	C	G	A	G	G
Negative	SDH007	D	D2	3.55	T	AG	C	T	GCAC	AAG	C	T	T	G	G	G	G
	SDH042	D	D1	<2.3	T	AG	C	T	GCAC	ATG	C	T	T	G	G	G	G
	SDH058	D	D6	2.33	G	TG	C	T	GCAC	ATG	C	T	T	G	G	G	G
	SDH061	D	D6	<2.3	G	TG	C	T	GCAC	ATG	C	T	T	G	G	G	G
	SDH126	A	A1	<2.3	T	AG	T	T	TCAT	ATG	C	A	C	G	A	G	G

Negative	SDH152	E	NA	4.62	T	AG	C	T	GCAC	ATT	C	T	T	G	G	G	G
	SDH172	A	A1	<2.3	T	AG	T	T	TCAT	ATG	C	A	C	T	A	G	G
	SDH174	E	NA	4.38	T	AG	C	T	GCAC	ATG	C	T	T	G	G	G	G
	SDH192	D/E	NA	2.78	T	AG	C	T	GCAC	ATG	C	T	T	G	G	A	A
	SDH197	A	A1	<2.3	T	AG	T	T	TCAT	ATG	C	A	C	G	A	G	G

S: S region, RT/POL: reverse transcriptase/ Polymerase, P: positive, N: Negative.

Table 3:S and polymerase mutations found in HBV isolated from HBV/HIV co-infected individuals.

Type of mutation	HBV region	Mutation	Genotype	HBsAg-status
Vaccine and immune escape	PreS1	I48E	A	Positive
	PreS2	Q10R	D/E	Positive
		R48T	E	Positive
			D	Positive
			D	Positive
			D	Negative
PreS2F22 mutation	PreS2	F22S	E	Positive
			D	Positive
		F22L	D/E	Positive
			E	Positive
			E	Positive
			D/E	Positive
a' determinant mutation	S	P120T	E	Positive
		M133T	D	Negative
		G145R	E	Positive
		E164G	D/E	Negative
			E	Negative
HBV "Reactivation" marker	S	V168G	E	Positive
			A	Positive
			D/E	Negative

HBV “Reactivation” marker	S	S174N*	D	Negative
			D/E	Negative
			D/E	Negative
Overlapping polymerase drug resistant mutations	RT/POL	A181	D/E	Negative
		Q215S	D	Positive
			D	positive
		I233	E	positive

NA: not applicable, * Significant association

Figure 1

A rooted phylogenetic tree of 31 HBV isolates from HBV/HIV co-infected individuals of the region 2624 – 1240 from *Eco*R1, covering the polymerase region and overlapping complete S (2848 to 835 from the *Eco*R1 site). The 30 isolates sequenced in the present study have the prefix **SDH** and are shown in **bold,** followed by the clinical status (HBsAg-positive, HBsAg-covert and OBI). These were compared using neighbour-joining with 30 HBV sequences from Sudanese liver disease patients[26] in *italic* and 55 reference sequences, labelled by their accession numbers and country of origin. Bootstrap statistical analysis was performed using 1000 datasets and the percentages are indicated on the nodes. Box A illustrates the clustering of SDH042-OBI relative to the subgenotypes of D when only the complete S (2848 to 835 from *Eco*R1 site) was analyzed. The different subgenotypes are represented by different shapes: Subgenotype D1 (*square*), D2 (*triangle*), D3 (*inverted triangle*) and D5 (*circle*).

Figure 2

The ratio of the distribution of genotypes in HBV in mono-infected liver disease patients [26] compared to HBV/HIV co-infected individuals.

Figure 3

Comparison of the genetic diversity of genotype D and E HBV isolated from HBV mono-infected liver disease patients and HBV/HIV co-infected individuals.

Figure 1

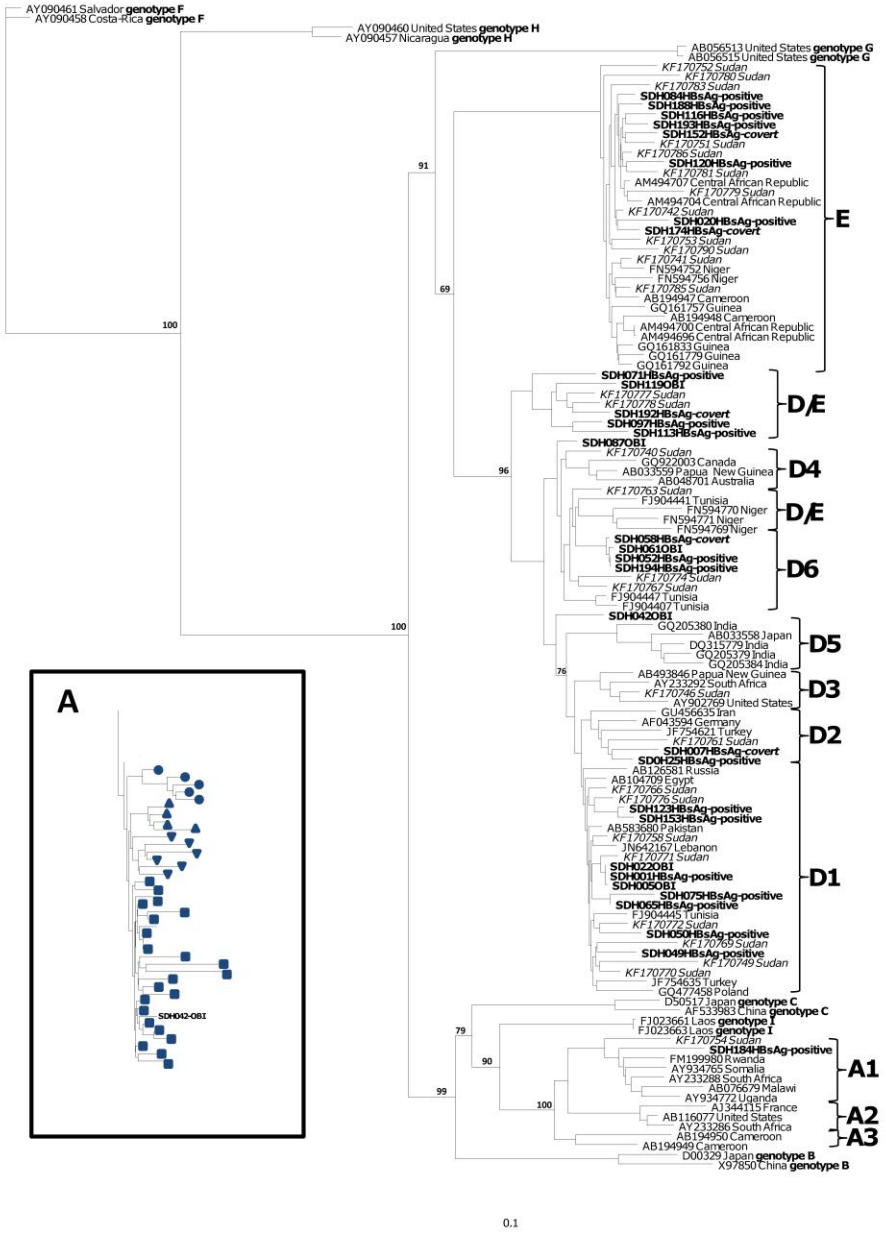

AY090461 Salvador genotype F
AY090458 Costa-Rica genotype F
AY090460 United States genotype H
AY090457 Nicaragua genotype H
AB056513 United States genotype G
AB056515 United States genotype G
KF170752 Sudan
KF170780 Sudan
KF170783 Sudan
SDH084HBsAg-positive
SDH188HBsAg-positive
SDH116HBsAg-positive
SDH193HBsAg-positive
SDH152HBsAg-covert
KF170751 Sudan
KF170786 Sudan
SDH120HBsAg-positive
KF170781 Sudan
AM494707 Central African Republic
KF170779 Sudan
AM494704 Central African Republic
KF170742 Sudan
SDH020HBsAg-positive
SDH174HBsAg-covert
KF170753 Sudan
KF170790 Sudan
KF170741 Sudan
FN594752 Niger
FN594756 Niger
KF170785 Sudan
AB194947 Cameroon
GQ161757 Guinea
AB194948 Cameroon
AM494700 Central African Republic
AM494696 Central African Republic
GQ161833 Guinea
GQ161779 Guinea
GQ161792 Guinea
E
SDH071HBsAg-positive
SDH119OBI
KF170777 Sudan
KF170778 Sudan
SDH192HBsAg-covert
SDH097HBsAg-positive
SDH113HBsAg-positive
D/E
SDH087OBI
KF170740 Sudan
GQ922003 Canada
AB033559 Papua New Guinea
AB048701 Australia
D4
KF170763 Sudan
FJ904441 Tunisia
FN594770 Niger
FN594771 Niger
FN594769 Niger
D/E
SDH058HBsAg-covert
SDH061OBI
SDH052HBsAg-positive
SDH194HBsAg-positive
KF170774 Sudan
KF170767 Sudan
FJ904447 Tunisia
FJ904407 Tunisia
D6
SDH042OBI
GQ205380 India
AB033558 Japan
DQ315779 India
GQ205379 India
GQ205384 India
D5
AB493846 Papua New Guinea
AY233292 South Africa
KF170746 Sudan
AY902769 United States
D3
GU456635 Iran
AF043594 Germany
JF754621 Turkey
KF170761 Sudan
SDH007HBsAg-covert
SDH025HBsAg-positive
D2
AB126581 Russia
AB104709 Egypt
KF170766 Sudan
KF170776 Sudan
SDH123HBsAg-positive
SDH153HBsAg-positive
AB583680 Pakistan
KF170758 Sudan
JN642167 Lebanon
KF170771 Sudan
SDH022OBI
SDH001HBsAg-positive
SDH005OBI
SDH075HBsAg-positive
SDH065HBsAg-positive
FJ904445 Tunisia
KF170772 Sudan
SDH050HBsAg-positive
KF170769 Sudan
SDH049HBsAg-positive
KF170749 Sudan
KF170770 Sudan
JF754635 Turkey
GQ477458 Poland
D1
D50517 Japan genotype C
AF533983 China genotype C
FJ023661 Laos genotype I
FJ023663 Laos genotype I
KF170754 Sudan
SDH184HBsAg-positive
FM199980 Rwanda
AY934765 Somalia
AY233288 South Africa
AB076679 Malawi
AY934772 Uganda
A1
AJ344115 France
AB116077 United States
AY233286 South Africa
A2
AB194950 Cameroon
AB194949 Cameroon
A3
D00329 Japan genotype B
X97850 China genotype B
0.1
A
SDH042-OBI

Figure 2

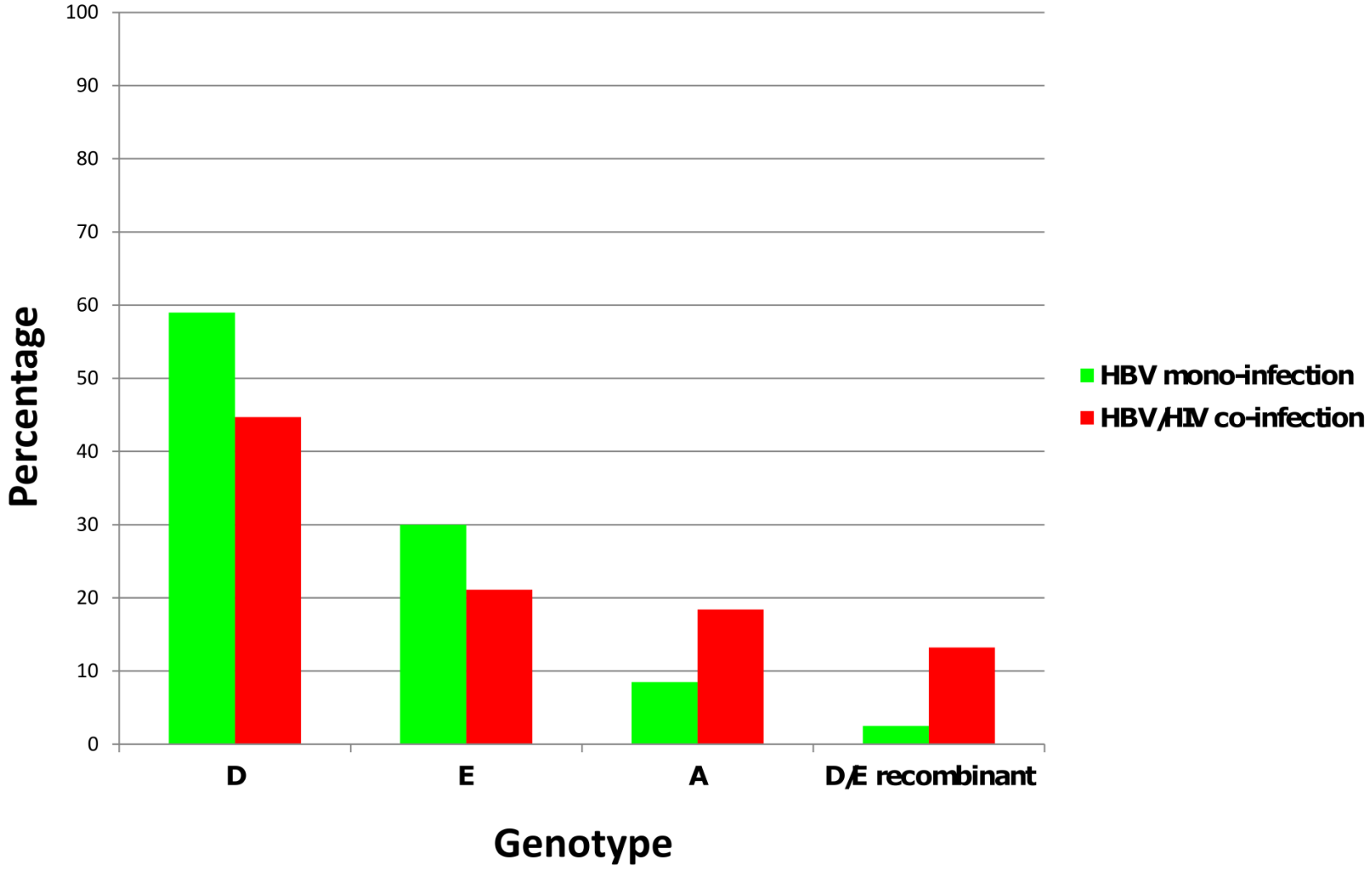
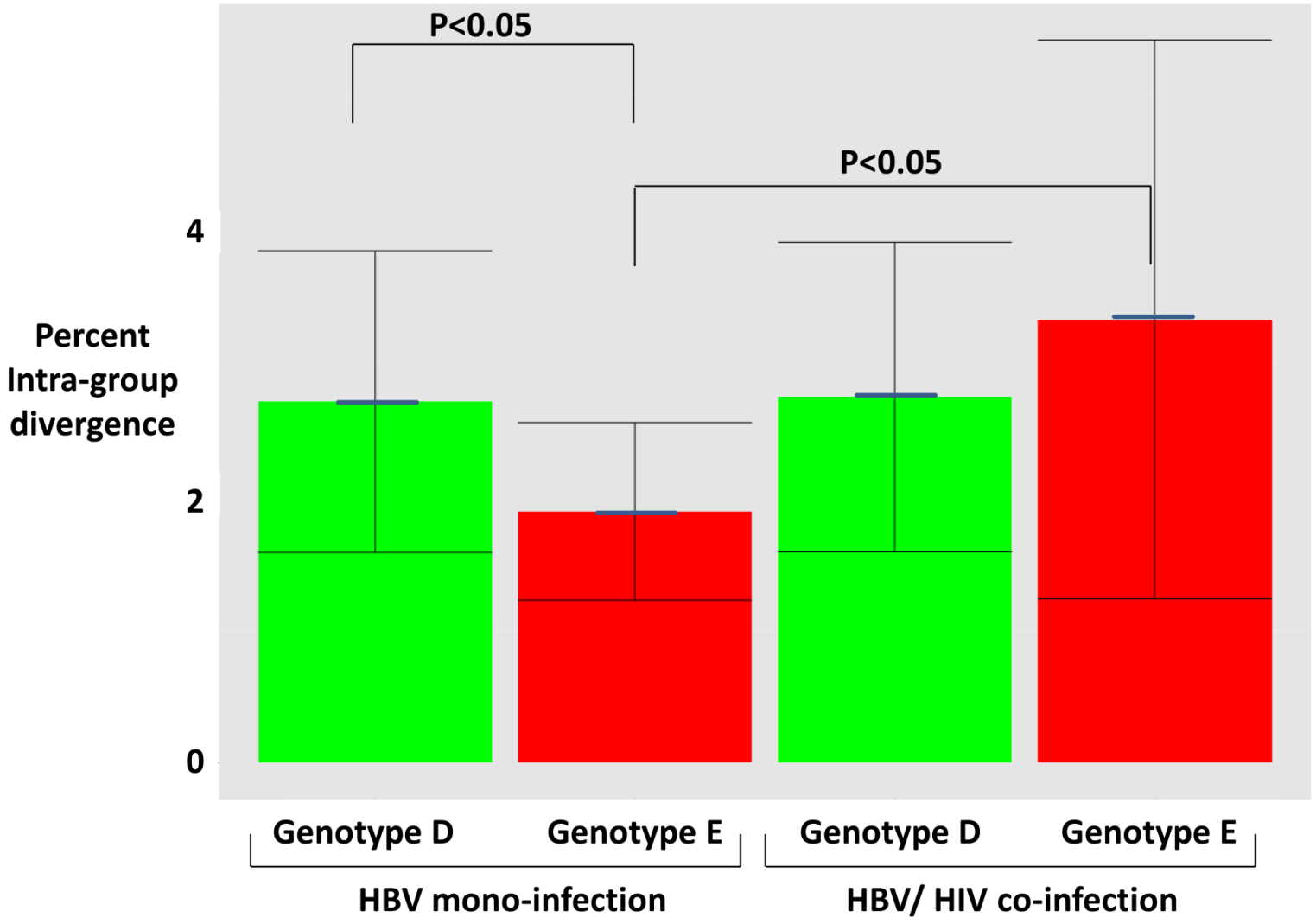

100
90
80
70
60
50
40
30
20
10
0
Percentage
D
E
A
D/E recombinant
Genotype
HBV mono-infection
HBV/HIV co-infection

Figure 3


P<0.05
P<0.05
4
2
0
Percent
Intra-group
divergence
Genotype D
Genotype E
Genotype D
Genotype E
HBV mono-infection
HBV/ HIV co-infection

Chapter 5: Paper IV

Analysis of Ultra-deep Pyrosequencing and Cloning Based Sequencing of the Basic Core Promoter/Precore/Core Region of Hepatitis B Virus Using Newly Developed Bioinformatics Tools

YOUSIF, M, Bell, T G, MUDAWI, H., GLEBE, D. & KRAMVIS, A (2014)

PLOSOne

Journal *PLOS ONE* is a peer-reviewed, ISI-accredited, open-access journal, with an impact factor of 4.04.

To whom it may concern

As joint first author of the paper described below, I hereby grant my permission for Mr Mukhlid Yousif to include the paper as a chapter in his PhD thesis. Details of the contributions made by each author are detailed within the paper.

- **YOUSIF, M., BELL, T. G.**, MUDAWI, H., GLEBE, D. and KRAMVIS, A. Analysis of Ultra-deep Pyrosequencing and Cloning Based Sequencing of the Basic Core Promoter/Precore/Core Region of Hepatitis B Virus Using Newly Developed Bioinformatics Tools. *Submitted* to PlosOne, 2013.

Trevor Bell, PhD
Post-Doctoral Research Fellow
University of the Witwatersrand

Anna Kramvis, PhD
Research Professor, Supervisor
University of the Witwatersrand

OPEN ACCESS Freely available online

PLOS ONE

Analysis of Ultra-Deep Pyrosequencing and Cloning Based Sequencing of the Basic Core Promoter/Precore/Core Region of Hepatitis B Virus Using Newly Developed Bioinformatics Tools

Mukhlid Yousif[1☯], Trevor G. Bell[1☯], Hatim Mudawi[2], Dieter Glebe[3], Anna Kramvis[1]*

1 Hepatitis Virus Diversity Research Programme, Department of Internal Medicine, University of the Witwatersrand, Johannesburg, Gauteng, South Africa, **2** Department of Medicine, Faculty of Medicine, University of Khartoum, Khartoum, Khartoum State, Sudan, **3** Institute of Medical Virology, National Reference Centre of Hepatitis B and D, Justus, Liebig-University of Giessen, Giessen, Hesse, Germany

Abstract

Aims: The aims of this study were to develop bioinformatics tools to explore ultra-deep pyrosequencing (UDPS) data, to test these tools, and to use them to determine the optimum error threshold, and to compare results from UDPS and cloning based sequencing (CBS).

Methods: Four serum samples, infected with either genotype D or E, from HBeAg-positive and HBeAg-negative patients were randomly selected. UDPS and CBS were used to sequence the basic core promoter/precore region of HBV. Two online bioinformatics tools, the ''Deep Threshold Tool'' and the ''Rosetta Tool'' (http://hvdr.bioinf.wits.ac.za/tools/), were built to test and analyze the generated data.

Results: A total of 10952 reads were generated by UDPS on the 454 GS Junior platform. In the four samples, substitutions, detected at 0.5% threshold or above, were identified at 39 unique positions, 25 of which were non-synonymous mutations. Sample #2 (HBeAg-negative, genotype D) had substitutions in 26 positions, followed by sample #1 (HBeAg-negative, genotype E) in 12 positions, sample #3 (HBeAg-positive, genotype D) in 7 positions and sample #4 (HBeAg-positive, genotype E) in only four positions. The ratio of nucleotide substitutions between isolates from HBeAg-negative and HBeAg-positive patients was 3.5:1. Compared to genotype E isolates, genotype D isolates showed greater variation in the X, basic core promoter/precore and core regions. Only 18 of the 39 positions identified by UDPS were detected by CBS, which detected 14 of the 25 non-synonymous mutations detected by UDPS.

Conclusion: UDPS data should be approached with caution. Appropriate curation of read data is required prior to analysis, in order to clean the data and eliminate artefacts. CBS detected fewer than 50% of the substitutions detected by UDPS. Furthermore it is important that the appropriate consensus (reference) sequence is used in order to identify variants correctly.

Citation: Yousif M, Bell TG, Mudawi H, Glebe D, Kramvis A (2014) Analysis of Ultra-Deep Pyrosequencing and Cloning Based Sequencing of the Basic Core Promoter/Precore/Core Region of Hepatitis B Virus Using Newly Developed Bioinformatics Tools. PLoS ONE 9(4): e95377. doi:10.1371/journal.pone.0095377

Editor: John E. Tavis, Saint Louis University, United States of America

Received December 21, 2013; **Accepted** March 26, 2014; **Published** April 16, 2014

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

Funding: The authors acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) of the special African-German collaborative programme, "Africa Initiative", (Grant GL595/3-1 to AK, HM and DG), Faculty of Health Sciences, University of the Witwatersrand. TB received post-doctoral funding from the University of the Witwatersrand (SPARC) and the National Research Foundation of South Africa (GUN #86215). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Anna Kramvis is a PLOS ONE Editorial Board member. This does not alter the author's adherence to PLOS ONE Editorial policies and criteria.

* E-mail: Anna.Kramvis@wits.ac.za

☯ These authors contributed equally to this work.

Introduction

The continued improvement of DNA sequencing technologies has led to the development of next generation sequencing (NGS) methods, including ultra-deep pyrosequencing (UDPS), which are capable of sequencing many thousands of nucleotides, quickly and at a low cost per nucleotide. These technologies have overcome the disadvantages of the traditional dye-terminating DNA sequencing technology developed by Frederick Sanger [1]. These disadvantages include the relatively high cost per nucleotide, in terms of money and time, and the fact that Sanger sequencing is only capable of detecting sequence variants, which are present in 20% or more of a quasispecies population [2,3]. Moreover, NGS methods also overcome several of the drawbacks of cloning based sequencing (CBS), such as the time, money and expertise required to prepare samples, especially when a large number of clones is required [4]. NGS methods are used primarily for *de novo* or ''shot-gun'' sequencing of new or known genomes. This produces a very

large number of short reads, which are then assembled to produce a complete sequence. Several algorithms and tools exist to process these short reads [5]. In addition to producing short reads, the pyrosequencing platform can be used for amplicon re-sequencing (UDPS). These longer reads are typically an amplicon covering a genomic region of interest. At present, the GS Titanium UDPS chemistry produces reads of approximately 400 bases in length.

Few bioinformatic tools, which are affordable and accessible to resource-constrained environments, are currently available to assist with the processing and analysis of amplicon re-sequencing data. The Roche AVA software (http://www.454.com/products/analysis-software/#amplicon-tabbing), although free of charge, can only be installed on a computer running a particular GNU/Linux distribution, and a number of commercial software packages cost several thousand US dollars for a single license. Alignment and visualization tools, which are used routinely for smaller datasets, are not suitable for datasets containing hundreds or thousands of reads. Additionally, many of these software solutions require a level of technical expertise, which many biological researchers may not possess.

Pyrosequencing is an error-prone technique [6]. Distinguishing between a true biological variant and an error (artefact) is a vital step in analysing pyrosequencing data. Although a number of studies discuss error correction in pyrosequencing data [6,7], there is currently no consensus regarding the error threshold, which should be applied. Knowledge of well-characterized regions of a genome is important in order to develop tools to examine pyrosequencing data and to distinguish between artefacts and true variations.

Hepatitis B virus (HBV) displays remarkable sequence heterogeneity, with 9 genotypes (named A to I) currently recognized [8,9]. The precore/core (PC/C) open reading frame (ORF) of HBV encodes for both the hepatitis B e antigen (HBeAg) and the core protein (HBcAg). This region is preceded by the basic core promoter (BCP) region, which controls transcription of both the PC/C mRNA and the pregenomic RNA (pgRNA) during the replication cycle [10]. The BCP/PC ORF overlaps the X ORF. HBeAg is a soluble, non-particulate protein that is secreted in the serum or expressed on the surface of the hepatocyte [11,12]. Conventionally, HBeAg expression is an indicator of active HBV infection and on-going viral replication [12]. However, HBeAg expression may be reduced or completely suppressed by various viral mutations, even in the presence of viral replication. Mutations in two regions may affect HBeAg expression: precore mutations (for example, G1896A) [13] and BCP mutations (for example, A1762T/G1764A) [14]. The viral capsid is composed of HBcAg [15]. Mutations may occur more frequently in N-terminal or central region of the core protein, which does not overlap other reading frames [16].

Using a segment of this well-characterized BCP/PC/C region of HBV as a model, the objectives of this study were to:

- develop bioinformatics tools to explore UDPS data,
- test and use them to determine the optimum error threshold, and
- compare results between UDPS and CBS using HBeAg-positive and –negative sera infected, with either genotype D or E.

Materials and Methods

Sample Selection

Written informed consent was obtained from all participants and the consent was approved by the Sudanese Ministry of Health, who gave permission for the sera to be used for research purposes. The Human Ethics Committees of the University of the Witwatersrand and the University of Khartoum approved the study. Four serum samples were selected from our previous study on HBV from monoinfected individuals, where the HBV genotype was determined using phylogenetic analysis [17]. Sample #1 was HBeAg-negative and infected with genotype E of HBV (GenBank, KF170783), sample #2 was HBeAg-negative, genotype D (KF170739), sample #3 was HBeAg-positive, genotype D (KF170740) and sample #4 was HBeAg-positive, genotype E (KF170788).

Wet Laboratory Work

Ultra-Deep Pyrosequencing (UDPS). A region of the HBV genome (1653–1959 from *Eco*R1 restriction site) was amplified using a slight modification of a previously described method [18]. Primers 1606 (+) and 1974 (−) were used for the first round PCR, and 1653 (+) and 1959 (−) for the second round PCR. The first round PCR was followed by gel-purification using Zymoclean Gel DNA Recovery Kit (Zymo Research Corp, Irvine, CA, USA). For the second round PCR, modified primers, which were ligated to adaptors and tags, were used (Table 1). Following second round PCR, the amplicons were gel-purified and subjected to UDPS in the forward direction on the Roche 454 GS Junior platform (454 Life Sciences, Roche Company, Switzerland), which provided reads covering the region of interest (coordinates 1653–1959). The UDPS sequencing data has been submitted to the GenBank SRA database, as BioProject accession: PRJNA239442 and the following are the BioSample accessions: SAMN02664575, SAMN02664576, SAMN02664577, SAMN02664578.

Cloning Based Sequencing (CBS). After nested PCR, the 307 nucleotide amplicon (1653–1959 from *EcoR*1 site) was gel-purified and cloned into pTZ57R/T vector (55 ng/μl) using Instaclone PCR Cloning Kit (Fermentas, Waltham, MA, USA), and transformed into TOP10 *Escherichia coli* (Invitrogen, Carlsbad, CA, USA). The transformants were grown on Ampicillin plates. Positive clones were identified by restriction fragment length polymorphism (RFLP) assay. At least 20 clones per sample were sequenced by direct sequencing, using a BigDye Terminator v3.0 Cycle Sequencing Ready Reaction Kit (Applied Biosystems., Foster City, USA) on an ABI 3130XL Genetic Analyzer (Applied Biosystems). The sequencing primer used was M13 forward (5′-GTAAAACGACGGCCAGT-3′). A phylogenetic tree was generated as described previously [17]. Clone sequences have been deposited in GenBank: Genotype D: KJ496256-KJ496297, Genotype E: KJ496206-KJ496255.

Dry Laboratory Work

Data pre-processing. UDPS data for three sequencing runs, for each of the four samples, was processed and analyzed as shown in the flow diagram (Figure 1). The data from each run, for each sample, was processed individually. Separate binary standard flowgram format (SFF) files were opened in the R statistical programming language [19], using the "raw" clip-mode parameter (which does not perform any clipping or trimming) of the "rSFFreader" library [20]. Sequence data were searched for the forward and reverse primer sequences and the adaptor sequence for verification. Sequence lengths in each file were plotted and examined statistically (*data not shown*).

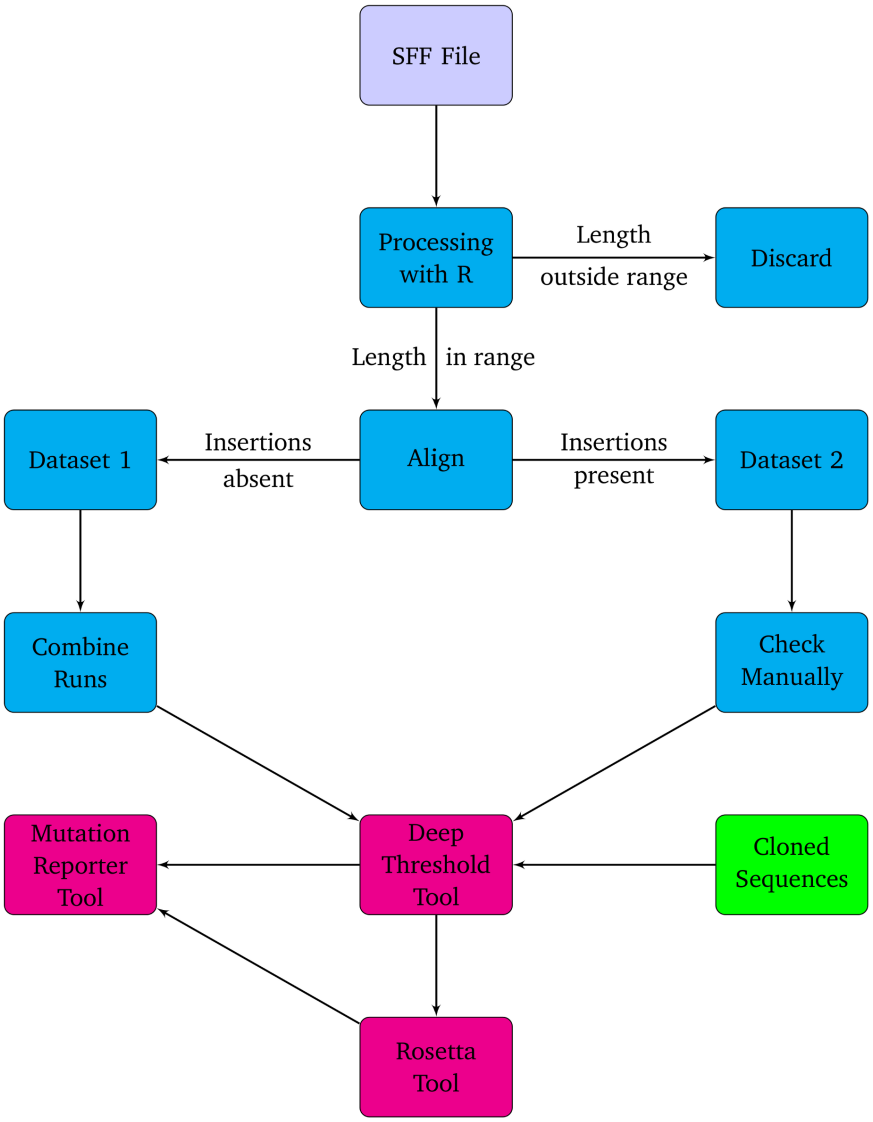


Figure 1. Flow diagram of the analysis procedure used for the UDPS data.
doi:10.1371/journal.pone.0095377.g001

Table 1. Sequences of HBV-specific primers, tags and adaptors used for ultra-deep pyrosequencing.

	Position from *Eco*RI site	HBV specific target sequence	Tag sequences[¶]	Adaptor sequences[€]
PCR 1	1606 (+) [1606–1625]	5′-GCATGGAGACCACCGTGAAC-3′	No tag	No adaptor
	1974 (−) [1974–1955]	5′GGAAAGAAGTCAGAAGGCAA-3′	No tag	No adaptor
PCR 2	1653 (+) [1653–1672]	5′-CGTATCGCCTCCCTCGCGCCATCAG-3′	ACACGACGACT[1]	CAT AAG AGG ACT CTT GGA CT
			ACACGTAGTAT[2]	
			ACACTACTCGT[3]	
			ACGACACGTAT[4]	
	1959 (−) [1959–1940]	5′-CTATGCGCCTTGCCAGCCCGCTCAG-3′	No tag	GGC AAA AAC GAG AGT AAC TC

[¶]Short specific sequences used to label the different samples: [1:] sample #1; [2:] sample #2; [3:] sample #3; [4:] sample #4.
[€]Short specific sequence used to ligated onto the ends of the fragments. These adaptors provide priming sequences for both amplification and sequencing of the sample-library fragments.
doi:10.1371/journal.pone.0095377.t001



Figure 2. The input pages of the bioinformatics tools (A) ''Deep Threshold Tool''. The first field specifies the input FASTA file. Fields are available for the user to specify the nucleotide offset mapping of the first position in the input file, the number of nucleotides (length) to process, the starting and ending probabilities of error to examine, and the probability of error increment (step) to use. (**B**) ''Rosetta Tool''. The first field specifies the input FASTA file. Fields are available for the user to specify the nucleotide offset mapping of the first position in the input file, the position of the first in-frame nucleotide of the coding region of interest, the last in-frame nucleotide of the coding region of interest, the amino acid offset of the first amino acid in the coding region of interest, and the probability of error to use.
doi:10.1371/journal.pone.0095377.g002

The distribution of all sequence lengths was examined and a length range was selected, which excluded reads with very low counts. Several Linux command-line BASH scripts and Python programming language scripts (*available on request*) were written to include only reads within a specified length range (between 330 to 360 nucleotides) for further processing. A genotype D reference sequence (GU456684) was then added to each dataset, and the file was aligned with the Muscle program [21]. Each alignment was then processed by a Python script, which scanned the reference sequence in the alignment and removed any reads from the alignment with an insertion (a residue aligned with a gap in the reference sequence). In the remaining alignment (excluding reads with insertions), positions (columns), containing only gaps, were collapsed and this alignment was ''Dataset 1''. The repeated runs for all ''Dataset 1'' sequences for each sample were then combined into one dataset, the final ''Dataset 1''. The file containing reads with insertions was ''Dataset 2'' for each run and these were processed individually because of variable read lengths, as a result of insertions at different positions in the reads.

Development of deep threshold tool. For pyrosequencing data of human immunodeficiency virus (HIV), a probability of error, ranging from 0.5% to 1%, has been used [6]. In the present study, using HBV data, a web-based tool (the ''Deep Threshold Tool'') (http://hvdr.bioinf.wits.ac.za/tools/) was developed to examine the number of errors in each position (column) in an alignment, depending on the probability of error value. In order to examine the number of errors, the tool requires an input alignment in FASTA format, the lower and upper bounds of the probability of error, and an increment value (Figure 2A). A nucleotide mapping offset can be specified, so that the resulting output coordinates reflect the correct position of the sequence in the entire genome. Potentially untidy ends of reads (such as the reverse primer region) can be excluded from the analysis by specifying a length shorter than the sequence length.

Statistical calculation of the threshold. A nucleotide was considered an ''error'' if its frequency in a column in the alignment was less than the threshold, which was determined as follows. An expected frequency of $\boldsymbol{E = probability\ of\ error \times read\ depth\ (R)}$ was used. A Pearson's χ^2 test statistic was calculated as follows:

$$M = \frac{(O-E)^2}{E} + \frac{((R-O)-(R-E))^2}{R-E}$$

with ***O*** being the observed value, starting at 1. If ***M*** was less than the χ^2 distribution (with $\alpha = 0.05$ and one degree of freedom), then ***O*** was incremented by a value of one and the test was repeated. The value for O at which the χ^2 distribution was exceeded, was considered the threshold value (count). This threshold was calculated for each position in the alignment. Any nucleotide with a frequency below this threshold was considered an error or artefact.

Development of rosetta tool. Amino acid data were examined using the newly-developed "Rosetta Tool". This tool requires the same input file as the "Deep Threshold Tool ". It also requires a nucleotide offset mapping and the start and end positions of a protein region. This does not have to include the position of the start or stop codon; any region of a protein can be processed, as long as the number of nucleotides specified by the range is a multiple of three. The probability of error at which the data must be analyzed is also required (Figure 2B).

Results

A total of 10952 reads were generated on the 454 GS Junior platform for the three runs for all four samples. Of these, 9738 reads (88.9%) were included in the study (2002, 3049, 1955 and 2732 reads for samples 1, 2, 3 and 4, respectively) and 1214 reads (11.1%), which were considered either too short or too long, were excluded. These 9738 reads were split into Dataset 1 (8967 reads, 92.1%) and Dataset 2 (771 reads, 7.9%) (Figure 1). Ninety-two clones were generated for all four samples: 23 clones for sample #1, 22 for sample #2, 20 for sample #3 and 27 for sample #4.

Deep Threshold Tool Output

The output page generated by the Deep threshold Tool includes a table for each increment of the probability of error (Figure 3), which shows the distribution of nucleotides at all columns at which at least one base can be considered an "error". Because a nucleotide was considered an "error", if its frequency in a column in the alignment was less than the threshold, any variation above the threshold was considered a legitimate variant for that probability of error (Figure 4). Figure 5 summarizes the results graphically. This summary table was consulted and the lowest probability of error, at which established, well-characterized variants are still detected, was selected. In the present study, the

Input file	Sample2.aa
Length	287
Probability of Error	0.0100
Expected	29
Threshold	40
Number of Interesting Columns	10
Interesting columns	[1727, 1739, 1742, 1748, 1753, 1772, 1775, 1909, 1912, 1913]

base/locus	1727	1739	1742	1748	1753	1772	1775	1909	1912	1913
A	63	16	13	41	6	2548	282	1	0	0
C	0	25	0	0	15	13	0	248	395	2698
G	2787	2746	2792	2805	40	1	2565	0	0	151
T	0	63	0	3	2789	274	1	2601	2455	1
M	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0
-	0	0	45	1	0	14	2	0	0	0

Figure 3. An example section of the output from the ''Deep Threshold Tool'', showing the two tables of output provided for each probability of error examined. The "expected" and "threshold" counts are shown in the top table, as well as the number of interesting columns (those columns containing at least one mutation at above-threshold frequency), and a list of the interesting columns. The bottom table provides detailed output, showing the number of each residue occurring in each interesting column.
doi:10.1371/journal.pone.0095377.g003

Probability of Error	Expected	Threshold	IC (#)	Interesting Columns
0.0050 (0.5%)	14	22	25	[1678, 1680, 1706, 1724, 1725, 1727, 1728, 1730, 1736, 1739, 1741, 1742, 1745, 1748, 1751, 1753, 1761, 1772, 1773, 1775, 1842, 1896, 1909, 1912, 1913]
0.0060 (0.6%)	17	26	21	[1678, 1680, 1724, 1725, 1727, 1728, 1736, 1739, 1741, 1742, 1745, 1748, 1753, 1761, 1772, 1775, 1842, 1896, 1909, 1912, 1913]
0.0070 (0.7%)	20	29	17	[1678, 1724, 1725, 1727, 1728, 1739, 1741, 1742, 1745, 1748, 1753, 1772, 1775, 1842, 1909, 1912, 1913]
0.0080 (0.8%)	23	33	15	[1678, 1724, 1727, 1728, 1739, 1741, 1742, 1748, 1753, 1772, 1775, 1842, 1909, 1912, 1913]
0.0090 (0.9%)	26	37	11	[1724, 1727, 1739, 1742, 1748, 1753, 1772, 1775, 1909, 1912, 1913]
0.0100 (1.0%)	29	40	10	[1727, 1739, 1742, 1748, 1753, 1772, 1775, 1909, 1912, 1913]

Figure 4. The first of two summary output tables provided by the ''Deep Threshold Tool''. For each probability of error in the range specified, the expected and threshold values are shown, the number of interesting columns (IC) and the list of these interesting columns. This table provides a summary of the output provided in the previous tables (Figure 3 *top*).
doi:10.1371/journal.pone.0095377.g004

lowest probability of error at which substitutions at positions 1753, 1773 and 1896 are still evident was 0.5%, and this was taken to be our probability of error value.

Rosetta Tool Output

Alignments generated from direct sequencing, UDPS or CBS can also be submitted to the Rosetta Tool. This would typically be done in order to make use of the nucleotide/amino acid alignment viewer component of the tool. The tool produces a number of output tables (Figures 6–8). Figure 6 is an alignment showing each codon followed by the amino acid. Amino acids have been colour-coded according six different categories: Aliphatic (Glycine, Alanine, Valine, Leucine and Isoleucine), Hydroxyl (Serine, Cysteine, Threonine and Methionine), Cyclic (Proline), Aromatic (Phenylalanine, Tyrosine and Tryptophan), Basic (Histidine, Lysine and Arginin) and Acidic (Aspartate, Glutamate, Asparagine and Glutamine). The display of nucleotides or amino acids can be toggled on or off for ease of reference. Figure 7 shows the distribution of each residue at each position at which at least one residue is considered an error. Such error residue counts are highlighted with a black background for reference. Figure 8 contains separate tables for each codon at which at least one residue is an ''error'', and shows the distribution of codons and amino acids at this position. Synonymous and non-synonymous mutations can be differentiated. Rows containing substitutions occurring below the threshold, ''error'' nucleotides are highlighted with a black background.

In order to analyze the data downstream, the Rosetta Tool produces a ''masked'' data file, which is generated by replacing all ''error'' residues in the nucleotide alignment, with an ''X'' character. This alignment is then be translated into amino acids, with an amino acid of ''X'' used whenever at least one ''X'' character per codon occurs. Both the nucleotide and amino acid masked files can be downloaded in FASTA format.

Using the selected probability of error of 0.5%, masked files were generated and the UDPS data were then analyzed using the two newly developed tools and the Mutation Reporter Tool [22].

Analysis of Pyrosequencing Reads

Each sample in Dataset 1 was then analysed using the newly developed ''Deep Threshold Tool'' and a probability of error of 0.5% was selected, because this was the lowest probability of error at which all three well characterized mutations (T1753G/C, T1773C and G1896A) were present. The resulting threshold (count) value will differ depending on the number of reads (depth) in each file, for a given probability of error. For each sample, output of the ''Deep Threshold Tool'' lists the loci detected at above threshold value and these were then analyzed using the Mutation Reporter Tool, with a reference motif being the corresponding consensus sequences for each genotype or sub-genotype. The distribution of substitutions at the nucleotide level in the BCP/PC/C region varied between samples, depending on the HBV genotype and HBeAg status (Figure 9). At 0.5% probability of error or above, substitutions were identified at 39 unique positions in the four samples:31 in the X region (1674 to 1838 from the *Eco*R1 site; 165 nucleotides), three in the PC region (1814 to 1900; 87 nucleotides) and five in the core region (1901 to 1939; 39 nucleotides) (Figure 9). Ten of the 39 positions were present in at least two samples.

Based on the fact that direct sequencing is capable of detecting substitutions occuring in ≥20%, of the quasispecies population substitutions were classified as high frequency (≥20%) and low frequency substitutions (<20%). High frequency substitutions were found at 11 positions and low frequency at 28 positions.

Probability of Error	1678	1680	1706	1724	1725	1727	1728	1730	1736	1739	1741	1742	1745	1748	1751	1753	1761	1772	1773	1775	1842	1896	1909	1912	1913
0.0100 (1.0%)						•				•		•		•		•		•		•			•	•	•
0.0090 (0.9%)				•		•				•		•		•		•		•		•			•	•	•
0.0080 (0.8%)	•			•		•	•			•	•	•		•		•		•		•	•		•	•	•
0.0070 (0.7%)	•			•	•	•	•			•	•	•	•	•		•		•		•	•		•	•	•
0.0060 (0.6%)	•	•		•	•	•	•		•	•	•	•	•	•		•	•	•		•	•	•	•	•	•
0.0050 (0.5%)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

Figure 5. The second of two summary output tables provided by the ''Deep Threshold Tool''. For each probability of error in the range specified (shown in reverse order in this table), a bullet is shown in the corresponding column of the table for each interesting column at which at least one mutation occurred at above-threshold frequency. This table can be consulted to determine the probability of error, which should be used on a given dataset. In this example, the well-characterized positions 1753, 1773 and 1896 are examined, and a probability of error of 0.005 selected, as this is the highest probability of error at which above-threshold mutations at the three positions are detected.
doi:10.1371/journal.pone.0095377.g005

Show/Hide Nucleotides | Show/Hide Amino Acids

ID	1814->	1	1817->	2	1820->	3	1823->	4	1826->	5	1829->	6
H6E7F4R01BUX0T	ATG	M	CAA	Q	CTT	L	TTT	F	CAC	H	CTC	L
H6E7F4R01AXWWO	ATG	M	CAA	Q	CTT	L	GTT	V	CAC	H	CTC	L
H6E7F4R01A9N1P	ATG	M	CAA	Q	CTT	L	TTT	F	CAC	H	CTC	L
H6E7F4R01CW0K1	ATG	M	CAA	Q	CTT	L	TTT	F	CAC	H	CTC	L
H6E7F4R01AKB00	ATG	M	CAA	Q	CTT	L	TTT	F	CAC	H	CTC	L
H6E7F4R01BNNCJ	ATG	M	CAA	Q	CTT	L	TTT	F	CAC	H	CTC	L
H6E7F4R01CBND9	ATG	M	CAA	Q	CTT	L	TTT	F	CAC	H	CTC	L
H6E7F4R01C3746	ATG	M	CAA	Q	CTT	L	TTT	F	CAC	H	CTC	L

Figure 6. The first output table of the ''Rosetta Tool'', showing codon (triplets), followed by single-letter translated amino acids, for each read in the input FASTA alignment. This alignment can be used to easily locate mutations of interest and to locate synonymous and non-synonymous mutations. The visibility of the nucleotide and/or amino acids columns can be toggled on or off, as required.
doi:10.1371/journal.pone.0095377.g006

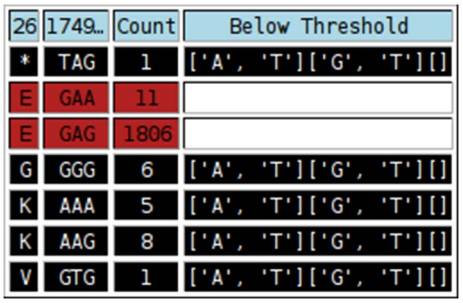
A consensus of genotype E was used to identify substitutions in genotype E (samples #1 and #4). The T1741C substitution was detected in both samples at a high frequency, regardless of the HBeAg status, while the following substitutions: A1757G, A1762T, G1764A, G1896A, G1937A/T and A1938C, were found at a high frequency in HBV from a HBeAg-negative patient (sample #1) (Figure 9). Substitutions A1735G, G1742A, A1747C, T1753C and T1909C were found at a low frequency in sample #1, and T1707C was found at a low frequency in sample #4 (Figure 9).

Similarly, when the genotype D sequences (samples #2 and #3) were compared to their corresponding consensus sequence, 1678T was found in sample #2 and 1678C in sample #3. The consensus of genotype D had 1678T. From phylogenetic analysis carried out in our previous study, HBV from sample #2 belongs to subgenotype D1 and from sample #3 to subgenotype D6 [17]. The consensus of subgenotype D1 has T at 1678 and that of

Number of reads: 2850
Threshold count: 167

AA	Codon	1814	1817	1820	1823	1826	1829	1832	1835	1838	1841	1844	1847	1850	1853
*	TAA		5												
*	TAG														
*	TGA												1		
-	-AG														
-	-GG														
-	-TC									1					
-	C-T			1											
-	CC-														
-	CT-								1						
-	T-G														
-	TC-														
-	TG-												1		
A	GCA													1	
A	GCC														
A	GCG														
A	GCT														
C	TGC							2824					14		5
C	TGT							7					2825		2830

Figure 7. The second output table of the ''Rosetta Tool'' (truncated), showing the frequency of each possible codon (triplet) and where it occurs in the alignment. Frequencies shown with a black background occurred at below-threshold levels and therefore can be disregarded.
doi:10.1371/journal.pone.0095377.g007



26	1749…	Count	Below Threshold
*	TAG	1	['A', 'T']['G', 'T'][]
E	GAA	11	
E	GAG	1806	
G	GGG	6	['A', 'T']['G', 'T'][]
K	AAA	5	['A', 'T']['G', 'T'][]
K	AAG	8	['A', 'T']['G', 'T'][]
V	GTG	1	['A', 'T']['G', 'T'][]

Figure 8. Examples of the final series of tables output by the "Rosetta Tool", showing details of the codons (triplets) and amino acids occurring at each position in the alignment. Cells with black backgrounds indicate where at least one nucleotide in the triplet occurred at below-threshold levels. These rows can be disregarded. The "Below Threshold" column lists the residues, for each position of the codon (indicated by the square brackets), which were below the. threshold.
doi:10.1371/journal.pone.0095377.g008

subgenotype D6 has C. Therefore, when sample #3 was compared to the consensus of subgenotype D6, only low frequency substitutions were detected (T1696C, G1733A, G1745A, G1748, G1751A, G1756A and T1765C) (Figure 9). When the reference sequence was changed from the D to D1, the mutation pattern of sample #2 (subgenotype D1), changed (Figures 9). Using either reference sequence D or D1, the following substitutions were detected with high frequency: A1727G, C1730A, A1761C, G1764A, A1775G and G1896A, whereas the frequency of 1773T and 1912T decreased when using D1 instead of D as the reference sequence (Figure 9). The following substitutions relative to D1, occurred in sample #2 at low frequency: T1678C, A1680C, C1706T, T1724C, A1725C, G1728A, G1736A, G1739C/T, T1741C, G1745A, G1748A, G1751, T1753G, A1772T, T1773C, T1842C, T1909C, T1912C and C1913G.

Summarizing the above, in the four samples substitutions were identified at 39 unique positions. Sample #2 (HBeAg-negative, genotype D) had substitutions in 26 positions, followed by sample #1 (HBeAg-negative, genotype E) in 12 positions, sample #3 (HBeAg-positive, genotype D) in 7 positions and sample #4 (HBeAg-positive, genotype E) in only four positions. The ratio of nucleotide substitutions between isolates from HBeAg-negative and HBeAg-positive patients was 3.5:1. Moreover, genotype D isolates showed greater variation in the X, PC and core regions, compared to genotype E isolates, with the two genotype D samples having 33 substitutions compared to the 16 detected in the genotype E samples.

The "Rosetta Tool", which was developed as part of this study, was used to analyze sequence data at the amino acid level. Substitutions identified at the nucleotide level were translated into amino acids and classified as synonymous or non-synonymous. Fourteen substitutions, 12 in the X region and 2 in the C region, were synonymous. Twenty-five, 19 in the X region, three each in the PC and C regions, were non-synonymous mutations. All non-synonymous mutations occurred within single, non-overlapping reading frames (1653 to 1814, and 1839 to 1939 from the *Eco*R1 restriction site), and the region between the start of the PC and the end of the X (1814 to 1838) was completely conserved in all ultradeep pyrosequences.

Most of the 21 insertions found in Dataset 2 occurred within homopolymeric regions and were therefore considered to be PCR or pyrosequencing artefacts [23].

Analysis of CBS and Comparison to UDPS

At least 20 clones were generated per sample. The BCP/PC region sequenced is relatively short and does not differentiate genotypes D and E following phylogetic analysis. Both identical and multiple clones were generated, with HBV from HBeAg-negative sera showing greater divergence (Figure 10).

CBS data was analyzed at the 39 loci, previously recognized by UDPS, using the Mutation Reporter Tool and a consensus sequence for each genotype/subgenotype as the reference sequence. In the four samples, substitutions at 18 of the 39 positions (46.2%) were detected by CBS (Table 2) (Figure 11). CBS detected all high frequency substitutions but only 25% (7/28) of the low frequency substitutions (Table 2). Moreover, the following nucleotide substitutions were detected in different samples by either UDPS or CBS, at position:

1. 1707 by CBS in sample #2 and by UDPS in sample #4,
2. 1775 by CBS in sample #1 and by UDPS in sample #2 and
3. 1842 by CBS in sample #3 and by UDPS in sample #2.

For the four samples, CBS detected only 14 of the 25 non-synonymous mutations, detected by UDPS (58.3%).

Discussion

The aims of this study were to build bioinformatics tools to assist in determining the threshold at which pyrosequencing data should be analyzed, and to compare quasispecies distributions obtained using UDPS and CBS, and compare results between UDPS and CBS using HBeAg-positive and –negative sera infected, with either genotype D or E.

Direct (Sanger) sequencing produces a single "read" for each sample. After curating the sequence and resolving ambiguous bases, the sequence is ready for further downstream processing. Whilst UDPS, which generates several thousand reads per sample, is a powerful technology, the analysis of the read data before downstream processing is critical. The depth of coverage provided by UDPS is also one of its shortcomings, as the data needs to be carefully curated for errors (artefacts), which may have been introduced by the PCR amplification and/or the sequencing process [2,23]. The increased sensitivity of the platform to detect thousands of reads also means that it may generate such artefacts.

A probability of error of between 0.5% and 1% for UDPS has been used previously for HIV samples [6]. Subsequent studies on HBV sequence data have either used the same probability of error, or have not reported details of this component of the analysis [2,3,24]. The probability of error, which is used, will influence the downstream detection of variants. As such, selecting an appropriate probability of error is an essential step in the analysis. In response to the lack of consensus in selecting a probability of error and determining a threshold, we developed an online bioinformatics tool to explore this aspect of the analysis. The "Deep Threshold Tool" provides the researcher with detailed output of variation at different probabilities of error. The analysis is objective and repeatable, and the selected probability of error can be reported and defended. Data for a project can be processed by the tool, so that a probability of error can be selected for that specific project, organism or assay. Using a fixed, predetermined probability of error for the UDPS platform as a whole is overly-broad and too general, as it is not possible to indicate how a particular probability of error would be applicable to a different organism, genomic region or investigation. Using the "Deep Threshold Tool" developed in the present study, a probability of

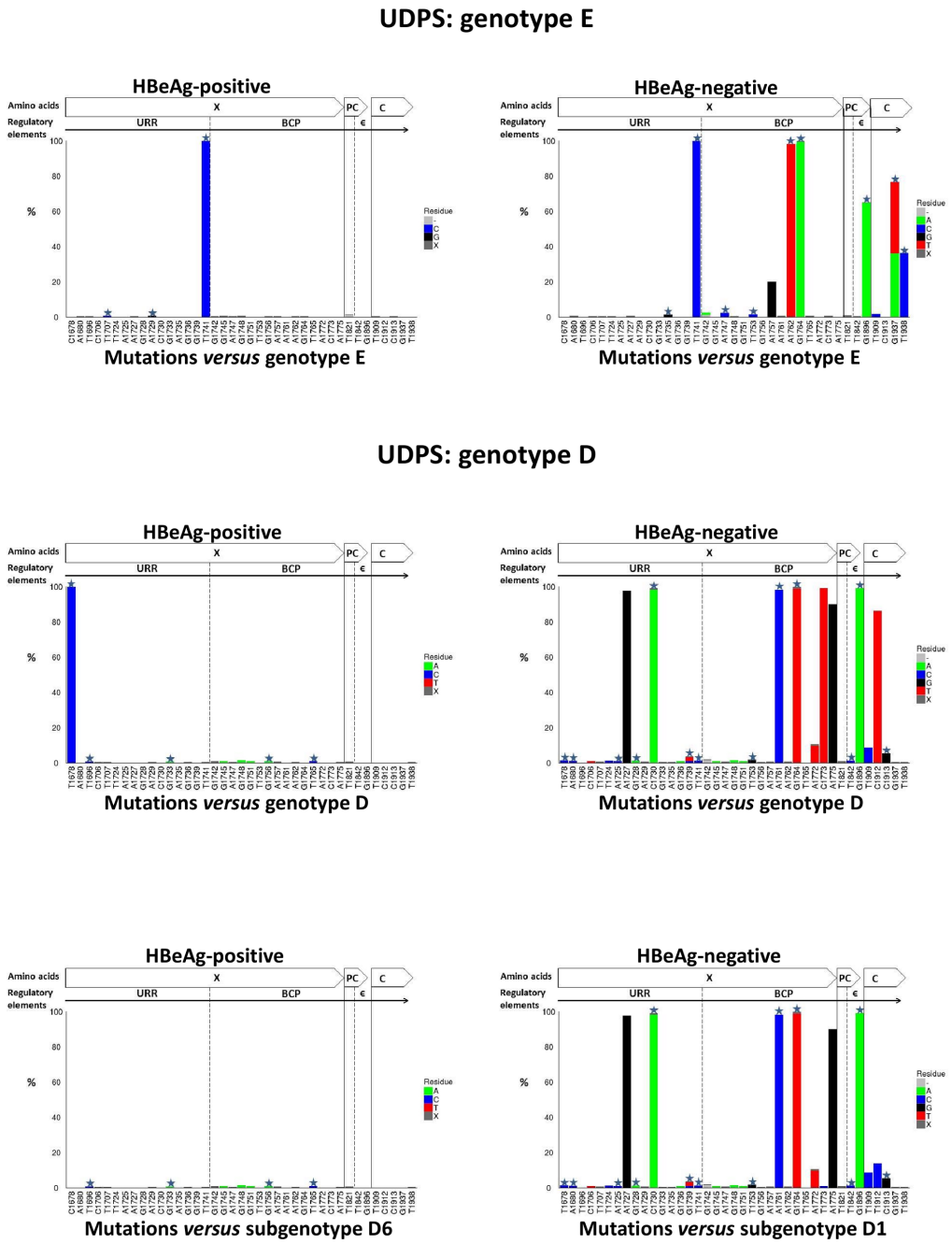


Figure 9. Graphs showing mutation distribution of the UDPS data at the nucleotide level using either genotype E or D consensus sequence as the reference. A star indicates a non-synonymous mutation. The graphs were built using the Mutation Reporter Tool [22].
doi:10.1371/journal.pone.0095377.g009

error of 0.5% was selected for the BCP/PC/C region of HBV, which agrees with previous reports for HIV [6].

The output must be interpreted in light of existing biological knowledge of the variation known to occur in the sequenced region. The tool is objective and outputs results for different probabilities of error "blindly". There is no "right answer" or absolute correct threshold, as we cannot possess complete knowledge of all the stochastic processes, from the sample to the PCR to the sequencing platform to the sequence results. Variation may be introduced at the various PCR stages, rather than by the sequencing hardware itself [23]. What we can do, however, is to interrogate these data at different probabilities of error, and make an informed decision on which value to select. It is important that the method used to process and curate the UDPS data, as well as any numerical values used (such as probability of error or threshold), be reported in all UDPS studies. Failure to provide this level of detail makes it difficult to accurately assess and relate any results reported.

The emergence of G1896A mutation in the PC region is known to be associated with HBeAg seroconverion [13]. The presence of wild-type (G) at 1896 in sample #1 and sample #2,which were isolated from HBeAg-negative patients, confirms the ability of UDPS to detect minor populations, which may not be detected by Sanger sequencing [25,26]. Similar results have been reported in

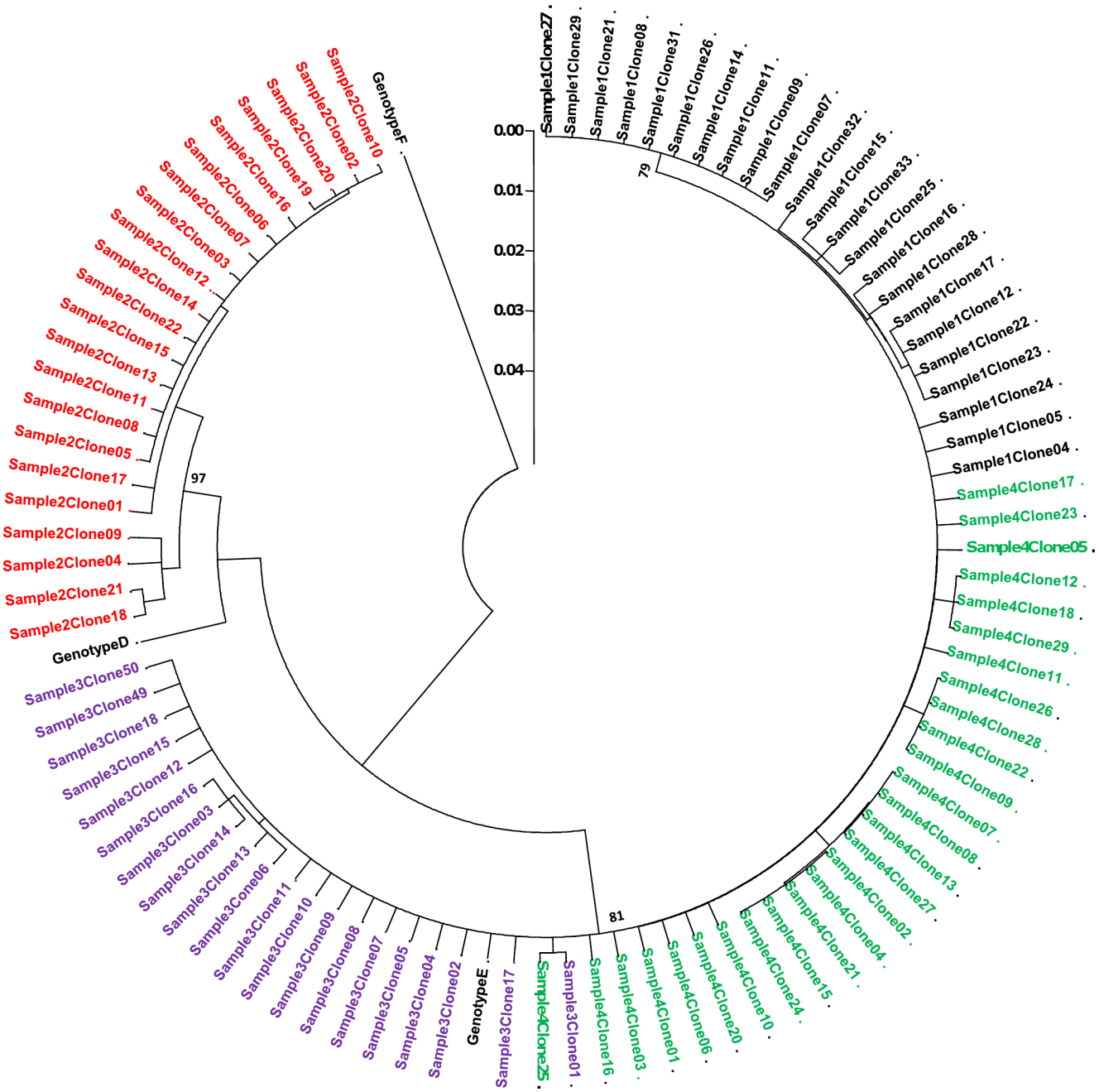


Figure 10. A rooted phylogenetic tree of 92 cloned BCP/PC sequences (position 1653 to 1939 from *EcoR*I site) from four serum samples. Sample #1 was HBeAg-negative and infected with genotype E of HBV, sample #2 was HBeAg-negative, genotype D, sample #3 was HBeAg-positive, genotype D and sample #4 was HBeAg-positive, genotype E. Bootstrap statistical analysis was performed using 1000 datasets, indicated as percentages on the nodes. The letters, D and E, represent the genotypes.
doi:10.1371/journal.pone.0095377.g010

more recent HBV studies. The HBV population from HBeAg-positive sera showed a high percentage of stop codon mutations in the precore region, while isolates from HBeAg-negative carriers had a low percentage of wild-type residues at codon 28 [24].

Although the selection of genotype D samples was random, we later discovered that sample #3 belonged to subgenotype D6, while sample #2 belonged to subgenotype D1. As illustrated in Figures 9 and 10, knowledge of the genotype and subgenotype of HBV is important when determining the presence of mutations. Depending on the reference or consensus sequence used, the variant at a particular position, may either represent the signature of a particular subgenotype or be a legitimate mutation. Therefore, where possible, a consensus sequence of the genotype or subgenotype should be used, to ensure that variants are examined in the appropriate context.

Six mutations (A1757G, A1762T, G1764A G1896A, G1937A/T and A1938C) were found in high frequency (>20%) in sample #1, genotype E isolated from a HBeAg-negative patient. The G1896A mutation is known to create the stop codon at amino acid 28 and to abrogate HBeAg expression [13], while the double

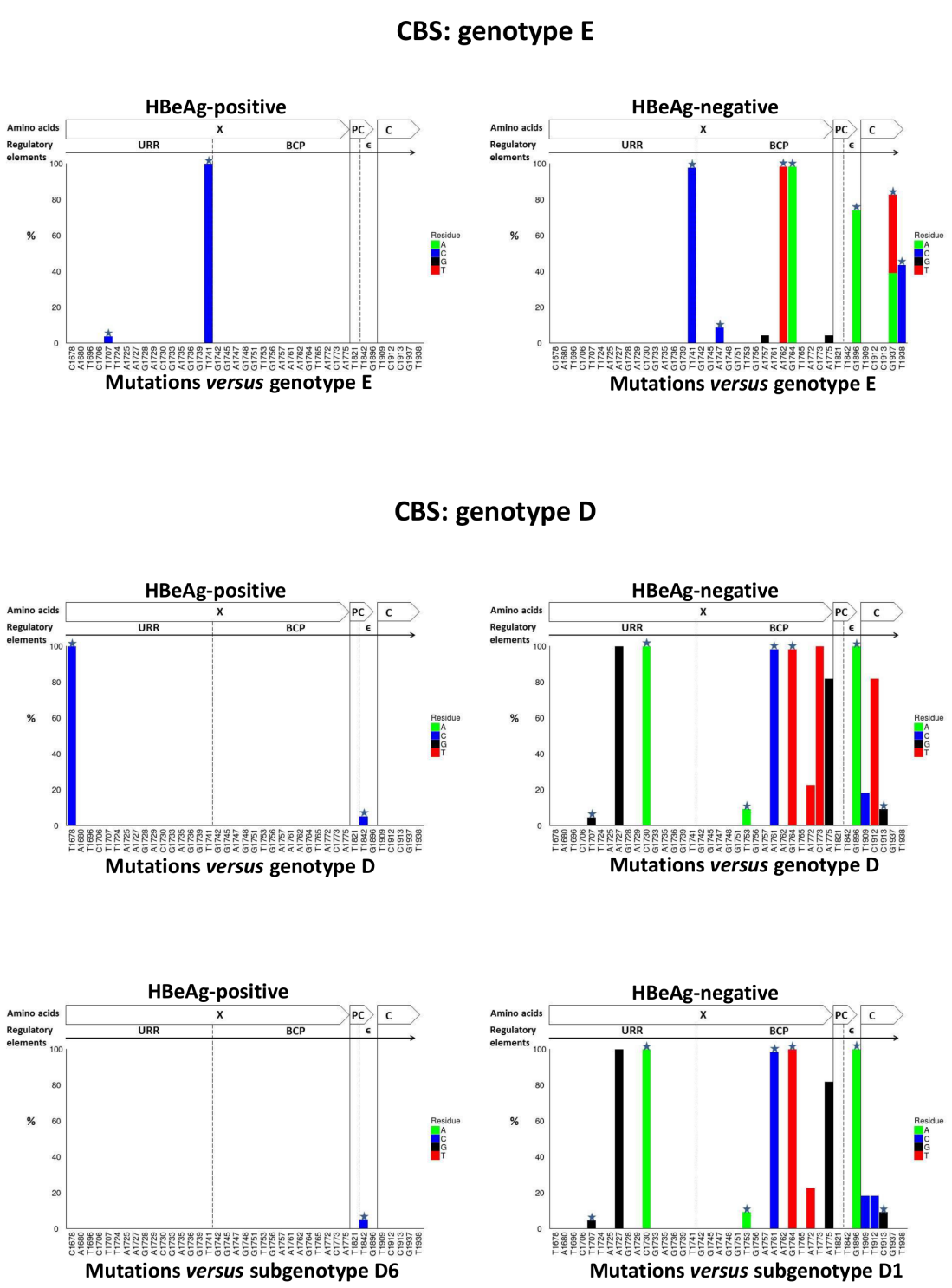


Figure 11. Graphs showing mutation distribution of the CBS data at the nucleotide level using either genotype E or D consensus sequence as the reference. A star indicates a non-synonymous mutation. The graphs were built using the Mutation Reporter Tool [22].
doi:10.1371/journal.pone.0095377.g011

mutation A1762T/G1764A is known to down-regulate the transcription of precore mRNA that is translated into HBeAg [14]. Although A1757G is a synonymous mutation and thus has no effect on the protein sequence, it overlaps *cis*-regulatory elements within the basic core promoter. In the present study, 1757G was found to be associated with A1762T/G1764A. This association has also been shown by others, who found that chronic hepatitis patients infected with HBV with 1757G/1762A1764A had higher HBV DNA levels compared to patients infected with the wild-type 1757A/1762T1764A [27]. Moreover, A1757G was found to in HCC patients infected by genotype C [28]. Non-synonymous mutations G1937A/T and T1938C within the core region occurred at a high frequency (Figure 9). These mutations are located within a T-cell epitope, which is an important

Table 2. Mutations detected by CBS and UDPS.

Frequency	Detected by	Mutations
High (≥20%)	UDPS and CBS	A1727G,C1730A,T1741C,A1757G,A1761C,A1762T,G1764A/T,A1775G,G1896A,G1937A/T,T1938C
Low (<20%)	UDPS and CBS	T1707C, A1747C, T1753G, A1772T, T1909C,T1912C, C1913G
Low (<20%)	UDPS	T1678C, A1680C, T1696C, C1706T, T1724C, A1725C, G1728A, A1729G, G1733A, A1735G, G1736A, G1739C/T, T1741C, G1742A, G1745A, G1748A, G1751A, T1753C, G1756A,T1765C, T1733C, T1842C

doi:10.1371/journal.pone.0095377.t002

component of the host's immune response to HBV infection [29]. These two mutations have recently been reported in strains of HBV genotype B isolated from Taiwanese patients [30]. Other substitutions (T1707C, A1735G, A1747C and T1909C) were found at low frequencies (<20%) and have not been reported in previous studies.

In sample #2 (genotype D isolated from HBeAg-negative), mutations A1727G, C1730A, A1761C, G1764A, A1775G and G1896A were detected at high frequency. A1727G and C1730A are located in the Enhancer II region and have been detected in cirrhotic patients [28] and are associated with reduced HBcAg expression and HBV DNA levels in the liver [31]. A1761C has previously been detected within a mutational motif (1761–1766) in isolates from patients with cirrhosis and chronic hepatitis [32]. The A1775G is associated with loss of HBeAg in Taiwanese children [33]. T1678C, G1753A and T1773C, which were found in the minority of the quasispecies population, have previously been associated with severity of HBV infection and progression to HCC [28,34].

The following substitutions were found as minor populations and have not previously been documented. In HBV from HBeAg-negative samples: A1735G, G1742A, A1747C and T1909C in genotype E and A1680C, C1706T, T1724C, A1725C, G1728A, G1736A, G1739C/T, G1751A, A1772T, T1842, T1909C, T1912C and C1913G in genotype D (Figure 9) and in HBeAg-positive samples: T1696C, G1733A and G1751A in genotype D and T1707C in genotype E. Mutations G1745A and G1748A were found in both HBeAg-negative and HBeAg-positive genotype D samples. It is possible that these have not previously been detected because direct (Sanger) sequencing can only detect variation that occurs in 20% or more of the population. More extensive studies may reveal the relevance of these minor variants.

The genotype E isolates were found to harbour fewer mutations in the X, PC and core regions compared to genotype D, which is in agreement with previous studies showing low genetic diversity of genotype E [35,36]. Furthermore, a greater number of mutations were found in HBeAg-negative samples of both genotype D and E compared to HBeAg-positive samples. It was reported that the frequency of HBV mutations is higher in HBeAg-negative patients, this is as a result the immune response of the host against the virus before the loss of HBeAg [37]. However, because only four samples, belonging to the two genotypes from HBeAg-positive and HBeAg-negative samples, were analyzed, additional samples would be required before any firm conclusions can be reached about the differences in nucleotide divergences between these genotypes from HBeAg-positive and –negative sera.

In this study, where 9738 sequence reads were generated by UDPS, 39 unique positions were detected by UDPS, while only 18 (46.2%) of these position were detected by CBS. High frequency substitutions were found in 11 positions and were all detected by CBS, whereas only 6/28 (25%) low frequency substitutions were detected by CBS ($p<0.05$) (Figures 9 and 10).

Although the testing of the tools was done on a small sample set and the findings cannot be generalized, it is evident that the data generated by the increased read-depth provided by UDPS should be approached with caution. Appropriate curation and examination of the reads are required to ensure that artefacts are not interpreted as variants. Moreover, identification of variants must be performed against a suitable reference or consensus sequence, as a "mutation" of interest may simply be a known signature or variant when examined in the correct genotypic or subgenotypic context. UDPS detected a greater number of substitutions than CBS. Relative to CBS, UDPS is cheaper to undertake, both in terms of time and expense. However, without rigorous and careful examination and interpretation of read data, the results generated by UDPS may be misleading. As illlustrated in the present study, a thorough knowledge of the genome of interest and its known variants is essential in order to accurately and reliably interpret the high resolution read data generated by UDPS.

Acknowledgments

We thank Justin Southey for guidance with some aspects of the statistical calculations.

Author Contributions

Conceived and designed the experiments: AK MY. Performed the experiments: MY. Analyzed the data: MY TB AK. Contributed reagents/materials/analysis tools: AK TB HM DG. Wrote the paper: MY TB AK. Read and approved the paper: MY TB HM DG AK. Designed the analysis algorithms: MY TB. Developed and wrote the tools and maintained the server: TB. Analysis the data: MY TB AK.

References

1. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74: 5463–5467.
2. Solmone M, Vincenti D, Prosperi MC, Bruselles A, Ippolito G, et al. (2009) Use of massively parallel ultradeep pyrosequencing to characterize the genetic diversity of hepatitis B virus in drug-resistant and drug-naive patients and to detect minor variants in reverse transcriptase and hepatitis B S antigen. J Virol 83: 1718–1726.
3. Margeridon-Thermet S, Shulman NS, Ahmed A, Shahriar R, Liu T, et al. (2009) Ultra-Deep Pyrosequencing of Hepatitis B Virus Quasispecies from Nucleoside and Nucleotide Reverse-Transcriptase Inhibitor (NRTI)–Treated Patients and NRTI-Naive Patients. Journal of Infectious Diseases 199: 1275–1285.
4. Ramírez C, Gregori J, Buti M, Tabernero D, Camós S, et al. (2013) A comparative study of ultra-deep pyrosequencing and cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. Antiviral Research.
5. Shendure J, Ji H (2008) Next-generation DNA sequencing. Nature biotechnology 26: 1135–1145.
6. Eriksson N, Pachter L, Mitsuya Y, Rhee S-Y, Wang C, et al. (2008) Viral population estimation using pyrosequencing. PLoS Computational Biology 4: e1000074.

7. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8: R143.
8. Kramvis A, Arakawa K, Yu MC, Nogueira R, Stram DO, et al. (2008) Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. J Med Virol 80: 27–46.
9. Kurbanov F, Tanaka Y, Mizokami M (2010) Geographical and genetic diversity of the human hepatitis B virus. Hepatology Research 40: 14–30.
10. Kramvis A, Kew MC (1999) The core promoter of hepatitis B virus. J Viral Hepat 6: 415–427.
11. Revill P, Yuen L, Walsh R, Perrault M, Locarnini S, et al. (2010) Bioinformatic analysis of the hepadnavirus e-antigen and its precursor identifies remarkable sequence conservation in all orthohepadnaviruses. J Med Virol 82: 104–115.
12. Jean-Jean O, Levrero M, Hans W, Perricaudet M, Rossignol J-M (1989) Expression mechanism of the hepatitis B virus (HBV) C gene and biosynthesis of HBe antigen. Virology 170: 99–106.
13. Carman W, Hadziyannis S, McGarvey M, Jacyna M, Karayiannis P, et al. (1989) Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. The Lancet 334: 588–591.
14. Okamoto H, Tsuda F, Akahane Y, Sugai Y, Yoshiba M, 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. Journal of Virology 68: 8102–8110.
15. Seeger C, Mason WS (2000) Hepatitis B virus biology. Microbiology and Molecular Biology Reviews 64: 51-+.
16. Mizokami M, Orito E, Ohba K, Ikeo K, Lau JY, et al. (1997) Constrained evolution with respect to gene overlap of hepatitis B virus. J Mol Evol 44 Suppl 1: S83–90.
17. Yousif M, Mudawi H, Bakhiet S, Glebe D, Kramvis A (2013) Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan. BMC Infectious Diseases 13: 328.
18. Takahashi K, Aoyama K, Ohno N, Iwata K, Akahane Y, et al. (1995) The precore/core promoter mutant (T1762A1764) of hepatitis B virus: clinical significance and an easy method for detection. J Gen Virol 76 (Pt 12): 3159–3164.
19. R Development Core Team (2005) R: A language and environment for statistical computing. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria, 2013. Available: http://www.R-project. org.
20. Settles M, Hunter S, Sarver B, Zhbannikov IaC, Kyu-Chul (2011) rSFFreader: rSFFreader reads in sff files generated by Roche 454 and Life Sciences Ion Torrent sequencers. R package version 080.
21. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32: 1792–1797.
22. Bell TG, Kramvis A (2013) Mutation Reporter Tool: An online tool to interrogate loci of interest, with its utility demonstrated using hepatitis B virus. Virology journal 10: 62.
23. Brodin J, Mild M, Hedskog C, Sherwood E, Leitner T, et al. (2013) PCR-Induced Transitions Are the Major Source of Error in Cleaned Ultra-Deep Pyrosequencing Data. Plos One 8: e70388.
24. Homs M, Buti M, Quer J, Jardí R, Schaper M, et al. (2011) Ultra-deep pyrosequencing analysis of the hepatitis B virus preCore region and main catalytic motif of the viral polymerase in the same viral genome. Nucleic acids research 39: 8457–8471.
25. Hoffmann C, Minkah N, Leipzig J, Wang G, Arens MQ, et al. (2007) DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. Nucleic acids research 35: e91.
26. Ijaz S, Arnold C, Dervisevic S, Mechurova J, Tatman N, et al. (2008) Dynamics of lamivudine-resistant hepatitis B virus during adefovir monotherapy versus lamivudine plus adefovir combination therapy. J Med Virol 80: 1160–1170.
27. Sendi H, Mehrab-Mohseni M, Zali MR, Norder H, Magnius LO (2005) T1764G1766 core promoter double mutants are restricted to Hepatitis B virus strains with an A1757 and are common in genotype D. Journal of General Virology 86: 2451–2458.
28. Yin J, Xie J, Liu S, Zhang H, Han L, et al. (2010) Association between the various mutations in viral core promoter region to different stages of hepatitis B, ranging of asymptomatic carrier state to hepatocellular carcinoma. The American journal of gastroenterology 106: 81–92.
29. Radecke K, Protzer U, Trippler M, Meyer zum Büschenfelde KH, Gerken G (2000) Selection of hepatitis B virus variants with aminoacid substitutions inside the core antigen during interferon-α therapy. Journal of medical virology 62: 479–486.
30. Wu J-F, Ni Y-H, Chen H-L, Hsu H-Y, Chang M-H (2013) The Impact of Hepatitis B Virus Precore/Core Gene Carboxyl Terminal Mutations on Viral Biosynthesis and the Host Immune Response. Journal of Infectious Diseases: jit638.
31. Zhu R, Zhang H-P, Yu H, Li H, Ling Y-Q, et al. (2008) Hepatitis B virus mutations associated with in situ expression of hepatitis B core antigen, viral load and prognosis in chronic hepatitis B patients. Pathology - Research and Practice 204: 731–742.
32. Veazjalali M, Norder H, Magnius L, Jazayeri S, Alavian S, et al. (2009) A new core promoter mutation and premature stop codon in the S gene in HBV strains from Iranian patients with cirrhosis. Journal of Viral Hepatitis 16: 259–264.
33. Ni Y-H, Chang M-H, Hsu H-Y, Tsuei D-J (2004) Longitudinal study on mutation profiles of core promoter and precore regions of the hepatitis B virus genome in children. Pediatric research 56: 396–399.
34. Ouneissa R, Bahri O, Alaya-Bouafif NB, Chouaieb S, Yahia AB, et al. (2012) Frequency and clinical significance of core promoter and precore region mutations in Tunisian patients infected chronically with hepatitis B. Journal of Medical Virology 84: 1719–1726.
35. Hübschen JM, Andernach IE, Muller CP (2008) Hepatitis B virus genotype E variability in Africa. Journal of Clinical Virology 43: 376–380.
36. Kramvis A, Restorp K, Norder H, Botha JF, Magnius LO, et al. (2005) Full genome analysis of hepatitis B virus genotype E strains from South-Western Africa and madagascar reveals low genetic variability. Journal of Medical Virology 77: 47–52.
37. Desmond CP, Gaudieri S, James IR, Pfafferott K, Chopra A, et al. (2012) Viral adaptation to host immune responses occurs in chronic hepatitis B virus (HBV) infection, and adaptation is greatest in HBV e antigen-negative disease. Journal of Virology 86: 1181–1192.

CHAPTER 6: GENERAL DISCUSSION AND CONCLUSION

It is surprising perhaps, that Sudan, being a HBV hyperendemic African country (>8% HBsAg-positivity), had no published data on the HBV genotypes prevailing in Sudanese liver disease and HIV patients, prior to the initiation of this study. A diversity of HBV genotypes is found in countries neighbouring Sudan (Kramvis and Kew, 2007): genotype D in Egypt to the north (Saudy *et al.*, 2003), genotype E in the Democratic Republic of Congo to the west (Mulders *et al.*, 2004) and genotype A in Kenya (Mwangi *et al.*, 2008, Ochwoto *et al.*, 2013) and Uganda to the south (Usuda *et al.*, 2000, Hannoun *et al.*, 2005, Forbi *et al.*, 2013). In view of Sudan's unique position and the movement of people across its borders, a description of the HBV genotypes prevailing in this country was important. HBV genotypes are relevant when considering both the clinical consequences of infection and response to antiviral treatment (Kramvis and Kew, 2005, Kao, 2002a). The first study describing HBV genotypes in Sudan, published after the initiation of the present study, was undertaken on blood donors, and reported that 57.5% were infected with genotype E, 40.5% with genotype D and 2% with subgenotype A2 (Mahgoub *et al.*, 2011). Subsequently, results of the present study showed that Sudanese liver disease patients were mainly infected with HBV genotypes D and E, with genotype D predominating (Yousif *et al.*, 2013).

Genotype D, which prevails in the Mediterranean, is the only genotype with a worldwide distribution. It is the genotype of HBV with the shortest genome (3182 nucleotides) characterized by a 33 nucleotide deletion in the preS1 region. The availability of a large number of sequences in the public databases, with diverse geographical distribution and the description of nine subgenotypes of D with close phylogenetic relationships, motivated us to carry out an extensive analysis of the sequencing data and to update the classification of

genotype D (Paper I, Chapter 2). This allowed for the accurate classification of the subgenotypes of D prevailing in Sudan.

Subgenotypes D1 to D4 were described previously, following phylogenetic analysis and based on an intergroup divergence of 4–8% and high bootstrap support, subgenotypes D1 to D4 were described (Kramvis *et al.*, 2008, Norder *et al.*, 2004). Later when the upper limit of intergroup divergence was lowered to 7.5% (Kramvis *et al.*, 2008), an additional five subgenotypes of D (D5 to D9) were identified (Banerjee *et al.*, 2006, Utsumi *et al.*, 2009, Meldal *et al.*, 2009, Abdou Chekaraou *et al.*, 2010, Ghosh *et al.*, 2012). However, our analysis of subgenotypes D1 to D8 (Paper I, Chapter 2) showed the separation of subgenotypes D6 (from Papua) and D3 to be uncertain. With the additional sequences analyzed, we found that the mean intergroup divergence $\pm$SD between these two subgenotypes was 2.6% $\pm$ 0.6, whereas it was reported to be more than 4% in an earlier study, when fewer sequences were compared (Utsumi *et al.*, 2009). The 2.6% nucleotide divergence is significantly lower than the 4% cut-off required for separation into subgenotypes (Kramvis *et al.*, 2005a, Kramvis *et al.*, 2008). Moreover, the bootstrap value in our phylogenetic analysis was low (56%), while it was reported as 88% in the earlier study (Utsumi *et al.*, 2009). Subgenotype D3 and "D6" were also found to share signature amino acids (Paper I, table 2). Therefore we reclassified subgenotype D6 as a clade of subgenotype D3, rather than a separate subgenotype. Although the intergroup divergence between subgenotypes D1 and D2 was only 3.1% $\pm$ 0.5, we proposed that this separation be retained because it is supported by high bootstrap values and the two subgenotypes have different serological subtypes, *ayw*2 in D1 and *ayw*3 in D2 and distinct geographical distributions (Paper I, figure 1).

We confirmed that the three strains proposed, as belonging to subgenotype D8 (GenBank Accession Numbers FN594769–FN594771) (Abdou Chekaraou *et al.*, 2010), are a

recombinant of genotypes D and E (Yousif and Kramvis, 2013). The recombination profile determined for the three strains did not differ significantly from that reported initially: FN594769– FN594771 clustered with genotype E within the regions ~1640–2430 and ~2830–3128, from the *Eco*R1 restriction site. We found that these isolates are a subgenotype D7/genotype E recombinant. Furthermore, subgenotype D8 has a three nucleotide deletion in the preS1, which is characteristic of genotype E. Therefore, we proposed that subgenotype "D8" should be renamed as a genotype D/E recombinant, rather than a subgenotype of D. Our updated analysis showed that there are in fact six and not eight subgenotypes.

Subsequent to our analysis, by using six complete HBV genomes sequences, Gosh and colleagues (2010) identified subgenotype "D9" from India, which is a genotype D/C recombinant (Ghosh *et al.*, 2012). Recombination events have been described between HBV genotypes that circulate in the same geographical region (Simmonds and Midgley, 2005). Genotypes C/D in Tibet, (Zhou *et al.*, 2011), E/A in Guinea (Garmiri *et al.*, 2009), and B/C in Asia (Sugauchi *et al.*, 2002). Although a number of researchers have suggested that recombinant strains should not be classified as novel subgenotypes but reported separately, these suggestions have not been followed widely (Shi *et al.*, 2013, Kramvis *et al.*, 2008, Pourkarim *et al.*, 2010a). Definition of subgenotypes should meet specific criteria (Kramvis *et al.*, 2005a, Kramvis *et al.*, 2008). These criteria include nucleotide divergence of between ~4% and 7.5%, monophyletic clustering and good bootstrap support to separate subgenotypes (Kramvis *et al.*, 2008). Distinct geographical separation (D1 *versus* D2) and/or different serological subtypes (D1 *versus* D2, and I1 *versus* I2), monophyletic clustering and good bootstrap support, can be used to separate subgenotypes, when the nucleotide divergence is less than 4% (Kramvis, 2014). The establishment of a scientific committee for controlling the definition of new HBV genotypes/subgenotypes has been proposed (Kramvis *et al.*, 2005a, Kurbanov *et al.*, 2010).

Genotypes D and E were found in Sudanese liver disease (Paper II, Chapter 3) and HIV-positive patients (Paper III, Chapter 4). This is in contrast to the predominance of genotype E found in Sudanese blood donors (Mahgoub *et al.*, 2011). The circulation of either genotype D or E in liver disease patients belonging to the same ethnic group, allowed for the comparison of patients infected with these two genotypes (Paper II, Chapter 3). No significant difference was observed in gender, age, ALT and clinical groups of patients infected with either genotype. When we compared the viral loads in the liver diseases patients infected with genotypes D and E, we found that viral loads were significantly higher in genotype E-infected patients compared to genotype D-infected patients. Moreover, patients infected with genotype E, showed a significantly higher frequency of HBeAg-positivity. The same results were reported by Mahgoub and colleagues (2011) in Sudanese blood donors (Mahgoub *et al.*, 2011) and in a Ghanaian study (Candotti *et al.*, 2006). The higher HBeAg-positivity seen in HBV genotype E mono-infected individuals with liver disease patients could confer tolerance resulting in less serious clinical manifestation when compared to infection with genotype D where HBeAg-positivity is lower (Chen *et al.*, 2005). This could explain why genotype E prevailed in the Sudanese blood donors (Mahgoub *et al.*, 2011), and genotype D prevailed in the liver disease patients in the our study. The genetic diversity of genotype E isolates was significantly lower than the genotype D isolates. Genotype E has been characterized as having a low diversity compared to other genotypes and subgenotypes of HBV (Kramvis and Paraskevis, 2012, Hübschen *et al.*, 2009, Kramvis *et al.*, 2005b).

Subgenotype D1 was found to predominate in the liver disease mono-infected patients (Paper II, Chapter 3) and HBV/HIV co-infected individuals (Paper III, Chapter 4). This is in agreement with the previous study of blood donors (Mahgoub *et al.*, 2011, Yousif *et al.*, 2013). Therefore the predominant subgenotype of D found in Sudan is D1. This is the subgenotype of D that prevails in the Mediterranean, Arabian subcontinent and in China

(Yousif and Kramvis, 2013). Moreover, our study is the first to report the presence of subgenotype A1 in Sudan. Subgenotype A1 predominates in southern and eastern Africa, including Kenya (Ochwoto *et al.*, 2013) and Uganda (Forbi *et al.*, 2013), which border with Sudan.

The Sudanese HCC patients infected with either genotype D or E were significantly older than asymptomatic carriers as well as the chronic, acute and cirrhotic hepatitis, with the acute hepatitis patients being significantly younger. This differs from southern African HCC patients, infected with genotype A, in particular subgenotype A1, who develop cancer at a significantly younger age (Kew *et al.*, 2005). Southern Indian HCC patients infected with subgenotype A1 have also been shown to be significantly younger than those infected with genotype D (Gopalakrishnan *et al.*, 2013). The Sudanese HCC group had the highest median viral load compared to the other disease groups. High viral loads have been implicated as a risk factor for HCC development (Liu *et al.*, 2008).

HBV and HIV share the same transmission route (Alter, 2006) and the prevalence of HIV among Sudanese adult ranges from 0.9% to 1.4% (UNAIDS., 2010). More than 60% of HIV patients were shown to have at least one marker of HBV infection, with over a quarter being HBV DNA positive, with ~12% overt infection and 15% OBI (Paper V, Appendix A). HBV isolated from 50 patients randomly sampled from our study were used to molecularly characterize HBV from Sudanese HIV-infected patients. Similar to HBV mono-infected patients, genotypes D and E were found in HBV/HIV co-infected individuals, with genotype D predominating (Paper III, Chapter 4). However, the frequency of genotypes A and D/E recombinant was relatively higher in HBV/HIV co-infected patients compared to HBV mono-infected liver disease patients (genotype A: 18.4% *versus* 8.6%, and D/E: 13.2% versus 2.5%). Moreover, there were a number of differences between the HBV isolates from mono-

infected compared to HBV/HIV co-infected individuals. Although the genotypes prevailing in the mono- and co-infected individuals did not differ significantly, a lower proportion of the mono-infected individuals were HBeAg-positive (12%) compared to the co-infected individuals (18%). It is possible that HIV, by causing immunosuppression may delay or prevent seroconversion to the inactive carrier phase in patients with chronic HBV infection (Bodsworth *et al.*, 1991). However, there is also the possibility that the HIV co-infected patients, who were significantly younger than the mono-infected liver disease patients, did not have enough time to seroconvert or were recently infected with HBV (Paper V, Appendix A). In contrast to mono-infected patients, HBeAg-expression and the viral loads in co-infected patients did not differ between those infected with genotype D and E. Furthermore, the intra-group divergence of genotype E in HBV/HIV co-infected individuals was significantly higher than that in HBV mono-infected liver disease patients. It is possible that the genetic diversity of genotype E is constrained in the immunocompetent mono-infected individuals but allowed to expand in HBV/HIVco-infected patients, who are immunocompromised.

Using the Taormina definition of OBI (Raimondo *et al.*, 2008), 36% of the 50 serum samples used in HBV/HIV study were HBsAg-negative, of which, 20% had viral loads of <200 IU/ml and therefore OBI. The absence of HBsAg and the presence of HBV DNA in these OBI cases is probably as a result of low viral loads, which prevent the detection of HBsAg by the routine serological techniques. Higher viral loads of >200 IU/ml HBV DNA, in the absence of HBsAg, were found in 16% of HBV/HIV co-infected individuals. This infection has been defined as "false occult" (Raimondo *et al.*, 2008) or "HBsAg-*covert"* (Bell *et al.*, 2012). A similar proportion of OBI and HBsAg-covert infection has also been found in southern African HBV/HIV co-infected individuals (Bell *et al.*, 2012), infected with HBV

subgenotype A1 (Makondo *et al.*, 2012). In order to explain the HBsAg-negativity seen in the HBsAg-covert individuals the preS/S region was sequenced (Paper III, chapter 4). Only one mutation, sS174N, was found that could differentiate between HBV isolated from HBsAg-negative and -positive individuals. This is in agreement with two previous studies (Makondo *et al.*, 2012, Henke-Gendo *et al.*, 2008), where patients were infected with subgenotype A1 and A2, respectively. sS174N has been described as a reactivation marker (Henke-Gendo *et al.*, 2008). Other mutations in the S region that could account for the HBsAg-negativity were also found in HBV isolated from HBV/HIV co-infected individuals including ps2R48T, sM133T, sE164G and sV168G, but these, in agreement with other studies in HBV/HIV co-infected individuals, could not differentiate between HBsAg-negative and HBsAg-positive infection (Pourkarim *et al.*, 2011b).

Eleven isolates from HBV mono-infected patients (paper II, chapter 3) and two isolates from HBV/HIV co-infected patients (paper III, chapter 4) had deletions in the preS region. The deletion mutants in HBV genotype E isolates from HBV mono-infected liver disease patients were mainly from HCC patients whereas this was not the case for genotype D isolates. The numbers, however, were too low to reach any firm conclusion. Deletion mutants have been reported to occur more frequently in HCC patients (Kew *et al.*, 2005) and HIV co-infected patients (Audsley *et al.*, 2010). Deletion mutants in the preS have been reported to enhance the progression of liver disease in genotypes B and C (Chen *et al.*, 2006). Similar deletion mutants were found in Sudanese and Guinean blood donors in genotype E (Mahgoub *et al.*, 2011, Garmiri *et al.*, 2009) and in South African HIV patients infected with subgenotype A1 (Makondo *et al.*, 2012). As noted previously, these deletion mutants may be a risk factor for the development of HCC in HBV/HIV co-infected individuals, if HBV infection is left untreated (Makondo *et al.*, 2012).

The genotyping and molecular characterization of HBV isolated from HBV-mono- and HBV/HIV co-infected Sudanese patients was carried out using Sanger nucleic acid sequencing, which can only detect variation that occurs in at least 20% of a viral population. Because of its complex replication cycle, high replication and mutations rates, HBV exists in patients as a heterogeneous viral population, composed of closely related, non-identical genomes, known as quasispecies (Rodriguez-Frias *et al.*, 2013). Traditionally, characterization of a viral quasispecies population has been carried out using amplification, cloning and sequencing (CBS: cloning based sequencing). This method is labour intensive and expensive, both in terms of time and cost. Recently developed next generation sequencing (NGS) produce a very high throughput of data and allow for the detection of mutations, even when present at very low levels in the quasispecies population (Hoffmann *et al.*, 2007, Ijaz *et al.*, 2008).

Bioinformatic tools were developed in order to analyze the data generated using UDPS and CBS of the BCP/PC region of four Sudanese serum samples, infected with either genotype D or E of HBV, from HBeAg-positive and HBeAg negative patients. This allowed for the identification of nucleotide substitutions and variants in the BCP/PC region and comparison between genotypes and different stages of HBV infection, namely, HBeAg-positive and negative phases. Using the newly developed 'Deep Threshold Tool', the probability of error that could detect biologically important mutations in the BCP/PC region was found to be 0.5%. Previously, a probability of error of between 0.5% and 1% for UDPS has been used for HIV samples (Eriksson *et al.*, 2008). Studies on HBV sequence data have either used the same probability of error, or have not reported details of this component of the analysis (Margeridon-Thermet *et al.*, 2009, Homs *et al.*, 2011, Solmone *et al.*, 2009). In agreement

with previous studies (Homs *et al.*, 2011), the presence of the wild type (G) at 1896 in sample #1 and sample #2, which were isolated from HBeAg-negative patients (Paper IV, Chapter 5), confirms the ability of UDPS to detect minor populations, which may not be detected by Sanger sequencing (Hoffmann *et al.*, 2007, Ijaz *et al.*, 2008). As shown in figures 9 and 10 (Paper IV, Chapter 5), the determination of mutation/variant depends on the reference used in the analysis. Thus knowledge of the genotype and subgenotype of HBV is important when determining the presence of mutations. Therefore, where possible, a consensus sequence of the genotype or subgenotype should be used, to ensure that variants are examined in the appropriate context.

UDPS was able to detect mutations that occurred both at high frequency ($\geq$ 20%) and at low frequency (< 20%), with more mutations being detected in samples from HBeAg-negative samples and genotype D samples showing greater variability compare to genotype E. HBV with mutations, previously shown to biologically important, including G1896A (Carman *et al.*, 1989) and A1762T/G1764A (Okamoto *et al.*, 1994), were detected in both HBeAg-positive and HBeAg-negative samples belonging to both genotypes. Moreover, substitutions not previously described in the literature were detected using UDPS, which can detect minor population of the quasispecies. It is possible that these have not previously been detected because direct (Sanger) sequencing can only detect variation that occurs in 20% of the population in the quasispecies. However, until more extensive analysis is undertaken on a larger number of samples, it cannot be determined whether these are true mutations or artefacts. Compared to CBS, UDPS identified 39 unique positions, of which only 18 (46.2%) were identified by CBS. As expected, CBS identified all high frequency substitutions, because these mutations were present in $\geq$20% of the population of the quasispecies, while only 6/28 (25%) of the low frequency mutations (< 20%) were identified by CBS,

confirming the limitations of the CBS.

Summary

Using the specific criteria, that have been suggested previously, to define genotypes/subgenotypes of HBV, we determined that genotype D has six and not eight subgenotypes. The importance of HBV genotypes in clinical consequences of infection and response to antiviral treatment has led us to characterize HBV genotypes circulating in Sudan. HBV mono-infected patients and HBV/HIV co-infected individuals, were mainly infected with genotype D or E. HBV mono-infected patients, infected with genotype E, had higher HBeAg-positivity and higher viral loads than those infected with genotype D. The ratio of genotype A to non-A, as well as the genotype E intra-group divergence were higher in HBV/HIV co-infected individuals compared to HBV mono-infected individuals. OBI was found in 15.1% HBV/HIV co-infected patients and its clinical relevance remains to be determined. In order to overcome the limitations of Sanger sequencing, which include its high cost and inability to detect minor populations in quasispecies, next generation sequencing techniques such as UDPS have been developed. We demonstrated that correct analysis of UDPS data requires appropriate curation of read data, in order to clean the data and eliminate artefacts and that the appropriate consensus (reference) sequence should be used in order to identify variants correctly. CBS detected fewer than 50% of the substitutions detected by UDPS. This new technology may allow the detection of minor variants between the different genotypes of HBV and provide biomarkers for the prediction of clinical manifestation of HBV and response to antiviral therapy.

Appendix A: Paper V

Overt and occult hepatitis B virus infection in adult Sudanese HIV patients.

MUDAWI, H., HUSSEIN, W., MUKHTAR, M., **YOUSIF, M**., NEMERI, O., GLEBE, D. & KRAMVIS, A (2014)

International Journal of Infectious Diseases

Journal: International Journal of Infectious Diseases, is a peer-reviewed, ISI-accredited, open access, with an impact factor of 3.03

Overt and occult hepatitis B virus infection in adult Sudanese HIV patients

Hatim Mudawi[1], Waleed Hussein[2], Maowia Mukhtar[2], Mukhlid Yousif[3], Omer Nemeri[4], Dieter Glebe[5], Anna Kramvis[3*]

[1]Department of Medicine, Faculty of Medicine, University of Khartoum, Khartoum, Sudan

[2]Institute of endemic diseases, University of Khartoum, Khartoum, Sudan

[3]Hepatitis Virus Diversity Research Programme, Department of Internal Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[4]Department of Medicine, College of Medicine, Bahri University, Khartoum, Sudan

[5]Institute of Medical Virology, Justus-Liebig-University of Giessen, Giessen, Germany

[*]**Corresponding author**: Professor Anna Kramvis

Hepatitis Virus Diversity Research Programme,

University of the Witwatersrand

7 York Road, Parktown,

Johannesburg, South Africa

Email: Anna.Kramvis@wits.ac.za

Telephone: +27 (0) 11 488 3100

Fax: +27 (0) 86 529 6806

Abstract

Objectives: Human immunodeficiency virus (HIV) in sub-Saharan Africa is complicated by co-infection with hepatitis B and C viruses (HBV, HCV) that share similar transmission routes**.** Our aims were to determine the prevalence of HBsAg-positive and HBsAg –negative HBV, and HCV infection among HIV-infected patients. *Methods:*A cross-sectional study was conducted among treatment-naïve HIV-positive adults in Khartoum state. HBV, HCV and HIV infection were detected using immune assays for HBsAg, anti-HBc, anti-HCV or anti-HIV, while real time polymerase chain reaction (PCR) was used to measure HBV DNA. *Results:* The age (mean±standard deviation (SD)) of 358 patients was 35.2±9.3 years, with a male: female ratio of 1.3:1.0 and mean±SD alanine aminotransferase (ALT) levels of 10.9±18.0 U/L. Evidence of current or past HBV infection was detected in 62.8% of the patients and HBV DNA in 96 patients (26.8%): 42 HBsAg-positive (11.7%) and 54 (15.1%) HBsAg-negative, indicating occult hepatitis B infection. Anti-HCV was detected in 1.7%. *Conclusion*: Evidence of HBV infection was detected in 26.8% of HIV patients with HBsAg-negative infection, with viremia, detected in 15.1% of the patients.. All HIV-infected patients should be carefully screened for HBV, preferably using PCR, prior to initiation of antiretroviral therapy.

Key words: polymerase chain reaction, HBV detection, HBsAg-negative, occult HBV infection, hepatitis C virus

INTRODUCTION

Human Immunodeficiency (HIV) is a serious health problem in sub-Saharan Africa [1]. HIV infection is complicated by co-infection with other pathogens including hepatitis B and C viruses (HBV and HCV) that share similar routes of infection. Worldwide, it is estimated that over 240 million people [2] and 150 million people [3] have chronic HBV or HCV infection, respectively. Globally there are 34 million people infected with HIV, 69% of them reside in sub-Saharan Africa [1]. Although the introduction of antiretroviral therapy (ART) has decreased the death rate and the incidence of new acquired immunodeficiency syndrome (AIDS) defining diseases among HIV infected individuals, liver disease, mainly as a result of HBV or HCV co-infection has emerged as one of the leading causes of non-AIDS related mortality and morbidity [4]. Co-infection with HBV or HCV and HIV is characterized by more rapid progression of liver disease including accelerated fibrosis, cirrhosis and hepatocellular carcinoma[5, 6].

Sudan, located in the north to east-central Africa is endemic for HBV infection, with HBV exposure rates from 47%[7], - 78%[8], prevalence of HBsAg 6.9%[7] - 18.7% [9], prevalence of HCV antibodies 2.2% [10]- 4.8% [11] and prevalence of HIV among adult population 0.9-1.4% [12]. However, there is no clear knowledge about HBV in the setting of HIV infection, and it would be important to determine this and select appropriate ART, specially as HIV patients may also develop hepatic toxicity and lactic acidosis under ART, which is based on a combination of drugs which includes Lamivudine, Stavudine, Efavirenz and Nevirapine[13].

No studies have been carried out to determine the presence of HBV DNA, in the absence of HBsAg, in HIV-positive Sudanese. Thus the aims of this study were to determine the prevalence of HBV and HCV infection and the prevalence of occult hepatitis B infection (OBI) among adult Sudanese HIV patients.

MATERIALS AND METHODS

Patients

A cross-sectional study was conducted during the period December 2010 to June 2012 in Khartoum, Sudan. All treatment-naïve adults (18 years and over), who tested positive for HIV antibodies using indirect enzyme linked immunosorbent assay (ELISA), attending Omdurman Medical AIDS care Unit (OMACU) and Bashair hospital HIV treatment center in Khartoum state, were included in the study. Children younger than 18 years and pregnant women were excluded. The study was approved by the Ethics committee of the Faculty of Medicine, University of Khartoum, the National Ethics committee, Federal Ministry of Health, Sudan and by the Human Ethics Committee of the University of the Witwatersrand, South Africa. Written consent was obtained from all participants before their enrollment.

Data collection

Data was collected using a specially designed questionnaire with basic demographic criteria including age, gender and potential risk factors for transmission of infection such as multiple sexual partners, intravenous drug use, previous history of jaundice and blood transfusion and history of hepatitis B vaccination. All patients were classified according to the WHO clinical staging [14]. All HIV-infected patients were classified according to WHO clinical staging[15]

Collection and storage of samples:

Ten milliliters (10 ml) venous blood sample was collected from each patient and divided in two blood collection tubes: 7 ml in ethylenediaminetetraacetic acid (EDTA) tubes and 3 ml in plain tubes for serum separation. Separated plasma and serum samples were stored at -30°C for further analysis.

Immune assays

Detection of anti-HIV was done using an indirect enzyme linked immunosorobet assay (ELISA) kit (Biorex, Antrium, United Kingdom). HBsAg, anti-HBc IgM and IgG, hepatitis B e antigen and antibodies (HBeAg, anti-HBe), and anti-HCV were detected using immune assays from Diasource (Nivelles, Belgium); the immune assay for hepatitis D virus antibodies (anti-HDV) came from Diagnostic Automation, Inc. (Calabasas, California). In all tests the instructions of the suppliers were strictly followed.

DNA extraction and quantitative real time-polymerase chain reaction (PCR)

HBV DNA was extracted with the SACACE Ribo virus kit (HBV Real-TM Quant, Sacace Biotechnologies, Como, Italy) as recommended by the supplier. An internal control (HBV IC) with known viral concentration (specific for each lot) was used to monitor the amplification and exclude any loss of genomic DNA or PCR inhibition and to enable the precise calculation of the HBV viral load in the samples. The extracted DNA was immediately used for PCR. Real Time PCR was done using the HBV Real-TM Quant. ® Kit (Sacace Biotechnologies, Como, Italy). 12.5 µL of the extracted HBV DNA was added to an equal volume of the reaction mixture provided by the kit. With each run, 3 standards known of the HBV DNA (QS1 HBV, QS2 HBV, and QS3 HBV) and the other three standards of the internal control (QS1 IC, QS2 IC, and QS3 IC) were used to calculate the viral copy numbers. Real Time PCR was done according to the manufacturer's instructions as follows: stage 1: 95° C for 15 minutes and in stage 2: a two temperature cycle with 95° C for 20 seconds and 60° C for 40 seconds which was repeated for 42 times. Cy3 dye was used for the sample and FAM for the internal control. The signals from these dyes were calculated and presented as numbers. The sensitivity of the assay is ~20 genome copies/ml, which is equivalent to 12 IU/ml (1.1 log10 IU/ml), conversion factor = 1,7.

Statistical analysis

Statistical analysis was done using SPSS software program to calculate frequencies, Chi-squared (χ^2) and t-values; significance was assumed at p-values < 0.05.

RESULTS

A total of 358 HIV- infected patients were enrolled in the study with a mean age ± standard deviation (SD) of 35.2 ± 9.3 years, with male: female ratio of 1.3:1.0. Mean alanine aminotransferase ± SD (ALT) levels of the study population was 10.9± 18.0 U/L. Demographic criteria of the study population and risk factors for acquiring HBV infection are shown (Table 1). All participants tested positive to HIV antibodies using indirect ELISA.

HBV serology

Of the 358 patients, 225 (62.8%) were anti- HBc IgG-positive indicating previous exposure to HBV infection. The mean ± SD, age and ALT levels of this subpopulation were 35.3 ± 9.4 years and 11.0 ± 20.9 U/L, respectively, and these values did not differ from those of the population as a whole (Table 1). Only 4 patients (1.1%) reported being vaccinated against HBV infection, however, anti- hepatitis B surface antibodies were not tested in this group of patients. Male gender and intravenous drug use were found to be risk factors for positivity for HBV markers ($p < 0.05$).

HBsAg-positive patients

Forty two patients (11.7%) were HBsAg-positive. Mean ± SD, age and ALT levels of the group was 34.7± 8.7 years and 9.8 ± 7.2 U/L respectively and these values did not differ from those of the population of 225 patients who were anti-HBc IgG-positive**.** History of jaundice was an indicator of being HBsAg-positive ($p < 0.05$) (Table 2). Eight patients (19.0%) were HBeAg-positive and anti-

HBe-negative, 24 patients (57.1%) were HBeAg negative and anti-HBe positive and 10 patients (23.8%) were negative for both HBeAg and anti-HBe. All patients were negative for anti-HDV.

Viral loads

HBV DNA was detected by PCR in 96 patients (26.8%); all were anti-HBc IgG–positive. Forty two patients (11.7%) were HBsAg-positive and 54 patients (15.1%) were HBsAg -negative indicating occult infection. Median viral loads, interquartile range (IQR) and demographic criteria of patients with detectable HBV DNA on PCR and risk factors are shown in tables 3, 4 and 5. Risk factors associated with HBV DNA-positivity included blood transfusion, previous dental procedures, history of jaundice and WHO clinical stages 3 and 4 ($p < 0.05$) (Table 3). Overall median [IQR] viral load was 2.9[1.7-4.1] log10 IU/ml, with HBsAg-positive individuals having significantly higher viral loads compared to HBsAg-negative patients (Table 5). Eight HBsAg-positive , HBeAg-positive, anti-HBe-negative patients had viral loads of 5.6 log10 IU/ml, 24 HBsAg-positive, HBeAg-negative, anti-HBeAg-positive patients had viral loads of 3.8 log10 IU/ml and 10 HBsAg-positive, HBeAg/anti-HBeAg-negative had viral loads of 3.9 log10 IU/ml. Thirty five (9.8%) HBsAg-negative patients had a viral load < 2.3 log10 IU/ml (< 200 IU/ml) (true OBI) and nineteen (5.3%) had viral loads >2.3 log10 IU/ml (>200 IU/ml) (false OBI or HBsAg-covert). One patient was reactive in both anti-HBc IgG and anti-HBc IgM, the patient was HBsAg-negative with normal ALT and had detectable HBV DNA on PCR indicating a recent HBV infection or a HBV reactivation in a patient with OBI.

HCV infection

Six patients (1.7%), of which five were female, tested positive for anti-HCV antibodies, mean age $\pm$ SD was 34.3$\pm$ 12.9 and mean$\pm$ SD ALT level was 6.0 $\pm$ 3.8U/L. None of them was an intravenous drug user and there were no significant risk factors for HCV infection. Five of these patients were anti-

HBc IgG-positive, with two having detectable HBV DNA on PCR. One patient was HBsAg positive, while the second patient was HBsAg negative indicating occult HBV infection.

DISCUSSION

The 62.8% exposure rate to HBV infection and prevalence of HBsAg of 11.7% detected in the HIV-positive Sudanese patients in the present study did not differ from HBV exposure rates found previously in HIV-negative patients where HBV exposure rates ranged from 47%[7], - 78% [8] and prevalence of HBsAg 6.9%[7] - 18.7%[9] . The prevalence of HCV antibodies of 1.7% was also similar to prevalence rates of HCV in HIV-negative patients of 2.2%[10] - 4.8%[11] from previous studies. Similar exposure rates to HBV and HCV among HIV-positive and HIV-negative individuals in sub-Saharan African countries have been reported [16]. In Ethiopia, in HIV-positive patients, the exposure to HBV and prevalence of HBsAg is 47.5% and 3.9%, respectively compared to 41.9% exposure to HBV and 5.4% HBsAg-positivity in HIV- negative patients [17]. In Uganda exposure to HBV and prevalence of HBsAg is 59% and 8.3%, respectively in HIV- positive patients compared to 51.9 % and 10.4% in HIV- negative patients [18]. A study in Mozambique in HIV-positive patients revealed exposure to HBV is 69.6%, HBsAg 10.6% and HCV antibodies 1.4% compared to 62.5%, 9.0% and 1.5%, respectively in HIV- negative individuals [19]. These findings support the hypothesis that in regions with high endemicity to HBV infection such as in sub-Saharan Africa, most HBV infections occur in childhood either perinatally or more often through horizontal transmission such as close contact within households, medical procedures using non-sterile equipment and through traditional medical practices long before exposure to HIV [20, 21]. This is in contrast to countries with low endemicity for HBV infection such as the United States, Europe and Australia. Here, although the HBV prevalence in HIV-positive patients may be similar to that found in high endemicity regions (with Soriano *et al* [22] , for example, reporting a 7.1% HBsAg-positiviity), it is relatively higher when compared to HIV-negative patients. This is because both HBV and HIV are usually acquired in adulthood through sexual

transmission or intravenous drug use [23]. Studies have also shown that following acute HBV infection, HIV patients are more likely to develop chronic infection than non-HIV patients and this may explain why HIV patients have a higher rate of chronic HBV infection in low endemicity regions [23, 24].

The present study showed a slightly higher prevalence of HBeAg (19%) among HBV/HIV co-infected patients when compared with our previous report on the prevalence of HBeAg among HIV-negative patients where only 12 % were HBeAg -positive[25]. This however was not significantly different. A similar trend was reported from Zambia where HBeAg was positive in 25% of HIV-positive patients compared to 8.5 % among HIV-negative patients [26]. It is possible that HIV, by causing immunosuppression may delay or prevent seroconversion to the inactive carrier phase in patients with chronic HBV infection [23]. It has been demonstrated that efficient restoration of immunity after the introduction of ART leads to high rates of seroconversion, especially in patients in the immunoactive phase of HBV infection[27]. However, there is also the possibility that the HIV co-infected patients, who were significantly younger than the mono-infected liver disease patients did not have enough time to seroconvert or were recently infected with HBV. Patients infected with HBV/HIV had higher HBV viral loads compared to HBV mono-infected patients ($p<0.05$) [25] and also lower ALT levels, within the normal range, compared to the patients infected with HBV only ($p<0.05$) [25]. The relatively low ALT levels have also been demonstrated in studies from South Africa where there was no significant difference in ALT levels among HIV patients serologically positive or negative for HBV infection [28]. The immunosuppression caused by the HIV infection is responsible for these differences [29] and can account for the low rate of clinical manifestation of liver disease in the study population.

Despite the lower ALT levels, these patients have a higher risk of progression to liver cirrhosis [30]. In addition to this risk, ART may cause further liver damage, either through direct toxicity or through idiosyncratic reactions [31]. HBV/HIV co-infection is known to increase the risk of hepatotoxicity from ART when compared to other HIV patients without HBV or HCV co infection[13, 32]. In sub-Saharan

Africa, however, this is risk is not entirely attributable to ART but may be as a result of elevated levels of HBV DNA in the serum[33].

HIV infection is also a known risk factor for development of OBI [34]; this is characterized by two main criteria: absence of HBsAg and low viral replication. This has been reported in several clinical contexts, the most common being chronic carriage with HBsAg too low to be detected by current serological tests and the presence of anti-HBc IgG as the only serological evidence of exposure to the virus, referred to as "isolated anti HBc" [35]. In this study, 96 patients (26.8%) had detectable HBV DNA on PCR, all of them were HBcIgG-positive, 11.7% were HBsAg-positive and 54 patients (15.1%) were HBsAg -negative indicating "occult" infection. OBI has been further defined by the Taormina consensus conference in 2008 as "presence of HBV DNA in the liver (with detectable or undetectable HBV DNA in serum) of individuals testing HBsAg- negative with currently available assays and a serum HBV DNA of < 2.3 log10 IU/ml (< 200 IU/ml) when detectable) [36]. In this study, 35 (9.8%) HBsAg-negative patients had a viral load < 2.3 log10 IU/ml (< 200 IU/ml) and 19 (5.3%) had viral loads > 2.3 log10 IU/ml (>200 IU/ml). Some individuals with chronic HBV infection may be HBsAg-negative but have HBV DNA levels comparable to patients with positive HBsAg (>2.3 log10 IU/ml ; >200 IU/ml). These individuals may infected with HBV variants with S gene escape mutants, which do not express HBsAg on serology as preS mutations may prevent HBsAg secretion[37]. In many HIV patients with OBI, the viral load may be higher than suggested by the Taormina conference, which did not deal explicitly with HIV co-infection. A similar study from South Africa has reported positive HBV DNA in 23.8% of HIV patients, 8.7% were HBsAg- positive and 15.7% were HBsAg-negative, with a viral load of < 2.3 log10 IU/ml (< 200 IU/ml) detected in 6.7% of HBsAg- negative patients[28]. OBI has been demonstrated in 21.3% of HIV infected patients from the Cote d' Ivoire [38] and in 10.5 % of cases from the United States [39]. There is evidence that OBI may be associated with progression to chronic liver disease, liver cirrhosis and hepatocellular carcinoma [40].

While HIV is treated with a combination of three drugs which includes Lamivudine, Stavudine , Efavirenz and Nevirapine[13], only Lamivudine is active against HBV, which may lead to rapid emergence of Lamivudine resistant HBV strains[41, 42.] This is particularly important with the cognition that many HIV patients may test negative on screening with HBsAg but still harbor occult HBV infection, as has been demonstrated in this and other studies[28, 34]. However, recent reports have suggested that although Lamivudine resistance, can occur in the setting of incomplete HBV suppression, this was infrequently observed among HBV/HIV co- infected patients in sub-Saharan Africa[43]. As of 2010 the World Health Organization has recommended introduction of ART which includes Tenofovir and Lamivudine (or emtricitabine) as first line therapy in all HBV/HIV co- infected patients needing treatment[44].

Interestingly, in this study, out of 96 patients with detectable HBV DNA on PCR, HIV/HBV/HCV tri-infection was detected in 2 patients (2%), with none having HIV/HBV/HDV tri- infection. This is different from other studies were HIV/HBV/ HCV was detected in 6% and HIV/HBV/HDV was detected in 4% of patients[45]. It is generally thought that HBV replication is significantly reduced in HCV and HDV tri- infection. The low HCV tri- infection in this study could be explained by the low HCV prevalence of only 2.2-4.8% found in Sudan[10, 11]. However, the absence of HDV tri-infection is difficult to explain considering that HDV antibodies were detected in 9.0 % of patients with chronic HBV mono- infection[46]. The determination of anti-HBs (to determine imunity to HBV in HBsAg-negative patients), CD4 counts, HIV viral loads and nucleic acid testing for HCV and HDV would have strenghtherned this study and their absence are thus limitations of the present study.

In conclusion in Sudan exposure rates to HBV infection and prevalence of HBsAg and HCV antibodies among adult HIV infected patients was similar to exposure rates in HIV-negative patients from previous studies. HBsAg-negative HBV infection was detected in 15.1% of patients. Prevention

of HBV and HIV through education and vaccination against HBV is essential in our setting. All HIV-infected patients should be carefully screened for HBV infection, with HBsAg and anti-HBc Ig G antibodies prior to starting ART, which should include Tenofovir and Lamivudine (or emtricitabine) as first line therapy. Although PCR testing would detect OBI, which would be missed by serological testing alone, it is too expensive to include as a routine screening test in Sudan.

ACKNOWLEDGMENTS

We acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) of the special African-German Collaborative programme, "Africa Initiative", (grant GL595/3-1 to HM, AK, and DG). The funding agency had no role in the study. We thank the staff at Omdurman AIDS care unit and Bashair hospital for their assistance with the study. The authors have no conflict of interest to declare. The study was approved by the Ethics committee of the Faculty of Medicine, University of Khartoum, the National Ethics committee, Federal Ministry of Health, Sudan and by the Human Ethics Committee of the University of the Witwatersrand, South Africa. Written consent was obtained from all participants before their enrollment.

REFERENCES

1.UNAIDS. UNAIDS Report on the global AIDS epidemic. Accessed July 2013. http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_en.pdf. . 2012.

2.WHO. Hepatitis B Fact sheet N°204. Accessed July 2013. http://www.who.int/mediacentre/factsheets/fs204/en/. 2013.

3.WHO. Hepatitis C Fact sheet N°164. Accessed July 2013. http://www.who.int/mediacentre/factsheets/fs164/en/. 2013.

4.Lewden C, Salmon D, Morlat P, Bévilacqua S, Jougla E, Bonnet F, Héripret L, Costagliola D, May T, Chêne G. Causes of death among human immunodeficiency virus (HIV)-infected adults in the era of potent antiretroviral therapy: emerging role of hepatitis and cancers, persistent role of AIDS. International journal of epidemiology 2005;34:121-30.

5.Thio CL. Hepatitis B and human immunodeficiency virus coinfection. Hepatology 2009;49:S138-S45.

6.Graham C, Baden L, Yu E, Mrus J, Carnie J, Heeren T, Koziel M. Influence of human immunodeficiency virus infection on the course of hepatitis C virus infection: a meta-analysis. Clinical Infectious Diseases 2001;33:562-9.

7.Mudawi HM, Smith HM, Rahoud SA, Fletcher IA, Saeed OK, Fedail SS. Prevalence of hepatitis B virus infection in the Gezira state of central Sudan. Saudi J Gastroenterol 2007;13:81-3.

8.McCarthy MC, Hyams KC, el-Tigani el-Hag A, el-Dabi MA, el-Sadig el-Tayeb M, Khalid IO, George JF, Constantine NT, Woody JN. HIV-1 and hepatitis B transmission in Sudan. AIDS 1989;3:725-9.

9.Hyams KC, al-Arabi MA, al-Tagani AA, Messiter JF, al-Gaali AA, George JF. Epidemiology of hepatitis B in the Gezira region of Sudan. Am J Trop Med Hyg 1989;40:200-6.

10.Mudawi HM, Smith HM, Rahoud SA, Fletcher IA, Babikir AM, Saeed OK, Fedail SS. Epidemiology of HCV infection in Gezira state of central Sudan. Journal of medical virology 2007;79:383-5.

11.Mudawi H, Smith H, Fletcher I, Fedail S. Prevalence and common genotypes of HCV infection in Sudanese patients with hepatosplenic schistosomiasis. Journal of medical virology 2007;79:1322-4.

12.UNAIDS. Report on the global AIDS epidemic. http://www.unaids.org/globalreport/documents/20101123_GlobalReport_full_en.pdf. Accessed July 2013 2010.

13.Sulkowski MS, Thomas DL, Mehta SH, Chaisson RE, Moore RD. Hepatotoxicity associated with nevirapine or efavirenz-containing antiretroviral therapy: role of hepatitis C and B infections. Hepatology 2002;35:182-9.

14.WHO. WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children. Accessed September 2013. World Health Organization; 2006.

15.HIV/AIDS programme- World Health Organization: WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children.

.

16.Barth RE, Huijgen Q, Taljaard J, Hoepelman AI. Hepatitis B/C and HIV in sub-Saharan Africa: an association between highly prevalent infectious diseases. A systematic review and meta-analysis. International Journal of Infectious Diseases 2010;14:e1024-e31.

17.Shimelis T, Torben W, Medhin G, Tebeje M, Andualm A, Demessie F, Mulu A, Tegbaru B, Gebre-Selassie S. Hepatitis B virus infection among people attending the voluntary counselling and testing centre and anti-retroviral therapy clinic of St Paul's General Specialised Hospital, Addis Ababa, Ethiopia. Sexually transmitted infections 2008;84:37-41.

18.Bwogi J, Braka F, Makumbi I, Mishra V, Bakamutumaho B, Nanyunja M, Opio A, Downing R, Biryahwaho B, Lewis RF. Hepatitis B infection is highly endemic in Uganda: findings from a national serosurvey. African health sciences 2009;9.

19.Cunha L, Plouzeau C, Ingrand P, Gudo JPS, Ingrand I, Mondlane J, Beauchant M, Agius G. Use of replacement blood donors to study the epidemiology of major blood-borne viruses in the general population of Maputo, Mozambique. Journal of medical virology 2007;79:1832-40.

20.Hoffmann CJ, Thio CL. Clinical implications of HIV and hepatitis B co-infection in Asia and Africa. The Lancet infectious diseases 2007;7:402-9.

21.Burnett RJ, Francois G, Kew MC, Leroux-Roels G, Meheus A, Hoosen AA, Mphahlele MJ. Hepatitis B virus and human immunodeficiency virus co-infection in sub-Saharan Africa: a call for further investigation. Liver Int 2005;25:201-13.

22.Soriano V, Mocroft A, Peters L, Rockstroh J, Antunes F, Kirkby N, de Wit S, Monforte A, Flisiak R, Lundgren J. Predictors of hepatitis B virus genotype and viraemia in HIV-infected patients with chronic hepatitis B in Europe. J Antimicrob Chemother 2010;65:548-55.

23.Bodsworth NJ, Cooper DA, Donovan B. The influence of human immunodeficiency virus type 1 infection on the development of the hepatitis B virus carrier state. Journal of Infectious Diseases 1991;163:1138-40.

24.Hadler SC, Judson FN, O'Malley PM, Altman NL, Penley K, Buchbinder S, Schable CA, Coleman PJ, Ostrow DN, Francis DR. Outcome of hepatitis B virus infection in homosexual men and its relation to prior human immunodeficiency virus infection. Journal of Infectious Diseases 1991;163:454-7.

25.Yousif M, Mudawi H, Bakhiet S, Glebe D, Kramvis A. Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan. BMC infectious diseases 2013;13:328.

26.Oshitani H, Kasolo F, Mpabalwani M, Mizuta K, Luo N, Suzuki H, Numazaki Y. Prevalence of hepatitis B antigens in human immunodeficiency virus type 1 seropositive and seronegative pregnant women in Zambia. Transactions of the Royal Society of Tropical Medicine and Hygiene 1996;90:235-6.

27.Matthews GV, Ali RJ, Avihingsanon A, Amin J, Hammond R, Bowden S, Lewin SR, Sasadeusz J, Littlejohn M, Locarnini SL, Ruxrungtham K, Dore GJ. Quantitative HBsAg and HBeAg predict hepatitis B seroconversion after initiation of HAART in HIV-HBV coinfected individuals. PLoS One 2013;8:e61297.

28.Bell TG, Makondo E, Martinson NA, Kramvis A. Hepatitis B virus infection in human immunodeficiency virus infected southern African adults: occult or overt–that is the question. PloS one 2012;7:e45750.

29.Thio CL. Hepatitis B in the human immunodeficiency virus-infected patient: epidemiology, natural history, and treatment.Seminars in liver disease: Copyright© 2002 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel.:+ 1 (212) 584-4662; 2003, pp. 125-36.

30.Colin JF, Cazals-Hatem D, Loriot MA, Martinot-Peignoux M, Pham BN, Auperin A, Degott C, Benhamou JP, Erlinger S, Valla D. Influence of human immunodeficiency virus infection on chronic hepatitis B in homosexual men. Hepatology 1999;29:1306-10.

31.Kaplowitz N. Drug-induced liver injury. Clinical infectious diseases 2004;38:S44-S8.

32.Puoti M, Torti C, Ripamonti D, Castelli F, Zaltron S, Zanini B, Spinetti A, Putzolu V, Casari S, Tomasoni L, Quiros-Roldan E, Favret M, Berchich L, Grigolato P, Callea F, Carosi G. Severe hepatotoxicity during combination antiretroviral treatment: incidence, liver histology, and outcome. J Acquir Immune Defic Syndr 2003;32:259-67.

33.Hoffmann CJ, Charalambous S, Martin DJ, Innes C, Churchyard GJ, Chaisson RE, Grant AD, Fielding KL, Thio CL. Hepatitis B virus infection and response to antiretroviral therapy (ART) in a South African ART program. Clin Infect Dis 2008;47:1479-85.

34.Mphahlele MJ, Lukhwareni A, Burnett RJ, Moropeng LM, Ngobeni JM. High risk of occult hepatitis B virus infection in HIV-positive patients from South Africa. Journal of clinical virology 2006;35:14-20.

35.Allain J-P. Occult hepatitis B virus infection. Transfusion clinique et biologique 2004;11:18-25.

36.Raimondo G, Allain J-P, Brunetto MR, Buendia M-A, Chen D-S, Colombo M, Craxì A, Donato F, Ferrari C, Gaeta GB. Statements from the *Taormina* expert meeting on occult hepatitis B virus infection. Journal of hepatology 2008;49:652-7.

37.Melegari M, Bruno S, Wands JR. Properties of hepatitis B virus pre-S1 deletion mutants. Virology 1994;199:292-300.

38.Attia KA, Eholié S, Messou E, Danel C, Polneau S, Chenal H, Toni T, Mbamy M, Seyler C, Wakasugi N. Prevalence and virological profiles of hepatitis B infection in human immunodeficiency virus patients. World journal of hepatology 2012;4:218.

39.Shire NJ, Rouster SD, Rajicic N, Sherman KE. Occult hepatitis B in HIV-infected patients. JAIDS Journal of Acquired Immune Deficiency Syndromes 2004;36:869-75.

40.Hu KQ. Occult hepatitis B virus infection and its clinical implications. Journal of viral hepatitis 2002;9:243-57.

41.Zollner B, Petersen J, Puchhammer-Stockl E, Kletzmayr J, Sterneck M, Fischer L, Schroter M, Laufs R, Feucht HH. Viral features of lamivudine resistant hepatitis B genotypes A and D. Hepatology 2004;39:42-50.

42.Benhamou Y, Bochet M, Thibault V, Di Martino V, Caumes E, Bricaire F, Opolon P, Katlama C, Poynard T. Long-term incidence of hepatitis B virus resistance to lamivudine in human immunodeficiency virus-infected patients. Hepatology 1999;30:1302-6.

43.Kim HN, Scott J, Cent A, Cook L, Morrow RA, Richardson B, Tapia K, Jerome KR, Lule G, John-Stewart G, Chung MH. HBV lamivudine resistance among hepatitis B and HIV coinfected patients starting lamivudine, stavudine and nevirapine in Kenya. J Viral Hepat 2011;18:e447-52.

44.HIV/AIDS programme- World Health Organization (WHO). Antiretroviral therapy for HIV infection in adults and adolescents. Recommendations for a public health approach, 2010 revision. .

45.Boyd A, Lacombe K, Miailhes P, Gozlan J, Bonnard P, Molina JM, Lascoux-Combe C, Serfaty L, Gault E, Desvarieux M, Girard PM. Longitudinal evaluation of viral interactions in treated HIV-hepatitis B co-infected patients with additional hepatitis C and D virus. J Viral Hepat 2010;17:65-76.

46.McCarthy MC, Burans JP, Constantine NT, el-Hag AA, el-Tayeb ME, el-Dabi MA, Fahkry JG, Woody JN, Hyams KC. Hepatitis B and HIV in Sudan: a serosurvey for hepatitis B and human immunodeficiency virus antibodies among sexually active heterosexuals. Am J Trop Med Hyg 1989;41:726-31.

Table 1: Demographic criteria and risk factors for HBV exposure in 358 HIV-positive Sudanese adults

Variable		Number of patients tested (358)	Number positive for anti-HBc IgG (225)	Number negative for Anti-HBc IgG	P value
Gender	Male	203 (56.7%)	139 (61.8%)	64(48.1%)	0.012*
	Female	155 (43.3%)	86(38.2%)	69(51.9%)	
Marital status	Married	214 (59.8%)	136(60.4%)	78(58.6%)	0.896
	Single	82(22.9%)	50(22.2%)	32(24.1%)	
	Divorced /widowed	62(17.3%)	39(18.8%)	23(17.3%)	
Multiple sexual partners	Yes	144 (40.2%)	95(42.2%)	49(36.8%)	0.316
	No	214(59.8%)	130(57.8%)	84(63.2%)	
Surgery	Yes	65(18.2%)	37(16.4%)	28(21%)	0.274
	No	293(81.8%)	188(83.6%)	105(79%)	
Blood transfusion	Yes	38(10.6%)	20(8.9%)	18(13.5%)	0.168
	No	320(89.4%)	205(91.1%)	115(86.5%)	
Dental procedures	Yes	92(25.7%)	54(24.0%)	38(28.6%)	0.339
	No	266(74.3%)	171(76.0%)	95(71.4%)	
Intravenous drug use	Yes	5(1.4%)	1(0.4%)	4(3%)	0.046*
	No	353(98.6%)	224(99.6%)	129(97%)	
History of Jaundice	Yes	111(31.0%)	78(34.7%)	33(24.8%)	0.51
	No	247(69.0%)	147(65.3%)	100(75.2%)	
HBV vaccination	Yes	4(1.1%)	2 (0.9%)	2(1.5%)	0.593
	No	354(98.9%)	223(99.1%)	131(98.5%)	
WHO classification	Stage 1&2	126(35.2%)	66(29.3%)	60(45.1%)	0.168
	Stage 3&4	232(64.8%)	159(70.7%)	73(54.9%)	

* Significant risk factor

Table 2: Comparison between patients who were HBV DNA positive and patients who were HBV DNA negative on PCR, among 358 HIV infected patients

Variable		Number of patients tested (358)	HBV DNA positive (96)	HBV DNA negative (262)	P value
Gender	Male	203 (56.7%)	59 (61.5%)	144(55%)	0.272
	Female	155 (43.3%)	37(38.5%)	118 (45%)	
Marital status	Married	214 (59.8%)	62 (64.6%)	152(58%)	0.342
	Single	82(22.9%)	24(25%)	58(22.1%)	
	Divorced/widowed	62(17.3%)	10(10.4%)	52(19.8%)	
Multiple sexual partners	Yes	144 (40.2%)	42(43.8%)	102(38.9%)	0.410
	No	214(59.8%)	54(56.2%)	160(61.1%)	
History of surgery	Yes	65(18.2%)	17(17.7%)	48(18.3%)	0.894
	No	293(81.8%)	79(82.3%)	214(81.7%)	
Blood transfusion	Yes	38(10.6%)	5(5.2%)	33(12.6%)	0.044*
	No	320(89.4%)	91(94.8%)	229(87.4)	
Dental procedures	Yes	92(25.7%)	16(16.7%)	76(29%)	0.018*
	No	266(74.3%)	80(83.3%)	186(71%)	
Intravenous drug use	Yes	5(1.4%)	0	5(1.9%)	0.173
	No	353(98.6%)	96 (100%)	257(98.1%)	
History of Jaundice	Yes	111(31.0%)	38(39.6%)	73(27.9%)	0.034*
	No	247(69.0%)	58(60.4%)	189(72.1%)	
HBV vaccination	Yes	4(1.1%)	1 (1%)	3(1.1%)	0.934
	No	354(98.9%)	95(99%)	259(98.9%)	

WHO classification	Stage 1 &2	126(35.2%)	23(24%)	103(39.3%)	0.037*
	Stage 3&4	232(64.8%)	73(76%)	159(60.7%)	

* Significant risk factor

Table 3: Comparison between HBsAg-positive and HBsAg-negative patients among 96 HIV infected patients with detectable HBV DNA on PCR

Variable		Number of patients tested (96)	HBsAg positive (42)	HBsAg negative (54)	P value
Gender	Male	59 (61.5%)	24(57.1%)	35(64.8%)	0.444
	Female	37(38.5%)	18(42.9%)	19(35.2%)	
Marital status	Married	62 (64.6%)	28(66.7%)	34(63%)	0.819
	Single	24(25%)	11(26.2%)	13(24%)	
	Divorced/widowed	10(10.4%)	3(7.2%)	7(13%)	
Multiple sexual partners	Yes	42(43.8%)	18(42.9%)	24(44.4%)	0.876
	No	54(56.2%)	24(57.1%)	30(55.6%)	
Surgery	Yes	17(17.7%)	8(19%)	9(16.7%)	0.762
	No	79(82.3%)	34(81%)	45(83.3%)	
Blood transfusion	Yes	5(5.2%)	3(7.1%)	2(3.7%)	0.452
	No	91(94.8%)	39(92.9%)	52(96.3%)	
Dental procedures	Yes	16(16.7%)	9(21.4%)	7(13%)	0.270
	No	80(83.3%)	33(78.6%)	47(87%)	
Intravenous drug use	Yes	0	0	0	-
	No	96 (100%)	42(100%)	54(100%)	
History of jaundice	Yes	38(39.6%)	23(54.8%)	15(27.8%)	0.007*
	No	58(60.4%)	19(45.2%)	39(72.2%)	
HBV vaccination	Yes	1(1%)	1(2.4%)	0	0.254
	No	95(99%)	41(97.6%)	54(100%)	

WHO classification	Stage 1 &2	23(24%)	11(26.2%)	12(22.2%)	0.947
	Stage 3 &4	73(76%)	31(73.8%)	42(77.8%)	

* Significant risk factor

Table 4: Median HBV viral loads in 96 HIV infected patients with detectable HBV DNA on PCR

HBsAg status	Number	Median viral load (log10IU/ml)	Interquartile range IQR (log10IU/ml)	P value
HBsAg-positive	42	3.2	3.3 - 6.6	0.000*
HBsAg-negative	54	1.7	1.3 -2.7	

* Significant risk factor

Appendix B

Ethical Clearances

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Mr Mukhlid Yousif

CLEARANCE CERTIFICATE	**M10914**
PROJECT	Molecular Characterization of Hepatits B Virus (HBV) Mono-Infected and from HBV/Human Immunodeficiency Virus (HIV) Co-Infected Individuals in Sudan
INVESTIGATORS	Mr Mukhlid Yousif.
DEPARTMENT	Department of Internal Medicine
DATE CONSIDERED	01/10/2010
DECISION OF THE COMMITTEE*	Approved unconditionally

Unless otherwise specified this ethical clearance is valid for 5 years and may be renewed upon application.

DATE 03/11/2010 **CHAIRPERSON** .....
(Professor PE Cleaton-Jones)

*Guidelines for written 'informed consent' attached where applicable
cc: Supervisor : Prof Anna Kramvis

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10004, 10th Floor, Senate House, University.
I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to a completion of a yearly progress report.**

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES...

Federal Ministry of Health		وزارة الصحة الاتحادية
General Directorate of Planning & Health Development		الإدارة العامة للتخطيط والتنمية الصحية
Research Directorate		إدارة البحوث

Date 11/11/2008 No: fmoh/rd/EC/75/9/08

ETHICAL CLEARANCE CERTIFICATE

This is to certify that the proposal **(No.75-9-08)** entitled **(Characterization of Hepatitis B virus genotypes in Sudan)** introduced by **Dr. Hatim M. Y. Mudawi** from **University of Khartoum-Faculty of Medicine - Department of Internal Medicine,** has been approved by the **National Health Research Ethics Committee, Federal Ministry of Health** to be carried out in the **Sudan.**

Dr. Iman Abdalla Mustafa
Reporter of the
National Research Ethics Review Committee

السودان الخرطوم. وزارة الصحة الاتحادية تقاطع شارع عثمان دقنة مع شارع النيل.الرمز البريدي: 11111. ص.ب 303
Tel: +249157845773 e-mail: fmohresearch@yahoo.com

Faculty of Medicine
University of Khartoum
Research Board
P. O. Box: 102
Khartoum, Sudan

بسم الله الرحمن الرحيم

كلية الطب
جامعة الخرطوم
محلس ابحاث الكلية
ص. ب: ١٠٢ الخرطوم - السودان

Date: 8/9/2008
Ref.: FOM/D/staff member

Dr. Hatim Mudawi,
Department of Medicine,

Dear Dr. Mudawi

Subject: Ethical Clearance of your project titled: Characterization of Hepatitis B genotypes in Sudan

The Ethical Committee had reviewed your proposal title "Characterization of Hepatitis B genotypes in Sudan" and following in careful review of the submitted proposal, the committee has approved the proposal based on its fulfillment of the ethical guideline of the Faculty.

I am pleased to inform yoou that your proposal is ethically approved to start your study.

Dr. Ahmed Mohamedain Eltom
Secretary, Faculty Research Board

UNIVERSITY OF KHARTOUM
FACULTY OF MEDICINE
- 8 SEP 2008
DEAN'S OFFICE

Appendix C

Informed Consents

Department of Internal Medicine, Faculty of Medicine

University of Khartoum

Study on Characterization of HBV genotyping in Sudanese patients

Consent form for participation in the study

We are Researchers from the department of Internal Medicine, Faculty of Medicine, University of Khartoum. We are conducting a study on the characterization of hepatitis B virus (HBV). We will collect a blood sample to assess whether you have been exposed to HBV infection in the past. All data will remain confidential and will not be accessible except for the main investigators and co- investigators. The samples will be analyzed in Sudan and in partner labs in South Africa. We will inform you about the result of the study after its completion .We will not use these samples in any other study without your agreement. You will not get any payment to participate in this study and your participation is completely voluntarily and you could drop out of the study at any time. If you have any questions please contact Dr. Hatim Mudawi, Faculty of Medicine, University of Khartoum on Tel. 0912202600

I…… read and understood the content of this form ,and voluntarily want to participate in this study

Date _____Signature _____

Interviewer ……..

Date _____Signature _____

Department of Internal Medicine, Faculty of Medicine

University of Khartoum

Study on Characterization of HBV genotyping in Sudanese subjects infected with the human immunodeficiency virus

Consent form for participation in the study

We are Researchers from the department of Internal Medicine, Faculty of Medicine, University of Khartoum. We are conducting a study on the characterization of hepatitis B virus (HBV) in patients with HIV infection. We will collect a blood sample to assess whether you have been exposed to HBV infection in the past. All data will remain confidential and will not be accessible except for the main investigators and co- investigators. The samples will be analyzed in Sudan and in partner labs outside Sudan. We will inform you about the result of the study after its completion .We will not use these samples in any other study without your agreement. You will not get any payment to participate in this study and your participation is completely voluntarily and you could drop out of the study at any time. If you have any questions please contact Dr. Hatim Mudawi, Faculty of Medicine, University of Khartoum on Tel. 0912202600

I…… read and understood the content of this form ,and voluntarily want to participate in this study

Date _____Signature _____

Interviewer ……..

Date _____Signature _____

Appendix D

Clinical Report Form

University Of Khartoum

Faculty of Medicine-Department of Medicine

Questionnaire

Project title: HBV genotyping in Sudan

Number: ☐☐☐

Sex: **Male** ☐ **Female** ☐

Age: ☐☐ **Occupation: ……**

Residence: ……

Tribe: ……

Marital status: ……

Phone No: ……..

Parental therapy for Malaria: **Yes** ☐ **No** ☐

Parental therapy for Schistosomiasis: **Yes** ☐ **No** ☐

History of hospitalization: **Yes** ☐ **No** ☐

(If yes) Date and condition of admission:……

History of surgery: **Yes** ☐ **No** ☐

(If yes) Date and condition of Surgery:……

History of haematemesis/maelana: Yes ☐ **No** ☐

History of blood transfusion: **Yes** ☐ **No** ☐

Previous dental treatment: **Yes** ☐ **No** ☐

History of Jaundice: **Yes** ☐ **No** ☐

(If yes) Date of incidence:……

History of Alcohol intake: **Yes** ☐ **No** ☐

(If yes)……

History of vaccination to HBV: **Yes** ☐ **No** ☐

(If yes) Date of vaccination:…….

Clinical examination:

Pale: **Yes** ☐ **No** ☐

Jaundice: **Yes** ☐ **No** ☐

Tattoos/Scarification: **Yes** ☐ **No** ☐

Clubbing: **Yes** ☐ **No** ☐

Palmar erythema: **Yes** ☐ **No** ☐

Dupuytren's contractures: **Yes** ☐ **No** ☐

Gynaecomastia: **Yes** ☐ **No** ☐

Liver:……cm **Enlarged:** ☐ **Not enlarged** ☐

Spleen:……..cm **Enlarged:** ☐ **Not enlarged** ☐

Ascitis: **Yes** ☐ **No** ☐

LL odema: **Yes** ☐ **No** ☐

***Investigation*:**

U/S:……

……

……

……

ALT:……

Albumin:……..

Bilirubin: ……

Platelet counts:……

PT:…… Control:…… INR:…….

Appendix E

1. Table 1: Primer sequences and PCR conditions used throughout the study

Name	Position	Nuclotide sequence	orientation	PCR round	Thermocycling conditions			
					Denaturation	Annealing	extension	cycles
2410(+)	2401-2439	5'-TCA ATC GCC GCG TCG CAG AAG ATC TCA ATC-3'	Forward	First	94 °C for 1 min	66 °C for 1 min	72 °C for 3 min	40
1314(-)	1314-1291	5'-TCC AGA CCK GCT GCG AGC AAA ACA-3'	Reverse					
2451(+)	2451-2482	5-'AAT GTT AGT ATT CCT TGG ACT CAT AAG GTG GG-3'	Forward	Second	94 °C for 1 min	65 °C for 1 min	72 °C for 3 min	40
1280(-)	1286-1254	5-'AGT TCC GCA GTA TGG ATC GGC AGA GGA-3'	Reverse					
1606(+)	1606-1625	5'-GCA TGG AGA CCA CCG TGA AC-3'	Forward	First	94 °C for 1 min	55 °C for 1 min	72 °C for 2 min	30
1974(-)	1955-1974	5'-GGA AAG AAG TCA GAA GGC AA-3'	Reverse					
1653(+)	1653-1672	5'-CAT AAG AGG ACT CTT GGA CT-3'	Forward	Second	94 °C for 1 min	55 °C for 1 min	72 °C for 2 min	30
1959(-)	1940-1959	5'-GGC AAA AAC GAG AGT AAC TC-3'	Reverse					
255(+)	253-275	5'-CTC GTG GTG GAC TTC TCT CAA TT-3'	Forward	First	94 °C for 30 sec	56 °C for 50 sec	72°C for 1 min	40
759(-)	759-739	5'-CCC CAA TAC CAC ATC ATC CAT-3'	Reverse					
459(+)	460-480	5'-TAT GTT GCC CGT TTG TCC TCT-3'	Forward	Second	94 °C for 30 sec	56 °C for 50 sec	72 °C for 50 sec	40
710(-)	711-691	5'-AGC CCT ACG AAC CAC TGA ACA-3'	Reverse					
P1	1821-1841	5'-TTT TTC ACC TCT GCC TAA TCA-3'	Forward	NA	98 °C for 10 sec	57 °C for 30 sec	72 °C for 1 min	35
P2	1825-1806	5-'AAA AAG TTG CAT GST GST GG-3'	Reverse					
230(+)	230 -249	5'-TCA CAA TAC CGC AGA GTC T-3'	Forward	First	94 °C for 1 min	55 °C for 1 min	72 °C for 2 min	40
800(-)	820-800	5'-AAC AGC GGT ATA AAG GGA CT-3'	Reverse					
P7	25 -278	5'-GTG GTG GAC TTC TCT CAA TTT TC-3'	Forward	Second	94 °C for 1 min	55 °C for 1 min	72 °C for 2 min	40
P8	796-776	5'-CGG TAT AAA GGG ACT CAC GAT-3'	Reverse					

2. Table 2: Sequences of sequencing Primers used throughout the study

Name	Position	Nucleotide sequence	Orientation
2497F	2497-2516	5'-TTC CTT GGA CTC ATA AGG TG-3'	Forward
3188F	3188-3206	5'-AGT CAG GAA GGC AGC CTA C-3'	Forward
591F	591-611	5'-ATT GCA CCT GTA TTC CCA TCC-3'	Forward
1069F	1069-1088	5'- TGT ATT CAA TCT AAG CAG GC-3'	Forward
2022F	2022-2041	5'-CCT TAG AGT CTC CTG AGC AT-3'	Forward
Post-Bgl III	2000-2017	5'-CCG ATA CAG AGC TGA GGC-3'	Reverse
2480R	2461-2480	5'-CAC CTT ATG AGT CCA AGG AA-3'	Reverse
BCPS1	1698-1717	5’-GAG GCA TAC TTC AAA GAC TG-3’	Forward
BCPS2	1928-1947	5’-AGT AAC TCC ACA GTA GCT CC -3’	Reverse

Appendix F

Laboratory Protocols

Complete genome amplification

The Expand High Fidelity PCR system (Roche Diagnostics, Mannheim, Germany) was used for the amplification of complete HBV genomes. The reaction mixture consisted of 10 µl of 5X phusion Expand High Fidelity (HF) buffer (7.5 mM $MgCl_2$, 1 µl of 10 mM dNTP mix), 3 µl of 10 µM primer P1 and P2, 2.5 µl of 100% Dimethyl Sulfoxide (DMSO) and 5 µl of extracted DNA and 45 µl with distilled water. The enzyme mixture constituted 4.5 µl 5X HF buffer mixed with 0. 5 µl of phusion Expand HF DNA polymerase enzyme. The reaction was preheated to 98° C for 1 minute and cooled to 80°C for 30 seconds at which time 5 µl of enzyme mixture was added. This was followed by heating to 98° C for 1 minute. Thirty five cycles were performed, with PCR conditions as indicated in Appendix E. Both HBV-positive and -negative controls were included, the latter consisting of best quality water (BQW) instead of DNA in the PCR reaction mixture. The reaction was carried out in a Techgene programmable thermo-cycler (Techne Inc., Princeton, NJ).

Complete S region amplification

Nested PCR was performed on the complete surface gene using HotStarTaq Master Mix Kit (QIAGEN-Netherlands). The reaction mixture for the first round consisted of 12.5 µl of HotStarTaq master mix (250 units HotStarTaq DNA Polymerase, PCR Buffer with 3 mM MgCl2, and 400 µM of each dNTP), 0.375 µl each of 10 µM primer 2410 (+) and 1314 (-), 2.5 µl of DNA, with the reaction topped up to 25 µl with distilled water. The reaction mixture

for the second round consisted of 25 µl HotStar Taq master mix, 0.75 µl of each 10 µM primer 2451 (+) and 1280 (-), 5 µl of first round PCR product and topped up to 50 µl with distilled water. For primer sequences and PCR conditions, see Appendix E.

BCP/PC region amplification

Nested PCR was performed on the BCP/PC region (1742 to 1900 from the *Eco*R1 site) using BIOLINE PCR kits (London, UK). The reaction mixture for the first round consisted of 2.5 µl of 10X NH4 PCR buffer, 0.75 µl of 50mM $MgCl_2$, 2 µl of 10 mM dNTPs, 1.25 µl of each 10 µM primer 1606 (+) and 1974 (-), 0.1 µl of Taq DNA polymerase, 2.5 µl of DNA and topped up to 25 µl with distilled water. The reaction mixture for the second round consisted of 5 µl of 10X NH4 PCR buffer, 1.5 µl of 50mM $MgCl_2$, 4 µl of 10mM dNTPs, 2.5 µl of each 10 µM primer 1606 (+) and 1974 (-), 0.2 µl of Taq DNA polymerase, 5 µl of DNA and topped up to 50 µl with distilled water. For primer sequences and PCR conditions, see Appendix E

Short S PCR

Nested PCR was performed on a short region in the S region (459-710 from *Eco*R1 site) using BIOLINE PCR kits. The reaction mixture for the first round consisted of 2.5 µl of 10X NH4 PCR buffer, 0.75 µl of 50mM $MgCl_2$, 2 µl of 10mM dNTPs, 1.25 µl of each 10 µM primer 255 (+) and 759 (-), 0.1 µl of Taq DNA polymerase, 2.5 µl of DNA and topped up to 25 µl with distilled water. The reaction mixture for the second round consisted of 5 µl of 10X NH4 PCR buffer, 1.5 µl of 50mM $MgCl_2$, 4 µl of 10mM dNTPs, 2.5 µl of each 10 µM primer 459 (+) and 710 (-), 0.2 µl of Taq DNA polymerase,5 µl of DNA and topped up to 50 µl with distilled water. For primer sequences and PCR conditions, see Appendix E

Detection of amplicons (Gel Electrophoresis)

A 10 µl aliquot of the amplicon was analyzed by agarose gel electrophoresis. The percentage of the agarose used varied between 0.8% and 3%, depending on the expected size of amplicon. Commercially available DNA ladders of 100 bp or 1 kb (Promega Madison, USA) were used to determine the size of the amplicon. HBV DNA from HBsAg-positive, DNA-positive patients was used as a positive control. BQW or DNA extracted from HBsAg-negative patients was used as a negative control. Gels were electrophoresed in 1X TBE (Appendix G) at 70 mA for 30-60 minutes. Ethidium bromide (Appendix G) was included in the preparation of the gel at a final concentration of 1 µg/ml. This allowed for viewing of DNA under UV illumination. The PCR products were stored at -20° C until sent for sequencing or further analysis.

Restriction Fragment Length Polymorphsim (RFLP) PCR

Nested PCR was performed to amplify the region 256 to 796 from the *Eco*R1 site, using BIOLINE PCR kits. The reaction mixture for the first round consisted of 2.5 µl of 10X NH4 PCR buffer, 0.75 µl of 50mM $MgCl_2$, 2 µl of 10mM dNTPs, 1.25 µl of each 10 µM primer 230(+) and 800(-), 0.1 µl of Taq DNA polymerase, 2.5 µl of DNA and topped up to 25 µl with distilled water. The reaction mixture for the second round consisted of 5 µl of 10X NH4 PCR buffer, 1.5 µl of 50mM $MgCl_2$, 4 µl of 10mM dNTPs, 2.5 µl of each 10 µM P7 and P8, 0.2 µl of Taq DNA polymerase, 5 µl of DNA and topped up to 50 µl with distilled water. For primer sequences and PCR conditions, see Appendix E

Digestion

RFLP analysis of the region 256 to 796 from *Eco*R1 site was undertaken on amplicons to determine the HBV genotype (Lindh *et al.,* 1997). Fifteen µl of the PCR product was

dispensed into two separate 0.5 ml PCR tubes. To one of these tubes, 0.5 µl of 5 U *Hinf*I restriction enzyme (New England Biolabs, Inc, Ipswich, Mass), 2.0 µl NEBuffer and 7.5 µl of BQW was added, while to the other, 0.5 µl of 5 U *Tsp*509I restriction enzyme (New England Biolabs), 2.0 µl NEBuffer and 7.5 µl BQW was added. Each reaction was overlaid with a thin layer of mineral oil before being incubated for 3 hours at 37 °C (for the *Hinf*I restriction) or 65 °C (for the *Tsp*509I restriction). The reaction was terminated by adding 4 µl gel loading dye (Promega). The entire restriction reaction was loaded on a 3% agarose gel containing 10 mg/ml ethidium bromide (Bio Basics Inc., Markham, Ontario) at ~3.5 volts/cm along with 10 µl of a 100 bp DNA ladder (Promega) and 10 µl of a 50 bp DNA ladder (Promega) until the bromophenol blue tracking dye had migrated across approximately 80% of the length of the gel. An image of the gel was recorded using the Bio-Rad Gel Doc XR Imaging System (Bio-Rad Laboratories Inc., Hercules, Calif.) and the sizes of the restricted fragments compared to those in table 3 of (Lindh *et al.*, 1997) to determine the genotype of the samples.

Cloning

The amplified region 1653 to 1959 from *Eco*R1 site was cloned with the InsTAclone™ PCR Cloning Kit. The final elongation step of the second round of the nested PCR amplification was increased from 10 minutes to 30 minutes to increase the number of 3'-A overhangs on the ends of the amplicons. The amplicons were gel-purified using the QIAGEN QIAquick Gel Extraction Kit according to the manufacturer's instructions. A 1.0 µl aliquot of this elute was resolved on a 1% agarose gel together with a O'GeneRuler 1 kb DNA ladder (Fermentas). This molecular weight marker is composed of fragments of known concentration and the concentration of the eluted DNA can be determined by comparing the intensity of the resolved fragment to the intensities of the fragments in the molecular weight marker. This concentration was used to determine the volume of extracted DNA equivalent to

the 0.52 pmol ends required for the cloning reaction which was set up according to the manufacturer's instructions and incubated overnight at 4° C. A 4.0 µl aliquot of the cloning reaction was transformed into One Shot® TOP10 chemically competent cells (Invitrogen, Life Technologies, United States of America), by adding by the whole reaction to 250 SOC medium (Sigma-Aldrich Inc., United States of America). Three different volumes of the transformation reaction (50 µl, 100 µl and 150 µl) were spread onto pre-warmed LB agar plates containing 50 µg/ml ampicillin (Appendix G) and X-gal and IPTG to differentiate between the positive and negative colonies (white for positive and blue for negative) and incubated overnight at 37° C. On the following day, individual colonies were selected and cultured overnight at 37° C in 5 ml Luria-Bertani (LB) medium containing either kanamycin or ampicillin, as required.

On the following day, the cracking procedure was undertaken, by centrifugation of the harvest of 500 µl of bacterial suspension with 100 µl cracking buffer, which was then loaded onto a gel. The presence of BCP and vector DNA (about 3.1 kb) indicated a positive clone. This was followed by Plasmid DNA extraction using the remaining 4 ml culture using the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich Inc., United States of America) according to the manufacturer's instructions. Finally, the RFLP technique was followed to separate the BCP product from the vector. A 3.0 µl aliquot of extracted plasmid DNA was restricted at 37° C for 1 hour with 20 U *Eco*R1 (New England BioLabs) with 1 µl NEBuffer *Eco*R1 and 4.0 µl of BQW. The restricted products were resolved overnight on a 1% agarose gel together with a 1 kb DNA ladder (Promega) to determine whether the plasmids contained the fragment of interest. The plasmid has a length of 2.8 kb and the BCP product has a length of 300 bases.

Appendix G

Solutions

10 x TBE buffer PH 8.3

108 g Tris base

55 g Boric acid

9,3 g EDTA

Adjust volume to 1000 ml with distilled water.

Sterilise by autoclaving and store at room temperature.

1% Agarose gel

1 g Agarose

100 ml 1 x TBE buffer

Dissolve agarose in 1 x TBE buffer by heating in a microwave oven for ~4 minutes, swirling occasionally, until all the agarose particles are completely dissolved. Care must be taken that especially the high percentage gels do not boil over.

Cool to ~60 °C.

Add 10µl mg / ml Ethidium Bromide and swirl to mix.

Pour into a prepared casting tray and allow to solidify completely at room temperature.

Luria-Bertani (LB) medium

10 g Tryptone

5 g Yeast

10 g Sodium chloride

Adjust volume to 1000 ml with distilled water.

Sterilize by autoclaving.

Add 1 ml of either 50 mg/ml ampicillin or 50 mg/ml kanamycin as required.

Store at 4 °C.

LB agar plates

10 g Tryptone

5 g Yeast

10 g Sodium chloride

15 g Bacteriological agar

Adjust volume to 1000 ml with distilled water.

Sterilize by autoclaving.

Cool to approximately 50° C before adding 1 ml of either 50 mg/ml ampicillin or 50 mg/ml kanamycin as required. After pouring, allow plates to solidify completely at room temperature before storing at 4° C.

75% Ethanol

To a RNase free, sterile bottle, add:

75 ml Absolute ethanol

25 ml sterile water

Store at -20° C.

X-gal

5-bromo-4-chloro-3-indolyl-_-D-galactoside (X-gal)

50 mg/ml in dimethylformamide stored at -20°C.

IPTG stock solution (0.5 M)

6 g isopropyl _-D-thiogalactopyranoside (IPTG) in 50 ml of deionised water.

Filter sterilized with a 0.2 μm filter store at -20°C.

Ampicillin

100 mg/ml in deionised water (filter sterilize)

1% cracking buffer

SDS (2.5 g)

2 mM EDTA Disodium Salt (0.19 g)

400 mM Sucrose (34.22 g)

50 mM Tris-HCl, pH 6.8 (125 ml)

100 μg/ml RNase A, DNase- and Protease-Free (2.5 μl)

0.01% (w/v) Bromophenol Blue Sodium Salt for Electrophoresis (0.025 g)

0.01% (w/v) Xylene Cyanol FF (0.025 g)

UltraPure™ distilled water, deoxyribonuclease (DNase)- and RNase free was added to make up to a final volume of 250 ml, after which it was stored at room temperature until further used.

5X loading dye

200 μl 5X loading buffer

100 μl _-mercaptoethanol

100 μl Bromophenol Blue (saturated solution)

10 mM dNTP Solution

10 mM dATP (10 μl), 10 mM dCTP (10 μl), 10 mM dGTP (10 μl), 10 mM dTTP (10 μl)

The solution was made up to a final volume of 100 µl with BQW. The solution was briefly vortexed, and centrifuged, after which the solution was stored at -20 °C until further used. All four reagents were purchased as a dNTP Set (Bioline Reagents Ltd., London, United Kingdom)

References

ABBAS, A. A., GABO, N. E. A. A. A., BABIKER, Z. O. E. & HERIEKA, E. A. M. 2010. Paediatric HIV in central Sudan: High sero-prevalence and poor performance of clinical case definitions. *Journal of Clinical Virology,* 47**,** 82-84.

ABDELRAHIM, M. S. 2010. HIV prevalence and risk behaviors of female sex workers in Khartoum, north Sudan. *AIDS,* 24**,** S55-S60.

ABDOU CHEKARAOU, M., BRICHLER, S., MANSOUR, W., LE GAL, F., GARBA, A., DÉNY, P. & GORDIEN, E. 2010. A novel hepatitis B virus (HBV) subgenotype D (D8) strain, resulting from recombination between genotypes D and E, is circulating in Niger along with HBV/E strains. *Journal of General Virology,* 91**,** 1609-1620.

ABOU, M., ELTAHIR, Y. M. & ALI, A. S. 2009. Seroprevalence of Hepatitis B virus and Hepatitis C virus among blood donors in Nyala, South Dar Fur, Sudan. *Virol J,* 6**,** 146.

AHMADIAN, A., EHN, M. & HOBER, S. 2006. Pyrosequencing: history, biochemistry and future. *Clinica chimica acta,* 363**,** 83-94.

AHN, S. H., KRAMVIS, A., KAWAI, S., SPANGENBERG, H. C., LI, J., KIMBI, G., KEW, M., WANDS, J. & TONG, S. 2003. Sequence variation upstream of precore translation initiation codon reduces hepatitis B virus e antigen production. *Gastroenterology,* 125**,** 1370-1378.

ALLAIN, J.-P. 2004. Occult hepatitis B virus infection. *Transfusion Clinique et Biologique,* 11**,** 18-25.

ALTER, M. J. 2003. Epidemiology of hepatitis B in Europe and worldwide. *J Hepatol,* 39 Suppl 1**,** S64-9.

ALTER, M. J. 2006. Epidemiology of viral hepatitis and HIV co-infection. *Journal of Hepatology,* 44**,** S6.

ANDERNACH, I. E., HUBSCHEN, J. M. & MULLER, C. P. 2009a. Hepatitis B virus: the genotype E puzzle. *Rev Med Virol,* 19**,** 231-40.

ANDERNACH, I. E., NOLTE, C., PAPE, J. W. & MULLER, C. P. 2009b. Slave trade and hepatitis B virus genotypes and subgenotypes in Haiti and Africa. *Emerg Infect Dis,* 15**,** 1222-1228.

ANDRE, F. 2000. Hepatitis B epidemiology in Asia, the Middle East and Africa. *Vaccine,* 18 Suppl 1**,** S20-2.

ARAUZ-RUIZ, P., NORDER, H., ROBERTSON, B. H. & MAGNIUS, L. O. 2002. Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. *Journal of General Virology,* 83**,** 2059-2073.

ARAUZ-RUIZ, P., NORDER, H., VISONA, K. A. & MAGNIUS, L. O. 1997. Genotype F prevails in HBV infected patients of hispanic origin in Central America and may carry the precore stop mutant. *Journal of Medical Virology,* 51**,** 305-312.

AUDSLEY, J., LITTLEJOHN, M., YUEN, L., SASADEUSZ, J., AYRES, A., DESMOND, C., SPELMAN, T., LAU, G., MATTHEWS, G. V., AVIHINGSANON, A., SEABERG, E., PHILP, F., SAULYNAS, M., RUXRUNGTHAM, K., DORE, G. J., LOCARNINI, S. A., THIO, C. L., LEWIN, S. R. & REVILL, P. A. 2010. HBV mutations in untreated HIV-HBV co-infection using genomic length sequencing. *Virology,* 405**,** 539-547.

BANCROFT, W. H., SNITBHAN, R., SCOTT, R. M., TINGPALAPONG, M., WATSON, W. T., TANTICHAROENYOS, P., KARWACKI, J. J. & SRIMARUT, S. 1977. Transmission of hepatitis B virus to gibbons by exposure to human saliva containing hepatitis B surface antigen. *J Infect Dis,* 135**,** 79-85.

BANERJEE, A., KURBANOV, F., DATTA, S., CHANDRA, P. K., TANAKA, Y., MIZOKAMI, M. & CHAKRAVARTY, R. 2006. Phylogenetic relatedness and genetic diversity of hepatitis B virus isolates in Eastern India. *Journal of Medical Virology,* 78**,** 1164-1174.

BAPTISTA, M., KRAMVIS, A. & KEW, M. C. 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**,** 946-53.

BAUMERT, T. F., MARRONE, A., VERGALLA, J. & LIANG, T. J. 1998. Naturally occurring mutations define a novel function of the hepatitis B virus core promoter in core protein expression. *J Virol,* 72**,** 6785-95.

BAUMERT, T. F., ROGERS, S. A., HASEGAWA, K. & LIANG, T. J. 1996. Two core promotor mutations identified in a hepatitis B virus strain associated with fulminant hepatitis result in enhanced viral replication. *J Clin Invest,* 98**,** 2268-76.

BEASLEY, R. P. & HWANG, L.-Y. 1983. Postnatal infectivity of hepatitis B surface antigen-carrier mothers. *Journal of Infectious Diseases,* 147**,** 185-190.

BECK, J. & NASSAL, M. 2007. Hepatitis B virus replication. *World J Gastroenterol,* 13**,** 48-64.

BELL, T. G., MAKONDO, E., MARTINSON, N. A. & KRAMVIS, A. 2012. Hepatitis B Virus Infection in Human Immunodeficiency Virus Infected Southern African Adults: Occult or Overt–That Is the Question. *Plos One,* 7**,** e45750.

BIGGAR, R. J., GOEDERT, J. J. & HOOFNAGLE, J. 1987. Accelerated loss of antibody to hepatitis B surface antigen among immunodeficient homosexual men infected with HIV. *The New England journal of medicine,* 316**,** 630.

BODSWORTH, N. J., COOPER, D. A. & DONOVAN, B. 1991. The influence of human immunodeficiency virus type 1 infection on the development of the hepatitis B virus carrier state. *Journal of Infectious Diseases,* 163**,** 1138-1140.

BONHOEFFER, S. & SNIEGOWSKI, P. 2002. Virus evolution: the importance of being erroneous. *Nature,* 420**,** 367-369.

BOSCH, V., BARTENSCHLAGER, R., RADZIWILL, G. & SCHALLER, H. 1988. The duck hepatitis B virus P-gene codes for protein strongly associated with the 5'-end of the viral DNA minus strand. *Virology,* 166**,** 475-85.

BOWYER, S. M., VAN STADEN, L., KEW, M. C. & SIM, J. G. 1997. A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa. *J Gen Virol,* 78 (Pt 7)**,** 1719-29.

BRÉCHOT, C., GOZUACIK, D., MURAKAMI, Y. & PATERLINI-BRÉCHOT, P. Year. Molecular bases for the development of hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC). *In:* Seminars in cancer biology, 2000. Elsevier, 211-231.

BRICHLER, S., LAGATHU, G., ABDOU-CHEKARAOU, M., LE GAL, F., EDOUARD, A., DÉNY, P., CESAIRE, R. & GORDIEN, E. 2013. African, Amerindian and European Hepatitis B Virus Strains Circulate in the Caribbean Island of Martinique. *Journal of General Virology*.

BUCKWOLD, V. E., XU, Z., CHEN, M., YEN, T. & OU, J.-H. 1996. Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication. *Journal of Virology,* 70**,** 5845-5851.

BURNETT, R. J., FRANCOIS, G., KEW, M. C., LEROUX-ROELS, G., MEHEUS, A., HOOSEN, A. A. & MPHAHLELE, M. J. 2005. Hepatitis B virus and human

immunodeficiency virus co-infection in sub-Saharan Africa: a call for further investigation. *Liver Int,* 25**,** 201-13.

BURNETT, R. J., KRAMVIS, A., DOCHEZ, C. & MEHEUS, A. 2012. An update after 16 years of hepatitis B vaccination in South Africa. *Vaccine,* 30, Supplement 3**,** C45-C51.

CANDOTTI, D., OPARE-SEM, O., REZVAN, H., SARKODIE, F. & ALLAIN, J. P. 2006. Molecular and serological characterization of hepatitis B virus in deferred Ghanaian blood donors with and without elevated alanine aminotransferase. *Journal of Viral Hepatitis,* 13**,** 715-724.

CAO, G.-W. 2009. Clinical relevance and public health significance of hepatitis B virus genomic variations. *World Journal of Gastroenterology: WJG,* 15**,** 5761.

CAO, J., YANG, E. B., SU, J. J., LI, Y. & CHOW, P. 2003. The tree shrews: adjuncts and alternatives to primates as models for biomedical research. *J Med Primatol,* 32**,** 123-30.

CAPOBIANCHI, M. R., GIOMBINI, E. & ROZERA, G. 2013. Next-generation sequencing technology in clinical virology. *Clinical Microbiology and Infection,* 19**,** 15-22.

CARMAN, W., HADZIYANNIS, S., MCGARVEY, M., JACYNA, M., KARAYIANNIS, P., MAKRIS, A. & THOMAS, H. 1989. Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. *The Lancet,* 334**,** 588-591.

CARMAN, W. F., ZANETTI, A. R., KARAYIANNIS, P., WATERS, J., MANZILLO, G., TANZI, E., ZUCKERMAN, A. J. & THOMAS, H. C. 1990. Vaccine-induced escape mutant of hepatitis B virus. *Lancet,* 336**,** 325-9.

CARMICHAEL, A., JIN, X., SISSONS, P. & BORYSIEWICZ, L. 1993. Quantitative analysis of the human immunodeficiency virus type 1 (HIV-1)-specific cytotoxic T lymphocyte (CTL) response at different stages of HIV-1 infection: differential CTL

responses to HIV-1 and Epstein-Barr virus in late disease. *The Journal of experimental medicine,* 177**,** 249-256.

CAVINTA, L., CAO, G. & SCHAEFER, S. 2009a. Description of a new hepatitis B virus C6 subgenotype found in the Papua province of Indonesia and suggested renaming of a tentative C6 subgenotype found in the Philippines as subgenotype C7. *Journal of Clinical Microbiology,* 47**,** 3068-3069.

CAVINTA, L., SUN, J., MAY, A., YIN, J., VON MELTZER, M., RADTKE, M., BARZAGA, N. G., CAO, G. & SCHAEFER, S. 2009b. A new isolate of hepatitis B virus from the Philippines possibly representing a new subgenotype C6. *Journal of Medical Virology,* 81**,** 983-987.

CHANG, S. F., NETTER, H. J., BRUNS, M., SCHNEIDER, R., FROLICH, K. & WILL, H. 1999. A new avian hepadnavirus infecting snow geese (Anser caerulescens) produces a significant fraction of virions containing single-stranded DNA. *Virology,* 262**,** 39-54.

CHEN, B. F., LIU, C. J., JOW, G. M., CHEN, P. J., KAO, J. H. & CHEN, D. S. 2006. High Prevalence and Mapping of Pre-S Deletion in Hepatitis B Virus Carriers With Progressive Liver Diseases. *Gastroenterology,* 130**,** 1153-1168.

CHEN, C. Y., CROWTHER, C., KEW, M. C. & KRAMVIS, A. 2008. A valine to phenylalanine mutation in the precore region of hepatitis B virus causes intracellular retention and impaired secretion of HBe-antigen. *Hepatol Res,* 38**,** 580-92.

CHEN, M., SÄLLBERG, M., HUGHES, J., JONES, J., GUIDOTTI, L. G., CHISARI, F. V., BILLAUD, J.-N. & MILICH, D. R. 2005. Immune tolerance split between hepatitis B virus precore and core proteins. *Journal of Virology,* 79**,** 3016-3027.

CHEN, P., GAN, Y., HAN, N., FANG, W., LI, J., ZHAO, F., HU, K. & RAYNER, S. 2013. Computational Evolutionary Analysis of the Overlapped Surface (S) and Polymerase

(P) Region in Hepatitis B Virus Indicates the Spacer Domain in P Is Crucial for Survival. *Plos One,* 8**,** e60098.

CHING-LUNG LAI, M.-F. Y. A. S. L. 2002. Treatment of chronic hepatitis B infection. In: Lai, C. And Locarnini, S. (Eds.), Hepatitis B virus, pp. 193–218. International Medical Press, London, UK, first edition.**,** 193-218.

CHISARI, F. V. & FERRARI, C. 1995. HEPATITIS-B VIRUS IMMUNOPATHOGENESIS. *Annual Review of Immunology,* 13**,** 29-60.

CHONGSRISAWAT, V., THAWORNSUK, N., THEAMBOONLERS, A., LOUISIRIROTCHANAKUL, S. & POOVORAWAN, Y. 2006. Hepatitis B virus DNA in unusual serological profiles of hepatitis B surface antigen-positive sera. *Viral Immunol,* 19**,** 623-9.

COLIN, J. F., CAZALS-HATEM, D., LORIOT, M. A., MARTINOT-PEIGNOUX, M., PHAM, B. N., AUPERIN, A., DEGOTT, C., BENHAMOU, J. P., ERLINGER, S. & VALLA, D. 1999. Influence of human immunodeficiency virus infection on chronic hepatitis B in homosexual men. *Hepatology,* 29**,** 1306-1310.

COOREMAN, M. P., VAN ROOSMALEN, M. H., TE MORSCHE, R., SÜNNEN, C. M., SCHOONDERMARK-VAN DE VEN, E. M., JANSEN, J. B., TYTGAT, G. N., DE WIT, P. L. & PAULIJ, W. P. 1999. Characterization of the reactivity pattern of murine monoclonal antibodies against wild-type hepatitis B surface antigen to g145r and other naturally occurring “a” loop escape mutations. *Hepatology,* 30**,** 1287-1292.

CUI, C. Y., SHI, J. X., HUI, L. J., XI, H. F., ZHUOMA, QUNI, TSEDAN & HU, G. X. 2002. The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. *Journal of General Virology,* 83**,** 2773-2777.

DANE, D. S., CAMERON, C. H. & BRIGGS, M. 1970. Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *Lancet,* 1**,** 695-8.

DE FRANCHIS, R., HADENGUE, A., LAU, G., LAVANCHY, D., LOK, A., MCINTYRE, N., MELE, A., PAUMGARTNER, G., PIETRANGELO, A., RODES, J., ROSENBERG, W. & VALLA, D. 2003. EASL International Consensus Conference on Hepatitis B. 13-14 September, 2002 Geneva, Switzerland. Consensus statement (long version). *J Hepatol,* 39 Suppl 1**,** S3-25.

DEPAMEDE, S. N., SURAYAH, K., TJAHYONO, A. A. H., NAGASHIMA, S., TAKAHASHI, M. & OKAMOTO, H. 2010. Identification and characterization of novel hepatitis B virus subgenotype C10 in Nusa Tenggara, Indonesia. *Archives of Virology,* 155**,** 705-715.

DESMET, V. J. & ROSKAMS, T. 2004. Cirrhosis reversal: a duel between dogma and myth. *Journal of Hepatology,* 40**,** 860-867.

DEVESA, M., RODRIGUEZ, C., LEON, G., LIPRANDI, F. & PUJOL, F. H. 2004. Clade analysis and surface antigen polymorphism of hepatitis B virus American genotypes. *J Med Virol,* 72**,** 377-84.

DI MARTINO, V., THEVENOT, T., COLIN, J. F., BOYER, N., MARTINOT, M., DEGOS, F., COULAUD, J. P., VILDE, J. L., VACHON, F. & DEGOTT, C. 2002. Influence of HIV infection on the response to interferon therapy and the long-term outcome of chronic hepatitis B. *Gastroenterology,* 123**,** 1812-1822.

DICKENS, C., KEW, M. C., PURCELL, R. H. & KRAMVIS, A. 2013. Occult Hepatitis B Virus Infection in Chacma Baboons, South Africa. *Emerg Infect Dis,* 19**,** 598-605.

DONATO, F., BOFFETTA, P. & PUOTI, M. 1998. A meta-analysis of epidemiological studies on the combined effect of hepatitis B and C virus infections in causing hepatocellular carcinoma. *Int J Cancer,* 75**,** 347-54.

EL-SONY, A. 2006. The cost to health services of human immunodeficiency virus (HIV) co-infection among tuberculosis patients in Sudan. *Health Policy,* 75**,** 272-279.

ELDUMA, A. & SAEED, N. 2011. La Revue de Santé de la Méditerranée orientale. *EMHJ,* 17**,** 474-478.

ELSHEIKH, R. M., DAAK, A. A., ELSHEIKH, M. A., KARSANY, M. S. & ADAM, I. 2007. Hepatitis B virus and hepatitis C virus in pregnant Sudanese women. *Virol J,* 4**,** 104.

EPSTEIN, O., AJDUKIEWICZ, A. B., DICK, R. & SHERLOCK, S. 1979. Hypertrophic hepatic osteoarthropathy: clinical, roentgenologic, biochemical, hormonal and cardiorespiratory studies, and review of the literature. *The American journal of medicine,* 67**,** 88-97.

ERIKSSON, N., PACHTER, L., MITSUYA, Y., RHEE, S.-Y., WANG, C., GHARIZADEH, B., RONAGHI, M., SHAFER, R. W. & BEERENWINKEL, N. 2008. Viral population estimation using pyrosequencing. *PLoS Computational Biology,* 4**,** e1000074.

FEITELSON, M. 1992. Hepatitis B virus infection and primary hepatocellular carcinoma. *Clin Microbiol Rev,* 5**,** 275-301.

FEITELSON, M. A. 1994. Biology of hepatitis B virus variants. *Lab Invest,* 71**,** 324-49.

FERRER-ORTA, C., ARIAS, A., ESCARMÍS, C. & VERDAGUER, N. 2006. A comparison of viral RNA-dependent RNA polymerases. *Current opinion in structural biology,* 16**,** 27-34.

FIRNHABER, C. & IVE, P. 2009. Hepatitis B and HIV co-infection in South Africa: just treat it! *Southern African Journal of HIV Medicine,* 10.

FORBI, J. C., BEN-AYED, Y., XIA, G.-L., VAUGHAN, G., DROBENIUC, J., SWITZER, W. M. & KHUDYAKOV, Y. E. 2013. Disparate distribution of hepatitis B virus genotypes in four sub-Saharan African countries. *Journal of Clinical Virology,* 58**,** 59-66.

GANEM, D. & PRINCE, A. M. 2004. Hepatitis B virus infection--natural history and clinical consequences. *N Engl J Med,* 350**,** 1118-29.

GANEM, D. & SCHNEIDER, R. J. 2001. Hepadnaviridae: the viruses and their replication. *Viral Hepatitis.*

GANEM, D. & VARMUS, H. E. 1987. The Molecular Biology of the Hepatitis B Viruses. *Annual Review of Biochemistry,* 56**,** 651-693.

GARMIRI, P., LOUA, A., HABA, N., CANDOTTI, D. & ALLAIN, J. P. 2009. Deletions and recombinations in the core region of hepatitis B virus genotype E strains from asymptomatic blood donors in Guinea, west Africa. *J Gen Virol,* 90**,** 2442-51.

GASIM, G. I., HAMDAN, H. Z., HAMDAN, S. Z. & ADAM, I. 2011. Epidemiology of hepatitis B and hepatitis C virus infections among hemodialysis patients in Khartoum, Sudan. *Journal of Medical Virology***,** n/a-n/a.

GASSMELSEED, D. E. A., NASR, A. M., HOMEIDA, S. M., ELSHEIKH, M. A. & ADAM, I. 2006. Prevalence of HIV infection among pregnant women of the central Sudan. *Journal of Medical Virology,* 78**,** 1269-1270.

GERLICH, W. H., HEERMANN, K. H. & LU, X. 1992. Functions of hepatitis B surface proteins. *Arch Virol Suppl,* 4**,** 129-32.

GHEIT, T., SEKKAT, S., COVA, L., CHEVALLIER, M., PETIT, M. A., HANTZ, O., LESENECHAL, M., BENSLIMANE, A., TREPO, C. & CHEMIN, I. 2002. Experimental transfection of Macaca sylvanus with cloned human hepatitis B virus. *J Gen Virol,* 83**,** 1645-9.

GHOSH, S., BANERJEE, P., DENY, P., MONDAL, R. K., NANDI, M., ROYCHOUDHURY, A., DAS, K., BANERJEE, S., SANTRA, A., ZOULIM, F., CHOWDHURY, A. & DATTA, S. 2012. New HBV subgenotype D9, a novel D/C

recombinant, identified in patients with chronic HBeAg-negative infection in Eastern India. *Journal of Viral Hepatitis***,** n/a-n/a.

GISH, R., JIA, J.-D., LOCARNINI, S. & ZOULIM, F. 2012. Selection of chronic hepatitis B therapy with high barrier to resistance. *The Lancet Infectious Diseases,* 12**,** 341-353.

GOPALAKRISHNAN, D., KEYTER, M., SHENOY, K. T., LEENA, K. B., THAYUMANAVAN, L., THOMAS, V., VINAYAKUMAR, K., PANACKEL, C., KORAH, A. T. & NAIR, R. 2013. Hepatitis B virus subgenotype A1 predominates in liver disease patients from Kerala, India. *World J Gastroenterol,* 19**,** 9294-9306.

GULUBE, Z., CHIRARA, M., KEW, M., TANAKA, Y., MIZOKAMI, M. & KRAMVIS, A. 2011. Molecular characterization of hepatitis B virus isolates from Zimbabwean blood donors. *Journal of Medical Virology,* 83**,** 235-244.

GÜNTHER, S., FISCHER, L., PULT, I., STERNECK, M. & WILL, H. 1999. Naturally occurring variants of hepatitis B virus. *Advances in virus research,* 52**,** 25-137.

GUO, W. T., BELL, K. D. & OU, J. H. 1991. Characterization of the hepatitis B virus EnhI enhancer and X promoter complex. *J Virol,* 65**,** 6686-92.

GUTIÉRREZ-GARCÍA, M. L., FERNANDEZ-RODRIGUEZ, C. M., LLEDO-NAVARRO, J. L. & BUHIGAS-GARCIA, I. 2011. Prevalence of occult hepatitis B virus infection. *World Journal of Gastroenterology: WJG,* 17**,** 1538.

HANNOUN, C., NORDER, H. & LINDH, M. 2000. An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam. *J Gen Virol,* 81**,** 2267-72.

HANNOUN, C., SODERSTROM, A., NORKRANS, G. & LINDH, M. 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**,** 2163-7.

HASEGAWA, I., TANAKA, Y., KURBANOV, F., YOSHIHARA, N., EL-GOHARY, A., LYAMUYA, E., MATEE, M., MAGESSA, P., FUJIWARA, K. & OZASA, A. 2006. Molecular epidemiology of hepatitis B virus in the United Republic of Tanzania. *Journal of Medical Virology,* 78**,** 1035-1042.

HASHIM, M., SALIH, M., HAG, A. E., KARRAR, Z., OSMAN, E., EL-SHIEKH, F., TILIB, I. E. & ATTALA, N. 1997. AIDS and HIV infection in Sudanese children: a clinical and epidemiological study. *AIDS Patient Care and STDs,* 11**,** 331-337.

HENKE-GENDO, C., AMINI-BAVIL-OLYAEE, S., CHALLAPALLI, D., TRAUTWEIN, C., DEPPE, H., SCHULZ, T. F., HEIM, A. & TACKE, F. 2008. Symptomatic Hepatitis B Virus (HBV) Reactivation despite Reduced Viral Fitness Is Associated with HBV Test and Immune Escape Mutations in an HIV-Coinfected Patient. *Journal of Infectious Diseases,* 198**,** 1620-1624.

HILDT, E., MUNZ, B., SAHER, G., REIFENBERG, K. & HOFSCHNEIDER, P. H. 2002. The PreS2 activator MHBs(t) of hepatitis B virus activates c-raf-1/Erk2 signaling in transgenic mice. *EMBO J,* 21**,** 525-35.

HODGSON, P. D. & MICHALAK, T. I. 2001. Augmented hepatic interferon gamma expression and T-cell influx characterize acute hepatitis progressing to recovery and residual lifelong virus persistence in experimental adult woodchuck hepatitis virus infection. *Hepatology,* 34**,** 1049-59.

HOFFMANN, C., MINKAH, N., LEIPZIG, J., WANG, G., ARENS, M. Q., TEBAS, P. & BUSHMAN, F. D. 2007. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic acids research,* 35**,** e91.

HOLLINGER, F. & SOOD, G. 2010. Occult hepatitis B virus infection: a covert operation. *Journal of Viral Hepatitis,* 17**,** 1-15.

HOMS, M., BUTI, M., QUER, J., JARDÍ, R., SCHAPER, M., TABERNERO, D., ORTEGA, I., SANCHEZ, A., ESTEBAN, R. & RODRIGUEZ-FRIAS, F. 2011. Ultra-deep pyrosequencing analysis of the hepatitis B virus preCore region and main catalytic motif of the viral polymerase in the same viral genome. *Nucleic acids research,* 39**,** 8457-8471.

HOMS, M., BUTI, M., TABERNERO, D., QUER, J., SANCHEZ, A., CORRAL, N., ESTEBAN, R. & RODRIGUEZ-FRIAS, F. 2012. Quasispecies dynamics in main core epitopes of hepatitis B virus by ultra-deep-pyrosequencing. *World Journal of Gastroenterology: WJG,* 18**,** 6096.

HOU, J., LIU, Z. & GU, F. 2005. Epidemiology and Prevention of Hepatitis B Virus Infection. *Int J Med Sci,* 2**,** 50-57.

HU, J., FLORES, D., TOFT, D., WANG, X. & NGUYEN, D. 2004. Requirement of heat shock protein 90 for human hepatitis B virus reverse transcriptase function. *J Virol,* 78**,** 13122-31.

HUANG, C. F., LIN, S. S., HO, Y. C., CHEN, F. L. & YANG, C. C. 2006. The immune response induced by hepatitis B virus principal antigens. *Cell Mol Immunol,* 3**,** 97-106.

HÜBSCHEN, J., MBAH, P., FORBI, J., OTEGBAYO, J., OLINGER, C., CHARPENTIER, E. & MULLER, C. 2011. Detection of a new subgenotype of hepatitis B virus genotype A in Cameroon but not in neighbouring Nigeria. *Clinical Microbiology and Infection,* 17**,** 88-94.

HÜBSCHEN, J. M., MUGABO, J., PELTIER, C. A., KARASI, J.-C., SAUSY, A., KIRPACH, P., ARENDT, V. & MULLER, C. P. 2009. Exceptional genetic variability of hepatitis B virus indicates that Rwanda is east of an emerging African genotype E/A1 divide. *Journal of Medical Virology,* 81**,** 435-440.

HUNT, C. M., MCGILL, J. M., ALLEN, M. I. & CONDREAY, L. D. 2000. Clinical relevance of hepatitis B viral mutations. *Hepatology,* 31**,** 1037-1044.

HUY, T. T., USHIJIMA, H., QUANG, V. X., WIN, K. M., LUENGROJANAKUL, P., KIKUCHI, K., SATA, T. & ABE, K. 2004. Genotype C of hepatitis B virus can be classified into at least two subgroups. *J Gen Virol,* 85**,** 283-92.

HUY, T. T., USHIJIMA, H., SATA, T. & ABE, K. 2006. Genomic characterization of HBV genotype F in Bolivia: genotype F subgenotypes correlate with geographic distribution and T(1858) variant. *Arch Virol,* 151**,** 589-97.

HUY, T. T. T., NGOC, T. T. & ABE, K. 2008. New Complex Recombinant Genotype of Hepatitis B Virus Identified in Vietnam. *Journal of Virology,* 82**,** 5657-5663.

HYAMS, K. C., AL-ARABI, M. A., AL-TAGANI, A. A., MESSITER, J. F., AL-GAALI, A. A. & GEORGE, J. F. 1989. Epidemiology of hepatitis B in the Gezira region of Sudan. DTIC Document.

IJAZ, S., ARNOLD, C., DERVISEVIC, S., MECHUROVA, J., TATMAN, N., TEDDER, R. & NAOUMOV, N. 2008. Dynamics of lamivudine-resistant hepatitis B virus during adefovir monotherapy versus lamivudine plus adefovir combination therapy. *J Med Virol,* 80**,** 1160 - 1170.

ILLUMINA 2013. HiSeq technology, Illumina, accessed from http://www.illumina.com/systems/sequencing.ilmn on 02/09/2013.

JEAN-JEAN, O., LEVRERO, M., HANS, W., PERRICAUDET, M. & ROSSIGNOL, J.-M. 1989. Expression mechanism of the hepatitis B virus (HBV) C gene and biosynthesis of HBe antigen. *Virology,* 170**,** 99-106.

JONAS, M. M. 2009. Hepatitis B and pregnancy: an underestimated issue. *Liver International,* 29**,** 133-139.

JUNKER-NIEPMANN, M., BARTENSCHLAGER, R. & SCHALLER, H. 1990. A short cis-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. *EMBO J,* 9**,** 3389-96.

KANN, M. 2002. Structural and molecular virology. *Hepatitis B virus, International Medical Press: 9-21*.

KANN, M., SCHMITZ, A. & RABE, B. 2007. Intracellular transport of hepatitis B virus. *World J Gastroenterol,* 13**,** 39-47.

KAO, J. H. 2002a. Clinical relevance of hepatitis B viral genotypes: a case of deja vu? *J Gastroenterol Hepatol,* 17**,** 113-5.

KAO, J. H. 2002b. Hepatitis B viral genotypes: clinical relevance and molecular characteristics. *J Gastroenterol Hepatol,* 17**,** 643-50.

KATO, H., ORITO, E., GISH, R. G., SUGAUCHI, F., SUZUKI, S., UEDA, R., MIYAKAWA, Y. & MIZOKAMI, M. 2002. Characteristics of Hepatitis B Virus Isolates of Genotype G and Their Phylogenetic Differences from the Other Six Genotypes (A through F). *Journal of Virology,* 76**,** 6131-6137.

KEW, M. C. 2010. Epidemiology of chronic hepatitis B virus infection, hepatocellular carcinoma, and hepatitis B virus-induced hepatocellular carcinoma. *Pathologie Biologie,* 58**,** 273-277.

KEW, M. C., KRAMVIS, A., YU, M. C., ARAKAWA, K. & HODKINSON, J. 2005. Increased hepatocarcinogenic potential of hepatitis B virus genotype A in Bantu-speaking sub-saharan Africans. *Journal of Medical Virology,* 75**,** 513-521.

KIMBI, G. C., KRAMVIS, A. & KEW, M. C. 2004. Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa. *Journal of General Virology,* 85**,** 1211-1220.

KOTTILIL, S., JACKSON, J. O. & POLIS, M. A. 2005. Hepatitis B & hepatitis C in HIV-infection. *Indian J Med Res,* 121**,** 424-50.

KOZAK, M. 1989. The scanning model for translation: an update. *J Cell Biol,* 108**,** 229-41.

KRAMVIS, A. 2014. Genotypes and Genetic Variability of Hepatitis B Virus. *Intervirology, submitted.*

KRAMVIS, A., ARAKAWA, K., YU, M. C., NOGUEIRA, R., STRAM, D. O. & KEW, M. C. 2008. Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. *J Med Virol,* 80**,** 27-46.

KRAMVIS, A., BUKOFZER, S., KEW, M. C. & SONG, E. 1997. Nucleic acid sequence analysis of the precore region of hepatitis B virus from sera of southern African black adult carriers of the virus. *Hepatology,* 25**,** 235-40.

KRAMVIS, A. & KEW, M. 1998. Structure and function of the encapsidation signal of hepadnaviridae. *Journal of Viral Hepatitis,* 5**,** 357-367.

KRAMVIS, A., KEW, M. & FRANCOIS, G. 2005a. Hepatitis B virus genotypes. *Vaccine,* 23**,** 2409-23.

KRAMVIS, A. & KEW, M. C. 1999. The core promoter of hepatitis B virus. *J Viral Hepat,* 6**,** 415-27.

KRAMVIS, A. & KEW, M. C. 2005. Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy. *J Viral Hepat,* 12**,** 456-64.

KRAMVIS, A. & KEW, M. C. 2007. Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes. *Hepatology Research,* 37**,** S9-S19.

KRAMVIS, A., KEW, M. C. & BUKOFZER, S. 1998. Hepatitis B virus precore mutants in serum and liver of Southern African Blacks with hepatocellular carcinoma. *Journal of Hepatology,* 28**,** 132-141.

KRAMVIS, A. & PARASKEVIS, D. 2012. Subgenotype A1 of HBV-tracing human migrations in and out of Africa. *Antiviral therapy,* 18**,** 513-521.

KRAMVIS, A., RESTORP, K., NORDER, H., BOTHA, J. F., MAGNIUS, L. O. & KEW, M. C. 2005b. Full genome analysis of hepatitis B virus genotype E strains from South-Western Africa and madagascar reveals low genetic variability. *Journal of Medical Virology,* 77**,** 47-52.

KRAMVIS, A., WEITZMANN, L., OWIREDU, W. K. & KEW, M. C. 2002. Analysis of the complete genome of subgroup A′ hepatitis B virus isolates from South Africa. *Journal of General Virology,* 83**,** 835-839.

KROGSGAARD, K., LINDHARDT, B. Ö., NIELSEN, J. O., ANDERSSON, P., KRYGER, P., ALDERSHVILE, J. & GERSTOFT, J. 1987. The influence of HTLV-III infection on the natural history of hepatitis B virus infection in male homosexual HBsAg carriers. *Hepatology,* 7**,** 37-41.

KURBANOV, F., TANAKA, Y., FUJIWARA, K., SUGAUCHI, F., MBANYA, D., ZEKENG, L., NDEMBI, N., NGANSOP, C., KAPTUE, L., MIURA, T., IDO, E., HAYAMI, M., ICHIMURA, H. & MIZOKAMI, M. 2005. A new subtype (subgenotype) Ac (A3) of hepatitis B virus and recombination between genotypes A and E in Cameroon. *J Gen Virol,* 86**,** 2047-56.

KURBANOV, F., TANAKA, Y., KRAMVIS, A., SIMMONDS, P. & MIZOKAMI, M. 2008. When should "I" consider a new hepatitis B virus genotype? *J Virol,* 82**,** 8241-2.

KURBANOV, F., TANAKA, Y. & MIZOKAMI, M. 2010. Geographical and genetic diversity of the human hepatitis B virus. *Hepatology Research,* 40**,** 14-30.

LADA, O., BENHAMOU, Y., POYNARD, T. & THIBAULT, V. 2006. Coexistence of hepatitis B surface antigen (HBs Ag) and anti-HBs antibodies in chronic hepatitis B virus carriers: influence of "a" determinant variants. *J Virol,* 80**,** 2968-75.

LAI & LOCARNINI 2002. Human Virus Guides:Hepatitis B Virus. *International Medical Press*.

LANFORD, R. E., CHAVEZ, D., BRASKY, K. M., BURNS, R. B., 3RD & RICO-HESSE, R. 1998. Isolation of a hepadnavirus from the woolly monkey, a New World primate. *Proc Natl Acad Sci U S A,* 95**,** 5757-61.

LASKUS, T., RAKELA, J., NOWICKI, M. J. & PERSING, D. H. 1995. Hepatitis B virus core promoter sequence analysis in fulminant and chronic hepatitis B. *Gastroenterology,* 109**,** 1618-1623.

LAUKAMM-JOSTEN, U., MÜLLER, O., BIENZLE, U., FELDMEIER, H., UY, A. & GUGGENMOOS-HOLZMANN, I. 1988. Sir, Decline of naturally acquired antibodies to hepatitis B surface antigen in HIV-1 infected homosexual men. *AIDS,* 2**,** 400.

LEVRERO, M., POLLICINO, T., PETERSEN, J., BELLONI, L., RAIMONDO, G. & DANDRI, M. 2009. Control of cccDNA function in hepatitis B virus infection. *J Hepatol,* 51**,** 581-92.

LI, J., TONG, S., WEN, Y., VITVITSKI, L., ZHANG, Q. & TREPO, C. 1993. Hepatitis B virus genotype A rarely circulates as an HBe-minus mutant: possible contribution of a single nucleotide in the precore region. *Journal of Virology,* 67**,** 5402-5410.

LIN, C.-L. & KAO, J.-H. 2010. Clinical implications of hepatitis B virus variants. *Journal of the Formosan Medical Association,* 109**,** 321-325.

LIN, C. G. & LO, S. J. 1992. Evidence for involvement of a ribosomal leaky scanning mechanism in the translation of the hepatitis B virus pol gene from the viral pregenome RNA. *Virology,* 188**,** 342-52.

LIN, C. L. & KAO, J. H. 2011. The clinical implications of hepatitis B virus genotype: recent advances. *Journal of Gastroenterology and Hepatology,* 26**,** 123-130.

LINDH, M., ANDERSSON, A.-S. & GUSDAL, A. 1997. Genotypes, nt 1858 Variants, and Geographic Origin of Hepatitis B Virus—Large-Scale Analysis Using a New Genotyping Method. *Journal of Infectious Diseases,* 175**,** 1285-1293.

LINDSTRÖM, A., ODEBERG, J. & ALBERT, J. 2004. Pyrosequencing for detection of lamivudine-resistant hepatitis B virus. *Journal of Clinical Microbiology,* 42**,** 4788-4795.

LIU, T. T., FANG, Y., XIONG, H., CHEN, T. Y., NI, Z. P., LUO, J. F., ZHAO, N. Q. & SHEN, X. Z. 2008. A case-control study of the relationship between hepatitis B virus DNA level and risk of hepatocellular carcinoma in Qidong, China. *World Journal of Gastroenterology: WJG,* 14**,** 3059.

LOCARNINI, S., MCMILLAN, J. & BARTHOLOMEUSZ, A. 2003. The hepatitis B virus and common mutants. *Semin Liver Dis,* 23**,** 5-20.

LOK, A. S., AKARCA, U. & GREENE, S. 1994. Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal. *Proc Natl Acad Sci U S A,* 91**,** 4077-81.

LU, C.-C., CHEN, M., OU, J.-H. & YEN, T. S. B. 1995. Key Role of a CCAAT Element in Regulating Hepatitis B Virus Surface Protein Expression. *Virology,* 206**,** 1155-1158.

MAHGOUB, S., CANDOTTI, D., EL EKIABY, M. & ALLAIN, J. P. 2011. Hepatitis B virus (HBV) infection and recombination between HBV genotypes D and E in asymptomatic blood donors from Khartoum, Sudan. *Journal of Clinical Microbiology,* 49**,** 298-306.

MAHMOUD, O. A. E. K., GHAZAL, A. A. E. R., METWALLY, D. E. S., ELNOUR, A. M. & YOUSIF, G. E. A. 2013. Detection of occult hepatitis B virus infection among blood donors in Sudan. *The Journal Of The Egyptian Public Health Association,* 88**,** 14-18.

MAKONDO, E., BELL, T. G. & KRAMVIS, A. 2012. Genotyping and Molecular Characterization of Hepatitis B Virus from Human Immunodeficiency Virus-Infected Individuals in Southern Africa. *Plos One,* 7**,** e46345.

MAKUWA, M., SOUQUIERE, S., TELFER, P., LEROY, E., BOURRY, O., ROUQUET, P., CLIFFORD, S., WICKINGS, E. J., ROQUES, P. & SIMON, F. 2003. Occurrence of hepatitis viruses in wild-born non-human primates: a 3 year (1998-2001) epidemiological survey in Gabon. *J Med Primatol,* 32**,** 307-14.

MARDIS, E. R. 2008. The impact of next-generation sequencing technology on genetics. *Trends in genetics,* 24**,** 133-141.

MARGERIDON-THERMET, S., SHULMAN, N. S., AHMED, A., SHAHRIAR, R., LIU, T., WANG, C., HOLMES, S. P., BABRZADEH, F., GHARIZADEH, B., HANCZARUK, B., SIMEN, B. B., EGHOLM, M. & SHAFER, R. W. 2009. Ultra-Deep Pyrosequencing of Hepatitis B Virus Quasispecies from Nucleoside and Nucleotide Reverse-Transcriptase Inhibitor (NRTI)–Treated Patients and NRTI-Naive Patients. *Journal of Infectious Diseases,* 199**,** 1275-1285.

MARION, P. L., OSHIRO, L. S., REGNERY, D. C., SCULLARD, G. H. & ROBINSON, W. S. 1980. A virus in Beechey ground squirrels that is related to hepatitis B virus of humans. *Proc Natl Acad Sci U S A,* 77**,** 2941-5.

MASON, W. S., SEAL, G. & SUMMERS, J. 1980. Virus of Pekin ducks with structural and biological relatedness to human hepatitis B virus. *J Virol,* 36**,** 829-36.

MAXAM, A. M. & GILBERT, W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences,* 74**,** 560-564.

MBAYED, V. A., BARBINI, L., LOPEZ, J. L. & CAMPOS, R. H. 2001. Phylogenetic analysis of the hepatitis B virus (HBV) genotype F including Argentine isolates. *Arch Virol,* 146**,** 1803-10.

MCCARTHY, M. C., BURANS, J. P., CONSTANTINE, N. T., EL-HAG, A. A. & EL-TAYEB, M. 1989a. Hepatitis B and HIV in Sudan: a serosurvey for hepatitis B and human immunodeficiency virus antibodies among sexually active heterosexuals. DTIC Document.

MCCARTHY, M. C., EL-TIGANI, A., KHALID, I. O. & HYAMS, K. C. 1994. Hepatitis B and C in Juba, southern Sudan: results of a serosurvey. *Trans R Soc Trop Med Hyg,* 88**,** 534-6.

MCCARTHY, M. C., HYAMS, K. C., EL-HAG, A. E.-T., EL-DABI, M. A., EL-TAYEB, M. E.-S., KHALID, I. O., GEORGE, J. F., CONSTANTINE, N. T. & WOODY, J. N. 1989b. HIV-1 and hepatitis B transmission in Sudan. *AIDS,* 3**,** 725-730.

MCCARTHY, M. C., KHALID, I. O. & TIGANI, A. E. 1995. HIV-1 infection in Juba, Southern Sudan. *Journal of Medical Virology,* 46**,** 18-20.

MCMAHON, B. J. 2009. The influence of hepatitis B virus genotype and subgenotype on the natural history of chronic hepatitis B. *Hepatology international,* 3**,** 334-342.

MCMAHON, B. J., ALWARD, W. L. M., HALL, D. B., HEYWARD, W. L., BENDER, T. R., FRANCIS, D. P. & MAYNARD, J. E. 1985. Acute Hepatitis B Virus Infection: Relation of Age to the Clinical Expression of Disease and Subsequent Development of the Carrier State. *Journal of Infectious Diseases,* 151**,** 599-603.

MCMAHON, B. J., HOLCK, P., BULKOW, L. & SNOWBALL, M. 2001. Serologic and clinical outcomes of 1536 Alaska Natives chronically infected with hepatitis B virus. *Ann Intern Med,* 135**,** 759-68.

MELDAL, B. H. M., MOULA, N. M., BARNES, I. H. A., BOUKEF, K. & ALLAIN, J.-P. 2009. A novel hepatitis B virus subgenotype, D7, in Tunisian blood donors. *Journal of General Virology,* 90**,** 1622-1628.

METZKER, M. L. 2010. Sequencing technologies [mdash] the next generation. *Nat Rev Genet,* 11**,** 31-46.

MOHAMMED, A. A., BABIKER, Z. O. E., ALI, A. K. M., ADAM, A. A. A. H., HASSAN, E. A., OSMAN, H. K. E. & HERIEKA, E. A. M. 2011. Seroprevalence of the human immunodeficiency virus (HIV) among pregnant women in eastern Sudan. *Journal of Infection and Public Health,* 4**,** 55-58.

MOOLLA, N., KEW, M. & ARBUTHNOT, P. 2002. Regulatory elements of hepatitis B virus transcription. *Journal of Viral Hepatitis,* 9**,** 323-331.

MORAES, M., GOMES, S. & NIEL, C. 1996. Sequence analysis of pre-S/S gene of hepatitis B virus strains of genotypes A, D, and F isolated in Brazil. *Archives of Virology,* 141**,** 1767-1773.

MPHAHLELE, M. J., LUKHWARENI, A., BURNETT, R. J., MOROPENG, L. M. & NGOBENI, J. M. 2006. High risk of occult hepatitis B virus infection in HIV-positive patients from South Africa. *Journal of Clinical Virology,* 35**,** 14-20.

MUDAWI, H. M. Y., SMITH, H. M., RAHOUD, S. A., FLETCHER, I. A., SAEED, O. K. & FEDAIL, S. S. 2007. Prevalence of hepatitis B virus infection in the Gezira state of central Sudan. *Saudi J Gastroenterol,* 13**,** 81-3.

MULDERS, M. N., VENARD, V., NJAYOU, M., EDORH, A. P., BOLA OYEFOLU, A. O., KEHINDE, M. O., MUYEMBE TAMFUM, J. J., NEBIE, Y. K., MAIGA, I., AMMERLAAN, W., FACK, F., OMILABU, S. A., LE FAOU, A. & MULLER, C. P. 2004. Low genetic diversity despite hyperendemicity of hepatitis B virus genotype E throughout West Africa. *J Infect Dis,* 190**,** 400-8.

MULYANTO, DEPAMEDE, S. N., WAHYONO, A., JIRINTAI, NAGASHIMA, S., TAKAHASHI, M. & OKAMOTO, H. 2011. Analysis of the full-length genomes of

novel hepatitis B virus subgenotypes C11 and C12 in Papua, Indonesia. *J Med Virol,* 83**,** 54-64.

MULYANTO, PANCAWARDANI, P., DEPAMEDE, S. N., WAHYONO, A., JIRINTAI, S., NAGASHIMA, S., TAKAHASHI, M., NISHIZAWA, T. & OKAMOTO, H. 2012. Identification of four novel subgenotypes (C13–C16) and two inter-genotypic recombinants (C12/G and C13/B3) of hepatitis B virus in Papua Province, Indonesia. *Virus Research,* 163**,** 129-140.

MULYANTO, D. S., SURAYAH, K., TSUDA, F., ICHIYAMA, K., TAKAHASHI, M. & OKAMOTO, H. 2009. A nationwide molecular epidemiological study on hepatitis B virus in Indonesia: identification of two novel subgenotypes, B8 and C7. *Arch Virol,* 154**,** 1047-1059.

MWANGI, J., NGANGA, Z., SONGOK, E., KINYUA, J., LAGAT, N., MURIUKI, J., LIHANA, R., KHAMADI, S., OSMAN, S., LWEMBE, R., KIPTOO, M., MWAU, M., CHIRCHIR, R., MPOKE, S., NYAMONGO, J., OKOTH, F., YAMADA, R., KAGEYAMA, S. & ICHIMURA, H. 2008. Molecular genetic diversity of hepatitis B virus in Kenya. *Intervirology,* 51**,** 417-21.

NAIL, A., ELTIGANNI, S. & IMAM, A. 2008. Seroprevalence of Hepatitis B and C among health care workers in Omdurman, Sudan. *History,* 11**,** 9.

NASSAL, M. 2008. Hepatitis B viruses: reverse transcription a different way. *Virus Res,* 134**,** 235-49.

NASSAL, M. & RIEGER, A. 1996. A bulged region of the hepatitis B virus RNA encapsidation signal contains the replication origin for discontinuous first-strand DNA synthesis. *Journal of Virology,* 70**,** 2764-2773.

NAUMANN, H., SCHAEFER, S., YOSHIDA, C. F., GASPAR, A. M., REPP, R. & GERLICH, W. H. 1993. Identification of a new hepatitis B virus (HBV) genotype

from Brazil that expresses HBV surface antigen subtype adw4. *J Gen Virol,* 74 (Pt 8)**,** 1627-32.

NORDER, H., ARAUZ-RUIZ, P., BLITZ, L., PUJOL, F. H., ECHEVARRIA, J. M. & MAGNIUS, L. O. 2003. The T1858 variant predisposing to the precore stop mutation correlates with one of two major genotype F hepatitis B virus clades. *Journal of General Virology,* 84**,** 2083-2087.

NORDER, H., COUROUCE, A. M., COURSAGET, P., ECHEVARRIA, J. M., LEE, S. D., MUSHAHWAR, I. K., ROBERTSON, B. H., LOCARNINI, S. & MAGNIUS, L. O. 2004. Genetic diversity of hepatitis B virus strains derived worldwide: Genotypes, subgenotypes, and HBsAg subtypes. *Intervirology,* 47**,** 289-309.

NORDER, H., COUROUCE, A. M. & MAGNIUS, L. O. 1992a. Molecular basis of hepatitis B virus serotype variations within the four major subtypes. *The Journal of general virology,* 73 (Pt 12)**,** 3141-5.

NORDER, H., COUROUCE, A. M. & MAGNIUS, L. O. 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.

NORDER, H., EBERT, J. W., FIELDS, H. A., MUSHAHWAR, I. K. & MAGNIUS, L. O. 1996. Complete sequencing of a gibbon hepatitis B virus genome reveals a unique genotype distantly related to the chimpanzee hepatitis B virus. *Virology,* 218**,** 214-23.

NORDER, H., HAMMAS, B., LEE, S. D., BILE, K., COUROUCE, A. M., MUSHAHWAR, I. K. & MAGNIUS, L. O. 1993. Genetic relatedness of hepatitis B viral strains of diverse geographical origin and natural variations in the primary structure of the surface antigen. *The Journal of general virology,* 74 (Pt 7)**,** 1341-8.

NORDER, H., HAMMAS, B., LOFDAHL, S., COUROUCE, A. M. & MAGNIUS, L. O. 1992b. Comparison of the amino acid sequences of nine different serotypes of

hepatitis B surface antigen and genomic classification of the corresponding hepatitis B virus strains. *J Gen Virol,* 73 (Pt 5)**,** 1201-8.

NURAINY, N., MULJONO, D. H., SUDOYO, H. & MARZUKI, S. 2008. Genetic study of hepatitis B virus in Indonesia reveals a new subgenotype of genotype B in east Nusa Tenggara. *Arch Virol,* 153**,** 1057-65.

OCHWOTO, M., CHAUHAN, R., GOPLAKRISHNAN, D., CHEN, C.-Y., NG'ANG'A, Z., OKOTH, F., KIOKO, H., KIMOTHO, J., KAIGURI, P. & KRAMVIS, A. 2013. Genotyping and molecular characterization of hepatitis B virus in liver disease patients in Kenya. *Infection, Genetics and Evolution*.

ODEMUYIWA, S. O., MULDERS, M. N., OYEDELE, O. I., OLA, S. O., ODAIBO, G. N., OLALEYE, D. O. & MULLER, C. P. 2001. Phylogenetic analysis of new hepatitis B virus isolates from Nigeria supports endemicity of genotype E in West Africa*. *Journal of Medical Virology,* 65**,** 463-469.

OKAMOTO, H., TSUDA, F., AKAHANE, Y., SUGAI, Y., YOSHIBA, M., MORIYAMA, K., TANAKA, T., MIYAKAWA, Y. & MAYUMI, M. 1994. Hepatitis B virus with mutations in the core promoter for an e antigen-negative phenotype in carriers with antibody to e antigen. *Journal of Virology,* 68**,** 8102-8110.

OKAMOTO, H., TSUDA, F., SAKUGAWA, H., SASTROSOEWIGNJO, R. I., IMAI, M., MIYAKAWA, Y. & MAYUMI, M. 1988. Typing Hepatitis B Virus by Homology in Nucleotide Sequence: Comparison of Surface Antigen Subtypes. *Journal of General Virology,* 69**,** 2575-2583.

OLINGER, C. M., VENARD, V., NJAYOU, M., OYEFOLU, A. O. B., MAIGA, I., KEMP, A. J., OMILABU, S. A., LE FAOU, A. & MULLER, C. P. 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.

ORITO, E., SUGAUCHI, F., TANAKA, Y., ICHIDA, T., SATA, M., TANAKA, E., OKANOUE, T., SAKUGAWA, H., WATANABE, H. & MIYAKAWA, H. 2005. Differences of hepatocellular carcinoma patients with hepatitis B virus genotypes of Ba, Bj or C in Japan. *Intervirology,* 48**,** 239-245.

OWIREDU, W. K., KRAMVIS, A. & KEW, M. C. 2001. Hepatitis B virus DNA in serum of healthy black African adults positive for hepatitis B surface antibody alone: possible association with recombination between genotypes A and D. *Journal of Medical Virology,* 64**,** 441-454.

PARASKEVIS, D., MAGIORKINIS, G., MAGIORKINIS, E., HO, S. Y., BELSHAW, R., ALLAIN, J. P. & HATZAKIS, A. 2013. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology*.

PARKIN, D. M. 2001. Global cancer statistics in the year 2000. *Lancet Oncol,* 2**,** 533-43.

PARKIN, D. M., PISANI, P. & FERLAY, J. 1999. Estimates of the worldwide incidence of 25 major cancers in 1990. *International journal of cancer,* 80**,** 827-841.

PARKIN, D. M., SITAS, F., CHIRENJE, M., STEIN, L., ABRATT, R. & WABINGA, H. 2008. Part I: Cancer in Indigenous Africans—burden, distribution, and trends. *The Lancet Oncology,* 9**,** 683-692.

PAWLOTSKY, J. M. 2005. The concept of hepatitis B virus mutant escape. *J Clin Virol,* 34 Suppl 1**,** S125-9.

POURKARIM, M. R., AMINI-BAVIL-OLYAEE, S., LEMEY, P., MAES, P. & RANST, M. V. 2011a. HBV subgenotype misclassification expands quasi-subgenotype A3. *Clinical Microbiology and Infection,* 17**,** 947-949.

POURKARIM, M. R., AMINI-BAVIL-OLYAEE, S., LEMEY, P., MAES, P. & VAN RANST, M. 2010a. Are hepatitis B virus “subgenotypes” defined accurately? *Journal of Clinical Virology,* 47**,** 356-360.

POURKARIM, M. R., LEMEY, P., AMINI-BAVIL-OLYAEE, S., HOUSPIE, L., VERBEECK, J., RAHMAN, M., MAES, P., VANWIJNGAERDEN, E., NEVENS, F. & VAN RANST, M. 2011b. Molecular characterization of hepatitis B virus strains circulating in Belgian patients co-infected with HIV and HBV: Overt and occult infection. *Journal of Medical Virology,* 83**,** 1876-1884.

POURKARIM, M. R., LEMEY, P., AMINI-BAVIL-OLYAEE, S., MAES, P. & VAN RANST, M. 2010b. Novel hepatitis B virus subgenotype A6 in African-Belgian patients. *Journal of Clinical Virology,* 47**,** 93-96.

PRASSOLOV, A., HOHENBERG, H., KALININA, T., SCHNEIDER, C., COVA, L., KRONE, O., FROLICH, K., WILL, H. & SIRMA, H. 2003. New hepatitis B virus of cranes that has an unexpected broad host range. *J Virol,* 77**,** 1964-76.

PULT, I., NETTER, H. J., BRUNS, M., PRASSOLOV, A., SIRMA, H., HOHENBERG, H., CHANG, S. F., FROLICH, K., KRONE, O., KALETA, E. F. & WILL, H. 2001. Identification and analysis of a new hepadnavirus in white storks. *Virology,* 289**,** 114-28.

RAIMONDO, G., ALLAIN, J. P., BRUNETTO, M. R., BUENDIA, M. A., CHEN, D. S., COLOMBO, M., CRAXÌ, A., DONATO, F., FERRARI, C. & GAETA, G. B. 2008. Statements from the Taormina expert meeting on occult hepatitis B virus infection. *Journal of Hepatology,* 49**,** 652.

RAIMONDO, G., CACCAMO, G., FILOMIA, R. & POLLICINO, T. Year. Occult HBV infection. *In:* Seminars in Immunopathology, 2013. Springer, 39-52.

RAIMONDO, G., POLLICINO, T., CACCIOLA, I. & SQUADRITO, G. 2007. Occult hepatitis B virus infection. *Journal of Hepatology,* 46**,** 160-170.

RAMÍREZ, C., GREGORI, J., BUTI, M., TABERNERO, D., CAMÓS, S., CASILLAS, R., QUER, J., ESTEBAN, R., HOMS, M. & RODRIGUEZ-FRÍAS, F. 2013. A comparative study of ultra-deep pyrosequencing and cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. *Antiviral Research*.

RODRIGUEZ-FRIAS, F., BUTI, M., TABERNERO, D. & HOMS, M. 2013. Quasispecies structure, cornerstone of hepatitis B virus infection: Mass sequencing approach. *World Journal of Gastroenterology: WJG,* 19**,** 6995.

RODRIGUEZ, C., CHEVALIEZ, S., BENSADOUN, P. & PAWLOTSKY, J.-M. 2013. Characterization of the dynamics of hepatitis B virus resistance to adefovir by ultra-deep pyrosequencing. *Hepatology***,** n/a-n/a.

RONAGHI, M. 2001. Pyrosequencing sheds light on DNA sequencing. *Genome research,* 11**,** 3-11.

ROYCHOUDHURY, S. & SHIH, C. 1990. cis rescue of a mutated reverse transcriptase gene of human hepatitis B virus by creation of an internal ATG. *J Virol,* 64**,** 1063-9.

SAKAMOTO, T., TANAKA, Y., ORITO, E., CO, J., CLAVIO, J., SUGAUCHI, F., ITO, K., OZASA, A., QUINO, A., UEDA, R., SOLLANO, J. & MIZOKAMI, M. 2006. Novel subtypes (subgenotypes) of hepatitis B virus genotypes B and C among chronic liver disease patients in the Philippines. *J Gen Virol,* 87**,** 1873-82.

SAKAMOTO, T., TANAKA, Y., SIMONETTI, J., OSIOWY, C., BORRESEN, M. L., KOCH, A., KURBANOV, F., SUGIYAMA, M., MINUK, G. Y., MCMAHON, B. J., JOH, T. & MIZOKAMI, M. 2007. Classification of hepatitis B virus genotype B into 2 major types based on characterization of a novel subgenotype in Arctic indigenous populations. *J Infect Dis,* 196**,** 1487-92.

SALLAM, T. A. & WILLIAM TONG, C. 2004. African links and hepatitis B virus genotypes in the Republic of Yemen. *Journal of Medical Virology,* 73**,** 23-28.

SÁNCHEZ, L. V., MALDONADO, M., BASTIDAS-RAMÍREZ, B. E., NORDER, H. & PANDURO, A. 2002. Genotypes and S-gene variability of Mexican hepatitis B virus strains. *Journal of Medical Virology,* 68**,** 24-32.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences,* 74**,** 5463-5467.

SANTANTONIO, T., JUNG, M.-C., SCHNEIDER, R., FERNHOLZ, D., MILELLA, M., MONNO, L., PASTORE, G., PAPE, G. R. & WILL, H. 1992. Hepatitis B virus genomes that cannot synthesize pre-S2 proteins occur frequently and as dominant virus populations in chronic carriers in Italy. *Virology,* 188**,** 948-952.

SAUDY, N., SUGAUCHI, F., TANAKA, Y., SUZUKI, S., AAL, A. A., ZAID, M. A., AGHA, S. & MIZOKAMI, M. 2003. Genotypes and phylogenetic characterization of hepatitis B and delta viruses in Egypt. *Journal of Medical Virology,* 70**,** 529-536.

SCHUPPAN, D. & AFDHAL, N. H. 2008. Liver cirrhosis. *The Lancet,* 371**,** 838-851.

SCOTT, R. M., SNITBHAN, R., BANCROFT, W. H., ALTER, H. J. & TINGPALAPONG, M. 1980. Experimental transmission of hepatitis B virus by semen and saliva. *J Infect Dis,* 142**,** 67-71.

SEEGER, C. & MASON, W. S. 2000. Hepatitis B virus biology. *Microbiology and Molecular Biology Reviews,* 64**,** 51-+.

SHEPARD, C. W., SIMARD, E. P., FINELLI, L., FIORE, A. E. & BELL, B. P. 2006. Hepatitis B virus infection: epidemiology and vaccination. *Epidemiol Rev,* 28**,** 112-25.

SHERLOCK, S. & DOOLEY, J. 2008. *Diseases of the liver and biliary system,* Wiley. com.

SHI, W., ZHANG, Z., LING, C., ZHENG, W., ZHU, C., CARR, M. J. & HIGGINS, D. G. 2013. Hepatitis B virus Subgenotyping: History, Effects of Recombination, Misclassifications, and Corrections. *Infection, Genetics and Evolution*.

SHI, W., ZHU, C., ZHENG, W., CARR, M. J., HIGGINS, D. G. & ZHANG, Z. 2012a. Subgenotype reclassification of genotype B hepatitis B virus. *BMC Gastroenterology,* 12**,** 116.

SHI, W., ZHU, C., ZHENG, W., ZHENG, W., LING, C., CARR, M. J., HIGGINS, D. G. & ZHANG, Z. 2012b. Subgenotyping of Genotype C Hepatitis B Virus: Correcting Misclassifications and Identifying a Novel Subgenotype. *Plos One,* 7**,** e47271.

SIMMONDS, P. & MIDGLEY, S. 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. *Journal of Virology,* 79**,** 15467-15476.

SNAP, S. N. A. C. P. 2012. Global AIDS response progress reporting 2010-2011.

SOLMONE, M., VINCENTI, D., PROSPERI, M. C., BRUSELLES, A., IPPOLITO, G. & CAPOBIANCHI, M. R. 2009. Use of massively parallel ultradeep pyrosequencing to characterize the genetic diversity of hepatitis B virus in drug-resistant and drug-naive patients and to detect minor variants in reverse transcriptase and hepatitis B S antigen. *J Virol,* 83**,** 1718-26.

SPRENGEL, R., KALETA, E. F. & WILL, H. 1988. Isolation and characterization of a hepatitis B virus endemic in herons. *J Virol,* 62**,** 3832-9.

STUYVER, L., DE GENDT, S., VAN GEYT, C., ZOULIM, F., FRIED, M., SCHINAZI, R. F. & ROSSAU, R. 2000. A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *Journal of General Virology,* 81**,** 67-74.

SUGAUCHI, F., KUMADA, H., ACHARYA, S. A., SHRESTHA, S. M., GAMUTAN, M. T., KHAN, M., GISH, R. G., TANAKA, Y., KATO, T., ORITO, E., UEDA, R., MIYAKAWA, Y. & MIZOKAMI, M. 2004. Epidemiological and sequence differences

between two subtypes (Ae and Aa) of hepatitis B virus genotype A. *J Gen Virol,* 85**,** 811-20.

SUGAUCHI, F., MIZOKAMI, M., ORITO, E., OHNO, T., KATO, H., SUZUKI, S., KIMURA, Y., UEDA, R., BUTTERWORTH, L. A. & COOKSLEY, W. G. E. 2001. A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. *Journal of General Virology,* 82**,** 883-892.

SUGAUCHI, F., ORITO, E., ICHIDA, T., KATO, H., SAKUGAWA, H., KAKUMU, S., ISHIDA, T., CHUTAPUTTI, A., LAI, C.-L., UEDA, R., MIYAKAWA, Y. & MIZOKAMI, M. 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.

SUGAUCHI, F., ORITO, E., KATO, H., SUZUKI, S., KAWAKITA, S., SAKAMOTO, Y., FUKUSHIMA, K., AKIBA, T., YOSHIHARA, N. & UEDA, R. 2003. Genotype, serotype, and phylogenetic characterization of the complete genome sequence of hepatitis B virus isolates from Malawian chronic carriers of the virus. *Journal of Medical Virology,* 69**,** 33-40.

SUMMERS, J. & MASON, W. S. 1982. Replication of the genome of a hepatitis B--like virus by reverse transcription of an RNA intermediate. *Cell,* 29**,** 403-15.

SUZUKI, S., SUGAUCHI, F., ORITO, E., KATO, H., USUDA, S., SIRANSY, L., ARITA, I., SAKAMOTO, Y., YOSHIHARA, N., EL-GOHARY, A., UEDA, R. & MIZOKAMI, M. 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**,** 459-65.

TAKAHASHI, K., AOYAMA, K., OHNO, N., IWATA, K., AKAHANE, Y., BABA, K., YOSHIZAWA, H. & MISHIRO, S. 1995. The precore/core promoter mutant (T1762A1764) of hepatitis B virus: clinical significance and an easy method for detection. *J Gen Virol,* 76 (Pt 12)**,** 3159-64.

TAKAHASHI, K., BROTMAN, B., USUDA, S., MISHIRO, S. & PRINCE, A. M. 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.

TANAKA, Y., HASEGAWA, I., KATO, T., ORITO, E., HIRASHIMA, N., ACHARYA, S. K., GISH, R. G., KRAMVIS, A., KEW, M. C. & YOSHIHARA, N. 2004. A case-control study for differences among hepatitis B virus infections of genotypes A (subtypes Aa and Ae) and D. *Hepatology,* 40**,** 747-755.

TATEMATSU, K., TANAKA, Y., KURBANOV, F., SUGAUCHI, F., MANO, S., MAESHIRO, T., NAKAYOSHI, T., WAKUTA, M., MIYAKAWA, Y. & MIZOKAMI, M. 2009. A Genetic Variant of Hepatitis B Virus Divergent from Known Human and Ape Genotypes Isolated from a Japanese Patient and Provisionally Assigned to New Genotype J. *Journal of Virology,* 83**,** 10538-10547.

TECHNOLOGIES, L. 2013. SOLiD, Life Technologies, accessed from http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems/solid-4-system.html?CID=FL-091411_solid4 on 02/09/2013.

TESTUT, P., RENARD, C. A., TERRADILLOS, O., VITVITSKI-TREPO, L., TEKAIA, F., DEGOTT, C., BLAKE, J., BOYER, B. & BUENDIA, M. A. 1996. A new hepadnavirus endemic in arctic ground squirrels in Alaska. *J Virol,* 70**,** 4210-9.

THAKUR, V., GUPTAN, R. C., KAZIM, S. N., MALHOTRA, V. & SARIN, S. K. 2002. Profile, spectrum and significance of HBV genotypes in chronic liver disease patients in the Indian subcontinent. *Journal of Gastroenterology and Hepatology,* 17**,** 165-170.

THEDJA, M., MULJONO, D., NURAINY, N., SUKOWATI, C., VERHOEF, J. & MARZUKI, S. 2011. Ethnogeographical structure of hepatitis B virus genotype distribution in Indonesia and discovery of a new subgenotype, B9. *Archives of Virology,* 156**,** 855-868.

THIO, C. L. 2009. Hepatitis B and human immunodeficiency virus coinfection. *Hepatology,* 49**,** S138-S145.

TIOLLAIS, P., POURCEL, C. & DEJEAN, A. 1985. The hepatitis B virus. *Nature,* 317**,** 489-95.

TONG, S., LI, J., WANDS, J. R. & WEN, Y.-M. 2013. Hepatitis B virus genetic variants: biological properties and clinical implications. *Emerging Microbes & Infections,* 2**,** e10.

TORBENSON, M. & THOMAS, D. L. 2002. Occult hepatitis B. *The Lancet Infectious Diseases,* 2**,** 479-486.

TORRESI, J. 2002. The virological and clinical significance of mutations in the overlapping envelope and polymerase genes of hepatitis B virus. *Journal of Clinical Virology,* 25**,** 97-106.

TRAN, T. T., TRINH, T. N. & ABE, K. 2008. New complex recombinant genotype of hepatitis B virus identified in Vietnam. *J Virol,* 82**,** 5657-63.

TSAI, A., KAWAI, S., KWEI, K., GEWAILY, D., HUTTER, A., TONG, D. R., LI, J., WANDS, J. R. & TONG, S. 2009. Chimeric constructs between two hepatitis B virus genomes confirm transcriptional impact of core promoter mutations and reveal multiple effects of core gene mutations. *Virology,* 387**,** 364-372.

UNAIDS. 2010. *2010 Global Report on the global AIDS epidemic*, World Health Organization.

URBAN, S., SCHULZE, A., DANDRI, M. & PETERSEN, J. 2010. The replication cycle of hepatitis B virus. *Journal of Hepatology,* 52**,** 282-284.

USUDA, S., OKAMOTO, H., TANAKA, T., KIDD-LJUNGGREN, K., HOLLAND, P. V., MIYAKAWA, Y. & MAYUMI, M. 2000. Differentiation of hepatitis B virus genotypes D and E by ELISA using monoclonal antibodies to epitopes on the preS2-region product. *J Virol Methods,* 87**,** 81-9.

UTSUMI, T., LUSIDA, M. I., YANO, Y., NUGRAHAPUTRA, V. E., AMIN, M., JUNIASTUTI, SOETJIPTO, HAYASHI, Y. & HOTTA, H. 2009. Complete Genome Sequence and Phylogenetic Relatedness of Hepatitis B Virus Isolates in Papua, Indonesia. *Journal of Clinical Microbiology,* 47**,** 1842-1847.

VILLENEUVE, J. P. 2005. The natural history of chronic hepatitis B virus infection. *J Clin Virol,* 34 Suppl 1**,** S139-42.

VOGT, V. 1997. Retroviral virions and genomes.

WAI, C. T., FONTANA, R. J., POLSON, J., HUSSAIN, M., SHAKIL, A. O., HAN, S. H., DAVERN, T. J., LEE, W. M., LOK, A. S. F. & THE, U. S. A. L. F. S. G. 2005. Clinical outcome and virological characteristics of hepatitis B-related acute liver failure in the United States. *Journal of Viral Hepatitis,* 12**,** 192-198.

WANG, G. & SEEGER, C. 1993. Novel mechanism for reverse transcription in hepatitis B viruses. *Journal of Virology,* 67**,** 6507-6512.

WANLESS, I. R., NAKASHIMA, E. & SHERMAN, M. 2000. Regression of human cirrhosis: morphologic features and the genesis of incomplete septal cirrhosis. *Archives of pathology & laboratory medicine,* 124**,** 1599-1607.

WASMUTH, J.-C. 2009. Hepatitis B-Epidemiology, transmission and natural history. *Hepatology. Dusseldorf: Flying Publisher***,** 25-36.

WATSON, J. D. & CRICK, F. H. 1953. Molecular structure of nucleic acids. *Nature,* 171**,** 737-738.

WHO 2006. *WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children*, World Health Organization.

WHO 2012. Hepatitis B virus Fact sheet N° 204.

XIE, Y., ZHAI, J., DENG, Q., TIOLLAIS, P., WANG, Y. & ZHAO, M. 2010. Entry of hepatitis B virus: mechanism and new therapeutic target. *Pathol Biol (Paris),* 58**,** 301-7.

Y LEONE, F. G. P., MBAYED, V. A. & CAMPOS, R. H. 2003. Evolutionary history of hepatitis B virus genotype F: an in-depth analysis of Argentine isolates. *Virus Genes,* 27**,** 103-110.

YAN, H., ZHONG, G., XU, G., HE, W., JING, Z., GAO, Z., HUANG, Y., QI, Y., PENG, B. & WANG, H. 2012. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *eLife Sciences,* 1.

YIM, H. J., HUSSAIN, M., LIU, Y., WONG, S. N., FUNG, S. K. & LOK, A. S. 2006. Evolution of multi-drug resistant hepatitis B virus during sequential therapy. *Hepatology,* 44**,** 703-12.

YIM, H. J. & LOK, A. S. 2006. Natural history of chronic hepatitis B virus infection: what we knew in 1981 and what we know in 2005. *Hepatology,* 43**,** S173-81.

YOKOSUKA, O. & ARAI, M. 2006. Molecular biology of hepatitis B virus: effect of nucleotide substitutions on the clinical features of chronic hepatitis B. *Med Mol Morphol,* 39**,** 113-20.

YOUSIF, M. & KRAMVIS, A. 2013. Genotype D of hepatitis B virus and its subgenotypes: An update. *Hepatology Research,* 43**,** 355-364.

YOUSIF, M., MUDAWI, H., BAKHIET, S., GLEBE, D. & KRAMVIS, A. 2013. Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan. *BMC Infectious Diseases,* 13**,** 328.

YU, H., YUAN, Q., GE, S.-X., WANG, H.-Y., ZHANG, Y.-L., CHEN, Q.-R., ZHANG, J., CHEN, P.-J. & XIA, N.-S. 2010. Molecular and Phylogenetic Analyses Suggest an Additional Hepatitis B Virus Genotype "I". *Plos One,* 5**,** e9297.

YUEN, M. F., TANAKA, Y., MIZOKAMI, M., YUEN, J. C., WONG, D. K., YUAN, H. J., SUM, S. M., CHAN, A. O., WONG, B. C. & LAI, C. L. 2004. Role of hepatitis B virus genotypes Ba and C, core promoter and precore mutations on hepatocellular carcinoma: a case control study. *Carcinogenesis,* 25**,** 1593-8.

YUH, C. H., CHANG, Y. L. & TING, L. P. 1992. Transcriptional regulation of precore and pregenomic RNAs of hepatitis B virus. *Journal of Virology,* 66**,** 4073-4084.

YUH, C. H. & TING, L. P. 1990. The genome of hepatitis B virus contains a second enhancer: cooperation of two elements within this enhancer is required for its function. *J Virol,* 64**,** 4281-7.

YUH, C. H. & TING, L. P. 1993. Differentiated liver cell specificity of the second enhancer of hepatitis B virus. *J Virol,* 67**,** 142-9.

ZHOU, B., XIAO, L., WANG, Z., CHANG, E. T., CHEN, J. & HOU, J. 2011. Geographical and ethnic distribution of the HBV C/D recombinant on the Qinghai-Tibet Plateau. *Plos One,* 6**,** e18708.

ZHOU, X., REN, L., LI, Y., ZHANG, M., YU, Y. & YU, J. 2010. The next-generation sequencing technology: a technology review and future perspective. *Science China Life Sciences,* 53**,** 44-57.

ZOULIM, F. & LOCARNINI, S. 2009. Hepatitis B virus resistance to nucleos(t)ide analogues. *Gastroenterology,* 137**,** 1593 - 1608.