

https://dspace.org/

# Genotyping and hepatitis B virus in liver disease patients from Kerala, India

**Keyter, Mark Andrew Robson**

**2014-04-04**
http://hdl.handle.net10539/14499

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

# GENOTYPING AND MOLECULAR CHARACTERISATION OF HEPATITIS B VIRUS IN LIVER DISEASE PATIENTS FROM KERALA, INDIA

Mark Andrew Robson Keyter

Dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Medicine

Johannesburg, 2013

**DECLARATION**

I, Mark Andrew Robson Keyter declare that this dissertation is my own work. It is being submitted for the degree of Master of Science in Medicine at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.

.....

..... day of ..... 2013

**PRESENTATIONS AND PUBLICATIONS**

1. Conference: The 21st Conference of the Asian Pacific Association for the Study of the Liver (APASL), Bangkok, Thailand. February 17-20, 2011.

   **Poster: Mark Keyter,** Deepak Gopalakrishnan, K.T. Shenoy, Leena Mohanan, Anna Kramvis. Molecular Characterisation of Hepatitis B Virus isolates from Chronically infected, Cirrhotic and Hepatocellular Carcinoma Patients from Kerala, India

2. Symposium: Hepatitis B Virus Genotypes ~ Global Diversity and Evolution, Johannesburg, South Africa. September 29-30, 2011.

   **Presentation: Mark Keyter**. Molecular Characterisation of Hepatitis B Virus isolates from Chronic, Cirrhotic and Hepatocellular Carcinoma Patients from Kerala, India

3. Vermeulen M, Dickens C, Lelie N, Walker E, Coleman C, **Keyter M**, Reddy R, Crookes R, Kramvis A. (2012). 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* 52: 880-92

4. Deepak Gopalakrishnan, **Mark Keyter**, Shenoy KT, Leena K B, Thayumanavan L, Varghese Thomas, Vinayakumar KR, Charles Panackal, Arun T Korah, Ramesh M, Anna Kramvis. Predominance of subgenotype A1

of hepatitis B virus in liver disease patients from the State of Kerala, Southern India. *Submitted.*

**ACKNOWLEDGEMENTS**

First and foremost, I would like to thank my supervisor, Prof Anna Kramvis. Thank you for your endless support and guidance, and the wealth of knowledge you have shared with me, not only about the hepatitis B virus, but about life in general. Your understanding of the molecular evolution of HBV and in particular, subgenotype A1 is surpassed by no other. Thank you for all the hours spent reviewing and editing my dissertation. You have taught me to critically analyse everything I do, and I know this will stand me in good stead throughout my career. I am honoured to have completed my Masters Research project under the auspices of the Hepatitis Virus Diversity Research Programme.

To Dr Deepak Gopalakrishnan, thank you for collecting all the serum samples and allowing me to use them for my research project. Thank you for all your support and your assistance in helping me put my thesis together. Your assistance with the statistical analysis has allowed me to complete my dissertation, and for this, I thank you.

To Trevor Bell, your bioinformatics tools have allowed me to analyse my sequences and data with ease, and for this I am very grateful. Nimisha Bhoola, thank you for teaching me how to perform an array of laboratory techniques, and especially for sharing your knowledge in the field of cloning with me. To all the other members of the HVDRP team, that you for making the 'Think Tank' such a wonderful learning place and experience, your assistance and advice is going to be sorely missed.

To my family, thank you for your love and support, and allowing me to follow my dreams. Andrea Julsing, thank you for believing in me and pushing me to achieve my goals, this dissertation would not have been possible without you.

For financial support, I thank the University of the Witwatersrand, the National Research Foundation, the Poliomyelitis Research Foundation, and the Ernst and Ethel Eriksen Trust.

**ABSTRACT**

India is classified as an intermediate hepatitis B virus (HBV) endemic zone, with an estimated 50 million HBV carriers, thereby forming the second largest pool of chronic HBV infection in the world. Kerala, in south-west India is the most densely populated state of India, with a population of 33 million. There is currently a paucity of information on the prevalence of HBV genotypes and the respective subgenotypes in Kerala. Therefore the aim of this study was to molecularly characterise HBV isolated from liver disease patients in Kerala, and to determine the clinical implications of HBV genotypes and mutations on liver disease progression. Seventy HBV DNA-positive serum samples belonging to the three disease groups: Chronic Hepatitis (CH), cirrhosis (CR) and hepatocellular carcinoma (HCC) were used in the present study. Analysis of 50 complete S region sequences, 57 BCP/PreC sequences and 9 representative complete HBV genome sequences was performed. Irrespective of disease status, the predominant genotype was A (64%), with 97% belonging to subgenotype A1, followed by genotype D (34%) and a single strain of genotype C. All subgenotype A1 isolates clustered within the 'Asian' clade as a separate branch, together with sequences from Asia, Africa and South America. A significantly higher prevalence of subgenotype A1 was found in HCC patients, with these patients being significantly younger than those infected with genotype D ($p<0.05$). Nine pre-S deletion mutants were detected (8 subgenotype A1; 1 subgenotype D2). Although a higher frequency of pre-S deletion mutants where found in subgenotype A1 isolates from HCC patients, this did not reach statistical significance ($p=0.4$). The ps2: M1I/T was found exclusively in subgenotype A1 isolates, and this was significantly associated with HCC as compared to other disease groups in patients infected with subgenotype A1($p=0.0072$). The ps2: F22L was found in genotypes A and D, and

was significantly associated with HCC in genotype D ($p=0.0098$). The G1862T was significantly associated with subgenotype A1 (90%), and in combination with the A1762T/G1764A double mutation, was significantly associated with HCC in subgenotype A1 ($p<0.05$). Subgenotype A1 was predominant in liver disease patients from Kerala, with a high prevalence in HCC patients. This study confirms the association shown from previous studies, that subgenotype A1 is significantly associated with HCC, and also showed a strong correlation between subgenotype A1 and the development of HCC at a younger age.

**TABLE OF CONTENTS**

DECLARATION .....i
PRESENTATIONS AND PUBLICATIONS .....ii
ACKNOWLEDGEMENTS.....iv
ABSTRACT.....vi
TABLE OF CONTENTS.....viii
LIST OF FIGURES.....xii
LIST OF TABLES.....xiv
LIST OF ABBREVIATIONS.....xv
**CHAPTER 1: INTRODUCTION.....1**
1.1THE HEPATITIS B VIRUS.....4
1.1.1 HISTORY OF THE HBV.....4
1.1.2 CLASSIFICATION: FAMILY HEPADNAVIRIDIAE.....5
1.1.3 STRUCTURE OF THE HBV.....6
1.1.4 HBV GENOME ORGANISATION.....8
1.1.5 THE HBV TRANSCIPTION PRODUCTS.....9
1.1.6. HBV VIRAL *cis*-ELEMENTS.....11
1.1.6.1 HBV PROMOTERS.....11
1.1.6.2 HBV ENHANCERS.....12
1.1.6.3 THE HBV ENCAPSIDATION SIGNAL: EPSILON.....13
1.1.6.4 THE POLYADENYLATION SIGNAL.....13
1.1.7 THE HBV GENE PRODUCTS.....14
1.1.7.1 PRECORE AND CORE GENE.....14
1.1.7.2 THE POLYMERASE PROTEIN.....15
1.1.7.3 THE HBX PROTEIN.....16
1.1.7.4 THE HBV SURFACE PROTEINS.....16
1.1.7.4.1 THE LARGE ENVELOPE PROTEIN.....17
1.1.7.4.2. THE MEDIUM ENVELOPE PROTEIN.....18
1.1.7.4.3 THE SMALL ENVELOPE PROTEIN.....18
1.1.8 THE HBV LIFE CYCLE.....21
1.1.9 CLINICAL SPECTRUM OF HBV INFECTION.....25
1.1.9.1 ACUTE HBV INFECTION.....25
1.1.9.2 CHRONIC HEPATITIS B VIRUS INFECTION.....25

1.1.9.3 THE LONG TERM SEQUELAE OF CHB INFECTION.....28
1.1.9.3.1 CIRRHOSIS.....28
1.1.9.3.2 HEPATOCELLULAR CARCINOMA.....28
1.2 HBV GENOTYPES.....30
1.2.1 MOLECULAR CHARACTERISATION AND GEOGRAPHICAL DISTRIBUTION.....30
1.2.3 SUBGENOTYPE A1 OF THE HEPATITIS B VIRUS.....32
1.2.3 EFFECT OF HBV ON DISEASE PROGRESSION AND CLINICAL OUTCOME.....33
1.3 HBV MUTANTS: DISTRIBUTION AND EFFECT ON DISEASE PROGRESSION.....35
1.3.1 THE BASIC CORE PROMOTER.....35
1.3.2 THE PRECORE/CORE REGION.....36
1.3.3 THE HBX PROTEIN.....37
1.3.4 THE POLYMERASE PROTEIN.....37
1.3.5 THE COMPLETE SURFACE REGION.....38
1.4 HBV INFECTION ON THE INDIAN SUBCONTINENT.....40
1.4.1 HBV GENOTYPES: DISTRIBUTION AND EFFECT ON DISEASE PROGRESSION.....40
1.5 AIMS AND OBJECTIVES OF THE STUDY.....44
**CHAPTER 2: MATERIALS AND METHODS.....46**
2.1 OVERVIEW OF STUDY DESIGN AND ANALYSIS.....46
2.2 STUDY SAMPLES AND PATIENT ALLOCATION.....47
2.3 SERUM SAMPLE COLLECTION AND STORAGE.....48
2.4 SERUM DNA EXTRACTION.....48
2.5 DETECTION OF BASIC CORE PROMOTER/PRECORE/CORE REGION MUTANTS.....48
2.5.1 POLYMERASE CHAIN REACTION OF THE BCP/PRE-C/C REGION.....48
2.5.2 SEQUENCING OF THE BCP/PRE-C/C AMPLICON.....50
2.6 COMPLETE S REGION ANALYSIS OF THE HBV GENOME.....50
2.6.1 POLYMEARSE CHAIN REACTION OF THE COMPLETE S REGION.....50
2.6.2 GEL EXTRACTION OF THE 2 KB PCR AMPLICON.....52

2.6.3 SEQUENCING OF THE COMPLETE S REGION.....52
2.7 MOLECULAR ANALYSIS OF THE COMPLETE HBV GENOME.....53
2.7.1 CLONING OF THE 3.2 KB HBV GENOME.....53
2.7.1.1 FULL LENGTH HEPATITIS B VIRUS GENOME AMPLIFICATION.....53
2.7.1.2 CLONING OF THE AMPLIFIED 3.2 KB PCR FRAGMENT.....54
2.7.1.3 SEQUENCING OF THE COMPLETE 3.2 KB HBV GENOME.56
2.7.2 COMPLETE HBV GENOME ANALYSIS USING AMPLIFICATION OF TWO SUBGENOMIC FRAGMENTS.....57
2.7.2.1 PCR AMPLIFICATION OF THE 2 KB AND 1.6 KB SUBGENOMIC FRAGMENTS.....57
2.7.2.2 SEQUENCING OF THE 1.6 KB SUBGENOMIC PCR AMPLICON.....58
2.8 BIOINFORMATICS AND PHYLOGENETIC ANALYSIS.....59
2.9 STATISTICAL ANALYSIS.....60
**CHAPTER 3: RESULTS.....61**
3.1 DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF PATIENTS.....62
3.2 COMPLETE SURFACE REGION AMPLIFICATION AND SEQUENCE ANALYSIS.....63
3.3 GENOTYPING USING PHYLOGENETIC ANALYSIS OF THE COMPLETE S REGION.....64
3.4. GENOTYPE DISTRIBUTION AMONG THE THREE DISEASE GROUPS.....70
3.5 PREVALENCE AND PATTERN OF COMPLETE S REGION MUTATIONS.....71
3.6 ANALYSIS OF OVERLAPPING POLYMERASE REGION.....78
3.7 BASIC CORE PROMOTER (BCP) AND PRECORE REGION ANALYSIS.....79
3.7.1 BCP/PREC REGION AMPLIFICATION AND SEQUENCE ANALYSIS.....79
3.7.2 BCP/PRECORE REGION MUTATION ANALYSIS.....80
3.8 COMPLETE HBV GENOME ANALYSIS.....90
3.8.1 OPTIMIZATION OF NEW PCR FOR COMPLETE GENOME ANALSYSIS.....91
3.8.2 PHYLOGENETIC ANALYSIS OF THE COMPLETE HBV GENOME.....95

3.8.3 COMPARATIVE BIOINFORMATIC ANALYSIS OF SUBGENOTYPE A1 ISOLATES FROM KERALA, INDIA.....97
**CHAPTER 4: DISCUSSION.....100**
4.1 DESIGN AND OPTIMIZATION OF EXPERIMENTS.....100
4.2 GENOTYPING OF HBV ISOLATES FROM KERALA.....103
4.3 GENOTYPE DISTRIBUTION AMONG DISEASE GROUPS AND INFLUENCE ON DISEASE PROGRESSION.....107
4.4 ANALYSIS OF COMPLETE S REGION AND THE OVERLAPPING POLYMERASE REGION.....108
4.5 BCP/PRE-C REGION ANALYSIS.....111
**CHAPTER 5: CONCLUSION.....117**
**APPENDIXES.....119**
APPENDIX A: SOLUTIONS AND REAGENTS.....119
**REFERENCES.....121**

**LIST OF FIGURES**

**Figure 1.1** Geographical distribution of chronic hepatitis B virus infection .....3

**Figure 1.2** Dane particle: The hepatitis B virion structure.....7

**Figure 1.3** Complete HBV virion and surrounding subviral particles.....8

**Figure 1.4** Genomic organisation of the HBV genome.....9

**Figure 1.5** The three domains of the complete surface gene.....17

**Figure 1.6** Topological models of the three HBV surface proteins.....20

**Figure 1.7** Lifecycle of the HBV.....24

**Figure 1.8** Clinical phases and disease progression associated with chronic HBV infection.....26

**Figure 1.9** Geographical distribution of the eight HBV genotypes.....32

**Figure 1.10** HBV genotype distribution on the Indian subcontinent.....42

**Figure 2.1** Overview of molecular characterisation of HBV isolates from Kerala, India.....46

**Figure 3.1** Overview of the study layout and representation of the results.....61

**Figure 3.2** Two of the three primers, 3188F and 2497F, used to sequence the complete S region.....64

**Figure 3.3** Phylogenetic tree comparing the genotype A HBV isolates from India with genotype A sequences obtained from Genbank (designated by accession numbers).....67

**Figure 3.4** Phylogenetic tree comparing the genotype D HBV isolates with genotype D sequences obtained from Genbank (designated by accession numbers).....68

**Figure 3.5** Phylogenetic tree comparing the genotype C HBV isolate with genotype C sequences obtained from Genbank (designated by accession numbers).....69

**Figure 3.6** Mapping of the pre-S2 deletions of the HBV isolates from Kerala, India.....73

**Figure 3.7** Detection of the amplified BCP/PreC region amplicon, nt 1653 to 1959 (numbering from the *EcoR*I site, Genbank # AY233289), visualized on an ethidium bromide stained 2% agarose gel.....79

**Figure 3.8** Chromatogram sequence showing subgenotype A1 Kozak sequence and the PreC start codon.....80

**Figure 3.9** Mutation distribution graphs comparing HBeAg-positive versus HBeAg-negative HBV Genotype A isolates.....82

**Figure 3.10** Mutation distribution graphs comparing HBeAg-positive versus HBeAg-negative HBV Genotype D isolates.....83

**Figure 3.11** Overlapping fragment technique used for complete HBV genome analysis.....91

**Figure 3.12** Optimization of the 1.6 kb PCR reaction using different annealing temperatures.....93

**Figure 3.13** Optimization of DNA polymerase enzymes: HotStarTaq vs BioTaqDNA polymerase.....94

**Figure 3.14** Phylogenetic analysis of the nine complete HBV sequences from Kerala, India.....96

**Figure 3.15** Phylogenetic relationship of the complete genomes of subgenotype A1.....98

**Figure 3.16** Comparison of the complete HBV genome subgenotype A1 phylogenetic tree with the phylogenetic trees of the polymerase and X ORF's.....99

**LIST OF TABLES**

**Table 2.1** Primers and PCR conditions used for amplification of 3.2 kb HBV genome.....54

**Table 2.2** Sequencing primers for compete 3.2 kb analysis.....56

**Table 3.1** Demographic and clinical characteristics of HBsAg-positive patients with different disease profiles.....62

**Table 3.2** Genotype and Subgenotype distribution among the disease groups...70

**Table 3.3** Analysis of the nine pre-S deletion mutants: Clinical and Virological characteristics.....72

**Table 3.4** Amino acid changes in the complete S region of genotype A HBV isolates and the single genotype C isolate.....75

**Table 3.5** Amino acid changes in the complete S region of genotype D HBV isolates.....77

**Table 3.6** BCP/PreC nucleotide changes in genotype A HBV isolates: HCC vs Non-HCC patients.....85

**Table 3.7** BCP/PreC nucleotide changes in genotype D HBV isolates: HCC vs Non-HCC patients.....87

**Table 3.8** Multiple logistic regression analysis for mutations independently associated with HCC development among chronic HBV infection.....88

## LIST OF ABBREVIATIONS

| | |
|---|---|
| aa | amino acid |
| ALT | Alanine aminotransferase |
| ASC | Asymptomatic chronic carrier |
| ASHV | Arctic Squirrel Hepatitis Virus |
| bp | base pair |
| BCP | Basic core promoter |
| BQW | Best quality water |
| cccDNA | Covalently closed circular DNA |
| CHB | Chronic Hepatitis B virus infection |
| DNA | Deoxyribonucleic acid |
| DNAML | DNA maximum likelihood |
| dNTP | Deoxyribonucleotide triphosphate |
| DR1 | Direct repeat one |
| DR2 | Direct repeat two |
| EDTA | Ethylenediamine tetra-acetic acid |
| ENHI | Enhancer one |
| ENHII | Enhancer two |
| ER | Endoplasmic reticulum |
| EtBr | Ethidium bromide |
| EtOH | Ethanol |
| GGH | Ground glass hepatocytes |
| GSHV | Ground Squirrel Hepatitis Virus |
| HAI | Histological Activity index |
| HBcAg | Hepatitis B core antigen |

| | |
|---|---|
| HBeAg | Hepatitis B e antigen |
| HBsAg | Hepatitis B surface antigen |
| HBV | Hepatitis B virus |
| HCC | Hepatocellular carcinoma |
| HIV | Human immunodeficiency virus |
| IFN | Interferon |
| kb | Kilobase |
| LHBs | Large surface protein |
| MHBs | Medium surface protein |
| mRNA | Messenger ribonucleic acid |
| NEIGHBOR | Neighbor-joining |
| nt | Nucleotide |
| NRE | Negative regulatory element |
| ORF | Open reading frame |
| PC | Precore gene |
| PCR | Polymerase chain reaction |
| pgRNA | Pregenomic RNA |
| PHYLIP | Phylogeny inference package |
| rcDNA | Relaxed circular DNA |
| RFLP | Restriction fragment length polymorphism |
| RNA | Ribonucleic acid |
| RT | Reverse transcriptase |
| S gene | Surface gene |
| SHBs | Small surface protein |
| TBE | Tris Boric acid EDTA |

| | |
|---|---|
| TBP | Terminal binding protein |
| TP | Terminal protein |
| URR | Upstream regulatory element |
| UV | Ultraviolet |
| WHO | World Health Organisation |
| WHV | Woodchuck Hepatitis Virus |
| WMHBV | Woolly Monkey Hepatitis B virus |
| YMDD | Tyrosine, Methionine, Aspartic acid, Aspartic acid |

# CHAPTER 1: INTRODUCTION

The hepatitis B virus (HBV) belongs to the family *Hepadnaviridae.* The 3.2 kb, partially double stranded DNA virus, is an extremely important preventable human pathogen, causing major health problems worldwide. An estimated 240 million people have chronic hepatitis B virus infection (CHB), which causes various long term sequelae such as cirrhosis, hepatocellular carcinoma and liver failure (WHO 2012). Annually, HBV infections account for 1 million deaths worldwide from cirrhosis, liver failure and HCC (Lok and MacMahon 2007).

HBV prevalence varies worldwide. Based on the prevalence of HBV surface antigen (HBsAg) carrier rates in the general population, areas are classified as high (HBsAg > 8%), intermediate (HBsAg 2-7%) and low (HBsAg < 2%) HBV endemic zones **(Figure 1.1)**. The high HBV carrier rates in endemic regions are due to the high rate of infection in infancy and early childhood, which occurs in most of these areas. Modes of HBV transmission vary between geographic regions, and hence also between certain genotypes. In Africa, horizontal transmission during early childhood predominates (Kramvis and Kew, 2007), whereas a perinatal transmission occurs predominantly in Asia. The clinical expression of HBV is somewhat dependent on the time of acquisition during life, for example in Asia perinatal transmission results in CHB infection in more than 90% of patients. This differs from infection during adulthood, which can present with a clinically apparent acute hepatitis, but progresses to chronic infection in only 10% of the cases.

Phylogenetic analysis of HBV full-length genomes has led to the classification of HBV into nine genotypes (A-I) (Norder *et al.*, 1994; Kramvis *et al.*, 2005; Stuyver *et*

*al.,* 2000; Hannon *et al*., 2000;), defined by an intergroup divergence of the complete HBV genome sequence of greater than 7.5 – 8% (Kramvis *et al.,* 2005). Based on a single isolate, a tenth genotype, genotype J, has been proposed (Tatematsu *et al.,* 2009). HBV has a well defined geographic distribution. In addition to their distinct geographical distribution, the HBV genotypes have been implicated in affecting disease progression (Kay and Zoulim, 2007).

The Indian subcontinent and southern Africa both have a high incidence of chronic HBV infection, and both have two predominant circulating genotypes, namely genotypes A and D (Kramvis and Kew, 2007; Datta, 2008). Despite the similarity in the circulation of HBV genotypes, the incidence of HCC is low in India (average annual incidence of ≤ 3.3 HCC cases/100 000 individuals), compared to sub-Saharan Africa (average annual incidence of ≤ 98.9 cases/100 000 individuals) (Hussain *et al.,* 2007). This is further supported by the fact that there is a high relative risk of a black African chronic HBV carrier developing HCC, with a relative risk range between 9 and 23 (Kew, 1992). The development of HCC is multifactorial, with both host and environmental factors contributing to its pathogenesis. An important factor involved in hepatocarcinogenesis is chronic HBV infection. However, the role of factors such as HBV genotypes, presence of BCP/PreC mutations, pre-S deletions, and other nucleotide and deletion mutants in the natural history of chronic HBV infection and liver disease progression, remain largely unknown.

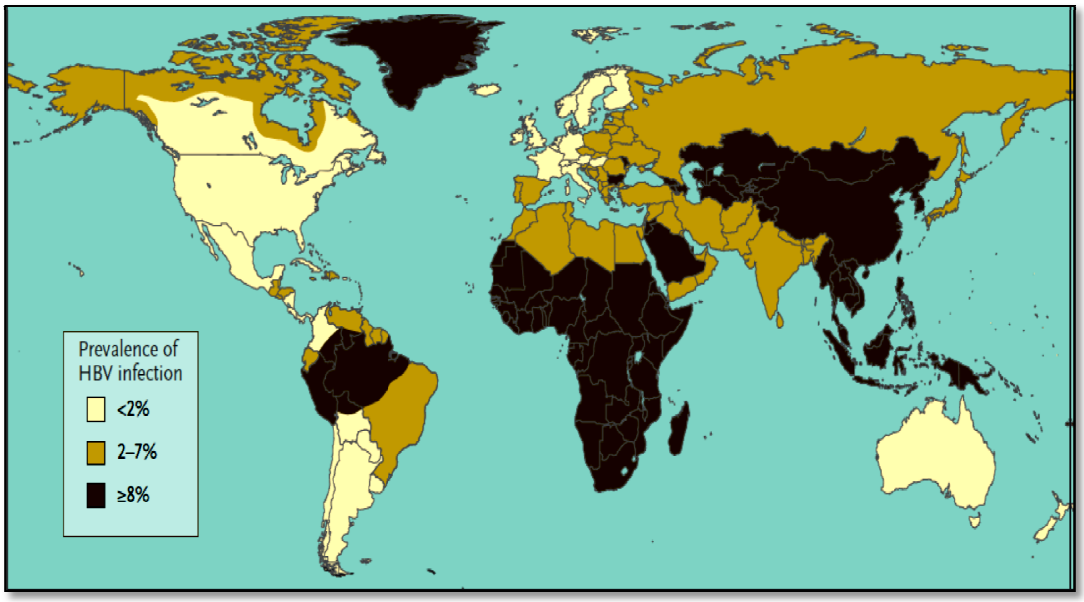



**Figure 1.1 Geographical distribution of chronic hepatitis B virus infection.** (Reproduced with permission from (Dienstag, 2008), Copyright Massachusetts Medical Society).

## 1.1 THE HEPATITIS B VIRUS

### 1.1.1 HISTORY OF THE HBV

Although hepatitis had been known for centuries, it was not until the decades following World War II, when doctors slowly began to differentiate the various types of acute hepatitis. FO MacCallum, a British doctor working on the yellow fever virus, was first to hypothesize that a virus carried in human blood was the causative agent for hepatitis. During the next decade and a half, little progress was made in isolating the possible causative agents thought to cause the two types of hepatitis.

In 1963, Baruch Blumberg, studying serum protein polymorphisms, detected an antigen rarely found in healthy patients' blood, but frequently detected in haemophiliacs and leukemia patients (Blumberg *et al.,* 1965). This protein was termed Australia antigen (Aa) because it was isolated from an Australian aborigine (Blumberg *et al.,* 1965). Although the Aa antigen had been identified, its relationship with hepatitis was unclear. Another group, led by Alfred Prince, would eventually confirm the link between Aa and hepatitis. Prince and co-workers, studying patients with transfusion-associated hepatitis, identified an antigen that appeared in the blood during the incubation period of the disease (Prince *et al*., 1964). This antigen was identical to Aa, and was later shown to represent the hepatitis B surface antigen (HBsAg) (Prince *et al.,* 1970). These seminal studies identified the causative agent for hepatitis B and opened up the field to investigations into HBV and its association with liver disease.

### 1.1.2 CLASSIFICATION: FAMILY HEPADNAVIRIDIAE

The family *Hepadnaviridae* is divided into two genera, the *Orthohepadnaviridae* (hepadnaviruses infecting mammals) and the *Avihepadnaviridae* (hepadnaviruses infecting birds). The prototype member of the orthohepadnaviruses is HBV. Other orthohepadnaviruses include the Woolly Monkey Hepatitis B virus (WMHBV) infecting the Woolley Monkey (*Lagothrix lagotricha*) (Lanford *et al*., 1998), Woodchuck Hepatitis Virus (WHV) infecting the Woodchuck (*Marmota monax*) (Summers *et al.,* 1978), Ground Squirrel Hepatitis Virus (GSHV) infecting the Ground Squirrel (*Spermophilus beecheyi*) (Marion *et al.,* 1980) an the Arctic Squirrel Hepatitis Virus (ASHV) infecting the Arctic Squirrel (*Spermophylus parryi kennicotti*) (Testut *et al.,* 1996). Non-human primates, including chimpanzees (*Pan troglodytes*), White handed gibbons (*Hylobates lar*), Orangutans (*Pongo Pygmaeus pygmaeus*) and Gorillas (*Gorilla gorilla*), are also infected with orthohepadnaviruses (Vaudin *et al.,* 1988; Norder *et al*., 1996; Warren *et al*., 1999; Grethe *et al*., 2000). Certain orthohepadnaviruses, namely the WHV and the GSHV, are closely related to HBV differing by only 17% (Schaefer 2007).

Avihepadnaviruses have been documented in ducks, herons, geese, storks and cranes. Although avihepadnaviruses show little nucleic acid sequence homology with HBV, the DHBV has been used as a model system to study hepadnavirus replication, and WHV infections have been used to study chronic liver disease (Seeger and Mason, 2000).

### 1.1.3 STRUCTURE OF THE HBV

Persistent infection with HBV is characterized by the presence of three distinct forms of the virus in the blood. The structure of a mature HBV virion (Dane particle) consists of an outer viral lipoprotein envelope embedded with three related envelope glycoprotein's (the hepatitis B surface antigens, denoted as small (S), middle (M) and large (L)) **(Figure 1.2)** (Heerman *et al*., 1984). Within the viral envelope is the electron dense inner core, or nucleocapsid. The viral core protein, also known as the core antigen (HBcAg), is a 183-amino-acid (183-aa) polypeptide composed of two moieties, the 140-aa N-terminal domain needed for core protein dimer formation and assembly, and the arginine rich C-terminal region needed for nucleic acid binding (Nassal, 1992). The nucleocapsid encloses the relaxed circular-partially double strand DNA (RC-dsDNA) of 3.2kb (Summers *et al.,* 1975). Covalently attached to the 5'-end of the (-) strand, is the viral polymerase (Summers *et al*., 1975). The size of a mature HBV virion is between 40 to 42 nm in diameter (Dane *et al*., 1970).

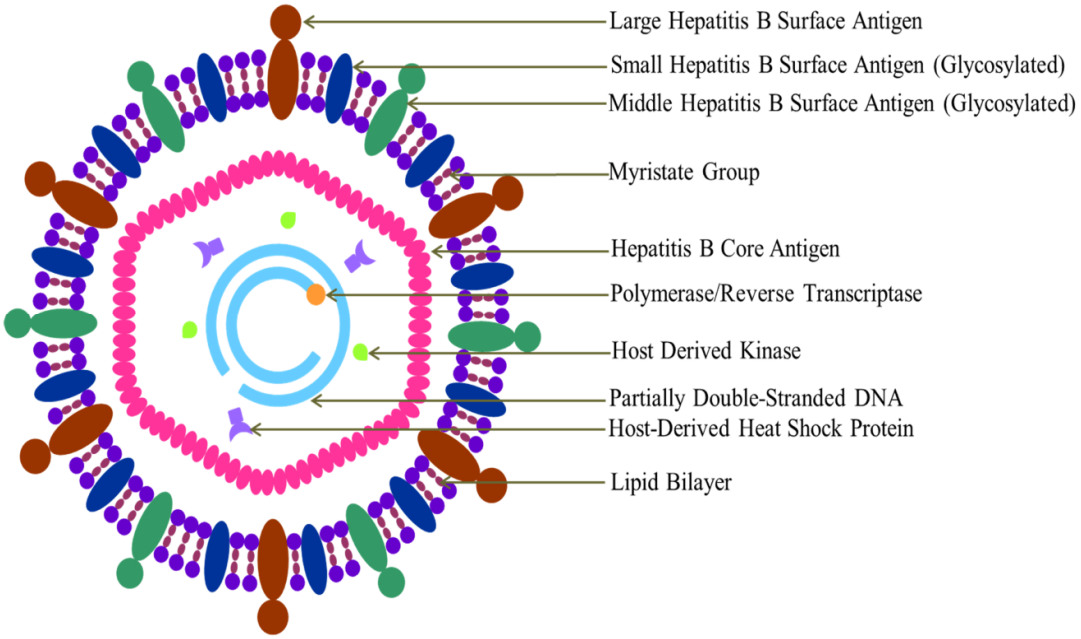



**Figure 1.2 Dane particle: The hepatitis B virion structure.** The complete HBV virion structure with a lipid bilayer embedded with small, middle and large HBV surface proteins. Enclosed in the lipid bilayer is the nucleocapsid of core proteins, and the partially double stranded HBV and the 5' attached polymerase protein. (Figure reproduced with permission from Nimisha Bhoola, ©).

HBV infected cells also produce two distinct subviral lipoprotein particles: the spherical form, about 20nm in diameter, and the filamentous form of similar diameter, but with varying length **(Figure 1.3)** (Ganem and Prince, 2004). These subviral HBsAg particles contain only the envelope glycoproteins and the host derived lipids. Subviral particles typically outnumber HBV virions by 1000:1 to 10 000:1 (Ganem and Prince, 2004).

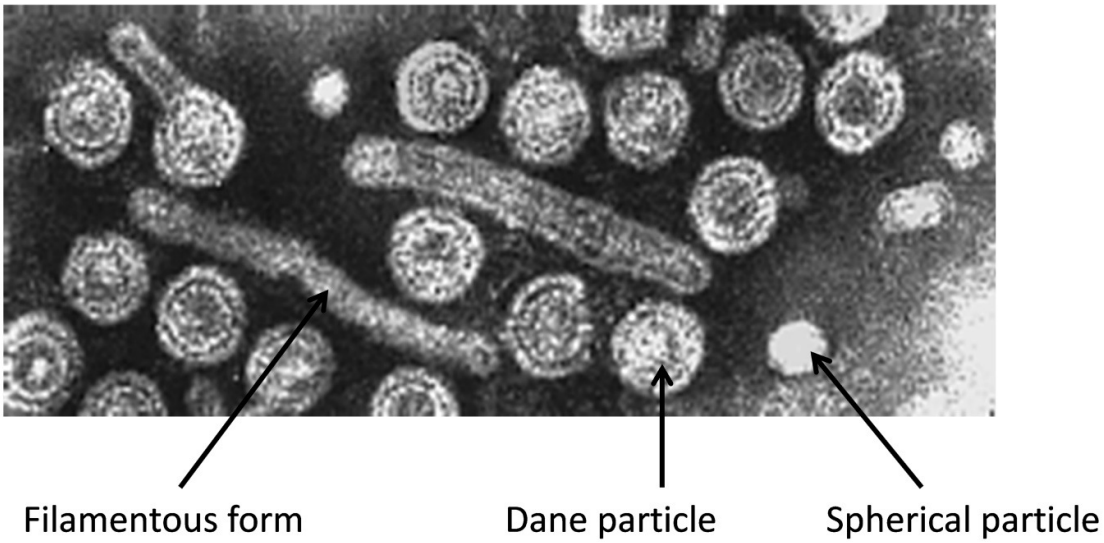



**Figure 1.3 Complete HBV virion and surrounding subviral particles**

### 1.1.4 HBV GENOME ORGANISATION

The 3.2 kb relaxed circular-doubled stranded DNA (RC-dsDNA) genome has a complete (-) strand with the viral polymerase attached to the 5' end and an incomplete (+) strand, about two thirds complete, with a short oligoribonucleotide at its 5' end (Nassal and Schaller, 1993). The (-) strand encodes for the viral proteins and is the template for the pgRNA, whereas the (+) strand is non-coding (Galibert *et al*., 1979). The relaxed circular nature of the HBV genome is maintained by via base pairing at the 5' ends of positive and negative stands (Toillais *et al.,* 1985). The HBV has four partially overlapping open reading frames (ORFs) (Toillais *et al*., 1985). The four open reading frames have been identified as: *S*, for the surface or envelope gene; *C*, for the core gene; *X*, for the *X* gene; and *P*, for the polymerase gene **(Figure 1.4)**.

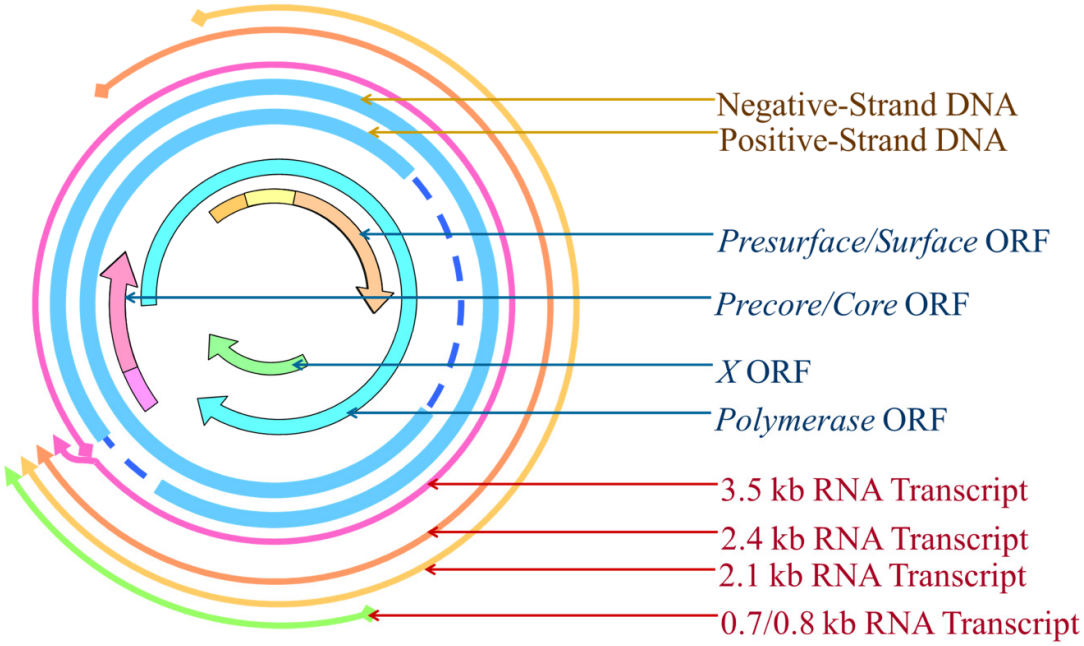



**Figure 1.4 Genomic organisation of the HBV genome.** The four viral transcripts (3.5-, 2.4-, 2.1-, and 0.7/0.8-kb) are shown as the four outermost circles. The four inner arrows represent the four ORF's; S ORF, PreC/C ORF, X ORF, Pol ORF. The relaxed circular dsDNA genome, shown between the viral transcripts and the ORF's consists of the complete (-) strand and the incomplete (+) strand. (Figure reproduced with permission from Nimisha Bhoola, ©).

### 1.1.5 THE HBV TRANSCRIPTION PRODUCTS

Upon infection of hepatocytes by HBV, the 3.2 kb RC-dsDNA genome is transferred to the cell nucleus, where it is converted into a covalently closed circular DNA (cccDNA) molecule (Tuttleman *et al.,* 1986). This process is thought to involve both host and viral factors (Levrero *et al.,* 2009), however the mechanisms and steps required for this conversion are as yet to be fully elucidated. Nonetheless, this process is vital to the HBV lifecycle and the formation of new mature HBV as the cccDNA molecule serves as the template for the transcription of all viral mRNAs.

Several genomic and subgenomic RNAs are transcribed by the host cellular RNA polymerase II, using the cccDNA as the template (Beck and Nassal, 2007).

Four major RNA transcripts, 3.5, 2.4, 2.1, and 0.7 kb in length, are transcribed from cccDNA. All viral mRNAs contain 5' cap structures and are 3' terminally poly-adenylated at a common site (Beck and Nassal, 2007).

There are two distinct 3.5 kb transcripts the precore (pre-C) mRNA and the pregenomic RNA (pgRNA) (Yaginuma *et al*., 1989). These are not only the longest transcripts, but also the most abundant and the most important. The functions of these two transcripts are distinct, in that the pgRNA serves as the template for genome replication by reverse transcription and encodes for the core and polymerase, and the pre-C mRNA encodes the precore-core protein, the precursor of HBeAg. The 5' ends of the pre-C mRNA are heterogenous and have been mapped to nucleotides (nts) 1785-1786 and 1791-1797 (Kramvis and Kew, 1999). The 5' end of the pgRNA has been mapped to a single nt region, 1815 ± 5 (Kramvis and Kew, 1999). The reason for the heterogenous nature of the 5' ends of the pre-C mRNA is the lack of a distinct TATA box in the BCP. Within the BCP, there are a total of four TATA-like boxes, three of which control the transcription of the pre-C mRNA, and the fourth controlling that of the pgRNA (Hiraga *et al.,* 1994).

Transcription of the 2.4 kb mRNA and the group of 2.1 kb mRNAs, with heterogenous 5' ends, is controlled by the S1 and S2 promoters, respectively (Moolla *et al.,* 2002). The LHBs protein is translated from the 2.4 kb mRNA (referred to as the pre-S1 transcript), and the MHBs and HBsAg are translated from the 2.1 kb

mRNAs (referred to as the pre-S2 transcripts). The 0.7 kb transcript encodes for the transactivating HBV X protein.

The presence of four major mRNA transcripts, transcribed from cccDNA, which are translated into 7 HBV proteins, makes the 3.2 kb HBV genome an extremely compact and efficient entity. Control of gene expression and the subsequent translation of HBV proteins must therefore be an organized and regulated process. Transcription, and the regulation thereof, has been extensively studied for HBV, with various elements in the HBV genome displaying promoter and/or enhancer activity (Seeger and Mason, 2000).

### 1.1.6. HBV VIRAL *cis*-ELEMENTS

#### 1.1.6.1 HBV PROMOTERS

Transcription of the four mRNAs is initiated and controlled by four promoters: core promoter (CP), large surface (S1 promoter), middle surface (S2 promoter) and X gene promoter. The core promoter controls the transcription of the pre-C mRNA and the pgRNA. The CP encompasses a region of at least 232 nt that includes two functional units, the basic core promoter (BCP) and the upstream regulatory region (URR). The BCP has been mapped to an essential sequence of 108bp (nt 1742-1849) (Yuh *et al*., 1992). The URR encompasses nucleotides 1613-1742. The BCP contains cis-elements that initiate transcription of the pre-C and pg RNAs (Yu and Mertz, 1996). The BCP lacks a distinctive TATA box, but contains four distinct TATA-like boxes, three of which initiate transcription of the pre-C mRNA, and the fourth controls transcription of the pgRNA (Kramvis and Kew, 1999). The URR can be further divided into two functional entities, the core upstream regulatory sequence

(CURS) and the negative regulatory element (NRE). The CURS, located immediately 5' to the BCP has been shown to activate the BCP 200-2000 fold, in an orientation and position dependent manner (Moolla *et al*., 2002). The NRE, located immediately upstream of the CURS, can suppress core promoter activity in an orientation independent manner (Moolla *et al.,* 2002).

The S1 and S2 promoters initiate the transcription of the 2.4 kb and the 5' heterogenous 2.1 kb transcripts, respectively (Moolla *et al.,* 2002). The S1 promoter, situated 5'-to the preS1 region, is the only HBV promoter with a classical TATA-box sequence (Siddiqui *et al.,* 1986). The S1 promoter contains sequence elements that positively regulate transcription of the 2.4 kb transcript. The S1 promoter is negatively regulated by the sequences in the S2 promoter (Bulla and Siddiqui, 1989). The S2 promoter, a distinctive TATA-less promoter, controls transcription of the 2.1 kb mRNA. The fourth HBV promoter, the X promoter, regulates transcription of the smallest 0.7 kb mRNA.

**1.1.6.2 HBV ENHANCERS**

Together with transcription control by promoter elements, two enhancer elements, I and II, have been identified that affect HBV gene expression. Enhancer I, spanning approximately 120 bp between the S and X ORFs, upregulates transcription of the pgRNA, pre-C mRNA and HBx mRNA in an orientation independent manner (Moolla *et al*., 2002). The enhancer II element, located upstream of the BCP, consists of two interacting sequences, a 23 bp box α and a 12 bp box β (Yuh *et al*., 1991). Enhancer II stimulates the two surface promoters (S1 and S2) and the X promoter from a

downstream position (Yuh *et al*., 1992), as well as activating the BCP in a position and orientation dependent manner (Yuh *et al*., 1991).

### 1.1.6.3 THE HBV ENCAPSIDATION SIGNAL: EPSILON

A vital step in the replication of HBV is the reverse transcription of pgRNA by the HBV polymerase. However, before the initiation of reverse transcription, the pgRNA together with polymerase must be encapsidated by the core protein to form subviral particles. This encapsidation of pgRNA is brought about by means of a stem-loop structure at the 5' end of the pgRNA. This stem-loop structure is known as the encapsidation signal or epsilon (ε). Epsilon (ε) consists of 70 nt (mapped to nt position 1846-1916) (Kawamoto *et al*., 1994), and is located near the 5' end of both the pre-C and pg RNAs, however only the slightly shorted pgRNA is encapsidated and reverse transcribed (Kramvis and Kew, 1998). Nuclear magnetic resonance (NMR) techniques have confirmed the presence of a 6 nt bulge and 3 nt apical loop structure, within the stem loop structure of ε (Flodell *et al*., 2002). This structure is as a result of inverted repeat sequences present in HBV ε (Kramvis and Kew, 1998). The sequence encoding ε is conserved among mammalian and avian hepadnaviruses, and between different HBV genotypes. Functions of ε include, activation of viral polymerase, encapsidation and initiation of (-) strand DNA synthesis (Kramvis and Kew, 1998).

### 1.1.6.4 THE POLYADENYLATION SIGNAL

A polyadenylation signal is shared by all HBV transcripts, and is located approximately 20 bp from the 3' end of the negative strand (Schaller and Fisher, 1991). A unique feature of the polyadenylation signal is that it is only recognized

after its second encounter with the transcription complex (Russnak, 1991). This phenomenon is as a result of the presence of two regulators. The positive and negative regulators of transcription (PET and NET, respectively), which ensure recognition of the polyadenylation signal at the second encounter (Beckel-Mitchener and Summers, 1997; Huang and Summers, 1994).

### 1.1.7 THE HBV GENE PRODUCTS

The 3.2 kb HBV genome is very compact and efficient, encoding for 7 proteins from four ORFs: (1) the precore/core gene encoding for the HBcAg and HBeAg, (2) the polymerase gene encoding for the Pol protein, (3) the L-, M-, and S-gene coding for the three envelope proteins, and (4), the X gene encoding for the HBx protein.

#### 1.1.7.1 PRECORE AND CORE GENE

The core (HBcAg) and precore (HBeAg) proteins are translated from two viral transcripts, namely the pregenomic and precore mRNA transcripts, respectively (Tiollais *et al*, 1985). The HBcAg and the HBeAg are derived from two alternative in frame AUG initiation codons, the internal AUG encodes the HBcAg and the upstream AUG the HBeAg (Ganem 1991). The HBcAg is a 183 aa polypeptide composed of two domains, the 140 aa N-terminal domain and the 43 aa C-terminal region (Nassal 1992). The HBcAg, which forms the viral nucleocapsid, self-assembles into a capsid-like structure and has a highly basic motif at its C terminus with RNA-binding activity (Hutton *et al*., 1992).

Translation from the upstream AUG results in the formation of a 210 amino acid (aa), HBeAg precursor protein, p25. The N terminal region of the precursor protein

encodes a hydrophobic signal sequence, which directs the p25 to the endoplasmic reticulum (ER). The precursor protein then undergoes N terminal cleavage in the ER, followed by C terminal cleavage in the Golgi apparatus to yield a 18-kDa mature HBeAg protein (McLachlan *et al*., 1987; Roossinck *et al*., 1986). The HBeAg is subsequently secreted into the bloodstream or expressed on the surface of hepatocytes. Although HBeAg is not required for assembly, infection or replication, various studies have shown a possible role as an immunoregulatory protein that down regulates the immune response to HBcAg (Milich *et al*., 1998; Chen *et al*., 2004) and is required for the establishment of a chronic infection. Its crystal structure has recently been determined (Dimattia *et al.,* 2013).

#### 1.1.7.2 THE POLYMERASE PROTEIN

The polymerase polypeptide is encoded by the polymerase ORF (P ORF) from an internal initiator codon on the pgRNA (Seeger and Mason, 2000). This internal start codon is located 600 nt down-stream of the 5' end of the pgRNA (Kann, 2002). The polymerase polypeptide of 95 kDa, consists of four domains, the terminal protein (TP), spacer domain, reverse transcriptase (RT) domain and the RNase H domain (Radziwill *et al*., 1990). The numbering of the polymerase protein is genotype specific for the RT and spacer domain, and genotype independent for the RT and RNase H domains (Stuyver *et al.,* 2000). The TP varies in length from 177-179 aa depending on the genotype (Stuyver *et al*., 2000). Functions of the TP include covalent attachment to the negative-strand DNA during the initiation of reverse transcription, priming and synthesis of the minus strand and packaging of pgRNA (Weber *et al*., 1994; Seeger and Mason 2000). The spacer domain varies in length from 158-169 aa depending on the genotype, and is also the domain that is very

flexible, tolerating mutations and deletions (Radziwill *et al*., 1990; Kim *et al*., 1999). The RT domain of 344 aa is a well conserved region of the P protein and functions in reverse transcription and priming. The C-terminal RNase H (RH) domain of 153 aa in length is the fourth and final domain of the P protein. The function of the RH domain is degradation the RNA strand in the RNA-DNA during the reverse transcription process (Beck and Nassal, 2007).

#### 1.1.7.3 THE HBX PROTEIN

The HBx (X protein) is a 17 kDa protein translated from the X ORF (Toillais *et al*., 1985). The exact function of the X protein is not known, however it has been shown to be essential for *in vivo* replication and spread of the virus (Zoulim et *al*., 1994). The X protein has a transactivating function, as well as stimulation of signal transduction and binding to various protein targets such as p53 and UV-damaged DNA binding protein (Benn *et al*., 1994; Feitelson *et al*., 1993; Lee *et al*., 1995).

#### 1.1.7.4 THE HBV SURFACE PROTEINS

Three surface proteins can be found embedded in the envelope of the mature HBV virion. The three surface proteins include LHBs, MHBs, and SHBs/HBsAg. The 2.4 kb pre-S1 mRNA transcript encodes for the LHBs and the shorter 2.1 kb pre-S2 transcript for both the MHBs and HBsAg. The three surface proteins are distinguished from each other by the presence of three domains: pre-S1 (108/119 aa) only in LHBs, pre-S2 (55 aa) in LHBs and MHBs, and S (226 aa) common to all three surface proteins **(Figure 1.5)**. The surface proteins constitute a major part of the envelope of mature virions, as well as forming part of nucleocapsid-free subviral particles (SVP). These SVP include spheres and filaments, of which the HBsAg is

the major component of the spheres and the mature virion. The surface proteins can exist either in a glycosylated or unglycosylated form (Seeger and Mason, 2000).

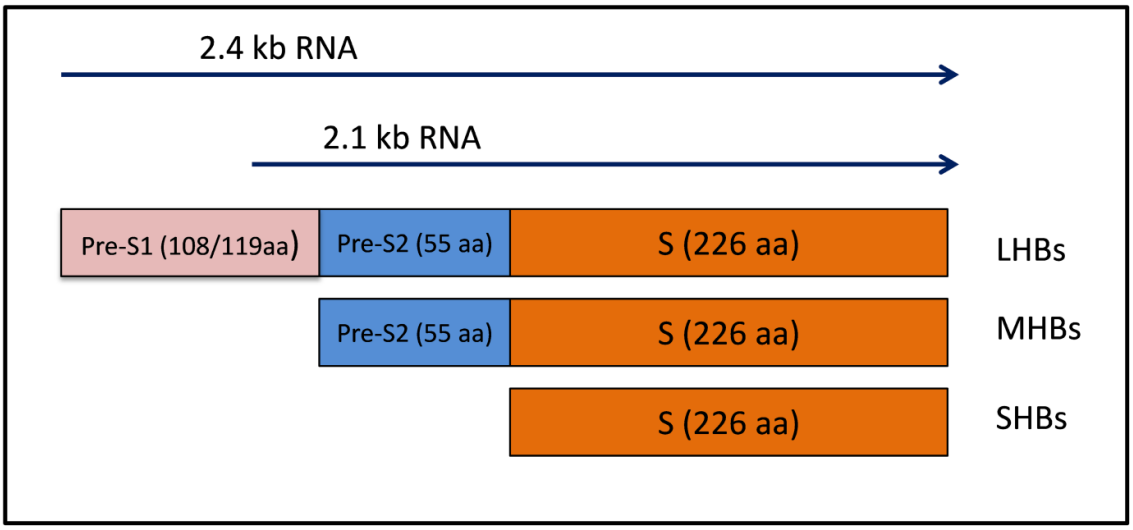



**Figure 1.5 The three domains of the complete surface gene.** The 2.4 kb mRNA translates into the LHBs, while the shorter 2.1 kb mRNA is used for the translation of both the MHBs and SHBs proteins.

#### 1.1.7.4.1 THE LARGE ENVELOPE PROTEIN

The LHBs protein is transcribed from the longer 2.4 kb pre-S1 mRNA **(Figure 1.6)**. The LHBs is the largest surface protein, ranging from 389-400 aa residues, depending on the length of the pre-S1 region, which varies between genotypes.

The pre-S1 domain, found only in the LHBs, is said to be the major role player in the initial entry of the virus into susceptible hepatocytes, via an unidentified receptor (Glebe and Urban, 2007; Xie *et al.*, 2010). Various amino acids of the pre-S1 domain and their post-translation modification can affect the attachment of the virion to the hepatocyte (Glebe and Urban, 2007; Xie *et al.*, 2010). The pre-S1 domain also

functions as a binding site for the core protein during viral particle assembly (Seeger and Masson, 2000). This binding property of the LHBs is dependent on localisation inside the virion (Kann, 2002). The LHBs can also be located in the endoplasmic reticulum and hence exposed to the surface of the HBV virion (Kann, 2002).

#### 1.1.7.4.2. THE MEDIUM ENVELOPE PROTEIN

The MHBs consists of the pre-S2 domain (55 aa) and the S region (226 aa) **(Figure 1.6)**. The pre-S2 region is N-glycosylated with a glycan at amino acid 4 and an additional O-glycan at position 37 (Kann, 2002). Additional acetylation of N-terminal methionines in the pre-S2 region also occurs (Kann, 2002). The MHBs is not required for HBV infectivity and its specific function is not known (Fernholz *et al*., 1993).

#### 1.1.7.4.3 THE SMALL ENVELOPE PROTEIN

The SHBs, also known as HBsAg protein, is composed of 226 aa **(Figure 1.6)**. The HBsAg is the most abundant surface protein in all three HBV particles, and is found either in the glycosylated or unglycosylated form. 50-60 % of the SHBs is folded into α-helices numbered I-V, by 4-5 lengths of hydrophobic amino acids (Kann, 2002).

The first helix (I) is located between aa 11-29, crosses the ER membrane and translocates the upstream sequence into the ER lumen (Kann, 2002). Helix II (aa 80-98), is inserted in the ER membrane to mediate translocation of downstream sequences and directed to the core particle (Kann, 2002). Between these first two α helixes, is the first polar hydrophilic domain, comprising aa residues 30-79.

Downstream of helix II is helix III, which passes through the ER. Helices IV and V span the ER membrane (Kann, 2002).

Located between helix II and III, is the second hydrophilic domain. This domain covers aa residues 99-168. Numerous cysteine residues are present in this domain. These cysteine residues are cross-linked to each other via disulfide bonds, creating a loop configuration, the major antigenic determinant of HBsAg, known as the 'a' determinant (Locarnini *et al*., 2003).

The 'a' determinant is common to all serological subtypes and is defined by a threonine or isoleucine at position 126 of HBsAg (Kramvis *et al*., 2005). Amino acid substitutions at positions 122 and 160 of HBsAg are used to distinguish between serotypes *d/y* and *w/r* (Kramvis *et al.,* 2005). During an infection, antibodies produced against the 'a' determinant, will provide protection against other HBV serological subtypes. This provides the basis for HBsAg as a immunogen used for vaccines against HBV (Torresi, 2002).

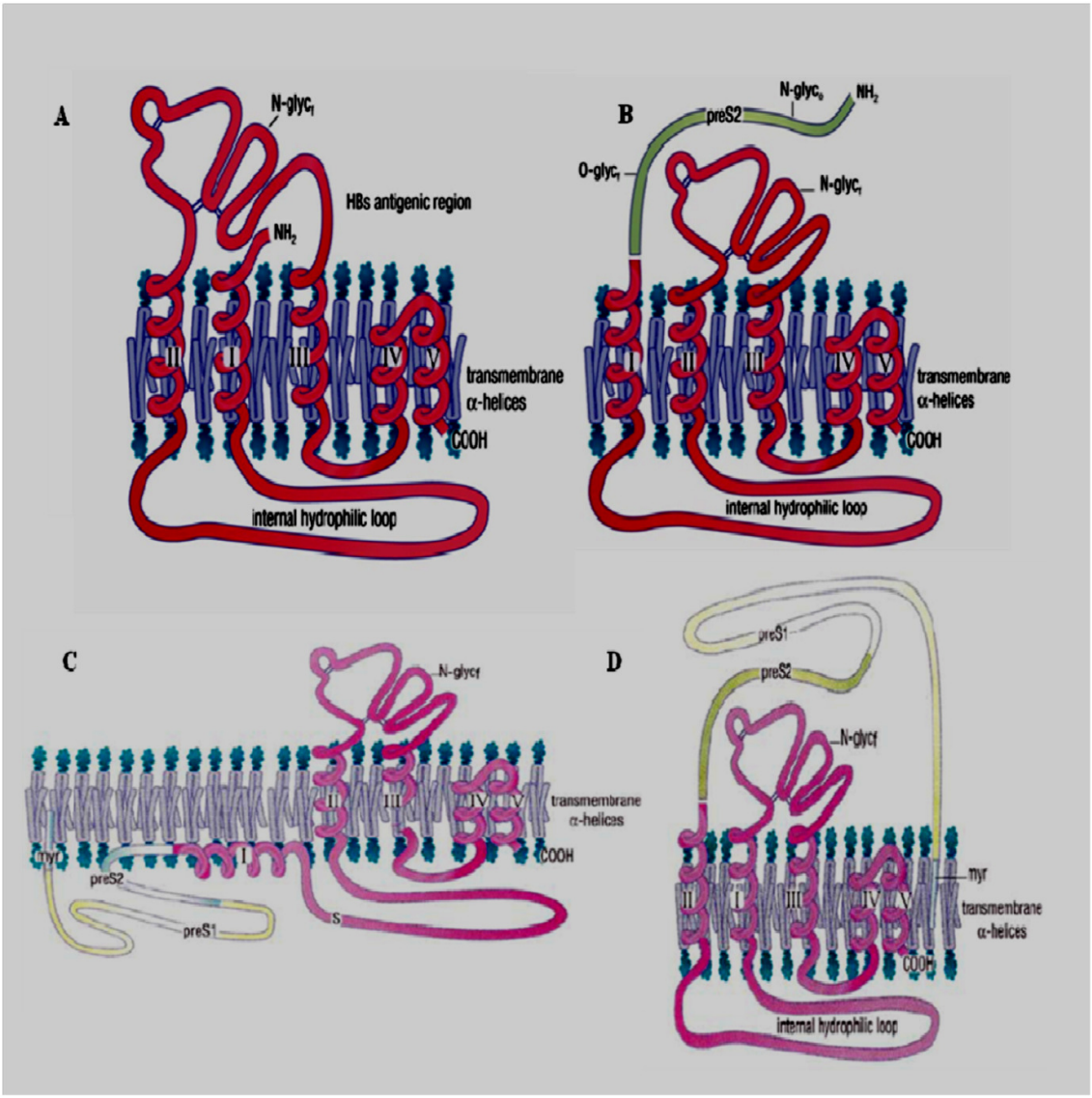



**Figure 1.6 Topological models of the three HBV surface proteins**. A. Small surface protein (SHBs), B. Middle surface protein (MHBs), C. Large surface protein (LHBs) with PreS domain localized in ER, D. LHBs with PreS domain located in cytoplasm. The top of each figure represents the lumen of the ER/pre Golgi compartment, and the bottom of each figure represents the hepatocyte cytosol or interior of the virus (Kann, 2002).

Reproduced with permission of International Medical Press, © 2008.

### 1.1.8 THE HBV LIFE CYCLE

Attachment of the HBV to hepatocytes is the very first step in HBV infection, however, the mechanisms of attachment and entry, are to date still enigmatic. The understanding of this crucial step has been hampered by the lack of feasible HBV *in vitro* infection systems (Glebe and Urban, 2007). Much of what is known about these early events has been elucidated using the duck hepatitis B virus model system. Although HBV surface proteins mediate early steps in the virus life cycle, the precise role for each glycoprotein in the entry of the virus remains controversial (Lu and Block, 2012). The current understanding of HBV attachment has revealed a specific domain within the pre-S1 region of the LHBs protein to be the major cell attachment epitope (Paran *et al.,* 2001). This domain has been functionally narrowed down to the amino acids 21–47 of the pre-S1region (Lu and Block, 2012). Other studies have also proposed a role for SHBs in the initial attachment and entry of the virus (Lu and Block, 2012). Although these proteins and specific regions within them may be involved in hepatocyte attachment, the receptors and/or binding partners to which they attach, still remain unknown. Numerous binding partners have been suggested, however only Carboxypeptidase D has been shown conclusively to play a role in the infection of avihepadnaviruses (Tong *et al,* 1995).

Once attached to the hepatocyte, a series of poorly understood events follow, possibly involving membrane fusion together with the release of the HBV virion into the cytoplasm (Cooper *et al*., 2003). The viral envelope is then thought to fuse with the endosome, resulting in the release of the nucleocapsid into the cytoplasm (Cooper *et al*., 2003). The core protein nucleocapsid disassembles, with the subsequent release of RC-dsDNA (Locarnini, 2004). Upon release from the

nucleocapsid, the RC-dsDNA of the virus is delivered to the host cell nucleus where the RNA primer and the 5' attached viral polymerase are removed, allowing for (+) strand completion and ligation. This (+) strand completion results in the formation of a circular ds-DNA molecule, termed covalently closed circular DNA (cccDNA) (Tuttleman *et al*., 1986). This episomal cccDNA serves as a nonreplicating template for viral transcription via cellular RNA polymerase II. Four viral RNA transcripts are transcribed from cccDNA: 3.5-, 2.4-, 2.1-, and 0.7-kb viral RNA transcripts. These four viral transcripts are transcribed from the enhancer II/basal core, large surface antigen (L), major surface antigen (S) and the enhancer I/X gene promoters, respectively.

The viral RNA transcript most relevant for HBV replication, is the greater than genome length (3.5 kb) pgRNA. The pgRNA functions as the mRNA for the core protein and the polymerase protein (Nassal and Beck, 2007), as well serving as the template for replication by reverse transcription (Summers and Mason, 1982). Viral mRNA transcripts are translated within the host cell cytoplasm, as this occurs; the polymerase protein binds to the pgRNA encapsidation signal, epsilon (ε) (Nassal and Beck, 2007). Binding of the polymerase protein is aided by two host chaperone proteins, p23 and heat shock protein 70 (Hsp70) (Kramvis and Kew, 1998; Seeger and Mason, 2000). This initial binding triggers the recruitment of core protein dimers, causing the formation of pgRNA-P protein containing nucleocapsids (Nassal and Beck, 2007). These immature nucleocapsids then undergo a process of maturation in which pgRNA is reversed transcribed by the polymerase protein to the mature RC-dsDNA.

A unique feature of this reverse transcription reaction is the polymerase protein primed initiation of minus (-) strand DNA synthesis (Gerlich and Robinson, 1980; Wang and Seeger, 1992). The polymerase protein binds to the bulge in ε, utilizing this region as a template for a short DNA oligonucleotide, which, at the 5' end becomes attached to a tyrosine (Y) residue in the terminal protein (TP) domain of the polymerase protein (Weber e*t al*., 1994; Zoulim and Seeger, 1994). The template oligonucleotide, attached to the polymerase protein, is then translocated to the 3' DR1, where it base pairs with the complementary sequence (Wang and Seeger, 1993). This binding acts as a primer for the completion of the (-) strand synthesis. As elongation of the (-) strand proceeds, simultaneous degradation of the pgRNA by the polymerase RNAse H domain occurs, leaving 18 bp including the DR1 motif (Seeger and Mason, 2000). This 18 bp sequence is subsequently translocated to the DR2 motif, to serve as the primer for (+) strand synthesis (Seeger and Mason, 2000). Positive strand elongation continues to the 5' end of the (-) strand. A 9 base terminal redundancy sequence complementary to the 3' end of the (-) strand is formed, resulting in the elongating 5' end of the (+) strand being transferred to the 3' end of the (-) strand, ensuring a circularized genome (Seeger and Mason, 2000).

These mature RC-dsDNA containing nucleocapsids, can either be re-directed into the host cell nucleus, allowing for increased cccDNA synthesis (Tuttleman *et al.,* 1986), or they can interact with envelope proteins in the ER, forming a mature infectious virion, and thus completing the life cycle of the HBV. The various steps involved in the HBV lifecycle are summarized in **Figure 1.7**.

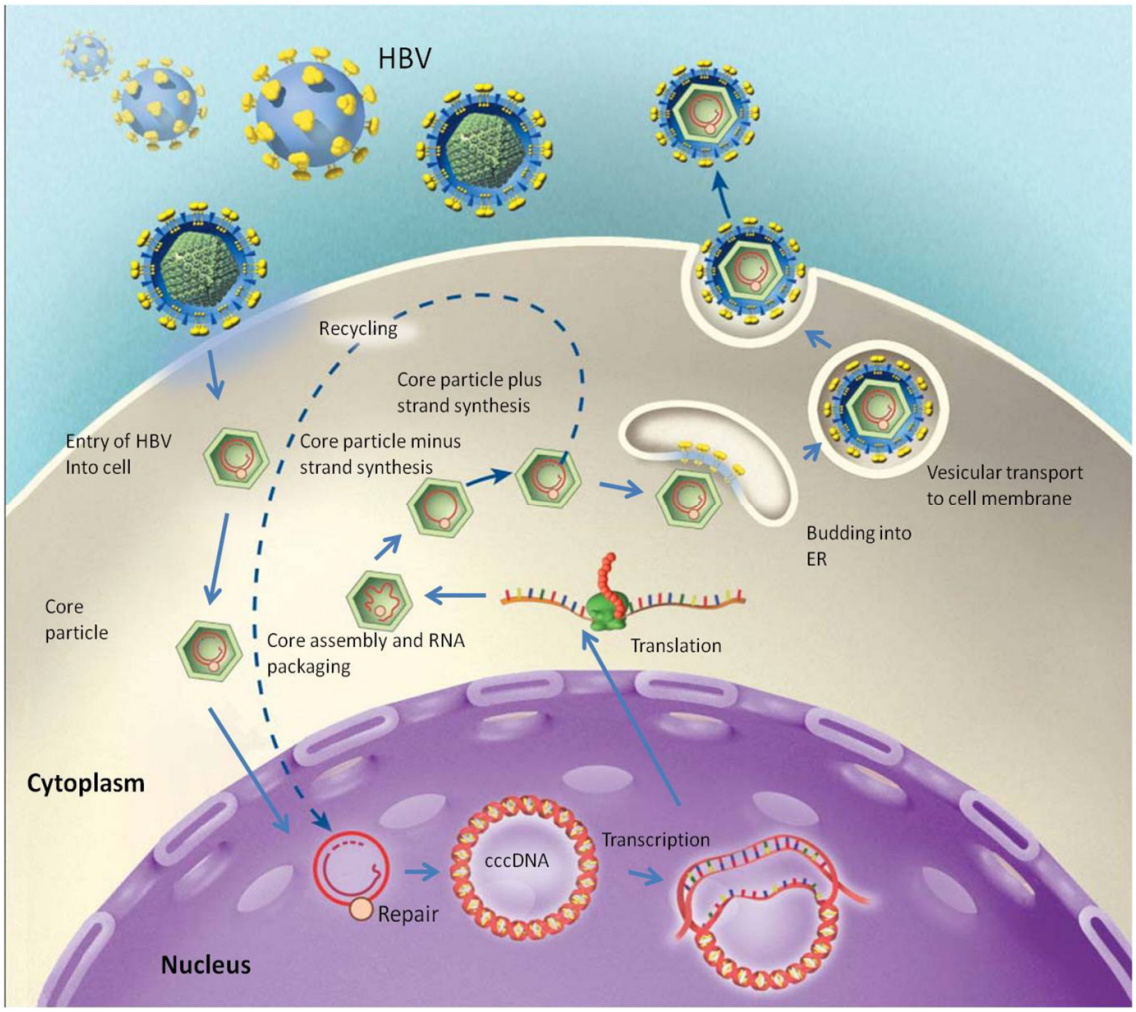



**Figure 1.7 Lifecycle of the HBV**. The lifecycle of the HBV begins with the attachment of the virus to unknown receptors present of the surface of the hepatocyte. Upon entry, the core particle with the RC-dsDNA is delivered to the nucleus, where the RC-dsDNA is converted to cccDNA. The cccDNA serves as the template for the four viral transcripts (3.5 kb, 2.4 kb, 2.1 kb and 0.8 kb). Viral mRNA is exported to the cytoplasm and translated into the viral proteins. The greater than genome length pgRNA also serves as the template for reverse transcription by the polymerase protein. Binding of the polymerase protein to ε initiates encapsidation, with the subsequent reverse transcription of pgRNA. The nucleocapsids are then enveloped via transport through the ER/Golgi complex and secreted via host cell secretory pathways as mature virions. (Reproduced with permission from (Ganem and Prince, 2004), Copyright Massachusetts Medical Society).

### 1.1.9 CLINICAL SPECTRUM OF HBV INFECTION

#### 1.1.9.1 ACUTE HBV INFECTION

Primary infection in susceptible hosts can either be asymptomatic or symptomatic, and in adults, this is usually self-limited with viral clearance and the development of immunity (Wright *et al*., 1993). If symptomatic disease occurs, clinical symptoms usually range from non-specific symptoms such as fatigue and nausea, to more specific symptoms of liver enlargement and right upper quadrant tenderness. The clinical incubation period of acute hepatitis B averages 2-3 months, followed by a prodromal period of symptoms such as fatigue, anorexia and nausea, there is also a rapid rise in serum ALT levels and HBsAg and HBV DNA levels are present at very high numbers (Liang, 2009). This phase is followed by a decrease in viral levels, together with a clearance of HBsAg and undetectable levels of HBV DNA.

In approximately 1 % of patients acute liver failure occurs (Berk and Popper 1978). The presence of fulminant hepatitis is marked by a sudden appearance of fever, abdominal pain, vomiting, jaundice, confusion and coma (Liang 2009).

#### 1.1.9.2 CHRONIC HEPATITIS B VIRUS INFECTION

Chronic HBV infection is defined as the presence of HBsAg in the serum for longer than 6 months. Over 90% of patients, who develop chronic hepatitis B virus infection, do so as a result of early perinatal transmission from infected mothers or horizontal transmission from siblings and playmates (Shi and Shi 2009). Only 5% of adults infected with HBV will develop chronic hepatitis B (CHB). The natural history of CHB is a very dynamic and complex process influenced by both viral and host factors. Four natural stages of CHB have been identified, immune tolerance stage, immune

clearance stage, inactive HBsAg carrier stage and the reactivation stage. The stages and the complexities associated with CHB are depicted in **Figure 1.8**

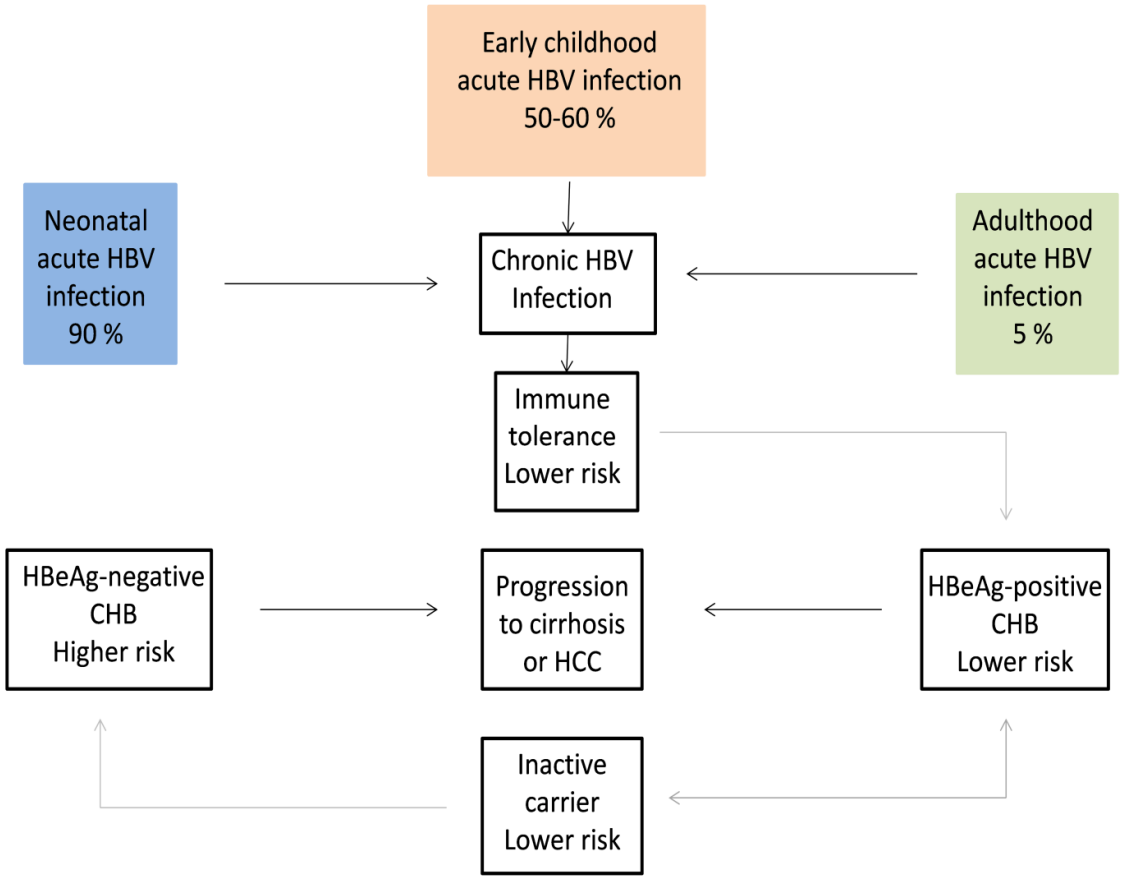



**Figure 1.8 Clinical phases and disease progression associated with chronic HBV infection.** (Figure adapted from Shi and Shi, 2009).

- **<u>Immune tolerance stage</u>**

Perinatal or early childhood infections with HBV have an immune tolerance stage characterized by HBeAg, high serum HBV DNA levels, normal ALT levels and minimal inflammation (Takashima *et al.*, 1992). These features are not present in patients infected in adulthood (Shi and Shi, 2009). Patients in the immune tolerance phase are said to be at a low risk of progression to cirrhosis or hepatocellular carcinoma (HCC) (Shi and Shi 2009).

- **Immune clearance stage**

This phase occurs as a result of maturation of the immune response. Patients with perinatally acquired infection, usually present with this stage of infection in the 2nd to 3rd decade of life (Shi and Shi 2009). This stage, also commonly referred to as the 'chronic hepatitis B phase', is associated with raised ALT levels, low levels of serum DNA and liver necroinflammation (Neuveut *et al*., 2010).

- **Inactive HBsAg carrier stage**

This phase is characterised by loss of HBeAg and anti-HBe seroconversion, a decline in serum HBV DNA levels, and a decrease in inflammation and fibrosis (Hoofnagle *et al*., 2007). The majority of CHB patients are in this inactive carrier stage.

- **Reactivation stage**

The patients who present in the HBeAg negative stage can be further divided into the CHB inactive HBsAg carriers and the CHB patients with biochemical and intermittent virological activity (Shi and Shi 2009). Patients who present in the latter group account for one third of CHB patients. These patients present with periodic stages of fluctuating ALT and HBV DNA levels and active hepatitis and fibrosis (Shi and Shi 2009). This transition to virological activity may be due to the occurrence of nucleotide substitutions in the BCP/precore region that decrease HBeAg expression (Carman *et al*., 1989). These patients are associated with active liver disease and an increased risk of cirrhosis and HCC.

#### 1.1.9.3 THE LONG TERM SEQUELAE OF CHB INFECTION

An estimated one-third of patients with CHB will develop long term complications of chronic infection. These long term consequences include cirrhosis, HCC and end-stage liver disease. Both viral and host factors are said to play a role in the outcome of CHB infection.

##### 1.1.9.3.1 CIRRHOSIS

Cirrhosis is said to have a twofold higher incidence in HBeAg negative CHB patients compared to HBeAg positive CHB patients (Fattovich *et al*., 2008). Although the effects of CHB on the development of cirrhosis are evident, other factors such as age, alcohol intake, other viral infections such as HCV, also contribute to its development.

##### 1.1.9.3.2 HEPATOCELLULAR CARCINOMA

A majority of liver cancers are said to occur in the background of cirrhosis, however, it has now become evident that a significant fraction of HBV-related HCCs occur in the absence of cirrhosis (Neuveut *et al.,* 2010). These findings suggest a more direct role of HBV in the turmorigenesis process and the development of HCC (Neuveut *et al.,* 2010). It has been estimated that over 50 % of HCC cases worldwide are due to HBV (Shi and Shi 2009). The exact mechanisms as to why there is this progression to HCC remain unknown.

The possible effects of HBV related tumorigenesis are numerous and include possible effects of host immune response to HBV, HBV DNA integration, high HBV replication rates, HBV genotypic variations and the effect of certain viral mutations.

With various reports of HBV integration into the DNA of infected hepatocytes, direct integration was originally thought to be a major contributing factor (Brechot *et al*., 1980). However, a recent study found the frequency and patterns of HBV insertions to be similar between tumor cells and the adjacent non-tumor cells, concluding that the majority of HBV DNA integration events were not associated with hepatocarcinogenesis (Jiang *et al.,* 2012).

It is well known that during the progression from CHB to HCC, various nucleotide changes occur in the viral genome. These changes can occur anywhere in the HBV DNA, however, particularly important changes have been noted in the pre-S region, the carboxy-terminal region of the X gene, which overlaps the enhancer II and BCP region and the in the preC/core region (Liu *et al*., 2009). The possible influence of different genotypes on HCC progression has also been raised; however conflicting data has been reported (Orito *et al*., 2003; Ni *et al*., 2004). The role of different HBV genotypes in liver disease progression is discussed further in section 1.2.

The effects of HBV on HCC development seem multifactorial and are most likely not due to a single mechanism, but rather involve various viral factors as well as certain host factors.

## 1.2 HBV GENOTYPES

### 1.2.1 MOLECULAR CHARACTERISATION AND GEOGRAPHICAL DISTRIBUTION

The term genotype, with regard to viruses, is referred to as the genomic sequence of the virus that has stabilized for an extended period of time (Francois *et al*., 2001). Nine genotypes of the HBV (A-I) have been classified based on an intergroup divergence of between 7.5 to 8 % in the complete genome sequence of the HBV (Norder *et al*., 1994; Kramvis *et al*., 2008; Stuyver *et al.,* 2000; Hannon *et al*., 2000;), and a greater than 4 % divergence within the S gene (Norder et al., 1992). Based on a single isolate, a tenth genotype, genotype J, has been proposed (Tatematsu *et al.,* 2009). The first four identified genotypes A-D (Okamoto *et al.,* 1988), were followed by the identification of four other genotypes E-H (Norder *et al*., 1994; Stuyver *et al*., 2000; Naumann *et al*., 1993; Arauz-Ruiz *et al*., 2002). Together with the identification of the various genotypes, subgenotypes (previously known as subgroups), based on a 4-8% intergroup nucleotide difference across the complete genome have also been identified (Kramvis *et al*., 2005). To date, subgenotypes for HBV genotypes, A, B, C, D, F and I have been identified.

Four major HBV serological subtypes have been described; *adw, adr, ayw, ayr.* These are based on amino acid variations in the surface open reading frame. The major antigenic domain is the common ‘a’ determinant, which is defined by the presence of a T or I at position 126. Two mutually exclusive sub-determinants at positions 122 and 160 specify whether the subtypes are *d/y* and *w/r* (Le Bouvier 1971; Bancroft *et al.,* 1972). These four subtypes can be further classified into ten serological subtypes, based on further variation at residues 122, 127 and 160. The

ten subtypes are as follows; *ayw*1, *ayw*2, *ayw*3, *awy*4, *ayr*, *adw*2, *adw*3, *adw*4, *adrq+* and *adrq-* (Norder *et al*., 1992; Magnuis and Norder 1995; Arauz-Ruiz *et al*., 2002). No definitive relationship between HBV genotypes and serological subtypes has been established.

It is well established that the HBV genotypes and subgenotypes have a particular geographical distribution **(Figure 1.9).** Genotype A and D have the widest distribution. Genotype A is found predominantly in northwestern Europe, North America, central and southern Africa, India and Brazil (Kramvis *et al*., 2005). Genotype D has a more global distribution; however its high prevalence in the Mediterranean area, India and North Africa has been noted. Genotypes B and C are found in central and eastern Asia, genotype E in Africa, and F in Central America (Kramvis *et al*., 2005). The exact distribution of genotypes G and H has not been extensively evaluated, however sequence analysis in France, Germany and the USA, have revealed the presence of genotypes G and H in these areas (Kramvis *et al*., 2005). The recently accepted ninth genotype I, has been found in Vietnam, South East Asia and North East India (Hannoun *et al.,* 2000; Huy *et al*., 2008; Arankalle *et al.,* 2010; Yu *et al.,* 2010). The proposed genotype J was isolated from a patient in Borneo (Tatematsu *et al.,* 2009).

Genotypes A, D and E are found in Southern Africa, with genotype A being the predominant genotype (Kew *et al.,* 2005). This is similar to the genotype distribution in the India subcontinent, which also have genotype A and D.

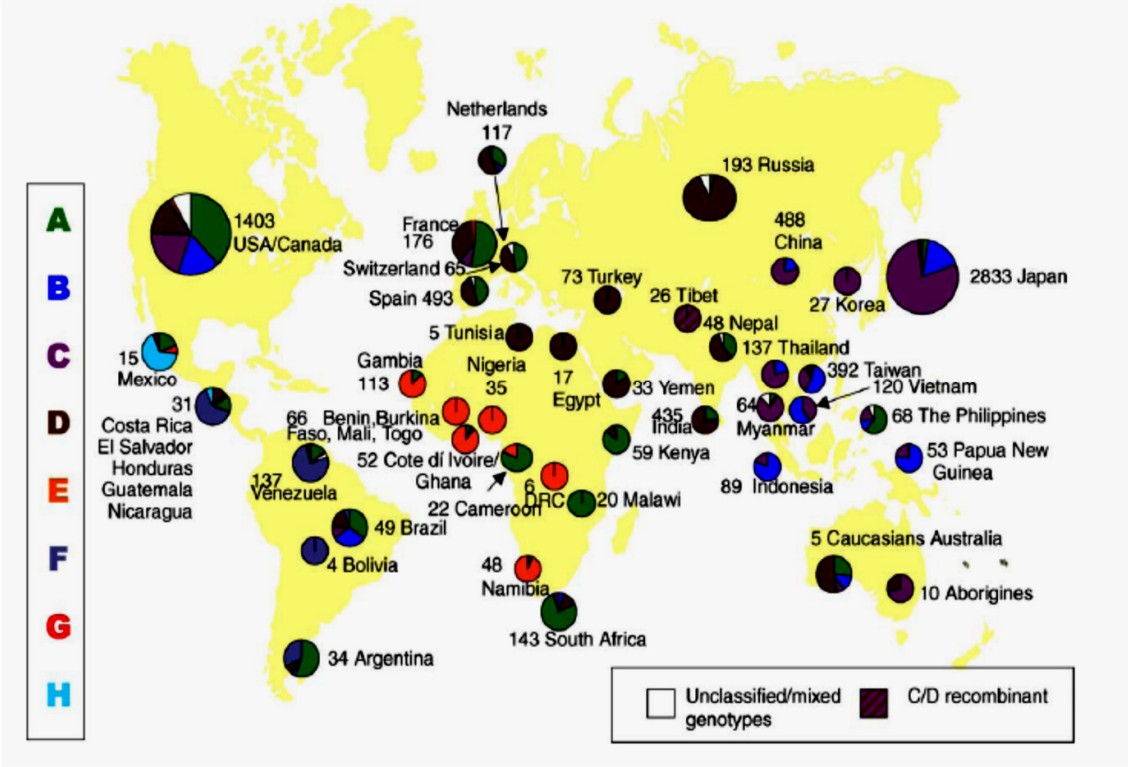



**Figure 1.9 Geographical distribution of the eight HBV genotypes** (Reproduced with permission from Prof A. Kramvis ©).

Genotypes B and C have a dominant distribution along the southeast regions of Asia. Africa has the presence of genotypes A, D and E, with A predominantly found in southern Africa. Genotype D has a more global distribution, with particular clustering in the Mediterranean. Genotypes F, G and H have been found in South and Central America (Figure adapted from Kramvis *et al.,* 2005).

### 1.2.2 SUBGENOTYPE A1 OF THE HEPATITIS B VIRUS

Subgenotype A1 of HBV was the first subgenotype to be recognized, with unique molecular characteristics that differentiate it from subgenotype A2 (Bowyer *et al.,* 1997; Kramvis *et al.,* 2002, Kramvis and Paraskevis, *in press*). The sequence alterations found in subgenotype A1 that differentiate it found subgenotype A2, occur

in the S, X, precore/core and polymerase ORF's as well as in the transcriptional regulatory elements (Bowyer *et al.,* 1997; Kramvis *et al.,* 2002; Kimbi *et al.,* 2004). Increasing evidence suggests that these molecular characteristics of subgenotype A1 account for the higher risk of developing HCC in patients infected with this subgenotype (Baptista *et al.,* 1999; Kew *et al.,* 2005).

Phylogenetic analysis of the complete genomes of subgenotype A1 isolates shows separation into two distinct clades, 'African' and 'Asian' (Kramvis and Kew 2007). The 'African' clade is found to have a greater nucleotide divergence than the 'Asian' clade, which may imply that the 'African' isolates predate the 'Asian' isolates (Kramvis and Kew, 2007). Subgenotype A1 is the dominant subgenotype A strain in Africa, and is also found in the Indian subcontinent and south America (Chandra *et al.,* 2007; Santos *et al.,* 2010). The high occurrence of subgenotype A1 isolates in India and South America may be as a result of the trans-oceanic slave trade, leading to dispersal of subgenotype A1 from Africa (Kramvis and Paraskevis, *in press*). South African patients infected with subgenotype A1 had a 4.5 fold increased risk of developing HCC compared to patients infected with non-A genotypes and also developed cancer 6.5 years earlier (Kew *et al.,* 2005)

### 1.2.3 EFFECT OF HBV ON DISEASE PROGRESSION AND CLINICAL OUTCOME

From various studies, is becoming more apparent that there is a correlation between the HBV genotypes and the severity of liver disease (Kao *et al.,* 2002; Sumi *et al.,* 2003). Most of these studies have focused on differences between genotypes B and C, as well as differences between genotypes A and D, since these respective genotypes often co-circulate. Findings from Asia have suggested that HBV genotype

C infections lead to a more rapid progression to cirrhosis and hepatocellular carcinoma (HCC) compared to genotype B infections (Ding *et al.,* 2001; Kao et al., 2000). However recent findings suggest, at age 45, the incidence of HCC is the same for genotype B and C (Sumi *et al*., 2003).

Most studies involving the analysis and comparison between genotype A and D, have been from Southern Africa and India, particularly with respect to subgenotype A1 in South Africa. A study from South Africa showed a 4.5-fold increased risk for HCC and an earlier age of cancer diagnosis in patients infected with subgenotype A1 as compared to non-A genotypes (Kew *et al.,* 2005). Characteristics of subgenotype A1, together with additional mutations, particularly in the BCP/PreC region may influence disease progression and hence differences in clinical outcome between subgenotype A1 and non-A genotypes (Baptista *et al.,* 1999; Kramvis and Kew, 2007).

## 1.3 HBV MUTANTS: DISTRIBUTION AND AFFECT ON DISEASE PROGRESSION

### 1.3.1 THE BASIC CORE PROMOTER

Naturally occurring mutations in the BCP tend to arise at the late HBeAg positive and HBeAg negative phases of infection (Tong, 2007). The most common BCP mutation is the BCP double mutant (A1762T/G1764A), which impairs production of HBeAg at the transcriptional level (Okamoto *et al.,* 1994).

Numerous other point mutations, including 1653T, 1753-1757, A1764T/G1766A double mutant, 1809T and 1812T, have been found within the BCP (Kramvis and Kew, 1999). These mutations can appear in isolation, together with the A1762T/G1764A double mutant, or in particular patterns that may lead to more severe liver disease and possibly HCC. Moreover, insertions, deletions and duplications may occur in the BCP, which may negatively or positively affect functioning of the BCP, by means of creating new transcription binding sites or deleting or altering existing binding sites (Kramvis and Kew, 1999).

The majority of mutations in the BCP affect transcription of preC mRNA and pgRNA, however mutations within the 1809-1812 regions, which are characteristic of subgenotype A1, affect the expression of HBeAg at the translational level. The 1809-1812 nucleotide sequence is situated adjacent to the precore initiation codon, 5'-AGCACCAUGC-3' (nt 1808-1817) (Tong *et al*., 1992). The presence of either double or triple mutations at the 1809-1812 region (positions -5, -3, and -2 to the AUG start codon) have been shown to effect the translation of HBeAg (Ahn *et al.,* 2003). The presence of these nucleotide changes, leading to an alteration of the Kozak sequence, result in impaired translation of HBeAg (Ahn *et al.,* 2003).

### 1.3.2 THE PRECORE/CORE REGION

The most prevalent mutation found in the precore region, is the G1896A, creating a premature stop codon. This nucleotide change completely abolishes HBeAg expression (Carman *et al.,* 1989). The presence of the G1896A mutant varies between genotypes, and hence between different geographic regions. The G1896A mutant is found in patients infected with HBV genotypes B, C, D and E that have a thymine at position 1858. The presence of C1858T allows base pairing with G1896A, which stabilizes ε, causing increased pregenomic encapsidation and initiation of DNA synthesis (Lok *et al*., 1994). In genotype A, in which a C occurs at position 1858, the presence of a G1896A mutation creates a 'wobble pair' (Fig 3) (Li *et al*., 1993). This 'wobble pair' causes destabilization of the stem loop structure, reducing the functional activity of ε (Li *et al*., 1993). For this reason, the G1896A mutation in genotype A is always accompanied by the C→T transition at 1858, allowing for Watson-Crick base pairing and stabilization of the stem loop structure (Kramvis *et al*., 1997).

The clinical significance of the G1896A mutation has not been conclusively established, with some reports showing a significant occurrence in chronic active hepatitis and fuliminant hepatitis, whereas others have shown a lower risk for liver disease progression (Chan, 2010). The reason for these findings could be due to the presence of various nucleotide changes in the BCP/PreC region having an influence on disease progression, and not only the presence of specific mutations. Nucleotide changes have also been found at positions 1846, 1850, 1862, 1888 and 1898. The majority of these nucleotide changes can influence HBeAg expression, be it by different mechanisms. The G1862T, found predominantly in subgenotype A1,

creates a valine to phenylalanine mutation in the precore region, which has shown to cause intracellular retention of HBeAg and hence decreased HBeAg secretion (Chen *et al*., 2008).

### 1.3.3 THE HBX PROTEIN

The X gene overlaps basic core promoter (nt 1742-1802) in the concomitant reading frame, hence nucleotide changes such as the A1762T and G1764A in the BCP cause two missense mutations in the HBx protein, K130M and V131I (Locarnini 2004). Another common BCP nucleotide change is the T1753C, which also causes an aa change at position 127 (I127T) within the HBx (Takahashi *et al.,* 1999). These mutations may be selected for during chronic infection due to immune pressure against HBx (Takahashi *et al.,* 1999).

### 1.3.4 THE POLYMERASE PROTEIN

With the introduction of nucleoside and nucleotide analogue treatment, pressure on the virus has resulted in the presence of selective mutations in the polymerase protein. These mutations have been mapped to certain domains within the reverse transciptase region (Zoulim and Locarnini 2009). The most commonly reported resistance mutations are those providing resistance to lamivudine. These mutations have been mapped to the YMDD locus of the catalytic (C) domain of the reverse transcriptase region (Stuyver *et al*., 2000). This major catalytic site in the palm of sub-domain of the polymerase overlaps the 'a' determinant of the S ORF (Torresi *et al*., 2002). Lamivudine mutations in the YMDD locus include rtM204I/V, as well as other mutations rtL180M and rtA181T/V, found outside the catalytic domain (Zoulim and Locarnini 2009). These resistance mutations are often accompanied by other

compensatory mutations such as rtL80V/I, rtI169T, rtV173L, rtT184S/G, rtS202I and rtQ215S, in other domains of the polymerase protein, which increase viral replication (Zoulim and Locarnini 2009).

Various other resistance mutations have been reported for other nucleoside and nucleotide analogues, which can also affect treatment and influence disease progression. These mutations may also provide cross resistance between various analogues, which complicates treatment further. The clinical impact of these mutations is because of the compensatory increase in viral replication and hence viral loads, which often accompanies these mutations. This leads to increased serum ALT levels and progression to liver disease (Lai *et al.,* 2003).

### 1.3.5 THE COMPLETE SURFACE REGION

Mutants within the pre-S regions are often found in chronic HBV patients, and have also been implicated in liver disease progression and HCC (Fan *et al*., 2001; Sugauchi *et al*., 2003). The pre-S1 region is present only in the LHBs protein and the pre-S2 region in both the LHBs and MHBs surface proteins. The pre-S regions provide important interactions with the immune response due to the presence of B and T-cell epitopes in the corresponding aa sequences (Kuroki *et al.,* 1990; Maeng *et al.,* 2000). This region also contains vital elements required for viral replication, such as the hepatocyte binding region, the CCAAT box and the S promoter sites (Cabrerizo *et al.,* 2002; Chen *et al.,* 2006). A correlation between the presence of pre-S mutations and deletions and HBV genotypes has been reported, with a higher frequency of pre-S deletions in patients infected with genotype C, as compared to other genotypes (Sugauchi *et al*., 2003).

Because the HBsAg protein also contains the highly immunogenic 'a' determinant at residues 124-147 (Carman *et al.,* 1990), mutations in the HBsAg may influence host immune response and also lead to vaccine escape. The most commonly described immune/vaccine escape mutants in the '*a*' determinant region include the G145R and Q129R amino acid substitutions (Carman *et al.,* 1990). Other common mutations include the K141E and T131I, and the 3 amino acid insertion between residues 123 and 124 (Seddigh-Tonekaboni *et al.,* 2000). These changes affect the antigenic structure of HBsAg. These vaccine escape mutants pose a major challenge in HBV prevention. HBsAg mutations also influence diagnostic assays used to detect HBsAg, and hence may lead to false negatives, which again poses a problem in the treatment and control of HBV infection.

## 1.4 HBV INFECTION ON THE INDIAN SUBCONTINENT

Infectious diseases are a major cause of deaths in the Indian subcontinent, of which HBV is one of the most important aetiological agents. The WHO report on HBV prevention in India estimates an HBsAg prevalence rate among the general population between 0.1% and 11.7%, with most studies between 2% and 8% (WHO, 2002). This places India in the intermediate (2-8%) zone of HBV endemicity. Together with these figures, the WHO estimated the HBV carrier rate as 50 million, forming nearly 15% of the entire pool of HBV carriers in the world (WHO, 2002).

In India, chronic HBV infection is the most important factor responsible for the development of HCC (Raza *et al.,* 2007), which has prompted an interest in the role and pathogenic mechanisms involved in HBV induced hepatocarcinogenesis in the Indian subcontinent.

### 1.4.1 HBV GENOTYPES: DISTRIBUTION AND EFFECT ON DISEASE PROGRESSION

The geographic location and social and cultural diversity of India, contribute to the diversity of HBV genotypes found in the Indian subcontinent. Genotype distribution in India varies between the north, south, east and west of the country. Although the prevalence of different genotypes varies between these areas, it is well documented that the predominant genotypes circulating in India are genotypes A and D (Datta 2008). Most of these studies have focused on the genotype distribution among patients from Northern and Eastern India (**Figure 1.10**). Studies from northern India have shown genotypes A and D to be prevalent, with a few cases of genotype C also reported. Studies have found genotypes A and D to be equally prevalent, whereas

others have shown a predominance of genotype D (Thakur *et al.,* 2002; Chattopahyay *et al.,* 2006a). Eastern India also have three genotypes present, A, D and C, which are found to be equally prevalent (Banerjee *et al.,* 2006a; Banerjee *et al.,* 2006b; Sugauchi *et al.,* 2004). Reports from Western and Southern India have shown a predominance of genotype D (Gandhe *et al.,* 2003; Vivekanandan *et al*., 2004). Although reports have shown a presence of genotype D, data from western and particularly southern India, remain limited compared to eastern and northern India. Differences in the geographic distribution and ethnicity of the individuals on the Indian subcontinent may influence HBV genotype prevalence among various regions of the country.

Although genotype distribution for most of India is well characterized, subgenotype characterisation is very limited. To date, the only comprehensive subgenotype data are from eastern India , with subgenotypes A1, C1, D1, D2, D3 and D5 been isolated from HBV infected patients (Banerjee *et al*., 2006; Chattopadhyay *et al.,* 2006).

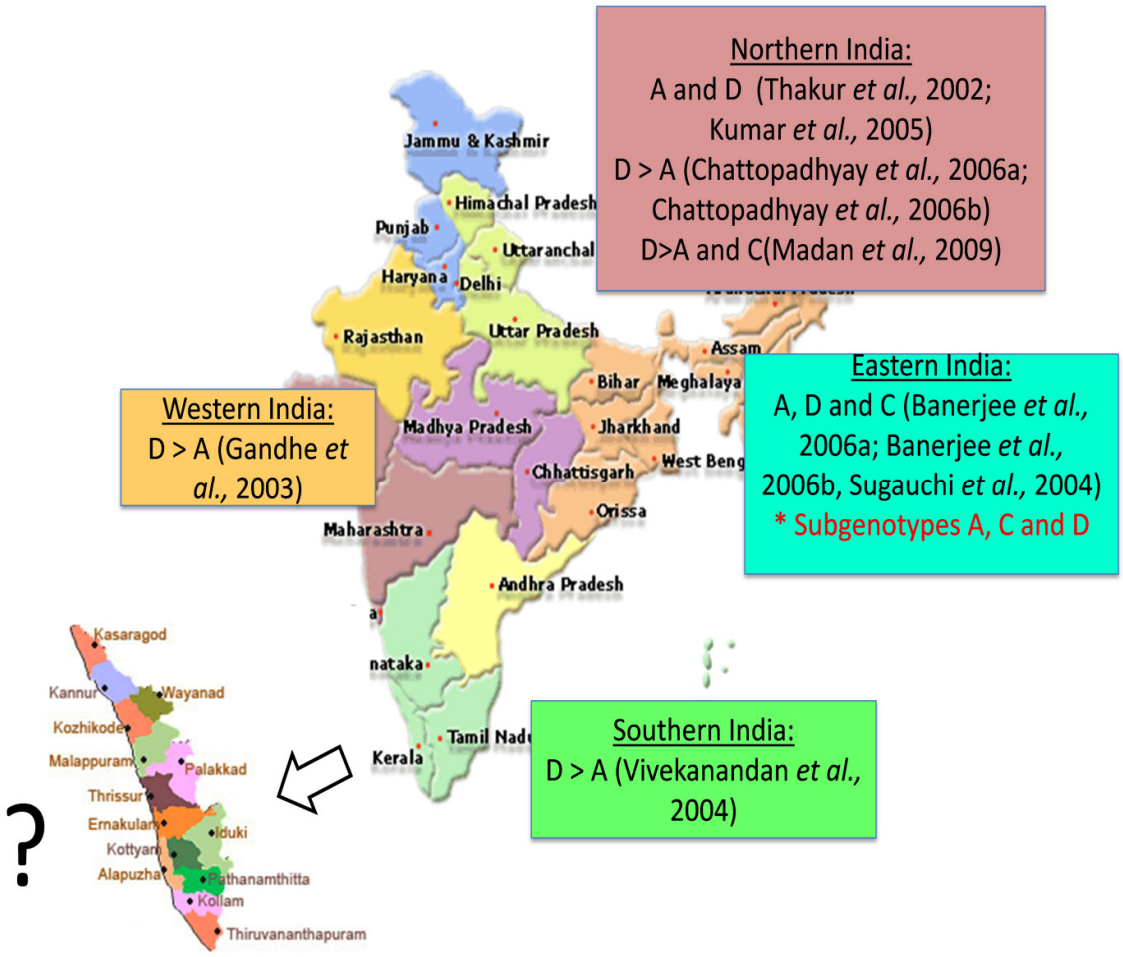



**Figure 1.10 HBV genotype distribution on the Indian subcontinent.** Genotype distribution varies between the four regions of the India subcontinent, with genotyped A and D being predominant. Subgenotypes have only been determined on HBV isolates from eastern India.

Various groups have tried to determine whether genotypes in India influence disease severity; however results remain somewhat conflicting. A recent study from India did not show any differences between genotypes A, D and C in terms of histological severity, biochemical severity and phenotypic disease status (Madan *et al.,* 2009). This is in contrast to another study from India, which showed a higher histological activity index in patients infected with genotype D than those with genotype A (Thakur *et al.*, 2002). Genotype D was also more commonly associated with HCC in

patients younger than 40 years of age, suggesting earlier progression to HCC in patients infected with genotype D (Thakur *et al.,* 2002). Another study from India showed that patients infected with genotype A had higher ALT levels and were more often positive for HBeAg as compared to genotype D (Kumar *et al*., 2005). The authors concluded that genotype A was more often associated with severe liver disease than is genotype D. These studies show conflicting results, and again show the difficulty in clarifying the true association between genotypes and disease progression.

Various groups have looked at the impact of certain HBV mutations, particularly those in the BCP/PreC region, which are known to influence disease progression (Chauhan *et al.,* 2006; Biswas *et al.,* 2011). To date, very limited data is available for the role of surface region mutations and deletions in HBV disease progression in Indian populations.

## 1.5 AIMS AND OBJECTIVES OF THE STUDY

India is classified as an intermediate HBV endemic zone, with an estimated 50 million HBV carriers, forming the second largest pool of chronic HBV infection in the world (WHO, 2002). Genotypes A and D are the predominant genotypes in India (Data 2008). However, these finding are mainly as a result of studies conducted in the eastern and northern regions of the country. Limited data is available from western and particularly southern regions of India.

Kerala, in south-west India is the most densely populated state of India, with a population of 33 million. An HBsAg prevalence of 0.5% in the normal population has been reported in northern Kerala, with an increased carrier rate of 1.5 % being detected among voluntary blood donors from Trivandrum, south Kerala (Sandesh *et al.,* 2006; Anjal *et al.,* 2012). There is currently a paucity of information on the prevalence of HBV genotypes and the respective subgenotypes in Kerala, together with limited data on the molecular characterisation of HBV isolates from Kerala.

The aim of this study was to investigate the distribution of HBV genotypes/subgenotypes in Kerala among patients with different clinical manifestations of HBV infection, and to molecularly characterise prevalent mutations.

<u>The objectives of the study included:</u>

1. PCR amplification and sequencing of the BCP/PreC region, S region, and complete HBV genome
2. Genotyping/subgenotypying by phylogenetic analysis of complete S region

3. Correlation of BCP/PreC mutants and S region mutants/deletions with genotypes and disease groups
4. Bioinformatic analysis of subgenotype A1 isolates

# CHAPTER 2: MATERIALS AND METHODS

## 2.1 OVERVIEW OF STUDY DESIGN AND ANALYSIS

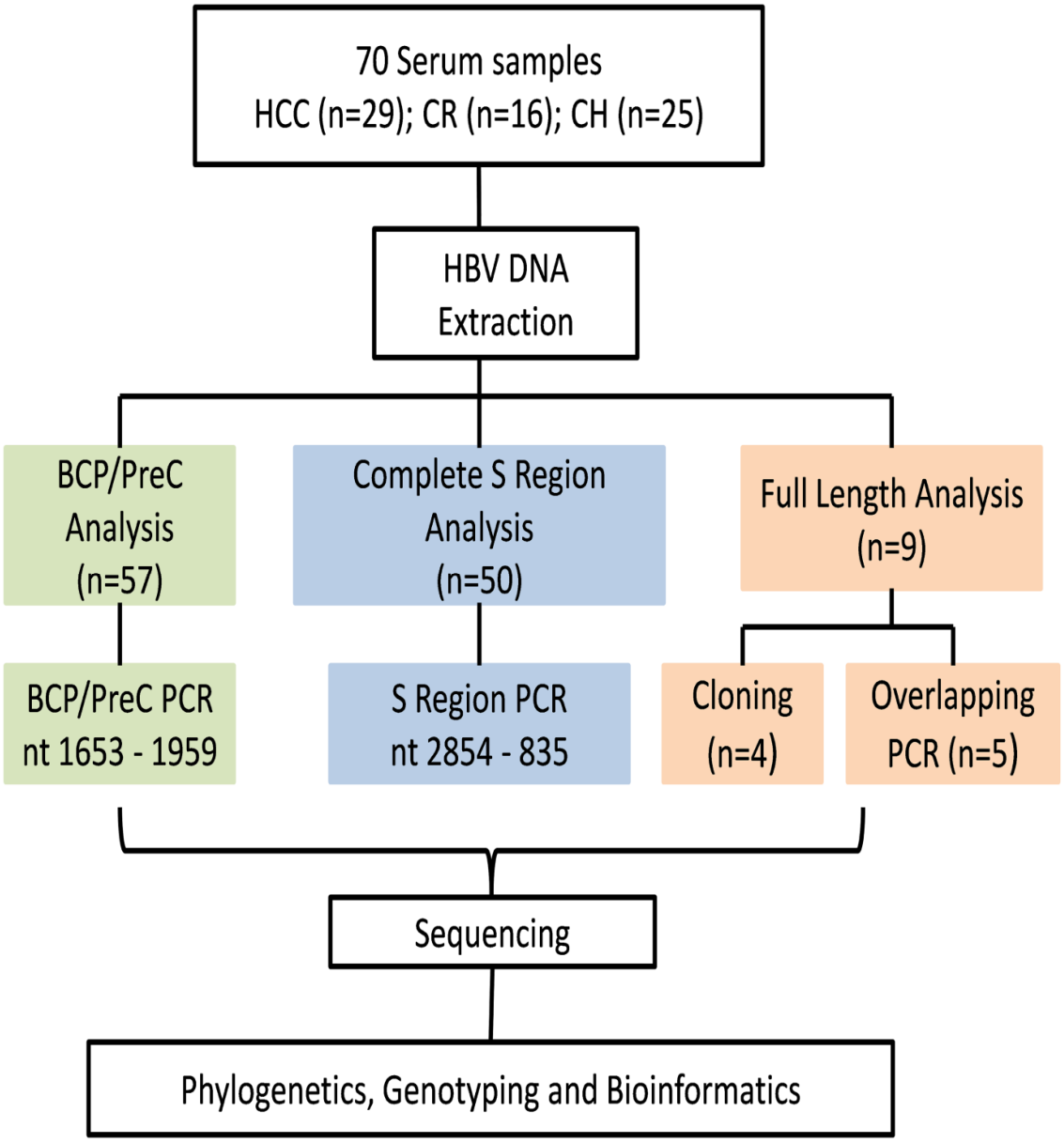



**Figure 2.1 Overview of the molecular characterisation of HBV isolates from Kerala, India**

## 2.2 STUDY SAMPLES AND PATIENT ALLOCATION

Patients from Kerala, India, who presented to the Department of Gastroenterology, and who met the criteria of being HBsAg positive for more than 6 months, were recruited for the study. A total of 249 patients were recruited, of which 100 were diagnosed as having CH, 100 had HCC, and 49 patients were diagnosed with CR.

The diagnosis of HBV related liver disease was based on clinical findings, laboratory tests, liver biopsies and imaging studies. The patients were subsequently classified into three groups:

- Group I (Hepatocellular carcinoma, HCC): serum α-fetoprotein > 400ng/mL and the presence of a lesion > 5 cm, on ultrasound scan
- Group II (Cirrhosis, CR): necro-inflammatory damage, fibrosis and nodule formation confirmed by liver biopsy, and ultrasonographic evidence of portal hypertension
- Group II (Chronic hepatitis, CH): HBsAg-positive for > 6 months, with normal or intermittently elevated ALT (1.5 times upper limit of normal). Patients in this group were considered for liver biopsy on the basis of elevated ALT levels and HBeAg negative status. Diagnosis of cirrhosis was made using Histological Activity index (HAI) and Fibrosis scores.

Of the 249 patients, serum samples from 70 patients were used in the study. These included 29 HCC patients, 25 CH, and 16 CR patients. Informed consent was obtained from each patient and the study was approved by the University of the

Witwatersrand Ethics Committee (# M090311), and the ethics committee of the Medical College Trivandrum, India.

### 2.3 SERUM SAMPLE COLLECTION AND STORAGE

Patient serum was collected by Dr Deepak Gopalakrishnan at the Thiruvandrum Medical College, Kerala. Serum aliquots of 1 ml were stored at - 80°C.

### 2.4 SERUM DNA EXTRACTION

HBV viral DNA was extracted from serum using the QIAmp MinElute virus kit (Qiagen Inc., Hilden, Germany). The method supplied by the manufacturer was used, with one modification. The final elution step was carried out in 75 µl best quality water (BQW) instead of 200 µl. The QIAmp MinElute virus kit (Qiagen) was used due to the lower volume of blood needed for the extraction process.

### 2.5 DETECTION OF BASIC CORE PROMOTER/PRECORE/CORE REGION MUTANTS

#### 2.5.1 POLYMERASE CHAIN REACTION OF THE BCP/PRECORE/C REGION

The region from nucleotide position 1653 to 1959 (numbering from the *EcoR*I site, Genbank # AY233289), which includes the entire BCP as well part of the PreC/C region, was amplified using a nested PCR. Published primers were used for the reaction (Takahashi *et al.*, 1995), with a single nucleotide change at position 10 of the second round antisense primer. Primers were aligned to genotype A and D sequences in Genbank. It was found that a single nucleotide change (G→C) at position ten of the inner antisense primer would be more suited to HBV genotypes A and D. First and second round reactions were performed in a 25 µl and 50 µl total

volumes respectively. The primers BCP1F (sense: 5'-GCATGGAGACCACCGTGAAC-3', nt 1606 – 1625) and BCP1R (antisense: 5'-GGAAAGAAGTCAGAGGGCAA-3', nt 1955 - 1974) were used for the first round reaction, followed by the primers BCP2F (sense: 5'-CATAAGAGGACTCTTGGACT-3', nt 1653-1672) and BCP2R (antisense: 5'-GGCAAAAAACAGAGTAACTC-3', nt 1940 - 1959) for the second round reaction. The first round reaction mixture consisted of 0.5 U of BIOTAQ™ DNA polymerase (Bioline, USA), 200 µM of each deoxynucleoside triphosphate (dNTP), 1 µM of each primer, 3 mM $MgCl_2$, 16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl (pH 8.8) and 0.01% Tween-20, to which 2.5 µl of DNA sample was added. Prior to the second round reaction, the first round amplicon was diluted 1: 100, and 5 µl of this was then reamplified using the inner set of primers. The second round reaction mixture contained 1U of BIOTAQ™ DNA polymerase (Bioline, USA), 200 µM of each deoxynucleoside triphosphate (dNTP), 1 µM of each primer, 1.5 mM $MgCl_2$, 16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl (pH 8.8) and 0.01% Tween-20, to which 5 µl of the 1:100 diluted first round amplicon was added. Both first and second round reactions were performed on the MyCycler thermal cycler™ PCR machine (BioRad, USA), using the following cycling conditions: initial denaturation at 94°C for 180 s, followed by 30 cycles of denaturation at 94°C for 60 s, annealing at 55°C for 60 s and elongation at 72°C for 120 s. A final elongation step at 72°C for 7 min was included. If amplification using the above mentioned protocol was unsuccessful, the same reaction and cycling conditions were performed, with the dilution step between first and second round reactions being omitted.

In order to determine successful amplification, a 5 µl aliquot of the second PCR product was resolved on a 2% agarose gel and visualized via ethidium bromide staining. Successful amplification yielded a 307 bp DNA fragment.

### 2.5.2 SEQUENCING OF THE BCP/PRE-C/C AMPLICON

The 307 bp PCR amplicon was sequenced in both the forward and reverse direction using the following primers: BCPS1 (sense: 5'-GAG GCA TAC TTC AAA GAC TG-3', nt 1698-1717) and BCPS2 (antisense: 5'-AGT AAC TCC ACA GTA GCT CC-3', nt 1928-1947). Sequencing was performed by the DNA Sequencing Facility, Stellenbosch University, South Africa, using the ABI 3130XL Genetic analyser (Applied Biosystems, Foster City, CA). Electropherograms of the forward and reverse sequences were then analysed and assembled using SeqMan software (Lasergene ®). A total sequence length of 247 bp from position 1698-1947 was assembled and analysed.

## 2.6 COMPLETE S REGION ANALYSIS OF THE HBV GENOME

### 2.6.1 POLYMEARSE CHAIN REACTION OF THE COMPLETE S REGION

The region from nucleotide position 2410 to 1314 (numbering from *EcoR*I site, Genbank accession # AY 233289) was amplified using a nested PCR. The final PCR product of 2050 bp included the entire S region (nt 2854 to nt 835) and the overlapping polymerase region. Both first and second round reactions were performed using hot-start PCR methodology. The first round primers were as follows, S1F (sense: TCA ATC GCC GCG TCG CAG AAG ATC TCA ATC, nt 2410 – 2439) and S1R (antisense: TCC AGA CCG/T GCT GCG AGC AAA ACA, nt 1291 – 1314). Second round primers included S2F (sense: AAT GTT AGT ATT CCT TGG ACT

CAT AAG GTG GG, nt 2451 – 2482) and S2R (antisense: AGT TCC GCA GTA TGG ATC GGC AGA GGA, nt 1254 – 1280). The first reaction mix consisted of 12.5 µl HotStarTaq Master Mix (Qiagen, Germany) with 5 U HotStarTaq DNA Polymerase (Qiagen, Germany), 1.5 mM $MgCl_2$ and 200 µM of each dNTP, 0.2 µl of each 20 µM primer and 9.6 µl of distilled water. To this reaction mix, 2.5 µl of HBV DNA was added for a total volume of 25 µl. Amplification was done using the Biorad MJ Mini Thermocycler (BioRad, USA) under the following cycling conditions: initial denaturation at 95°C for 15 minutes followed by 40 cycles of 94 °C for 1 minute, 66°C for 3 minutes and 72°C for 3 minutes. A final extension step at 72°C for 10 minutes was also included.

The second reaction mix consisted of 25 µl HotStarTaq Master Mix (Qiagen, Germany) with 5 U HotStarTaq DNA Polymerase, 1.5 mM $MgCl_2$ and 200 µM of each dNTP, 0.4 µl of each 20 µM primer and 19.2 µl of distilled water. To this reaction mix, 5 µl of the first round PCR was added for a total volume of 50 µl. Second round cycling conditions were the same as for those used in the first round amplification, with one exception, that is the annealing temperature was changed to 68°C for the second round PCR.

Confirmation of correct PCR amplification was performed using an ethidium bromide stained 1% agarose gel. A 5 µl aliquot of the second PCR product was run alongside a 1000 bp marker and visualized as red-orange band under a UV transiluminator. Successful amplification yielded a single 2050 bp fragment, which was subsequently sequenced. If multiple bands were present on the gel, the correct 2050 bp fragment was first gel extracted and then sequenced.

### 2.6.2 GEL EXTRACTION OF THE 2 KB PCR AMPLICON

Gel extraction was performed prior to sequencing on all samples in which multiple bands were visualized on ethidium bromide stained agarose gels. All samples with multiple bands were again run on a 1% agarose gel. No ethidium bromide was added to the gel. A 40 µl aliquot of the PCR reaction mixture was loaded into the gel with 10 µl Loading Buffer containing GelRed™ nucleic acid gel stain (Biotium, Hayward, CA, USA). The GelRed™ nucleic acid gel stain was used as this minimized DNA denaturation, which would affect the sequencing reactions. The bands on the agarose gel were visualized using UV luminescence, and the correct 2050 bp fragment was excised from the gel using a scalpel. The excised DNA band was then purified using the Zymoclean Gel DNA Recovery Kit™ (Zymoresearch, CA, USA) as per manufacturer's instructions. After the final step, the Zymo-Spin column was placed onto a clean 1.5 ml eppindorf tube, and to this, 10 µl of water was added to column matrix and spun down. A 10 µl aliquot of ultra pure DNA, at a minimum concentration of 20ng/µl was attained, and subsequently sequenced.

### 2.6.3 SEQUENCING OF THE COMPLETE S REGION

The correctly amplified PCR product of the S region was sequenced using a series of three overlapping primers. All primers were in the forward direction and allowed sufficient overlap between fragments to ensure the correct dominant population was assembled as the complete S gene region. Sequencing primers were as follows; 2497F (TTC CTT GGA CTC ATA AGG TG, nt 2461 – 2480), 3188F (AG TCA GGA AGG CAG CCT AC, nt 3152 – 3170) and 591F (ATT GCA CCT GTA TTC CCA TCC, nt 591 – 611). Sequencing was performed by the DNA Sequencing Facility, Stellenbosch University, South Africa, using the ABI 3130XL Genetic analyser

(Applied Biosystems, Foster City, CA). Electropherograms of the individual sequencing fragments were analysed and assembled into the 2 kb fragment using SeqMan software.

## 2.7 MOLECULAR ANALYSIS OF THE COMPLETE HBV GENOME

### 2.7.1 CLONING OF THE 3.2 KB HBV GENOME

#### 2.7.1.1 FULL LENGTH HEPATITIS B VIRUS GENOME AMPLIFICATION

The complete 3.2 kb genome of the HBV was amplified according to the methodology used by Gunther *et al*. (1995). The single round PCR was done using the Expand High Fidelity System (Roche), with the following primers; P1 (sense: 5'-TTT TTC ACC TCT GCC TAA TCA-3', nt 1821-1841) and P2 (antisense: 5'-AAA AAG TTG CAT GA/GT GA/CT GG-3', nt 1806-1825).

The reaction mix consisted of 2 µl of 10X Expand High Fidelity (HF) buffer with 15 mM MgCl2, 0.2 µl of 25 mM dNTP mix, 1.25 µl of each 20 µM primer P1 and P2 and 10.3 µl of distilled water, for a total volume of 15 µl. To this reaction mix, 5 µl of extracted HBV DNA was added. A second enzyme mix of 0.5 µl of Expand HF buffer, 0.75 µl of Expand HF enzyme and 3.75 µl of distilled water was also prepared. The reaction mixture was preheated in the thermocycler to 94°C for 2 minutes and then cooled done to 57°C at which point 5 µl of the enzyme mix were added. This was then followed by 40 cycles on the Biorad MJ Mini Thermocycler (BioRad, USA), using the cycling conditions provided below **(Table 2.1).** PCR amplification used a step up method of increasing the extension time. A series of 4 conditions each with 10 cycles made for a complete cycling profile of 40 cycles. An

initial denaturation set of 94℃ for 2 minutes was included prior to the cycling conditions. Both positive and negative controls were used in all reactions

**Table 2.1 Primers and PCR conditions used for amplification of 3.2 kb HBV genome**

| Primers | Position[a] | Primer sequence 5'-3' | Denaturation | Annealling | Extension | Cycles |
|---|---|---|---|---|---|---|
| P1 | 1821-1841 | 5'-TTT TTC ACC TCT GCC TAA TCA-3' | 94℃ for 40 sec | 57℃ for 90 sec | 68℃ for 3 min | 10 |
| P2 | 1806-1825 | 5'-AAA AAG TTG CAT GA/GT GA/CT GG-3' | | | | |
| P1 | 1821-1841 | 5'-TTT TTC ACC TCT GCC TAA TCA-3' | 94℃ for 40 sec | 57℃ for 90 sec | 68℃ for 5 min | 10 |
| P2 | 1806-1825 | 5'-AAA AAG TTG CAT GA/GT GA/CT GG-3' | | | | |
| P1 | 1821-1841 | 5'-TTT TTC ACC TCT GCC TAA TCA-3' | 94℃ for 40 sec | 57℃ for 90 sec | 68℃ for 7 min | 10 |
| P2 | 1806-1825 | 5'-AAA AAG TTG CAT GA/GT GA/CT GG-3' | | | | |
| P1 | 1821-1841 | 5'-TTT TTC ACC TCT GCC TAA TCA-3' | 94℃ for 40 sec | 57℃ for 90 sec | 68℃ for 9 min | 10 |
| P2 | 1806-1825 | 5'-AAA AAG TTG CAT GA/GT GA/CT GG-3' | | | | |

[a] numbering from *EcoR*I site

In order to determine successful amplification, a 5 µl aliquot of the PCR product was resolved on a 1% agarose gel and visualized via ethidium bromide staining and a UV transilluminator. Successful amplification yielded a complete HBV genome fragment of 3.2 kb. The size of the PCR product was estimated using a 1 kb marker that was run alongside all PCR products.

### 2.7.1.2 CLONING OF THE AMPLIFIED 3.2 KB PCR FRAGMENT

To ensure that sufficient DNA was available for sequencing and for further downstream analysis, the PCR product was cloned into the pSMART-cDNA vector as per instructions in the cSMART-cDNA cloning kit (Lucigen Corporation, Greenview Drive, Middleton, WI).

The 3.2 kb PCR product was first gel extracted to ensure that a single fragment free of any contaminants was used for the cloning reaction. The gel extracted PCR

product was processed to generate blunt phosphorylated ends using the PCRTerminator End Repair Kit (Lucigen). The blunted ended PCR product was then ligated into the pSMART vector using CloneSmart DNA Ligase as per the reaction conditions set out in the manufactures protocol. Transformation of the ligated PCR product into *E. Cloni* chemically competent cells (Lucigen) using heat shock transformation methodology was then performed. The transformed *E. Cloni* cells were then plated onto kanamycin enriched YT agar plates and incubated overnight at 37°C. Individual colonies were then picked and replated onto the kanamycin enriched YT agar plates and grown up overnight at 37°C.

Individual colonies were then inoculated into 5 ml of TB medium containing kanamycin. The inoculated TB medium was grown up overnight at 37°C shaking at 200 rpm. This process enabled amplification of correctly inserted HBV genomes and thus a sufficient quantity of DNA available for sequencing and downstream processes.

To ensure that the complete HBV genome had been correctly ligated into the pSMART vector, the plasmid DNA was extracted from the overnight TB medium and digested with the *EcoR*V enzyme. Plasmid DNA was extracted using the GeneAid High Speed Plasmid Mini Kit (GeneAid) as per manufacturer's instructions.

Digestion of plasmid DNA was done using *EcoR*V restriction enzyme which yielded a 3.2 kb HBV fragment and a 1.8 kb pSMART vector fragment. The presence of these two bands on gel electrophoresis confirmed successful cloning of the complete HBV genome into the pSMART vector.

Clones that were positive by restriction digestion were again plated on YT agar enriched with kanamycin 50 mg/ml (BioBasic Inc., Markham, Ontario, Canada) and grown up overnight at 37℃. Plates were then sent for sequencing using the sequencing primers in table 2.2.

### 2.7.1.3 SEQUENCING OF THE COMPLETE 3.2 KB HBV GENOME

Sequencing of the 3.2 kb HBV genome was done using a series of HBV specific primers as well as vector specific primers. Primers used in the sequencing reaction are provided **(Table 2.2).** Sequencing was performed using by Inqaba Biotechnical Industries Pty. Ltd., South Africa, using the ABI 3130XL Genetic analyser (Applied Biosystems, Foster City, CA). Electropherograms for both forward and reverse primers were assembled and analysed used SeqMan software.

**Table 2.2 Sequencing primers for compete 3.2 kb analysis**

| Primer | Primer sequence 5' – 3' | Orientation | Nucleotide position |
|---|---|---|---|
| CL3 | GTT GAT TGC AGT CCA GTT ACG CT | - | Vector specific |
| SR2 | GGT CAG GTA TGA TTT AAA TGG TCA GT | - | Vector specific |
| Kram1 | CTT CTG ACT TCT TTC CTT C | Sense | 1959-1977 |
| Kram2R | CCA AGA ATA TGG TGA CC | Antisense | 2821-2837 |
| Kram3 | GTT AGT ATT CCT TGG ACT | Sense | 2453-2471 |
| Kram4R | GTC CTA GGA ATC CTG ATG | Antisense | 168-185 |
| Kram5 | CTG GTG GCT CCA GTT C | Sense | 60-75 |
| Kram6R | CTG AAA GCC AAA CAG T | Antisense | 719-734 |
| Kram7 | CAC CTG TAT TCC CAT C | Sense | 595-610 |
| Kram8R | GAC GTA AAC AAA GGA CG | Antisense | 1246-1261 |
| Kram9 | CTG CCG ATC CAT ACT G | Sense | 1258-1273 |
| Kram10R | CAA TTT ATG CCT ACA GCC TC | Antisense | 1777-1796 |

### 2.7.2 COMPLETE HBV GENOME ANALYSIS USING AMPLIFICATION OF TWO SUBGENOMIC FRAGMENTS

#### 2.7.2.1 PCR AMPLIFICATION OF THE 2 KB AND 1.6 KB SUBGENOMIC FRAGMENTS

In order to overcome the difficulties encountered with the full length PCR (particularly when viral loads were low), and with the cloning process, an overlapping subgenomic fragment PCR was designed and optimized. Two HBV fragments of 2 kb and 1.7 kb, respectively, were amplified and the sequences assembled to obtain the full length 3.2 kb HBV genome.

The first 2 kb PCR fragment, which included the complete S region as well as part of the overlapping polymerase region was amplified and sequenced according to the methodology described earlier under complete S gene analysis. The second 1.6 kb fragment from nucleotide position 1096 to 2663 was amplified using a nested PCR. This fragment ensured sufficient overlap with the ends of the 2 kb HBV fragment, allowing for the assembly of a complete 3.2 kb HBV genome that represented the dominant quasispecies for each sample.

Primers used for the nested PCR included: 991F (sense: 5'-CA AA/CG AAT TGT GGG TCT TTT GG-3', nt 991-1012) and 2878R (antisense: 5'-T GTT CCC AAG AAT ATG GTG ACC-3', nt 2821-2842) for the first round, followed by 1096F (sense: 5'-TTC TCG CCA ACT TAC AAG GC-3', nt 1096-1115) and 2699R (antisense: 5'-AG GAT AAA AT/CC TAG CAG GCA-3', nt 2644-2663) for the second round. The first reaction mix consisted of 12.5 µl HotStarTaq Master Mix (Qiagen) with 5 U HotStarTaq DNA Polymerase, 1.5 mM $MgCl_2$ and 200 µM of each dNTP, 0.2 µl of

each 20 µM primer and 9.6 µl of distilled water. To this reaction mix, 2.5 µl of HBV DNA was added for a total volume of 25 µl. Amplification was performed using the Biorad MJ Mini Thermocycler (BioRad, USA), programmed with the following cycling conditions: 95°C for 15 minutes, followed by 40 cycles of 94°C for 1 minute, 61°C for 2 minutes and 72°C for 3 minutes. A final extension step of 10 minutes at 72°C was carried out following the completion of the 40 cycles.

The second round PCR reaction mix consisted of 25 µl HotStarTaq Master Mix (Qiagen) with 5 U HotStarTaq DNA Polymerase, 1.5 mM $MgCl_2$ and 200 µM of each dNTP, 0.4 µl of each 20 µM primer and 19.2 µl of distilled water. 5 µl of the first round PCR product was then added to the reaction mix for a total reaction volume of 50 µl. The second round PCR cycling conditions were as follows: 95°C for 15 minutes, followed by 40 cycles of 94°C for 1 minute, 58°C for 2 minutes and 72°C for 3 minutes. A final extension step of 10 minutes at 72°C was carried out following the completion of the 40 cycles.

A 5 µl aliquot of the second PCR product was resolved on a 1 % agarose gel and visualized via ethidium bromide staining using a UV transilluminator. The correct size of the PCR product was determined using a 1 kb molecular marker which was run alongside the PCR product. A final PCR amplicon of 1.6 kb confirmed a positive result.

#### 2.7.2.2 SEQUENCING OF THE 1.6 KB SUBGENOMIC PCR AMPLICON

The 1.6 kb PCR amplicon was sequenced using a series of three overlapping primers. The following were used for the sequencing reaction: 1096F (sense: 5'-TTC

TCG CCA ACT TAC AAG GC-3', nt 1096-1115), 1580F (sense: 5'-TGC ACT TCG CTT CAC CTG TG-3', nt 1096-1115), 1580F (sense: 5'-TGC ACT TCG CTT CAC CTC TG-3', nt 1580-1659) and 2058F (sense: 5'-CCT TAG AGT CTC CTG AGC AT-3', nt 2022-2041). All sequencing was done in the forward direction with sufficient overlap of the fragments to enable correct assembly of the 1.6 kb HBV sequence. All sequencing was performed by the DNA Sequencing Facility, Stellenbosch University, South Africa, using the ABI 3130XL Genetic analyser (Applied Biosystems, Foster City, CA). Electropherograms of the individual sequencing fragments were analysed and assembled into the 1.6 kb fragment using SeqMan software.

In order to obtain the sequence of the complete 3.2 kb HBV genome, the sequences of the 2 kb and the 1.6 kb fragments were assembled. The sufficient overlap of the two fragments enabled correct assembly of the two fragments when the overlapping sequences were homologous.

### 2.8 BIOINFORMATICS AND PHYLOGENETIC ANALYSIS

All sequences generated, were compared with the sequences of HBV genotypes (A-I) from Genbank. All multiple sequence alignments and nucleotide divergence measurements were performed using ClustalW (Thompson *et al.,* 1994). Multiple sequence alignments were used for phylogenetic analysis. Alignments were fed into PHYLIP (Phylogeny inference package) version 3.5c (Page, 1996). DNAML (maximum likelihood) alone and DNADIST, consecutively with NEIGHBOR (neighbour-joining) were used to generate dendograms. For bootstrapping, using 1000 data sets, SEQBOOT, DNADIST and NEIGHBOR were used. Trees were then

constructed using CONSENSE, and subsequently visualized using TreeView Win 32 version 1.6.1 software program (Page, 1996).

## 2.9 STATISTICAL ANALYSIS

Data are represented as means +/- (SD) and n %. Continuous variables were compared using an independent Student's t test. The $\chi^2$ test or the Fisher's exact test was used to compare categorical variables. Odds ratio was calculated to assess the risk of HCC. All p-values were two-sided, and the difference was considered statistically significant for $p<0.05$. The analysis was performed using Statistical package for Social Sciences (SPSS 15) program (SPSS Inc., Chicago, IL, USA).

# CHAPTER 3: RESULTS

A total of seventy HBV DNA-positive serum samples from liver disease patients from Kerala, India were used to molecularly characterize HBV (**Figure 3.1**), by amplifying and sequencing the BCP/PreC region and/or the complete S region. In addition, nine complete HBV genomes were sequenced and analysed.

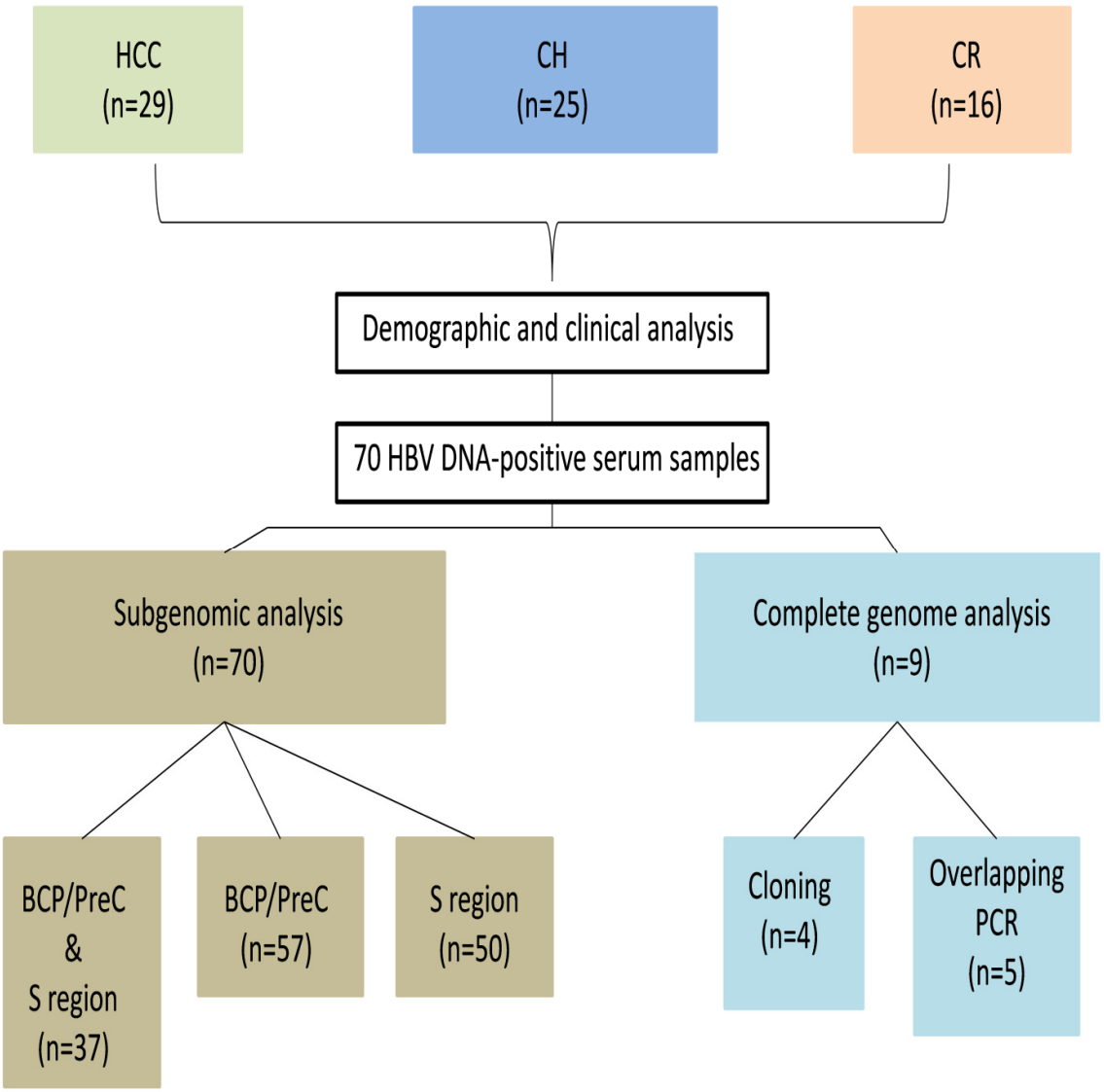



**Figure 3.1 Overview of the study layout and representation of the results.**

## 3.1 DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF PATIENTS

The demographic and clinical characteristics of the patients from which HBV was amplified and sequenced are shown in **Table 3.1**. The frequency of males was significantly higher in all three disease groups and patients with HCC were significantly older than those with CH ($p<0.05$). Thirty one percent of the cohort was HBeAg-positive, with no significant difference in the frequency of HBeAg-positivity between the three disease groups. HBeAg-positive patients, were significantly younger than HBeAg-negative individuals (32.1±17.9 vs 39.6±11.1 years, $p<0.05$). Another significant difference with respect to HBeAg-positivity was noted, that of ALT levels being higher in HBeAg-positive compared to HBeAg-negative patients (52.3 vs 37.4 IU/L, $p<0.05$).

**Table 3.1 Demographic and clinical characteristics of HBsAg-positive patients with different disease profiles**

| | Characteristic | Disease group | | | |
|---|---|---|---|---|---|
| | | Group I (HCC) n=29 | Group II (CR) n=16 | Group III (CH) n=25 | Total n=70 |
| Demographic and clinical data | Gender (M/F) | 21//8 | 13//3 | 17//8 | 52/18 |
| | Age(years)(Mean ±SD) | 48.70±10.94[a] | 40.68±11.52 | 26.28±11.10 | 40.87±14.66 |
| | ALT, IU/L | 73.50±11.65 | 70.50±41.09 | 56.32±30.01 | 67.45±37.74 |
| | HBeAg positive[b] | 7/22(32%) | 3/15 (20%) | 9/24 (37%) | 19/61 (31%) |

**HCC: Hepatocellular carcinoma; CR: Cirrhosis; CH: Chronic hepatitis**

[a] $p<0.05$ vs CH

[b] because of depletion of serum HBeAg status could only be determined for 22 HCC, 15 CR, 25 CHB

## 3.2 COMPLETE SURFACE REGION AMPLIFICATION AND SEQUENCE ANALYSIS

The region from nucleotide position 2410 to 1314 (numbering from *EcoR*I site, Genbank accession # AY 233289) was amplified using a nested PCR. This yielded a 2050 bp fragment, which included the entire S region (nt 2854 to 835). Primers were designed to ensure best possible amplification of both genotypes A and D. For both genotypes A and D, well conserved regions were found around positions 2410 and 1314 respectively, and this allowed for a longer fragment to be amplified. This longer fragment, together with the 1.6 kb fragment (nt 1096 to 2663 numbering from *EcoR*I site, Genbank accession # AY 233289), was subsequently also used to obtain full length HBV sequences.

Three overlapping primers, which were designed to ensure optimum binding to both genotypes A and D allowed the sequencing of three overlapping regions, with sufficient overlap to allow for the detection of deletion mutants. As illustrated in **Figure 3.2,** when sequenced using primer 3188F, sample INCHB170, showed the presence of a mixed population in the chromatogram (A), and when sequenced using the primer 591F, further upstream, a single population was present (B). The presence of the mixed population was as a result of a 21 nucleotide pre-S2 deletion mutant in the region sequenced by primer 3188F, nt position 35 – 55 (numbering from the *EcoR*I site).

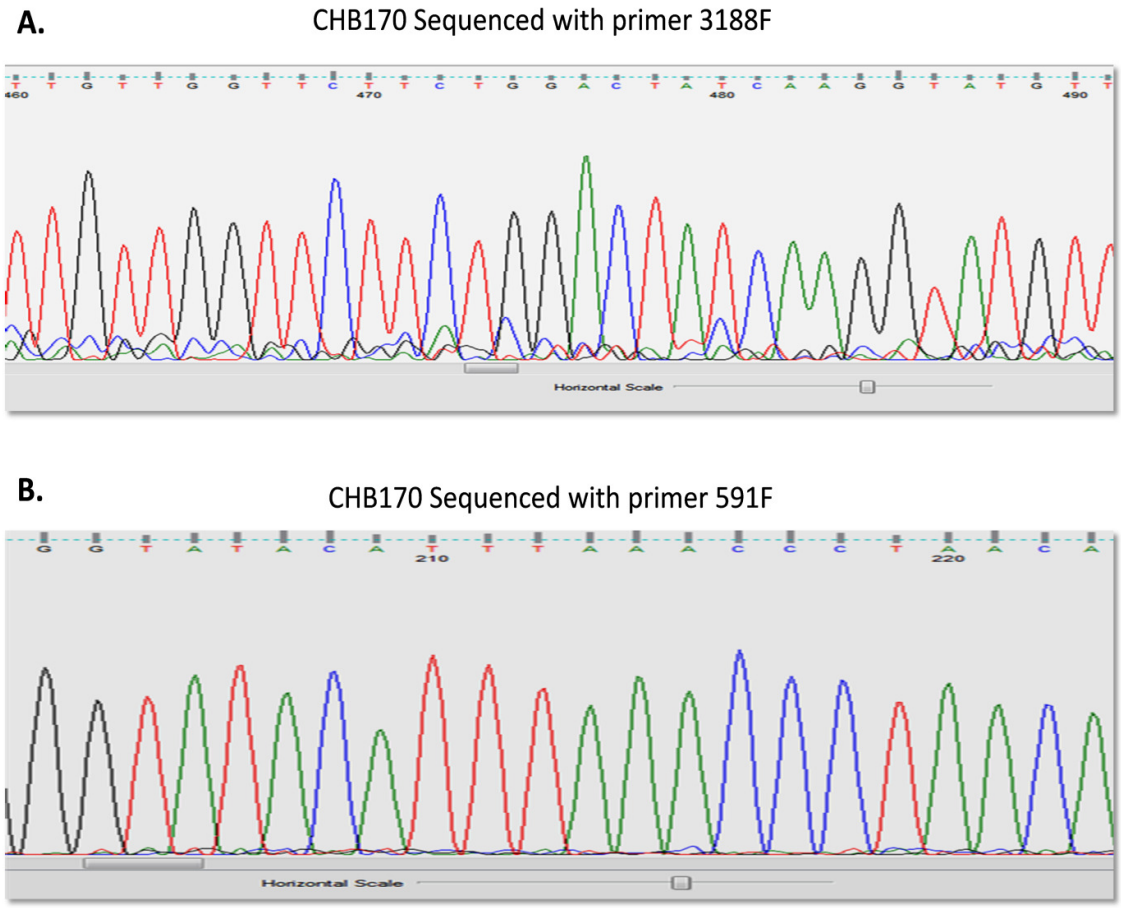



**Figure 3.2 Two of the three primers, 3188F and 2497F, used to sequence the complete S region.**

## 3.3 GENOTYPING USING PHYLOGENETIC ANALYSIS OF THE COMPLETE S REGION

The genotype and subgenotype of fifty HBV isolates were determined by phylogenetic analysis of the complete S region sequence: 32 belonged to genotype A (subgenotype A1:A2, 31:1) (**Figure 3.3**), 17 to genotype D (subgenotype D1:D2:D3, 4:12:1) (**Figure 3.4**), and 1 genotype C (subgenotype C1) (**Figure 3.5**). The genotype A strains belonged to serotype *adw*2 (84.4%) and *ayw*1 (15.6%). The genotype D strains were serotype *ayw*3 (58.8%), *ayw*2 (35.3%) and *adw*3 (5.9%). The single subgenotype C1 strain was *adr*. The serological subtypes of the HBV are

determined by analysis of amino acid substitutions at positions 122, 126 and 160 of the HBsAg. The 'a' determinant is common to all serological subtypes and is defined by a threonine or isoleucine at position 126 of HBsAg (Kramvis *et al.,* 2005). Amino acid substitutions at positions 122 and 160 of HBsAg are used to distinguish between serotypes *d/y* and *w/r,* respectively (Kramvis *et al.,* 2005).

The subgenotype A1 HBV isolates split into an 'African' and an 'Asian' cluster (Kramvis and Kew, 2007; Kramvis and Paraskevis, *in press*). The 31 subgenotype A1 isolates from Kerala clustered within the 'Asian' clade as a separate branch, together with sequences from Japan, India and the Philippines. The 31 isolates encoded the distinct subgenotype A1 amino acids, ps1: Q54, ps1: V74, ps1: A86, and ps1: V91 in the pre-S1 region, and ps2: L32 in the preS2 region (Bowyer *et al.,* 1997; Kimbi *et al.,* 2004). The majority of the isolates in the 'Asian' clade, including the Kerala isolates had [ps1: S5, ps1: S6]. The isolates in the 'African' clade displayed greater variation with one of the following [ps1:S5, ps1: S6], [ps1: S5, ps1: A6] or [ps1: 5L, ps1: P6].

A number of amino acids in the pre-S1 and pre-S2 regions differentiated the Kerala clade from the other Asian strains. The majority of the isolates from Kerala, 22/31 (71%), had ps1: V48 in the pre-S1 whereas V/I/N/T was found in the other clades (**Figure 3.3**). In the pre-S2, 28/31 (90%), had ps2:T7, whereas the other Asian strains had either ps2: T7 or ps2: A7. In contrast to the other Asian strains that had ps2: T37, the Kerala strains had ps2:N37, as did the strains in the 'African' clade. With regard to the pre-S2 region, the ps2: P54 was found in 90% of the Kerala strains, whereas the other clades had a higher diversity at this position. To define the

consensus sequence within the clades, a cut-off value of 60% amino acid sequence identity was used, as per Kramvis *et al.,* (2008).

Phylogenetic analysis of the 17 genotype D isolates is represented in **Figure 3.4.** The Keralite genotype D isolates had the characteristic 33 nucleotide deletion in the pre-S1 region. The pre-S2 signature amino acids, ps2: I42, ps2: S43, ps2: L54, were found in the Keralite strains belonging to subgenotype D1 (Yousif and Kramvis, 2012), which however, differed by having ps2: V39 instead of ps2: A39. With regard to the polymerase region, overlapping the pre-S1/pre-S2/S region, it was found to be relatively well conserved in genotype D isolates, as compared to subgenotype A1 isolates, which showed a higher diversity in this region.

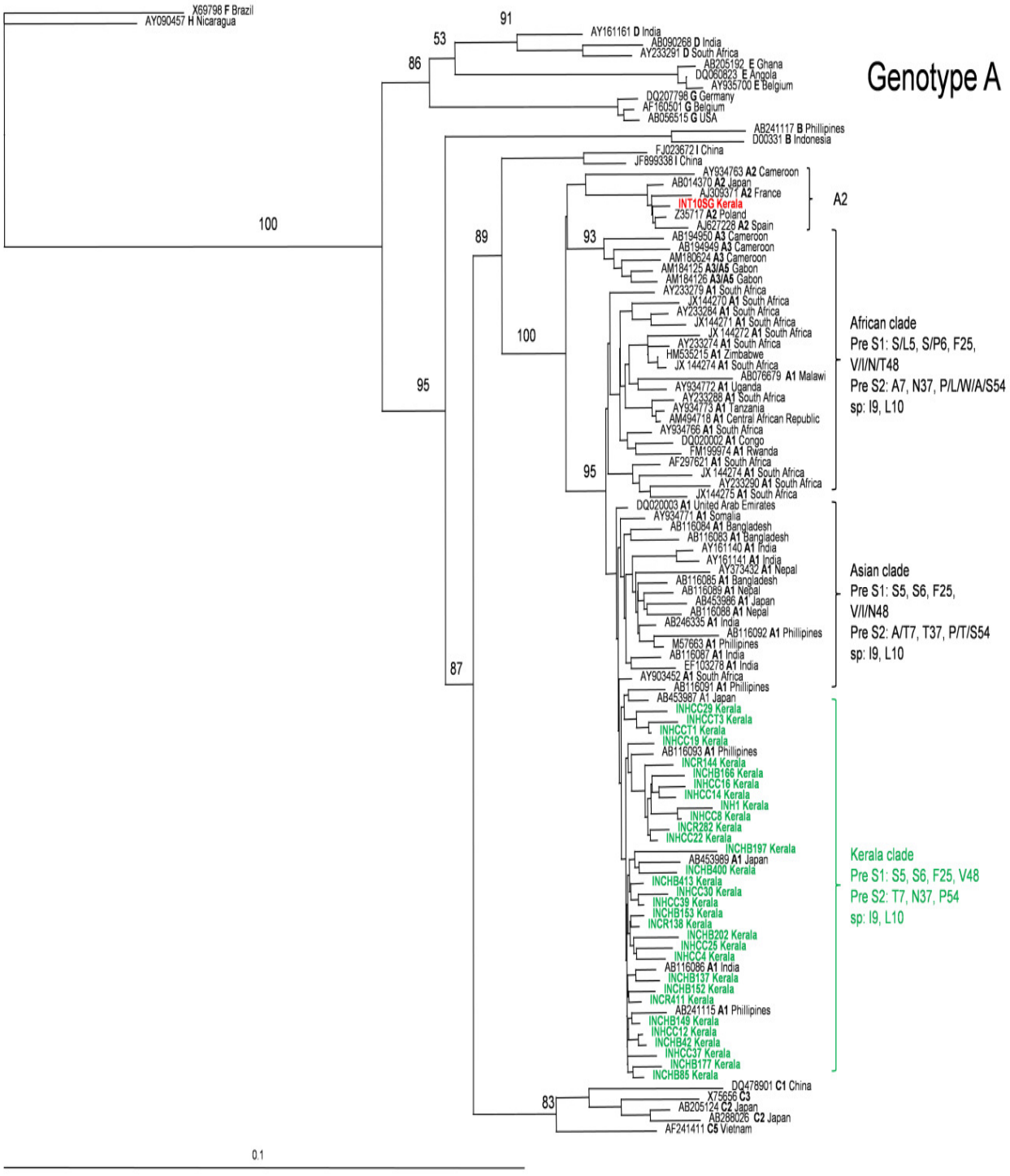



**Figure 3.3 Phylogenetic tree comparing the genotype A HBV isolates from India with genotype A sequences obtained from GenBank (designated by accession numbers)**

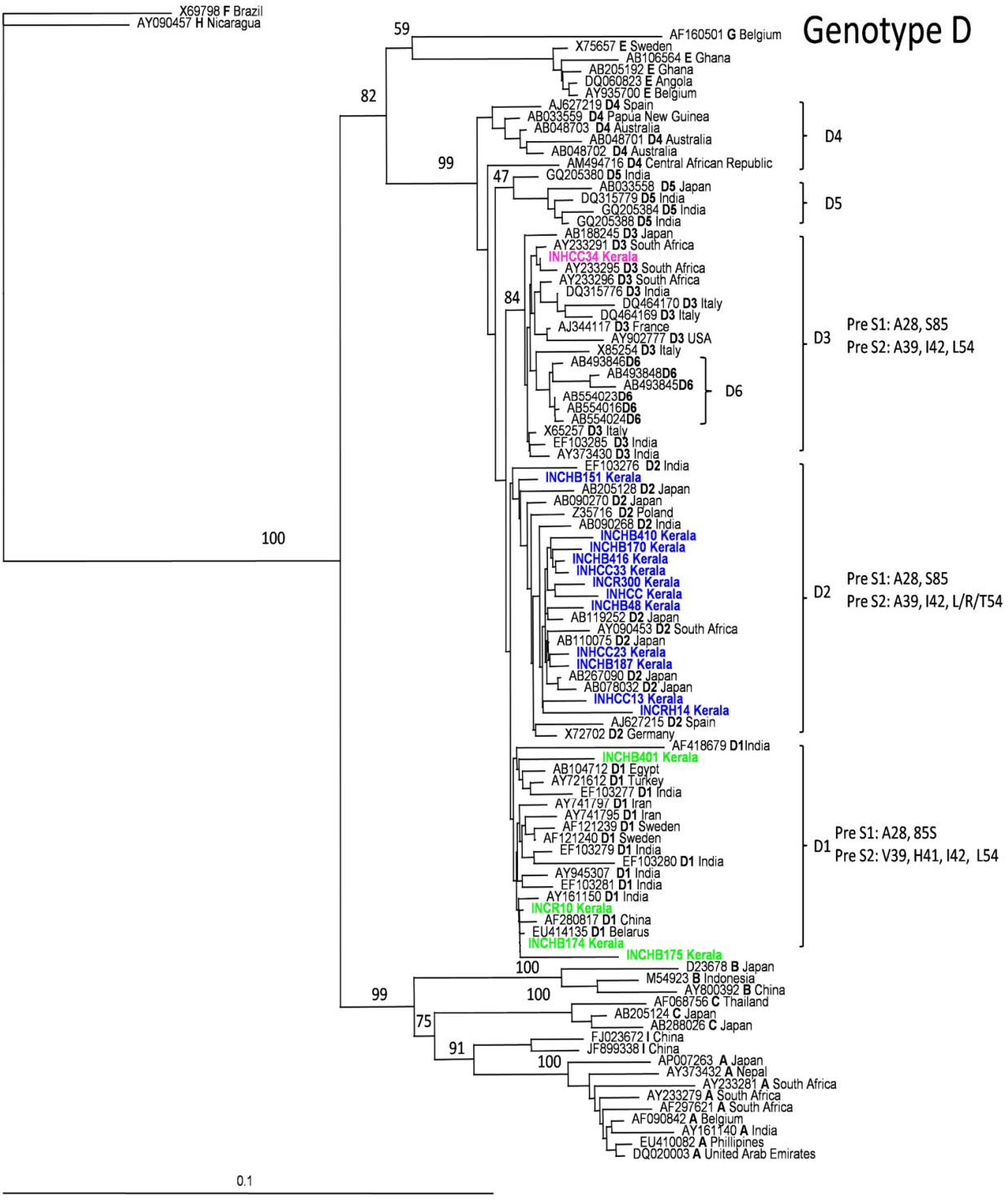



**Figure 3.4 Phylogenetic tree comparing the genotype D HBV isolates with genotype D sequences obtained from GenBank (designated by accession numbers)**

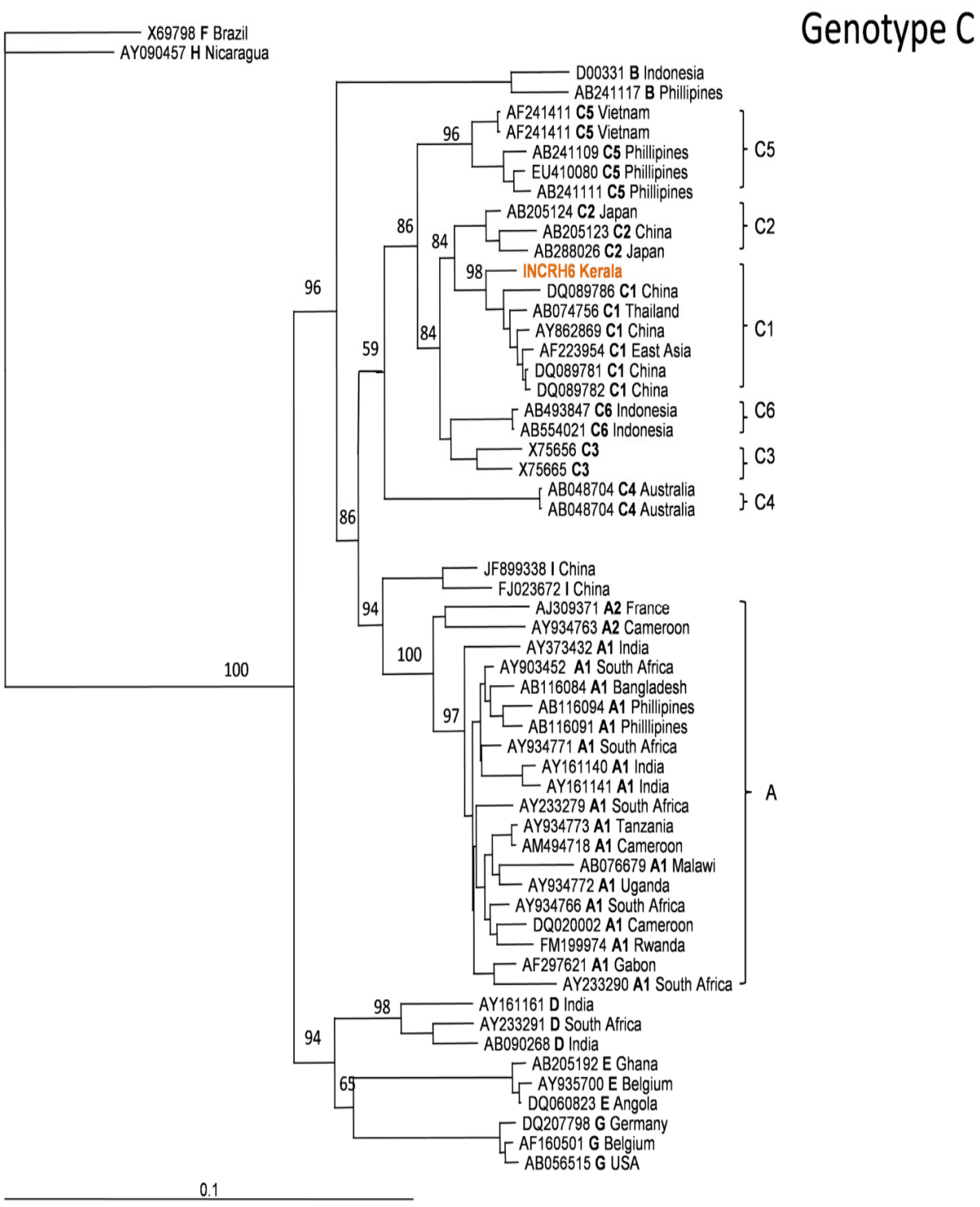



**Figure 3.5 Phylogenetic tree comparing the genotype C HBV isolate with genotype C sequences obtained from GenBank (designated by accession numbers)**

### 3.4. GENOTYPE DISTRIBUTION AMONG THE THREE DISEASE GROUPS

The majority of the patients were infected with genotype A (64%) (**Table 3.2**). A total of 17 (34%) patients were infected with genotype D, and a single patient was infected with genotype C (**Table 3.2**). In genotype A, subgenotype A1 was predominant (95.5%), with only a single subgenotype A2 isolate. Compared to other disease groups, HCC patients were predominantly infected with subgenotype A1 ($p<0.05$) **(Table 3.2)**. HCC patients infected with subgenotype A1, were also significantly younger (44.7±7.9 years), than HCC patients infected with genotype D (56.1±10.3 years) ($p<0.05$). This significance was also present when comparing only HBeAg-negative HCC patients, in which it was once again found that those patients infected with subgenotype A1 (44.1±8 years), were significantly younger than those infected with genotype D (53.0±8.8 years) ($p<0.05$). Among the genotype D isolates, three subgenotypes were identified, D1 (23.5%), D2 (70.6%), and D3 (5.9%), with subgenotype D2 prevailing across all disease groups.

**Table 3.2 Genotype and subgenotype distribution among the disease groups**

| | | Disease group | | | |
|---|---|---|---|---|---|
| | | **Group I (HCC) n=20** | **Group II (CR) n=9** | **Group III (CH) n=21** | **Total n=50** |
| **Genotype** | **Genotype A** | 15 (75%)[a] | 5 (55.6%) | 12 (57.1%) | 32 (64%) |
| | Subgenotype A1 | 14 (93.3%) | 5 (100%) | 12 (100%) | 21 (95.5%) |
| | Subgenotype A2 | 1 (6.7%) | 0 | 0 | 1 (4.5%) |
| | **Genotype C** | 0 | 1 (11.1%) | 0 | 1 (2%) |
| | **Genotype D** | 5 (25%) | 3 (33.3%) | 9 (42.9%) | 17 (34%) |
| | Subgenotype D1 | 0 | 1 (11.1%) | 3 (33.3%) | 4 (23 .5%) |
| | Subgenotype D2 | 4 (75%) | 2 (22.2%) | 6 (66.6%) | 12 (70.6%) |
| | Subgenotype D3 | 1 (25%) | 0 | 0 | 1 (5.9%) |

**HCC: Hepatocellular carcinoma; CR: Cirrhosis; CH: Chronic hepatitis**

[a] $p<0.05$ vs CH

### 3.5 PREVALENCE AND PATTERN OF COMPLETE S REGION MUTATIONS

In the present study, eight pre-S2 deletion mutants (16%) were detected. In addition, a single pre-S1 deletion (nucleotide position 2854-2872), which obliterated the pre-S1 start codon, was found (**Table 3.3**). The mean age of the patients, with and without pre-S deletions did not differ significantly (35.4±5 vs 37.0±15.8 years, p=0.78). No difference between ALT levels and the presence or absence of deletions was noted (58.2±20.8 vs 60.0±28.2, p=0.88). There was also no significant association between the presence of pre-S deletions and the HBeAg status. A higher frequency of pre-S deletion mutants was observed in HCC patients infected with subgenotype A1, although this did not reach statistical significance (p=0.4). All the pre-S2 deletions spanned the region, amino acid position 4 to 22 (nucleotide position 2-56, numbering from *EcoR*I) (**Table 3.3 and Figure 3.6**), which is a highly immunogenic region, and harbours both B and T cell epitopes (Park *et al.,* 2000).

## Table 3.3 Analysis of the nine pre-S deletion mutants: Clinical and Virological characteristics

| | | Clinical Characteristics | | | | | Virologic Characteristics | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sample # | Age/Sex | Clinical status | Genotype | HBeAg | ALT | Region | Nucleotide position [a] | Deletion size | Position [c] | Functional implication [b] |
| Genotype A | CHB137 * | 16/F | CHB | A1 | pos | 28 | pre-S1 | 2854-2872 (18) | 6 | 1 to 6 | pre-S2 initiation codon abolished |
| | CHB202 | 33/M | CHB | A1 | neg | 55 | pre-S2 | 28-51 (24) | 8 | 13 to 21 | B and T cell epitopes, pHSA binding site |
| | CHB413 * | 35/M | CHB | A1 | pos | 67 | pre-S2 | 50-55 (6) | 2 | 21 to 22 | B and T cell epitopes |
| | HCC30 * | 32/M | HCC | A1 | neg | - | pre-S2 | 30-53 (24) | 8 | 13 to 21 | B and T cell epitopes, pHSA binding site |
| | HCC37 * | 30/F | HCC | A1 | pos | 72 | pre-S2 | 24-56 (33) | 11 | 11 to 22 | B and T cell epitopes, pHSA binding site |
| | HCC39 | 55/M | HCC | A1 | - | 63 | pre-S2 | 24-56 (33) | 11 | 11 to 22 | B and T cell epitopes, pHSA binding site |
| | HCCT3 | 38/M | HCC | A1 | - | - | pre-S2 | 2-55 (54) | 18 | 4 to 22 | B and T cell epitopes, viral secretion, pHSA binding site |
| | HCC29 * | 50/F | HCC | A1 | neg | - | pre-S2 | 22-54 (33) | 11 | 11 to 22 | B and T cell epitopes, pHSA binding site |
| Genotype D | CHB170 * | 29/F | CHB | D2 | neg | 44 | pre-S2 | 35-55 (21) | 7 | 16 to 22 | B and T cell epitopes |

[a] nucleotide numbering from *EcoR*I site

[b] impact of deletion on viral function or host immune response; information compiled from Chen *et al.,* 2012

[c] amino acid position

* BCP/PreC region sequenced

pHSA, polymerised serum albumin

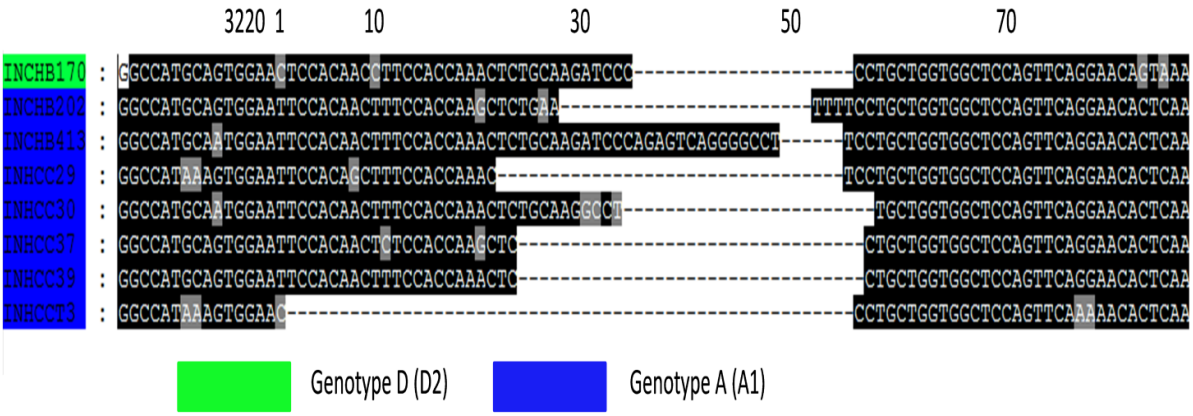



**Figure 3.6 Mapping of the pre-S2 deletions of the HBV isolates from Kerala, India.** Position 1 corresponds to the *EcoR*I cleavage site

In addition, amino acid changes in the complete surface region were analysed to determine any correlation between the presence of mutations and liver disease progression. Amino acid changes in the pre-S1, pre-S2 and surface antigen were detected (shown in red in **Tables 3.4 and 3.5**), with the majority showing no association with the different disease groups. However, two important amino acid substitutions were noted, the ps2: M1I/T and the ps2: F22L. The pre-S2 start codon mutation M1I/T, which prevents initiation of translation of the MHBs protein, was found exclusively in subgenotype A1 isolates from HCC patients. A total of six subgenotype A1 (20%) isolates had the start codon mutation, and this was significantly associated with HCC as compared to other disease groups (p=0.0072) (**Table 3.4**). The ps2:F22L was detected in 11/50 cases (22%), with an equal distribution in genotypes A and D. In genotype D isolates, ps2: F22L occurred at a higher frequency in HCC patients compared to genotype D, non-HCC patients ($p<0.05$) (**Table 3.5**). Although the F22L also occurred at a higher frequency in genotype A from HCC patients compared to non-HCC patients, this did not reach statistical significance (p=0.6).

The HBsAg region remained the most well conserved region, with very few amino acid substitutions occurring, particularly in subgenotype A1 isolates. The amino acid substitutions in the pre-S region were mainly variations, which differentiated between subgenotypes and in some cases, differentiated between clades within subgenotypes, such as the 'African' and 'Asian' clades of subgenotype A1 (**Figure 3.3**).

**Table 3.4 Amino acid changes in the complete S region of genotype A HBV isolates and the single genotype C isolate**

| | | Surface Region | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PRE-S1 | | | | | | PRE-S2 | | | | | | S GENE | | | | |
| | | S5 | S6 | F25 | I48V | A86T/V | T91V/A | M1T/I | A7T | F22L | N37 | A53V | P54 | S45P/T/A | L49R/H | I68T | K122R | Y161F |
| Subgenotype A1 | INCHB137SG_Kerala | S | S | F | D | A | V | M | T | F | N | A | P | S | L | I | K | Y |
| | INCHB149SG_Kerala | S | S | F | V | A | V | M | T | F | N | A | P | P | L | I | K | Y |
| | INCHB152SG_Kerala | S | S | F | V | A | V | M | T | F | N | A | P | S | L | I | K | Y |
| | INCHB153SG_Kerala | S | S | F | V | A | V | M | T | F | N | A | L | S | L | I | R | Y |
| | INCHB166SG_Kerala | S | S | F | I | A | A | M | T | F | N | A | P | P | L | I | K | Y |
| | INCHB177SG_Kerala | S | S | F | V | A | V | M | T | F | N | A | P | S | L | I | K | Y |
| | INCHB197SG_Kerala | S | S | F | V | A | V | M | T | F | N | A | P | P | R | I | K | Y |
| | INCHB202SG_Kerala | S | S | F | V | A | V | M | T | F | N | A | P | S | H | I | R | Y |
| | INCHB400SG_Kerala | S | S | F | V | A | V | M | T | L | N | A | P | P | R | I | K | Y |
| | INCHB85SG_Kerala | S | S | F | V | A | V | M | T | F | N | A | P | S | L | I | K | Y |
| | INCR138SG_Kerala | S | S | F | V | A | V | M | T | F | N | A | P | T | L | I | R | Y |
| | INCR144SG_Kerala | S | S | F | I | T | A | M | T | L | N | A | P | T | L | I | K | Y |
| | INCR282SG_Kerala | S | S | F | I | A | A | M | T | F | N | A | P | S | L | I | K | Y |
| | INCR411SG_Kerala | S | S | F | V | A | V | M | T | F | N | A | P | S | L | I | K | Y |
| | INH1SG_Kerala | S | S | F | I | A | A | M | T | L | N | V | P | S | L | T | K | Y |
| | INHCC12SG_Kerala | S | S | F | V | T | V | T | T | F | N | A | P | S | L | I | K | Y |
| | INHCC16SG_Kerala | S | S | F | I | T | A | M | T | F | N | A | P | P | L | I | K | Y |
| | INHCC19SG_Kerala | S | S | F | V | A | V | M | T | L | N | V | P | S | L | I | K | Y |
| | INHCC14SG_Kerala | S | S | F | I | A | A | M | T | I | N | A | P | P | L | I | K | Y |
| | INHCC22SG_Kerala | S | S | F | I | A | A | M | T | F | N | A | P | P | L | I | K | Y |
| | INHCC25SG_Kerala | S | S | F | V | A | V | T | T | L | N | A | P | S | L | I | R | Y |
| | INHCC29FL_Kerala | S | S | F | V | A | V | I | A | DEL | N | A | P | A | L | I | K | F |
| | INHCC30SG_Kerala | S | S | F | V | A | V | M | T | DEL | N | A | P | S | L | I | K | Y |
| | INHCC37SG_Kerala | S | S | F | V | V | V | M | T | DEL | N | A | P | S | R | I | K | F |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | INHCC39SG_Kerala | S | S | F | V | A | V | M | T | DEL | N | A | P | A | L | I | K | F |
| | INHCC42SG_Kerala | S | S | F | V | T | V | T | T | F | N | A | P | S | L | I | K | Y |
| | INHCC4SG_Kerala | S | S | F | V | A | V | M | T | L | N | A | P | A | L | I | R | Y |
| | INHCC8SG_Kerala | S | S | F | I | A | A | M | T | L | N | V | P | S | L | T | K | Y |
| | INHCCT3SG_Kerala | S | S | F | V | A | V | I | ? | DEL | N | A | P | S | L | I | K | F |
| | INHCCT1_Kerala | S | S | F | V | A | V | I | A | F | N | A | P | S | L | I | K | Y |
| | INCHB413SG | S | S | F | V | A | V | M | T | DEL | N | A | P | S | L | T | K | Y |
| Genotype C Subgenotype A2 | INT10SG_Kerala | S | S | F | I | T | I | M | A | S | N | V | T | S | L | I | K | Y |
| | INCRH6SG_Kerala | S | S | F | N | T | A | M | T | F | T | A | P | A | P | I | K | F |

Chronic Hepatitis B virus infection

Cirrhosis

Hepatocellular carcinoma

## Table 3.5 Amino acid changes in the complete S region of genotype D HBV isolates

| | | Surface Region | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | PRE-S1 | | PRE-S2 | | S GENE | | | |
| | | D42E | N102D | F22L | A39V | T118V | P127T | A128V | Y134A |
| Genotype D | INCHB174SGKerala D1 | D | N | F | V | T | P | A | Y |
| | INCHB175SGKerala D1 | E | N | F | V | T | P | A | A |
| | INCHB401SGKerala D1 | D | D | F | V | T | P | A | Y |
| | INCR10SGKerala D1 | D | D | F | V | T | P | A | Y |
| | INCHB170SGKerala D2 | D | N | DEL | A | V | T | V | Y |
| | INCHB410SGKerala D2 | D | N | F | A | V | T | V | Y |
| | INCHB416SGKerala D2 | D | N | F | A | V | T | V | Y |
| | INCHB48SGKerala D2 | E | N | F | A | V | P | V | A |
| | INCR300SGKerala D2 | D | D | F | A | V | T | V | Y |
| | INHCC23SGKerala D2 | D | N | L | A | V | T | V | Y |
| | INHCC32SGKerala D2 | D | D | L | A | V | T | V | A |
| | INHCC13SGKerala D2 | D | D | L | A | V | T | V | Y |
| | INHCC33SGKerala D2 | D | N | L | A | V | T | V | Y |
| | INCRH14SGKerala D2 | D | N | F | A | V | T | V | A |
| | INCHB151SGKerala D2 | D | N | F | A | T | T | V | Y |
| | INCHB187SGKerala D2 | E | N | F | A | V | T | V | Y |
| | INHCC34SGKerala D3 | D | N | F | A | T | P | A | Y |

### 3.6 ANALYSIS OF OVERLAPPING POLYMERASE REGION

The polymerase protein, which overlaps the complete surface gene, was also analysed for specific amino acid changes, particularly those that have been shown to confer resistance to the antiviral agents used to treat HBV infection. Only a single HBV isolate, INHCCT10 had the lamivudine resistance mutations, L180M, which is in the YMDD motif of the C domain of the reverse transcriptase, and the M204V,

The spacer domain, which overlaps both the pre-S1 and pre-S2 regions, was also analysed. The spacer domain, is the most tolerant to the presence of mutations in the polymerase protein, and hence the ability of the virus to withstand deletion mutations in the pre-S1 and pre-S2 regions. Amino acid changes and pre-S deletion mutants in the spacer domain were analysed, however no correlation with HBV genotypes or disease status was found.

### 3.7 BASIC CORE PROMOTER AND PRECORE REGION ANALYSIS

The BCP and the PreC region play a vital role in the replication of the HBV (Summers and Mason, 1982). Thus specific nucleotide changes in these two regions have been shown to influence HBV disease progression, as well as having a possible role in hepatocellular carcinogenesis.

#### 3.7.1 BCP/PREC REGION AMPLIFICATION AND SEQUENCE ANALYSIS

The region from nucleotide position 1653 to 1959 (numbering from the *EcoR*I site, Genbank # AY233289), which includes the entire BCP as well part of the PreC/C region, was amplified using a nested PCR. This yielded a 307 bp fragment, which was detected via ethidium bromide stained 2% agarose gel (**Figure 3.7**).

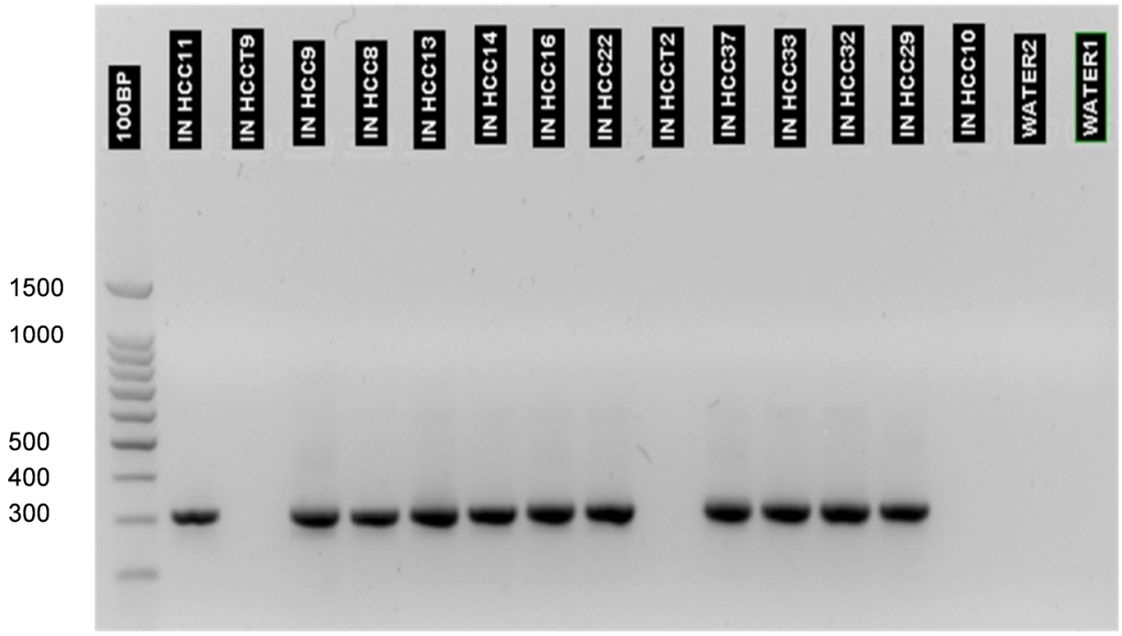



**Figure 3.7 Detection of the amplified BCP/PreC region amplicon, nt 1653 to 1959 (numbering from the *EcoR*I site, Genbank # AY233289), visualized on an ethidium bromide stained 2% agarose gel.**

A total of 57 HBV isolates were successfully amplified and sequenced. The chromatograms (**Figure 3.8**) and were manually aligned and edited, and further analysed using the Mutation Reporter Tool (Bell and Kramvis, *submitted*)

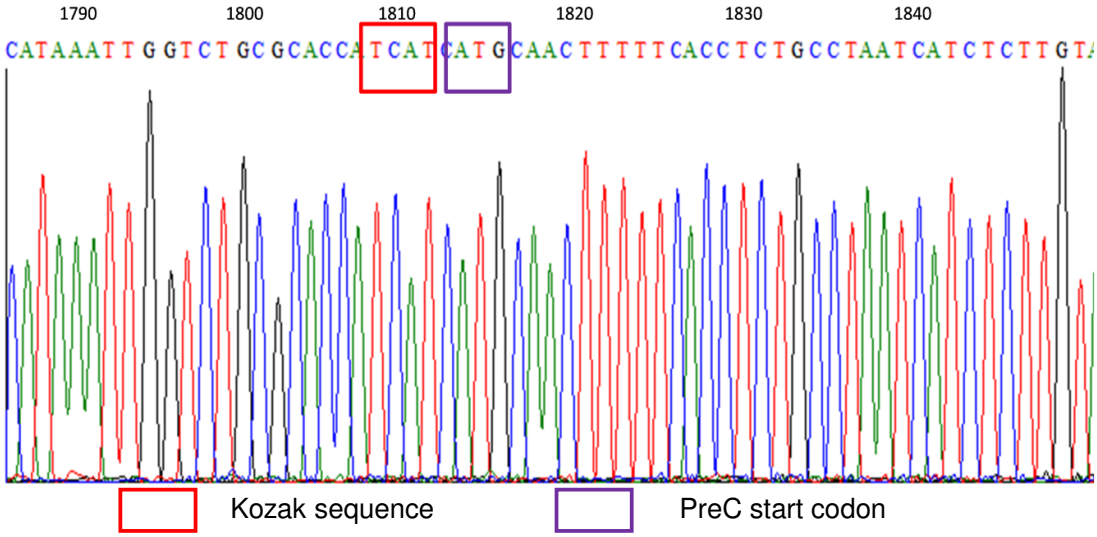



**Figure 3.8 Chromatogram sequence showing subgenotype A1 Kozak sequence and the PreC start codon**. Chromatograms were aligned and manually edited prior to analysis using the Mutation Reporter Tool (Bell and Kramvis, unpublished data).

### 3.7.2 BCP/PRECORE REGION MUTATION ANALYSIS

Detection of mutations in the BCP/PreC region was performed for 57 HBV isolates (40 Genotype A vs 17 Genotype D). There was no specific correlation between the presence of BCP/PreC mutations and either age or gender. Eighteen (33%) samples were HBeAg-positive, and 37 (67%) samples were HBeAg negative. The HBeAg status was not available for two samples. There was no significant difference in the frequency of HBeAg-negativity between genotypes, as well as between disease groups. HBeAg-positive and HBeAg-negative patients were analysed for the presence of BCP/PreC mutations, which can influence HBeAg expression and

synthesis. This may occur at three levels, transcriptional, translational, and post translational level.

Analysis of the fasta sequences was performed using the Mutation Reporter Tool (Bell and Kramvis, *submitted*). The Mutation Reporter Tool was used to generate the mutation distribution graphs, encompassing the BCP/PreC region (**Figures 3.9 and 3.10**). The mutation distribution graphs represent the percentage of mutant residues relative to the reference motif at fifteen loci of interest (1753, 1762, 1764, 1766, 1773, 1809-1812, 1814, 1858, 1862, 1888, 1896, 1899). The reference motifs used were TAGCTTCATACGGGG (Genotype A) and TAGCTGCACATGGG (Genotype D). The specific loci included in the analysis are the same as those described by Kramvis *et al.,* (2008). Nucleotide changes at these loci have an important role in genotype specific disease progression, as well as a possible role in liver disease progression and the development of HCC (Kramvis *et al.,* 2008; Baptista *et al*., 1999; Kao *et al*., 2003). The relative frequencies of nucleotide changes at these positions were determined, as well as specific nucleotide changes occurring in HBV isolates belonging to different genotypes and disease groups.

Examples of how the mutation distribution graphs are displayed using the Mutation Reporter tool are shown in **Figures 3.9 and 3.10**. The various nucleotide changes that may influence HBeAg status where analysed for both genotype A and genotype D isolates (**Figures 3.9 and 3.10**). Genotype D isolates, from HBeAg-negative patients had a higher occurrence of the G1896A as compared to isolates from HBeAg-positive patients (66% vs 0%) **(Figure 3.10).** This nucleotide change is known to abolish HBeAg expression, particularly in genotype D isolates, however

this is not the case in genotype A isolates. The G1896A does not occur in genotype A isolates, and thus other nucleotide changes such as the A1762T/G1764A double mutation and the G1862T may influence HBeAg expression in patients infected with genotype A isolates **(Figure 3.9).**

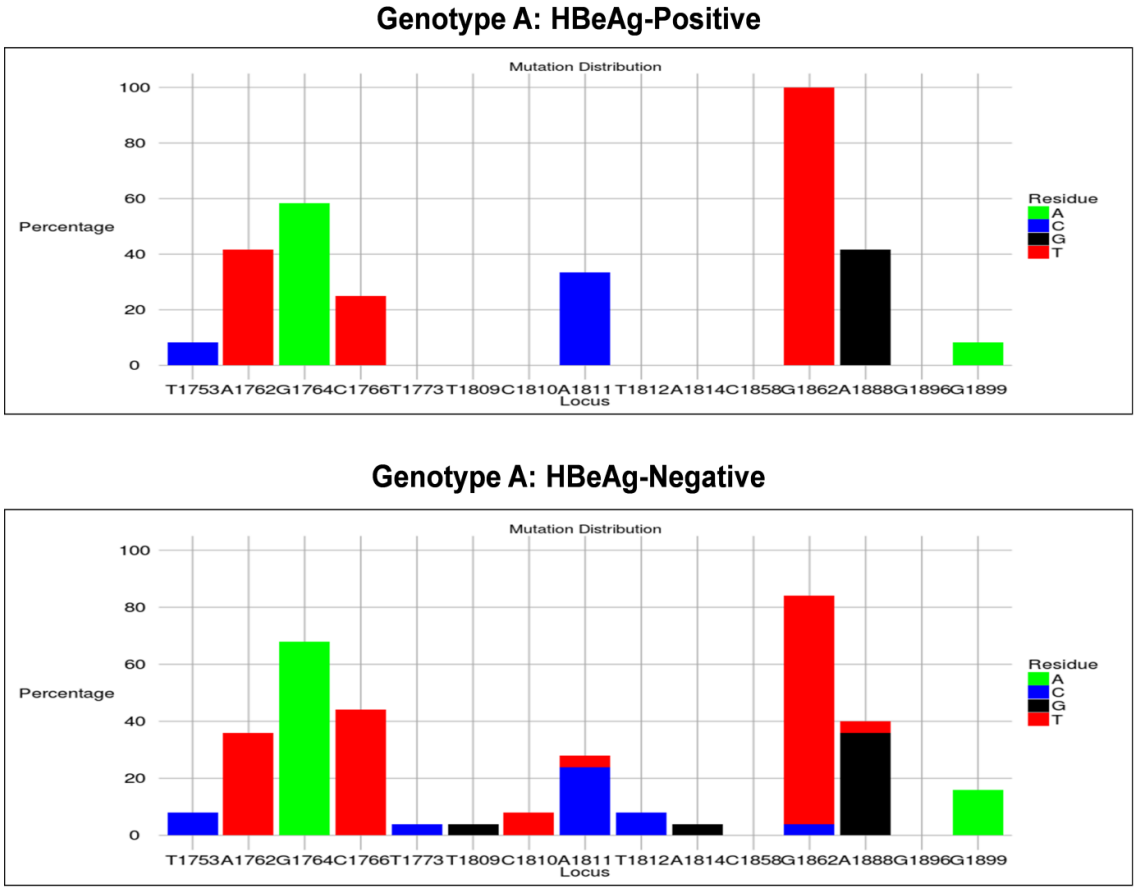
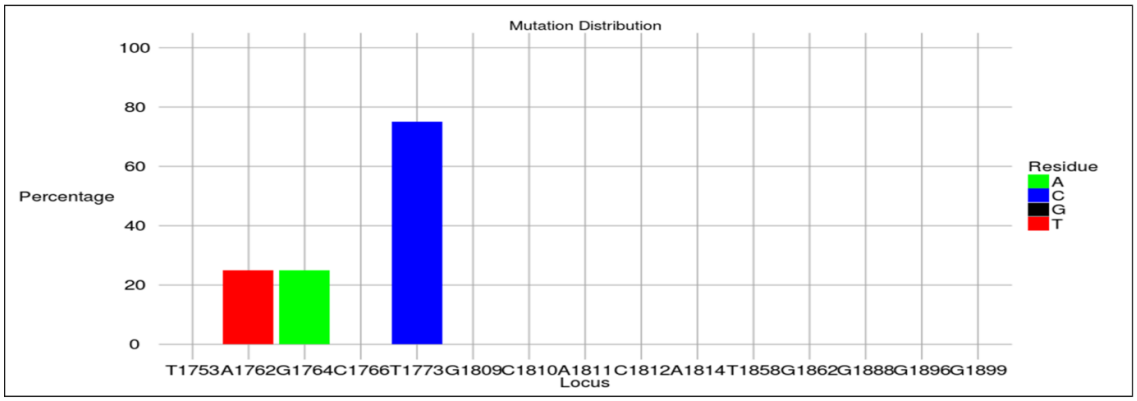



**Figure 3.9 Mutation distribution graphs comparing HBeAg-positive versus HBeAg-negative HBV Genotype A isolates.** The reference motif used was TAGCTTCATACGGGG (Genotype A)

**Genotype D: HBeAg-Positive**




**Genotype D: HBeAg-Negative**

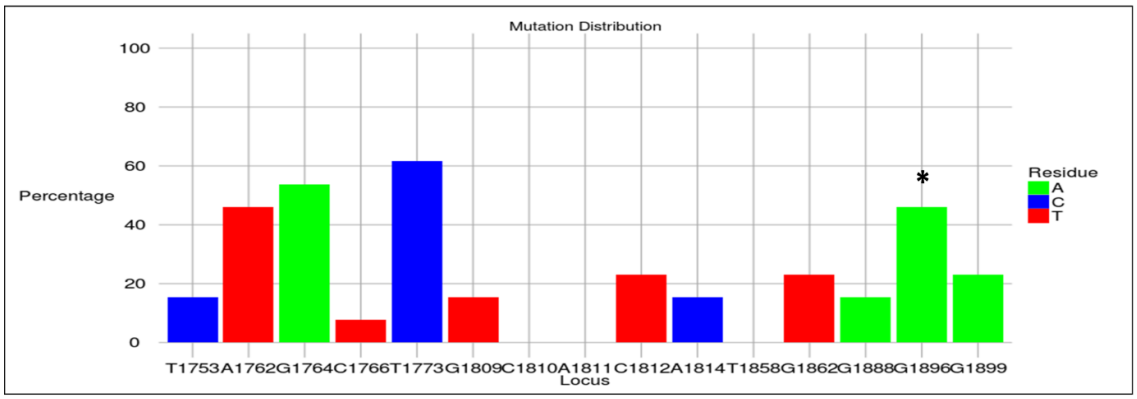



**Figure 3.10 Mutation distribution graphs comparing HBeAg-positive versus HBeAg-negative HBV Genotype D isolates.** The reference motif used was TAGCTGCACATGGG (Genotype D). (* $p<0.05$; HBeAg-negative vs HBeAg-positive)

The A1762T/G1764A double mutation had a higher occurrence in HBeAg-negative patients, irrespective of genotype or disease group, but this did not reach statistical significance. The double mutation A1762T/G1764A was found in 23 (46%), with the single mutation, G1764A, occurring in 10 (20%) of patients (**Table 3.8**). There was no significant difference between the frequency of A1762T/G1764A in genotypes A and D, as well as no significant difference between the disease groups. When combined, the single G1764A and the double A1762T/G1764A BCP mutations were

significantly associated with HCC in both genotypes A and D ($p<0.05$) (**Table 3.8**). The C1766T/T1768A found predominantly in subgenotype A1 was also significantly associated with HCC ($p<0.05$) (**Table 3.8**).

**Table 3.6 BCP/PreC nucleotide changes in genotype A HBV isolates: HCC vs Non-HCC patients**

| | | | | | BCP/PreC Region | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Sex | Age | HBeAg | 1753 | 1762/1764 | 1766 | 1768 | 1773 | 1809-1812 | 1814 | 1858 | 1862 | 1888 | 1896 | 1899 |
| HCC Genotype A | HCC12 | M | 60 | Neg | T | AG | T | A | T | TCAT | A | C | T | G | G | G |
| | HCC14 | M | 42 | Neg | T | AG | T | A | T | TCAT | A | C | T | G | G | G |
| | HCC15 | M | 45 | Pos | T | AG | T | A | T | TCCT | A | C | T | G | G | G |
| | HCC16 | F | 41 | Neg | T | AA | T | A | T | TCAT | A | C | T | A | G | G |
| | HCC17 | M | 32 | Pos | T | TA | C | T | T | TCAT | A | C | T | G | G | G |
| | HCC19 | M | 42 | Pos | T | AG | T | A | T | TCAT | A | C | T | G | G | G |
| | HCC22 | M | 32 | Neg | T | AA | T | A | T | TCAT | A | C | T | A | G | G |
| | HCC25 | M | 55 | Neg | T | AA | T | A | T | TCTT | A | C | T | T | G | G |
| | HCC29 | F | 2 | Neg | C | TA | C | T | T | TCAC | A | C | T | G | G | G |
| | HCC30 | M | 32 | Neg | T | AA | C | T | T | TCAT | A | C | G | A | G | A |
| | HCC36 | M | 65 | - | T | TA | C | T | T | TTAT | A | C | G | A | G | G |
| | HCC37 | F | 30 | Pos | C | TA | C | T | T | TCAT | A | C | T | A | G | G |
| | HCC6 | M | 44 | Neg | T | TA | C | T | T | TCCT | A | C | T | G | G | A |
| | HCC7 | M | 43 | Neg | C | TA | C | T | T | TCAT | A | C | T | G | G | G |
| | HCC8 | M | 57 | Neg | T | AA | T | A | T | TCCT | A | C | T | G | G | A |
| | HCC9 | F | 44 | Neg | T | TA | C | T | T | TCAT | A | C | T | A | G | G |
| | INHCC24 | F | 42 | Neg | T | AA | T | A | T | TCAT | A | C | T | A | G | G |
| | HCC3 | M | 62 | Pos | T | TA | C | T | T | TCCT | A | C | T | A | G | G |
| Non-HCC genotype A | CHB149 | M | 12 | Pos | T | TA | C | T | T | TCAT | A | C | T | A | G | G |
| | CHB137 | F | 16 | Pos | T | TA | C | T | T | TCAT | A | C | T | A | G | G |
| | CHB152 | F | 8 | Pos | T | AG | C | T | T | TCAT | A | C | T | A | G | G |
| | CHB153 | M | 9 | Pos | T | AG | C | T | T | TCAT | A | C | T | A | G | G |
| | CHB165 | M | 35 | Neg | T | AG | C | T | T | TCAT | A | C | T | A | G | G |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | CHB166A | M | 40 | Neg | T | AA | T | A | T | TTCT | A | C | T | A | G | G |
| | CHB177 | F | 10 | Pos | T | AG | C | T | T | TCAT | A | C | T | A | G | G |
| | CHB182 | M | 33 | - | T | AG | C | T | T | TCAT | C | C | T | A | G | G |
| | CHB241 | M | 48 | Pos | T | AG | C | T | T | TCAT | A | C | T | A | G | G |
| | CHB400 | M | 26 | Neg | T | AA | T | A | T | TCAT | A | C | T | A | G | G |
| | CHB413 | M | 35 | Pos | T | TA | C | T | T | TCCT | A | C | T | G | G | G |
| | CHB42A | M | 40 | Neg | T | AG | T | A | T | TCAT | A | C | T | G | G | G |
| | CR138 | M | 43 | Neg | T | TA | C | T | T | TCAT | A | C | T | A | G | G |
| | CR144 | M | 46 | Pos | T | TA | C | T | T | TCAT | A | C | T | A | G | G |
| | CR282 | F | 21 | Neg | T | AA | T | A | T | TCCT | A | C | T | A | G | G |
| | CR294 | M | 35 | Neg | T | TA | C | T | T | TTCT | A | C | T | A | G | G |
| | CR298 | F | 52 | Neg | T | AG | C | T | T | TCAT | A | C | C | A | G | G |
| | CR311 | M | 52 | Neg | T | AG | C | T | T | TCAT | G | C | G | A | G | G |
| | CR382 | M | 31 | Neg | T | TA | C | T | T | TCAT | A | C | G | A | G | G |
| | CR399 | M | 46 | Neg | T | TA | C | T | T | TCCT | A | C | T | G | G | G |
| | CRH1 | M | 57 | Pos | T | AA | T | A | T | TCCT | A | C | T | G | G | A |
| | CR411 | M | 55 | Neg | T | AG | C | T | C | GCAC | A | C | G | G | G | A |

## Table 3.7 BCP/PreC nucleotide changes in genotype D HBV isolates: HCC vs Non-HCC patients

| | | | | BCP/PreC Region | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Sex | Age | HBeAg | 1753 | 1762/1764 | 1766 | 1768 | 1773 | 1809-1812 | 1814 | 1858 | 1862 | 1888 | 1896 | 1899 |
| HCC Genotype D | HCC11 | M | 43 | Neg | T | AA | T | A | T | GCAT | A | T | T | A | A | G |
| | HCC13 | M | 65 | Neg | C | TA | C | T | C | GCAC | A | T | G | G | A | A |
| | HCC23 | M | 55 | Pos | T | TA | C | T | C | GCAC | A | T | G | G | G | G |
| | HCC32 | M | 55 | Neg | T | TA | C | T | C | GCAC | A | T | G | G | A | G |
| | HCC33 | F | 46 | Neg | T | AG | C | T | C | GCAC | A | T | G | G | G | G |
| | HCC34 | F | 56 | Neg | T | TA | C | A | T | TCAT | A | T | T | A | G | G |
| Non-HCC genotype D | CHB151 | M | 12 | Pos | T | AG | C | T | C | GCAC | A | T | G | G | G | G |
| | CHB170 | F | 29 | Neg | T | TA | C | T | C | GCAC | A | T | G | G | G | G |
| | CHB174 | F | 24 | Pos | T | AG | C | T | C | GCAC | A | T | G | G | G | G |
| | CHB180 | M | 29 | Neg | T | AG | C | T | C | GCAC | A | T | G | G | G | G |
| | CHB187 | M | 22 | Neg | T | AG | C | T | C | GCAC | A | T | G | G | A | A |
| | CHB401 | M | 25 | Neg | T | AG | C | T | T | GCAC | C | T | G | G | G | G |
| | CHB48 | F | 19 | Neg | T | AG | C | T | T | GCAC | A | T | G | G | G | G |
| | CR10 | M | 22 | Pos | T | AG | C | T | T | GCAC | A | T | G | G | G | G |
| | CR224 | M | 26 | Neg | C | TA | C | T | C | GCAC | A | T | G | G | A | A |
| | CR300 | M | 31 | Neg | T | TA | C | T | C | GCAC | A | T | G | G | A | G |
| | CR306 | F | 42 | Neg | T | AG | C | T | T | TCAT | C | T | T | G | G | G |

**Table 3.8 Multiple logistic regression analysis for mutations independently associated with HCC development among chronic HBV infection**

| | | HCC | | Non HCC (CR & CH) | | OR(95% CI) | *P* |
|---|---|---|---|---|---|---|---|
| | | Gen A N=18 | Gen D N=6 | Gen A N=22 | Gen D N=11 | | |
| Basic Core Promoter / Pre core Region | A1762T/G1764A | 8 (44) | 4 (66) | 8 (36) | 3 (27) | - | NS |
| | A1762T/G1764A and G1764A only | 14 (78) | 5(83) | 12 (54) | 3(27) | 4.56(1.2-18.23)[a] | 0.01[a] NS[d] |
| | C1766T/T1768A | 9(50) | 1(5) | 5(23) | 0 | 4(0.99-16.96)[a] 8.62 (1-192)[d] | 0.02[a] 0.02[d] |
| | 1773T | 16 (89) | | 22(100) | | | |
| | 1773C | | 4 (66) | | 7 (63) | | |
| | G1862T | 16(89) | 2(33) | 18 (82) | 1(9) | 26.44(4.87-168.89)[d] | NS[a] 0.0001[d] |
| | G1896A | 0 | 3(50) | 0 | 3(27) | 11(0.79-245) | NS[a] 0.02[c] 0.0002[d] |
| | A1762T/G1764A + G1862T | 12 (67) | 1(2) | 11 (50) | 0 | 4(1.02-16.39)[b] | NS[a] 0.02[b] |
| | G1764A only + G1862T | 5 (27.8) | 1 (16.7) | 3 (13.6) | 0 | - | NS |

[a] comparison between HCC and non-HCC, all genotypes
[b] comparison between HCC and non-HCC, restricted to genotype A isolates
[c] comparison between HCC and non-HCC, restricted to genotype D isolates
[d] comparison between genotype A and genotype D isolates

The G1862T, which interferes with post-translational modification of the HBeAg-precursor and reduces HBeAg expression, was significantly associated with subgenotype A1 (occurring 90% in HCC, and 82% in non-HCC) (**Table 3.6**). When looking at subgneotype A1, there was no significant difference between the occurrence of the G1862T from HBeAg-positive versus HBeAg-negative patients (100% vs 80%) (**Table 3.6**). Taken together with the A1762T/G1764A, again no specific correlation between HBeAg status could be made. There was no significant difference between the frequency of this nucleotide change and the different disease groups (HCC vs non-HCC). However, in combination with A1762T/G1764A, G1862T was significantly associated with HCC in patients infected with subgenotype A1 ($p<0.01$) (**Table 3.8**).

The 1773T was characteristic of genotype A, and the 1773C of genotype D, with no significant difference in the frequency of these nucleotide changes in the three disease groups. These specific nucleotide changes at position 1773, are thus genotype specific, and have no influence on disease progression.

Twenty isolates (16 genotype A; 4 genotype D) in which the BCP/PreC region was analysed, but were not successfully genotyped using the complete S region phylogenetic analysis, were genotyped/subgenotyped according to criteria of Kramvis *et al.,* (2008).

### 3.8 COMPLETE HBV GENOME ANALYSIS

Full length sequences of HBV isolated from nine patients were obtained: five from HCC patients, three from CHB patients, and a single full length HBV sequence from a CR patient. The full length sequences provide a few representative full length sequences from Kerala. No complete HBV genome sequences are currently available from this region. Two methods were used to generate the full length sequences; full length amplification followed by cloning, and a newly designed overlapping subgenomic fragment PCR technique.

The cloning method was performed by PCR amplification of the complete 3.2 kb genome of the HBV according to the methodology used by Gunther *et al*. (1995), with the PCR product being cloned into the pSMART-cDNA vector. Four full length sequences generated using the cloning method included, INCHB153, INCHB170, INHCC29 and INHCC23.

Because of the difficulty of the full length PCR technique and the low yield of successful PCR amplification, a new overlapping PCR fragment technique was designed. This included the amplification and sequencing of an additional region of the HBV genome (nt 1096 to 2660 from *EcoR*I), not covered by the ~ 2 kb amplicon (nucleotide position 2410 to 1314 from *EcoR*I site), used for the complete S region analysis. Sequencing of these two amplicons allowed for the complete genome to be sequenced. HBV full genomes generated by this technique were, for isolates INCH177, INCR10, INHCCT1, INHCC12, and INHCC13, which had relatively lower HBV DNA levels.

### 3.8.1 OPTIMIZATION OF NEW PCR FOR COMPLETE GENOME ANALSYSIS

A new 1.6 kb fragment PCR was designed to amplify the subgenomic region not covered by ~ 2 kb PCR amplicon (nt 2410 to nt 1314), which included the complete surface region. The two PCR amplicons used for the overlapping fragment technique are shown in **Figure 3.11**.

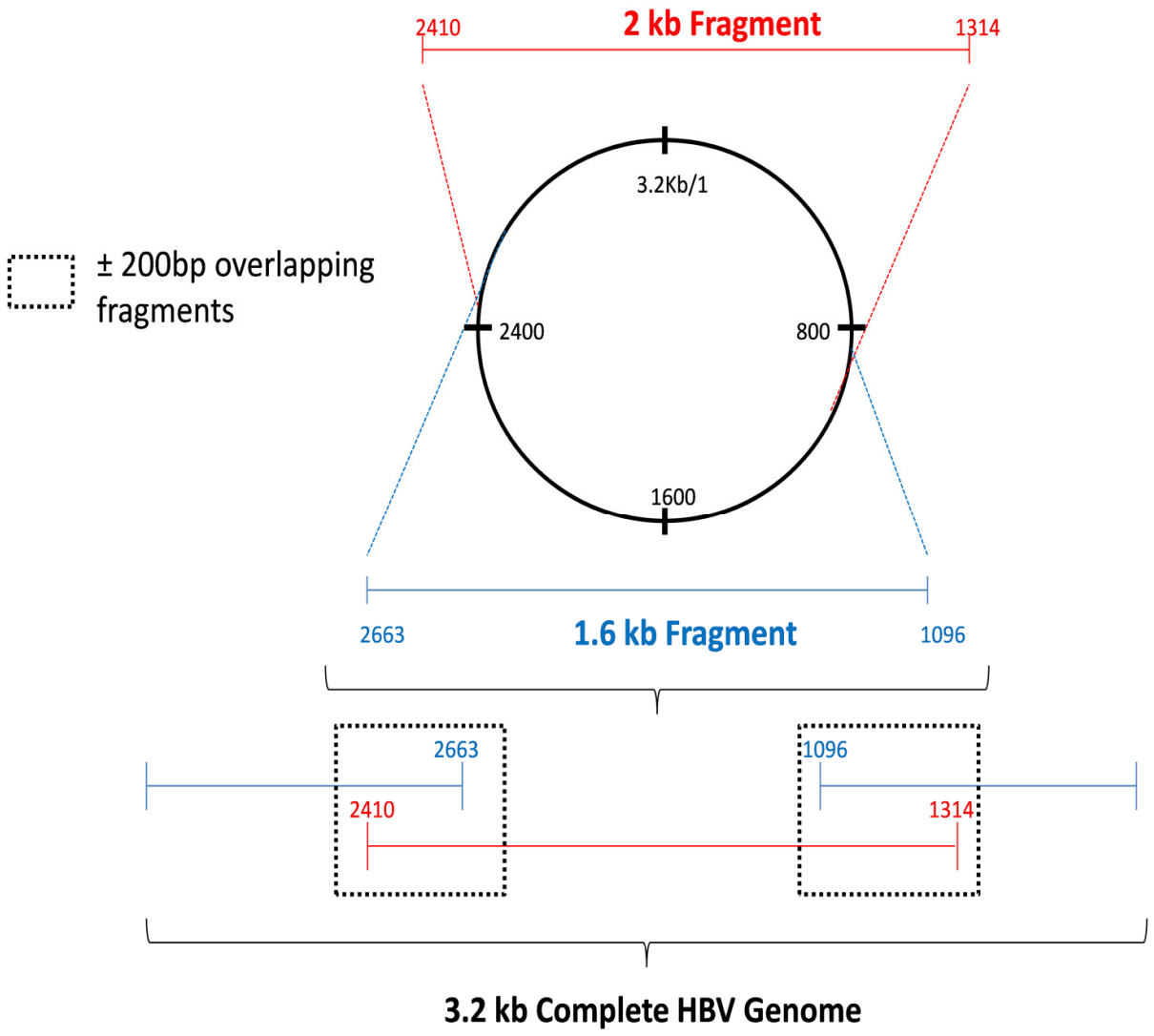



**Figure 3.11 Overlapping fragment technique used for complete HBV genome analysis**

The 1.6 kb fragment was designed to ensure sufficient overlap with the PCR amplicon used for the complete S region analysis. Successful amplification of the 1.6 kb PCR fragment involved optimization of five important steps; design of specific PCR primers, single vs nested PCR, annealing temperatures, polymerase enzyme analysis, and sequencing primer design.

Newly designed primers specific for both genotypes A and D were used for the amplification. This was done using published sequences from GenBank, in which conserved regions of both genotypes A and D were analysed for sequence similarity. Primer lengths were also optimized to ensure similar annealing temperatures. Mixed populations of sequences were used for 991F and 2699R PCR primers.

A nested PCR was used, as this allowed for greater amplification and ensured sufficient DNA for sequence analysis. A nested PCR was used, however during optimization, a semi-nested PCR technique was also explored, using the primer 2878R for both first and second round reactions. This was done, as this primer had the most specific sequence alignment for genotypes A and D. The nested PCR however had better amplification than the semi-nested reaction.

Annealing temperature was optimized using an annealing temperature gradient. This was done for both first and second round reactions. The annealing temperature gradient used for the first round PCR amplification is shown in figure. The best amplification was at an annealing temperature of 60.7℃, this amplicon was then subsequently used for the second round reaction (**Figure 3.12**). An annealing temperature gradient was also done for the second round reaction, whereby the best

annealing temperature was found to be 58℃. The annealing temperatures for the nested PCR were thus optimized at 60.7℃ and 58℃, respectively.

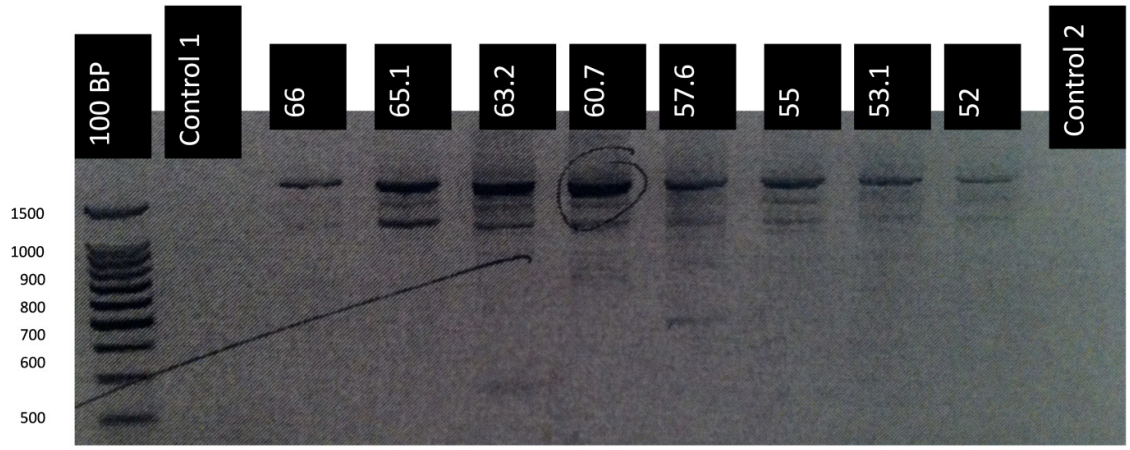



**Figure 3.12 Optimization of the 1.6 kb PCR reaction using different annealing temperatures.**

Different Taq polymerase enzymes and reagents were also compared during the optimization process. Three different enzymes where analysed, HotStarTaq DNA polymerase (Qiagen, Germany), BioTaqDNA polymerase (Biogen, USA), and the Expand High Fidelity PCR system (Roche), together with different $Mg^{2+}$ ion concentrations. Comparison between HotStarTaq DNA polymerase (Qiagen, Germany) and BioTaqDNA polymerase (Biogen, USA) is shown in **Figure 3.13**. A similar result was obtained when comparing the Expand High Fidelity PCR system (Roche), and hence it was found the HotStarTaq DNA polymerase (Qiagen, Germany) was the most efficient enzyme for the PCR amplification, and was used for all subsequent reactions (**Figure 3.13**).

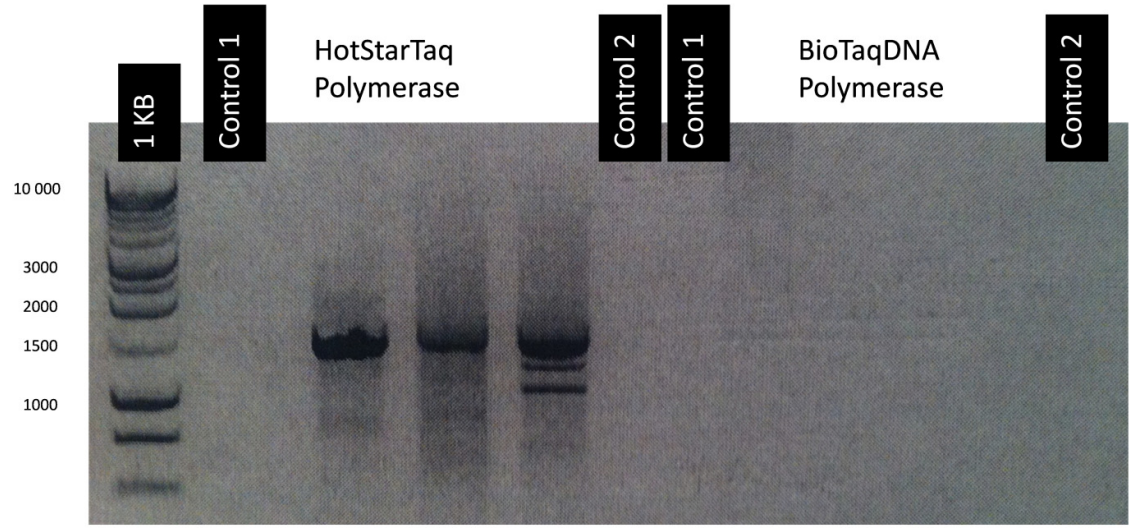



**Figure 3.13 Optimization of DNA polymerase enzymes: HotStarTaq vs BioTaqDNA polymerase.** Controls 1 and 2 represent negative controls for both first and second round amplifications.

Primers used for sequencing were also designed in the same manner as the primers for the PCR amplification, with specificity for both genotypes A and D. Three primers were used to sequence the 1.6 kb amplicon, 1096F, 1580F, 2058F (numbering from the *EcoR*I site), which ensured sufficient overlap to allow merge the three fragments into the complete 1.6 Kb sequence. This was then used in conjunction with the ~ 2 kb sequence, to generate the full length HBV genome (**Figure 3.11**).

### 3.8.2 PHYLOGENETIC ANALYSIS OF THE COMPLETE HBV GENOME

The nine complete HBV genomes together with complete HBV sequences from Genbank were used to generate the full length phylogenetic tree. Four HBV isolates belonged to genotype D (D1:1; D2:3), and five to genotype A (subgenotype A1) (**Figure 3.14**). The same genotype distribution was obtained when the four open reading frames were analyzed individually. Moreover, this genotype distribution is the same obtained using phylogenetic analysis of the corresponding complete S region (**Figures 3.3 and 3.4**).

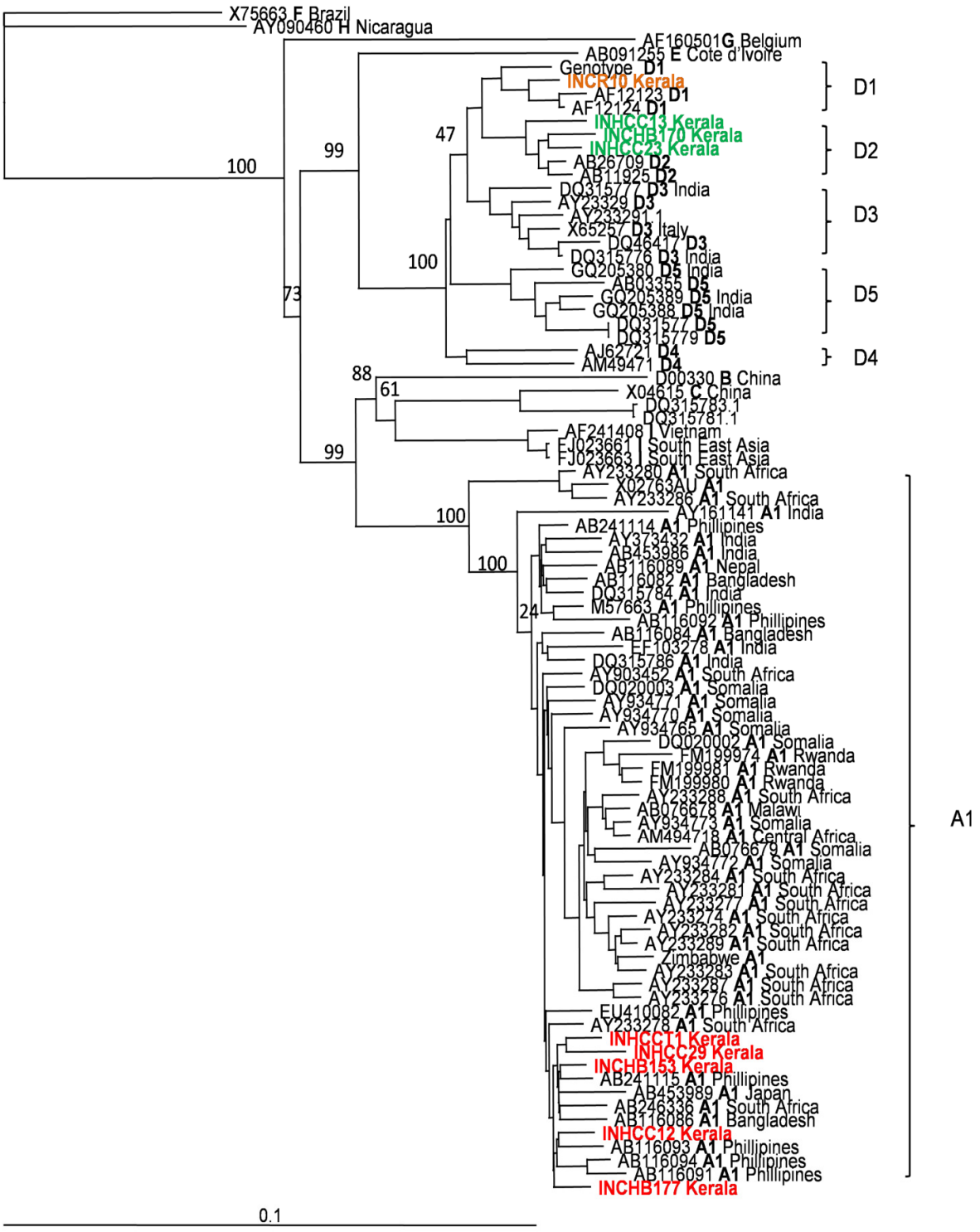



**Figure 3.14. Phylogenetic analysis of the nine complete HBV sequences from Kerala, India.**

### 3.8.3 COMPARATIVE BIOINFORMATIC ANALYSIS OF SUBGENOTYPE A1 ISOLATES FROM KERALA, INDIA

Five isolates, which had complete HBV genome sequences belonged to subgenotype A1, (INHCCT1, INHCC29, INCHB153, INHCC12, and INCHB177). The subgenotype A1 isolates were analyzed further using bioinformatics, to compare them to all subgenotype A1 complete HBV genome sequences available from Genbank.

Phylogenetic analysis of the complete genome sequences of the five Keralite subgenotype A1 isolates showed clustering within the 'Asian' clade (**Figure 3.15**). The 'Asian' clade separated into three branches, with the first branch exclusively containing sequences from Asian countries (*blue)*, a second branch including isolates from Asia, Africa and South America regions *(red)* and a third branch containing Somalian sequences *(green)* (**Figure 3.15**). The Kerala isolates were shown to cluster within the second branch with isolates from Asia, Africa and South America (**Figure 3.15**). Distinction between the subgenotype A1 clades was seen when complete HBV genome sequences, S ORF sequences, and Pol ORF sequences were used for phylogenetic analysis **(Figures 3.3, 3.15 and 3.16).** However, when the X ORF sequences **(Figure 3.16)** and the C ORF sequences (figure not shown) were analysed, this distinction into subgenotype A1 clades could not be determined.

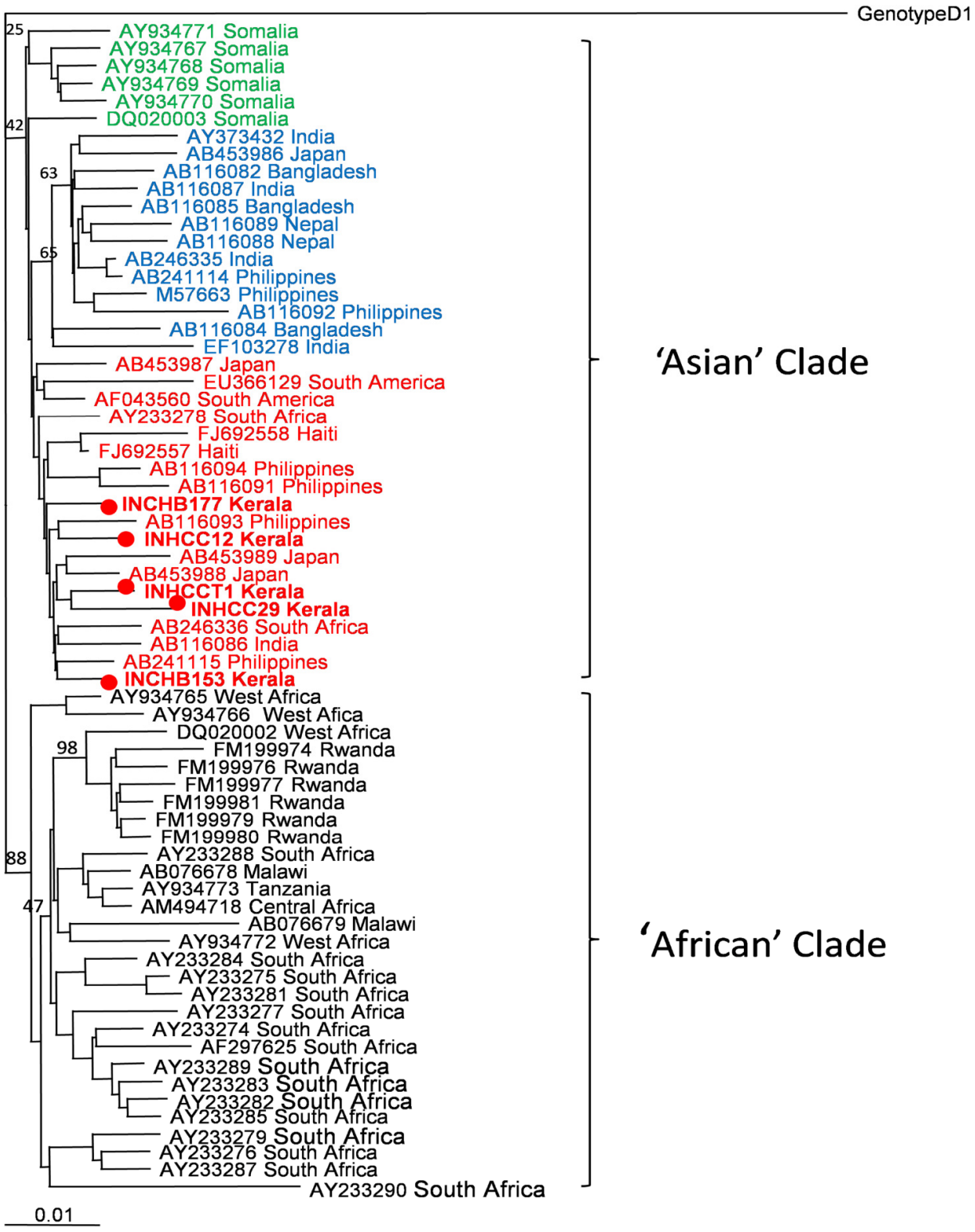



**Figure 3.15 Phylogenetic relationship of the complete genomes of subgenotype A1**

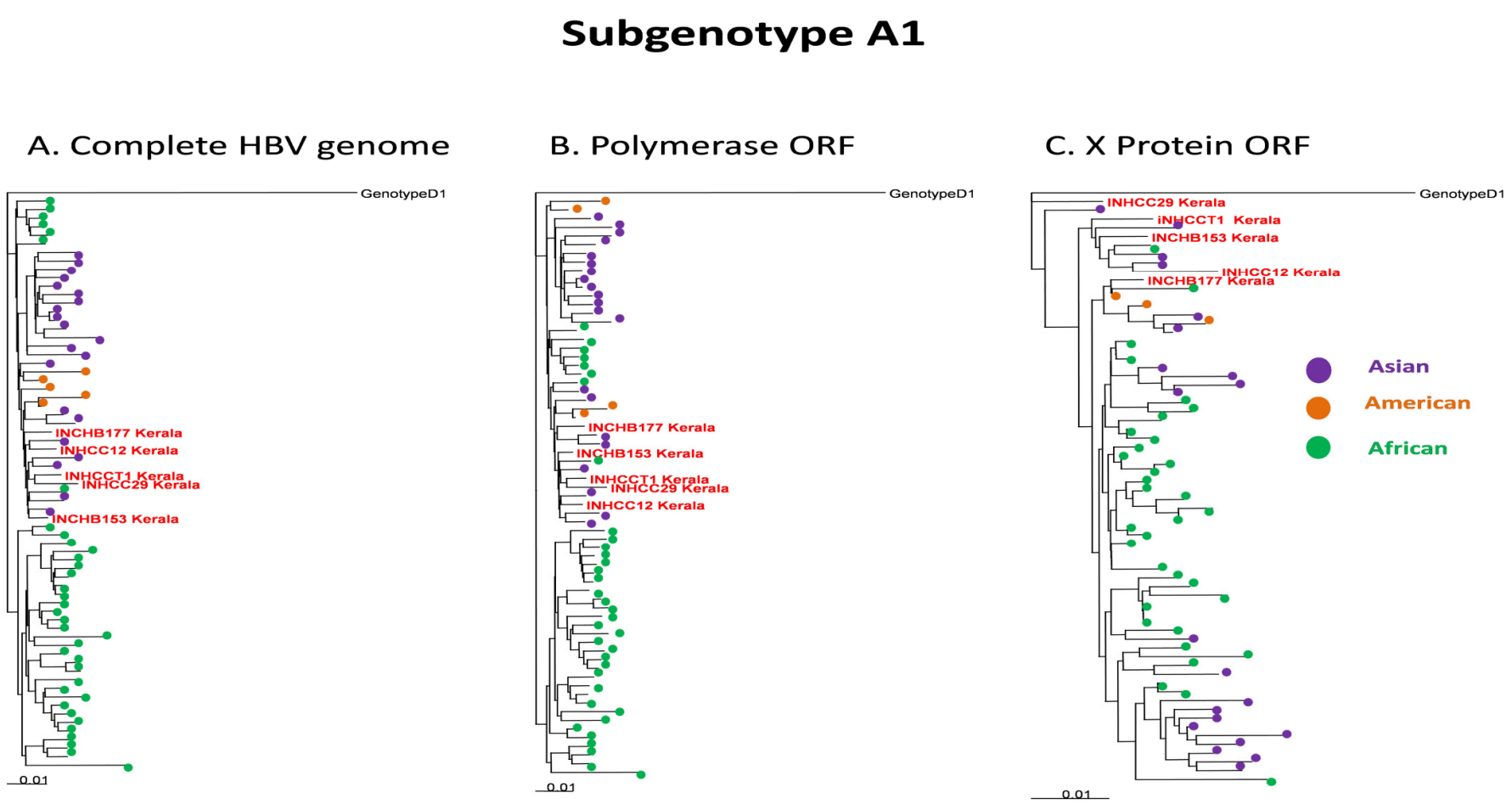



**Figure 3.16 Comparison of the complete HBV genome subgenotype A1 phylogenetic tree with the phylogenetic trees of the polymerase and X ORF's.**

# CHAPTER 4: DISCUSSION

Genotypes A and D are the predominant genotypes in India (Datta 2008). However, these findings were mainly based on studies conducted in the eastern and northern regions of the country (Datta 2008). Limited data is available from western and particularly southern regions of India. Thus in the present study, HBV isolates from Kerala, southern India, were used to investigate the distribution of HBV genotypes/subgenotypes from this region. Molecular characterisation of the HBV isolates, with a focus on the BCP/PreC and S regions was also performed. The genotypes/subgenotypes, BCP/PreC and S region mutations and complete genomes were then compared to those reported in other areas of India, as well as isolates from southern Africa and South America.

## 4.1 DESIGN AND OPTIMIZATION OF EXPERIMENTS

Two experiments were designed and optimized for the present study. These included the PCR amplification, for the sequencing of the complete S region, and the amplification of an overlapping fragment in order to obtain the sequence of the complete HBV genome.

The S region PCR was designed to serve a two-fold function: amplification and sequencing of the complete S region of genotype A and D isolates; and to be of sufficient length ~2 kb to ensure overlap with the 1.6 kb fragment, allowing for generation of complete HBV genome sequences (**Figure 3.11**). The use of this overlapping technique served as a means of overcoming the difficulty associated with the full length amplification described by Gunther *et*

*al*. (1995) and downstream cloning process. Complete genome amplification and cloning was used to generate full length sequences for four isolates; INCHB153, INCHB170, INHCC29 and INHCC23. The inability to amplify additional isolates using this method could be as a result of low titre and/or the presence of specific HBV variants, which escape detection (Gunther *et al.,* 1995).

The use of an overlapping fragment method to generate complete HBV genome sequences has been described (Takahashi *et al.,* 1998), however this method does not allow for analysis of the complete S region with a single PCR amplicon. A disadvantage of using the overlapping fragment method for complete genome analysis is that the sequences do not necessarily represent those of a single HBV isolate, but rather represent the dominant quasispecies.

The ~2 kb and 1.6 kb PCR's involved optimization of five steps; design of specific PCR primers, single versus nested PCR, annealing temperatures, polymerase enzyme analysis, and sequencing primer design. Correct and efficient PCR amplification requires specific primer binding to the sequence of interest. Reports have shown that different primer pairs for the same gene can exhibit up to 1000-fold differences in sensitivity (Sachse, 2004). Primers designed for both PCRs were aligned to genotype A and D sequences in Genbank to ensure greater specificity.

The importance of ensuring the correct annealing temperature is highlighted in **Figure 3.12**, whereby it was found that the using a temperature of ~ 60°C

enabled optimum amplification of the 1.6 kb fragment. The correct annealing temperature is the most important factor influencing the specificity of the PCR, and should be set within a few degrees of the melting temperature ($T_m$) (Sachse, 2004). Reactions run at temperature increments (2°C-5°C) starting at 5°C below the calculated $T_m$ will allow an approximation of the optimum annealing temperature (Roux, 2003).

Low yield and often no PCR amplicons were obtained when using BioTaqDNA polymerase (Biogen, USA) (**Figure 3.13**). HotStarTaq DNA polymerase (Qiagen, Germany), provided with an optimized master mix, showed a much higher yield (**Figure 3.13**) and was subsequently used for all reactions. The HotStarTaq DNA polymerase (Qiagen, Germany) utilizes the hot-start methodology, in which an initial denaturation step at 95°C for 15 minutes is performed. The hot-start PCR method prevents primer dimer formation and nonspecific priming, which can occur at incubations of a PCR mixture at temperatures below the $T_m$ (Roux, 2003), and thus improves the specificity of the PCR.

The sensitivity of the newly designed complete S region amplification was demonstrated when we used it to report the first known case of transfusion-transmitted HBV infection by blood screened using individual-donation nucleic acid testing (ID-NAT) by the South African National Blood Service (Vermeulen *et al.,* 2012). The viral burden in the infectious red blood cell unit was estimated at 32 (22 – 43) HBV DNA copies/ 20 ml of plasma (Vermeulen *et al.,* 2012). Moreover, it has recently been used by our group to detect occult

HBV infection in HIV infected South Africans (Bell *et al.,* 2012), as well as for the molecular characterization of isolates from these individuals (Makondo *et al.,* 2012)

## 4.2 GENOTYPING OF HBV ISOLATES FROM KERALA

Over the years numerous genotyping methods have been developed, including restriction fragment length polymorphism (RFLP), PCR using genotype specific primers, multiplex-PCR, hybridisation, serology based techniques, and the use of various bioinformatic tools and nucleotide analysis (Lindh *et al.,* 1997; McIver *et al.,* 2005, Laperche *et al.,* 2001; Usuda *et al.,* 1999; Kramvis *et al.,* 2008). Although these newer methods have become increasingly popular because of the ability to genotype a large number of samples in a relatively short duration, the classical method of complete HBV genome PCR amplification, sequencing and phylogenetic analysis, still remains the most specific method of HBV genotyping.

In the present study, complete S region PCR, sequencing and phylogenetic analysis was performed to determine the genotypes and subgenotypes. Of the 70 HBV DNA-positive serum samples, 50 HBV isolates were successfully genotyped and subgenotyped using complete S region sequencing and phylogenetic analysis. Phylogenetic analysis is used to assess the relative evolutionary relatedness of sequences to each other (McCormack and Clewley, 2002). The use of phylogenetic analysis for the genotyping and subgenotyping of HBV isolates is based on the use of specific algorithms used to determine the relatedness of sequences. These algorithms, which are

used in various phylogenetic software packages, enable the construction of phylogenetic trees, which display the evolutionary relatedness of the sequences (Bartholomuesz and Schaefer, 2004). The application of phylogenetics can be used in the analysis of subgenomic sequences of specific genes or regions, as well as complete genome sequences (McCormack and Clewley, 2002). Although subgenomic sequences are often used for genotyping of the HBV and in general correlate with complete genome sequence analysis (**Figure 3.3 and Figure 3.15**), the subgenomic region compared must be of sufficient length and diversity.

Phylogenetic analysis of the complete S region showed genotype A was the most prevalent genotype infecting liver disease patients in Kerala. The high prevalence of genotype A in our cohort, differs from previous studies from other geographical regions of India, which found genotype D to be the predominant genotype (Chattopadhyay *et al.,* 2006a; Vivekanandan *et al.,* 2004; Gandhe *et al.,* 2003), or occur at an equal prevalence as genotype A (Thakur *et al.,* 2002; Kumar *et al.,* 2005). Thirty two genotype A isolates were successfully subgenotyped. Thirty one (97%) isolates were subgenotype A1, and a single isolate was subgenotype A2. The high prevalence of subgenotype A1 is in keeping with other studies from India, which also found subgenotype A1 to be the predominant subgenotype of A (Sugauchi *et al.,* 2004; Tanaka *et al.,* 2004; Banerjee *et al.,* 2006). Subgenotype A2 which is found mainly outside Africa (Kramvis and Paraskevis, *in press*) has previously been described in eastern Indian blood donors and found to be restricted to the peripheral blood lymphocytes (Datta *et al.,* 2009). The subgenotype

distribution of HBV isolates has only been studied in eastern and northern regions of India. This study is the first to report the subgenotype distribution in Kerala, southern India, and is the first to sequence the complete HBV genome of subgenotype A1 isolates from this region.

The complete genome sequences and the complete S region sequences of the subgenotype A1 isolates from Kerala were found to cluster within the 'Asian' clade as a separate branch together with isolates from Asian, African and South American regions (**Figures 3.3 and 3.15**). Subgenotype A1 of HBV was the first subgenotype to be recognized, with unique molecular characteristics that differentiate it from subgenotype A2 (Bowyer *et al.,* 1997; Kramvis *et al.,* 2002). Subgenotype A1 is the dominant subgenotype A strain in Africa, and is also found in the Indian subcontinent and south America (Kramvis and Paraskevis, *in press*). These areas both have a history of migration events out of Africa, which could account for the high occurrence of subgenotype A1 in these regions. The major sea ports at Calicut and Cochin on the west coast of Kerala, were frequented by both the Dutch East India Company and the Portuguese colonists (Kurien, 1994), and this could explain the higher occurrence of genotype A isolates in Kerala, compared to other regions of India.

Genotype D has the shortest genome of 3182 nt, with a characteristic 33 nt deletion at the beginning of the Pre-S1 region (Kramvis *et al.,* 2005). Although previously, genotype D has been classified as having 8 subgenotypes, D1 to D8 (Abdou Chekaraou *et al.,* 2010), a recent analysis has suggested a

subgenotype classification of only six subgenotypes, which are geographically distributed (Yousif and Kramvis, 2012).

In the present study, 17/50 (34%) HBV isolates were genotype D. This is significantly lower than a previous study from Delhi, northern India, which reported genotype D as the most prevalent genotype at 68.3%, followed by genotype A 25.7% and genotype C 5.9% (Madan *et al.,* 2009). Studies from eastern, western and southern regions of India have also reported a predominance of genotype D (Biswas *et al.,* 2008; Vivekanandan *et al.,* 2004; Gandhe *et al.,* 2003).

The predominant subgenotype of D was D2, found in 70% of the patients, followed by D1 and D3. This is the first study to report the subgenotype distribution of genotype D from southern India, to report the relatively high prevalence of subgenotype D2 and to sequence the complete genome of subgenotype D2 from India. A recent study from western India, found D1 in 17%, D2 in 29%, D3 in 34% and D5 in 20% of the cases (Chandra *et al.,* 2009). Another study, which only had 5 genotype D isolates, found a single D1 isolate, two subgenotype D3 and two subgenotype D5 isolates (Banerjee *et al*., 2006). In a recent analysis of the geographical distribution of the six subgenotypes of D, subgenotype D2 was found to prevail in Malaysia, Japan, Russia, Estonia, Poland and the United Kingdom but not India (Yousif and Kramvis, 2012). When the subgenotype D2 isolates from Kerala were analyzed they were found to cluster with isolates from Japan (**Figure 3.4**) and to have amino acids residues characteristic of subgenotype D2 (**Table 3.5**).

Analysis of a larger cohort from southern India would be required to confirm the prevalence of D2 in southern India.

A single genotype C isolate was identified, and was subgenotyped as C1. The sequence was closely related to C1 isolates from China and Thailand, in agreement with a previous study from eastern India (Banerjee *et al*, 2006a) (**Figure 3.5**). Genotype C is found mainly in the eastern regions of India, with comparable proportions of genotypes A, D and C (Banerjee *et al.,* 2006a; Banerjee *et al.,* 2006b). The high sequence similarity with the South East Asian subgenotype C1 isolates could suggest a recent introduction into the eastern Indian population from Asia (Banerjee *et al.,* 2006a).

The HBV serological subtypes were determined via analysis of the translated HBsAg region using the amino acids at positions 122, 127, 134 and 160 (Kramvis *et al.,* 2008). The genotype A strains belonged to serotype *adw2* (84.4%) and *ayw*1 (15.6%), which is comparable to other studies (Kramvis and Kew, 2007). The genotype D strains were serotype *ayw*3 (58.8%), *ayw2* (35.3%) and *adw3* (5.9%), in agreement with previous findings from India (Chandra *et al.,* 2009). The single subgenotype C1 strain was *adr*.

## 4.3 GENOTYPE DISTRIBUTION AMONG DISEASE GROUPS AND INFLUENCE ON DISEASE PROGRESSION

A correlation between the HBV genotypes and severity of liver disease is becoming increasingly apparent. In the present study, patients infected with genotypes A and D did not differ from each other in terms of age, frequency of HBeAg-positivity and ALT levels. However, compared to cirrhotic and CHB

patients, a significantly higher proportion of HCC patients were infected with subgenotype A1. HCC patients infected with subgenotype A1 were also significantly younger than those infected with genotype D. This is in agreement with a previous study which showed that South African patients infected with subgenotype A1 had a 4.5 fold increased risk of developing HCC compared to patients infected with non-A genotypes and also developed cancer 6.5 years earlier (Kew *et al.,* 2005). The findings from this study and those from South Africa may mean an increased hepatocarcinogenic potential of subgenotype A1, regardless of host ethnicity.

## 4.4 ANALYSIS OF COMPLETE S REGION AND THE OVERLAPPING POLYMERASE REGION

A total of eight pre-S2 deletion mutants were detected, together with a single pre-S1 deletion (nucleotide position 2854-2872), which obliterated the start codon (**Table 3.3**). The pre-S deletion mutants were found predominantly in subgenotype A1 (88%), with a single subgenotype D2 isolate also having a pre-S deletion. Moreover, in the subgenotype A1 isolates, a higher frequency of pre-S deletion mutants was observed in HCC patients compared to other disease groups, although this did not reach statistical significance. This is the first study to describe the pre-S deletion mutants from Indian HCC patients infected with subgenotype A1 and is agreement with other studies that showed a strong correlation between genotype B and C pre-S mutants and the development of HCC (Takahashi *et al.,* 1998; Huy *et al.,* 2003; Sugauchi *et al.,* 2003). Another study from India found pre-S deletion mutants predominantly in genotype A and C, compared to genotype D, however they

did not find any correlation with the clinical outcome, with equal occurrence of pre-S deletion mutants in ASC, CLD and cirrhotic patients (Biswas *et al.,* 2012). Research from South Africa found 72% of HCC patients had HBV isolates with pre-S deletions mutants, and these mutants were found to be significantly associated with subgenotype A1 isolates (Skelton,2008). A recent study from South Africa also showed the presence of pre-S deletion mutants in subgenotype A1 isolates from patients with HBV-HIV co-infection (Makondo *et al.,* 2012).

All pre-S2 deletions occurred within the highly immunogenic region, which harbours both B and T cell epitopes (**Table 3.3**), and were comparable to findings from South Africa, where 48% of HBV subgenotype A1 isolates from HCC patients, also harboured deletions within this immunogenic region (Skelton, 2008). The T-cell-mediated immune response is critical in the control of HBV infection (Chang and Lewin, 2007), and hence deletions in this region may have an advantage for the viral pool to evade immune pressure and persist (Sobotta *et al.,* 2000). Other possible mechanisms that may be involved in pre-S deletions contributing to advanced liver disease, such as HCC, include, retention of mutant LHBs in the endoplasmic reticulum (ER), resulting in ER stress, and up-regulation of COX-2 and cyclin A, which induces uncontrolled cell cycle progression (Wang *et al.,* 2006). A recent study also found pre-S2 deletion mutants to be more oncogenic than pre-S1 deletion mutants (Lin *et al.,* 2012).

Nucleotide changes and amino acid substitutions in the complete surface region and overlapping polymerase were also analysed to determine any correlation with the disease groups and possible liver disease progression. Two important amino acid substitutions were noted, the ps2: M1I/T and the ps2: F22L. The pre-S2 start codon mutation M1I/T, which prevents initiation of the MHBs protein, was found exclusively in subgenotype A1 isolates from HCC patients. A total of six subgenotype A1 isolates had the start codon mutation, and this was significantly associated with HCC as compared to other disease groups ($p=0.0072$). The pre-S2 start codon mutation was found to be the predominant pre-S region mutation present in HBV sequenced in a recent study from eastern India (Biswas *et al.,* 2012), and was also more frequently found among HBV genotype A and C isolates as compared to genotype D isolates (Biswas *et al.,* 2012). The pre-S2 start codon mutation may thus have an important role in HBV induced liver disease and possibly the development of HCC, particularly in subgenotype A1 isolates. A possible mechanism in liver disease progression includes the altered ratio of the three surface protein expression caused by defective MHBs expression, which might lead to an increased accumulation of surface proteins, leading to increased cellular stress (Raimondo *et al.,* 2004; Choi *et al.,* 2007).

The ps2: F22L was found in both genotype A (22%) and D (24%) isolates, with the ps2: F22L being significantly associated with HCC in patients infected with genotype D ($p<0.0098$). Recent studies indentified the ps2: F22L as a risk factor for HCC among patients infected with genotype C (Mun *et al.,* 2011). Another study from India also showed a significant association of the

ps2: F22L with cirrhosis, and hence postulated a possible role in clinical liver disease progression (Biswas *et al.,* 2012). The possible mechanisms associated with this pre-S2 mutation and liver disease remain unknown, yet it would seem to have a possible role in hepatocarcinogenesis, particularly in genotype D.

The HBsAg region remained well conserved in all isolates examined, with few amino acid changes noted between different genotypes and disease groups. No vaccine escape mutants were detected, with the 'a' determinant being very well conserved. The overlapping polymerase region also remained highly conserved, with only a single HBV isolate, INHCCT10 having the lamivudine resistance mutations L180M and M204V. The former is in the YMDD motif of the C domain of the reverse transcriptase. These drug resistance mutations develop when the virus is placed under 'stress' as occurs during treatment regimens used in the treatment of HBV infection (Zoulim and Locarnini 2009).

**4.5 BCP/PREC REGION ANALYSIS**

Genotype distribution of the 57 BCP/PreC positive isolates was, 40 genotype A and 17 genotype D, and the majority were HBeAg-negative. The BCP/PreC mutations responsible for the higher rate of HBeAg-negativity differed between the two genotypes.

The higher frequency of HBeAg-negativity is an important feature of HBV carriers in South Africa (Dusheiko *et al.,* 1985), where subgenotype A1 predominates (Bowyer *et al.,* 1997; Kimbi *et al.,* 2004). In subgenotype A1,

nucleotide changes in the BCP/PreC region, including A1762T/G1764A, 1809-1812 Kozak sequence variants, and the G1862T, have possible roles in earlier HBeAg seroconversion (Kramvis and Kew, 2007). In subgenotype A1, no correlation between the A1762T/G1764A double mutation and HBeAg-negativity was found, which is in agreement with a previous study from South Africa (Baptista *et al.,* 1999). Previous reports from India also showed that the A1762T/G1764A was more common in HBeAg negative patients, but did not differentiate between genotypes (Chauhan *et al.,* 2006). The A1762T/G1764A BCP mutation may lead to reduced transcription of the precore mRNA, resulting in reduced HBeAg synthesis (Buckwold *et al.,* 1997).

G1764A occurred alone 20% of the isolates. When combining the single G1764A and the double A1762T/G1764A, in agreement with others (Baptista *et al.,* 1999; Kao *et al.,* 2003), a significant association with HCC was found (**Table 3.8**). However, studies from India have shown conflicting findings regarding the role of the A1762T/G1764A double mutation in liver disease progression and the development of HCC. One study from western India reported that the clinical manifestations were independent of the A1762T/G1764A mutations (Gandhe *et al.,* 2003), whereas another group from eastern India showed a significant association between the A1762T/G1764A mutations and patients with chronic liver disease (Biswas *et al.,* 2011). The A1762T/G1764A double mutation with the T1753C was found in 5 isolates, 3 genotype A and 2 genotype D (INHCC29, INHCC37, INHCC7, INHCC13, and INCR224). This combination of mutations has been described as a marker of HCC (Tanaka et al, 2003; Yeh *et al.,* 2010; Zhang *et al.,* 2010),

and was recently found to be more common in HBV isolates from cirrhotic patients in a study from eastern India (Biswas *et al.,* 2011). Although the A1762T/G1764A double mutation with the T1753C were found predominantly in HCC patients, and a single cirrhotic patient, no significant correlation with clinical disease was found. Analysis of the single T1753C mutation also showed no correlation with clinical liver disease, in keeping with a previous study from eastern India (Biswas *et al.,* 2011).

The C1766T/T1768A was found at a significantly higher frequency in subgenotype A1 isolates, however was not significantly associated with HCC in subgenotype A1 (**Table 3.8**). Analysis of the C1766T/T1768A across both genotypes A and D, found a significant association with HCC compared to non-HCC patients ($p<0.05$) (**Table 3.8**). No significant association between the C1766T/T1768A mutation and HBeAg-negativity was found. The C1766T/T1768A has been reported as an independent predictor of cirrhosis in HBeAg-negative patients, and is also associated with higher HBV replication rates by increasing the encapsidation of the pgRNA (Guo *et al.,* 2008; Baumert *et al.,* 1996). The T1768A mutation results in F132Y in the HBx and may play a synergetic role with K130M and V131I, introduced by A1762T/G1764A, possibly leading to increased hepatocarcinogenesis (Guo *et al.,* 2008).

The 1773T was characteristic of genotype A, and the 1773C of genotype D, with no significant difference in the frequency of these nucleotide changes in

the three disease groups. These specific nucleotide changes at position 1773, are thus genotype specific, and have no influence on disease progression.

In subgenotype A1, the consensus sequence of the Kozak sequence (positions nt 1809-1812), preceding the start codon of the precore, is TCAT (Baptista *et al.,* 1999; Kramvis *et al,* 2008). TCAT at this position has been shown to impair the translation of the HBeAg precursor, possibly by a leaky scanning mechanism (Ahn *et al.,* 2003). Further mutations within this region, such as TC**C**T, TC**T**T, T**T**AT and T**TC**T, which severely impair translation, were found predominantly in HBeAg-negative individuals (**Table 3.6**). When these mutations occur together with A1762T/G1764A they have been shown to reduce HBeAg to undetectable levels (Ahn *et al.,* 2003).

The G1862T was found in 90% of subgenotype A1 sequences analysed, which is much higher than the frequency of this mutation in South African subgenotype A1 isolates (Kramvis *et al*., 1997; Kramvis *et al.,* 1998; Tanaka *et al,* 2004). This high prevalence of G1862T has previously been described in subgenotype A1 isolates from India (Tanaka *et al.,* 2004; Chauhan *et al.,* 2006). The G1862T mutation results in a valine to phenylalanine substitution at the - 3 position to the signal peptide cleavage site, at the amino end of the precursor protein of HBeAg, leading to impaired processing of the HBeAg in the endoplasmic reticulum (Chen *et al.,* 2008). This provides a further mechanism by which HBeAg expression is reduced by subgenotype A1 (Kramvis *et al,* 2007). There was no association between the G1862T and HCC, with equal frequencies in all disease groups (**Table 3.8**). The

combination of the G1862T and the A1762T/G1764A, was however, significantly associated with HCC in patients infected with subgenotype A1 (Table 3.8).

The variation G1888A is unique to subgenotype A1, and rarely occurs in other genotypes (Kramvis *et al.,* 2008) The G1888A occurred in 62.5% of the subgenotype A1 isolates, 44% from HCC patients, and 77% from non-HCC patients. The G1888A mutation may interfere with initiation at the downstream 1901 core AUG, decreasing core protein translation (Kimbi *et al.,* 2012).

The G1896A mutation, which converts the codon TGG for tryptophan to a stop codon leading to truncation of the HBeAg precursor (Carman *et al.,* 1989), was responsible for the HBeAg-negativity in 76% of HBeAg-negative patients infected with genotype D. In genotype D, 1858T, allows stable T-A Watson Crick bond in the encapsdation signal. In contrast, genotype A isolates have 1858C and thus the G1896A would disrupt the G-C Watson Crick bond, destabilizing ε and compromising viral replication (Lok *et al.,* 1994). Thus, as has been discussed above, subgenotype A1 has alternative mechanisms of reducing HBeAg expression at the transcriptional, translational and post-translational levels. It is interesting to note that these mechanisms do not occur in subgenotype A2 (Kramvis *et al.,* 2007).

The presence of nucleotide changes in the BCP/PreC region as well as pre-S deletion mutants has been shown to play a role in liver disease progression (Fan *et al.,* 2001; Kao *et al.,* 2003). Although numerous studies have looked

at these regions independently, the combined effect of changes in these two regions, and the impact on disease progression has not been extensively evaluated, particularly for genotypes A and D. The combination of pre-S deletion mutants, in genotypes B and C, together with the A1762T/G1764A double mutation, may increase liver disease progression and ultimately increase the risk for HCC (Chen *et al.,* 2008). Of the nine isolates in which pre-S deletion mutants were detected, only six had successful amplification and analysis of the corresponding BCP/PreC region (**Table 3.3**). Five out of the six isolates had the 1762T/1764A double mutation present, with a single isolate, INHCC30, having 1764A. Four of the 5 isolates with 1762T/1764A also had G1862T. Thus it is possible that the combined effects of BCP/PreC mutations and S region mutations in HBV may play a role in liver disease progression and needs to be explored further.

# CHAPTER 5: CONCLUSION

This is the first study to analyse the genotype/subgenotype distribution of HBV isolates from Kerala, southern India. Genotype A (predominantly subgenotype A1) was found to be the predominant genotype in Kerala, differing from other regions of India where genotype D was found to predominate, or occur at a similar frequency to genotype A.

This study represents the first to sequence the complete genome of subgenotype D2 from India, and showed predominance of subgenotype D2 in Kerala, which differs from other areas of India.

The relationship between HBV genotypes and three disease groups, revealed a higher prevalence of subgenotype A1 in HCC patients and HCC patients infected with subgenotype A1 were significantly younger compared to those infected with genotype D. This concurs with the findings in southern African patients infected with subgenotype A1 (Kew *et al,* 2005).

Molecular characterisation of HBV isolates, included sequencing of the BCP/PreC region, and the S region, together with a few representative complete genome sequences, revealed a number of important nucleotide changes that could contribute to the higher hepatocarcinogenic potential of subgenotype A1. The association of A1762T/G1764A with HCC was not found to be as strong as shown previously in a South African study (Baptista *et al,* 1999). However, G1862T, which occurs mainly in subgenotype A1, was significantly associated with HCC, when it occurred together with

A1762T/G1764A. The pre-S deletion mutants were found predominantly in subgenotype A1, with the pre-S2 start codon mutation found exclusively in HCC patients infected with subgenotype A1. The pre-S deletion mutants occurred together with A1762T/G1764A and G1862T. It is possible that these nucleotide changes have an important function in HCC progression in patients infected with subgenotype A1 and this need to be studied using larger cohorts and functional studies.

# APPENDICES

## APPENDIX A: SOLUTIONS AND REAGENTS

### A1. 10X TBE Buffer

108 g Tris base

55 g Boric acid

9.3 g EDTA

Make up to 1 L with distilled water and autoclave.

### A2. 1X TBE Buffer

1 ml of 10X TBE

Make up to 1 L with distilled water and autoclave.

### A3. 1% Agarose

1 g agarose

100 ml 1X TBE

Microwave until completely dissolved.

Add ethidium bromide and add to gel trays with gel combs.

### A4. 50 mg/ml Kanamycin

0.5 g kanamycin sulphate

10 ml SABAX water for injections

The solution was filter-sterilised using a 0.2 µm Minisart® NML single-use syringe filter

1 ml of solution was aliquoted into foil-wrapped sterile 1.5 ml microcentrifuge tubes and stored at -20℃

**A5. 2X YT Agar Plates supplemented 50 µg/ml Kanamycin**

16 g Bacto-Tryptone

10 g Bacto-Yeast Extract

5 g NaCl

15 g Agar Bacteriological

Distilled water was added to make up to 1 L.

Autoclaved at 121℃ for 30 minutes after which it was cooled at room temperature for 1 hour after which 1 ml 50 mg/ml kanamycin was added and mixed.

50 ml of this agar and kanamycin containing medium was poured into 10 cm petri dishes and allowed to dry overnight.

The medium containing petri dishes were incubated at 37 ℃ for 30 minutes and stored at 4 ℃ until further use.

**A6. TB Medium**

12g Tryptone

24g Yeast extract

4ml Glycerol

Make up to 1 L with sterilised water

Autoclaved at 121 ℃ for 30 minutes

Allow to cool to room temperature and store at 4 ℃ until further use.

# REFERENCES


Abdou Chekaraou M, Brichler S, Mansour W. (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. *J Gen Virol* 91: 1609–20.

Ahn SH, Kramvis A, Kawai S, Spangenberg HC, 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: 1370-1380.

Anjali H, Issac A, Anjali M, Anish T. (2012). Transfusion-transmissible infections among voluntary blood donors at Government Medical College Thiruvananthapuram, Kerala, India. *Asian J Trans* Sc 6: 55.

Arankalle VA, Gandhe SS, Borkakoty BJ, Walimbe AM, Biswas D, Mahanta J. (2010). A novel HBV recombinant (genotype I) similar to Vietnam/Laos in a primitive tribe in eastern India. *J of Viral Hep* 17: 501-510.

Arauz-Ruiz P, Norder H, Robertson BH, Magnius LO. (2002). Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. *J Gen Virol* 83: 2059-2073.

Bancroft WH, Mundon FK, Russlee PK. (1972). Detection of additional antigenic determinants of hepatitis B antigen. *J Immunol* 109: 842-848.

Banerjee A, Datta S, Chandra PK, Roychowdhury S, Panda CK, Chakravarty R. (2006a). Distribution of hepatitis B virus genotypes: phylogenetic analysis and virological characteristics of genotype C circulating among HBV carriers in Kolkata, Eastern India. *World J Gastroenterol* 12: 596405971.

Banerjee A, Kurbanov F, Datta S, Chandra PK, Tanaka Y, Mizokami M, Chakravarty R. (2006b). Phylogenetic relatedness and genetic diversity of hepatitis B virus isolates in Eastern India. *J Med Virol* 78: 1164-1174.

Baptista M, Kramvis A, Kew MC. (1999). High prevalence of 1762(T)1764(A) mutations in the basic core promoter of hepatitis B virus isolated from black Africans with hepatocellular carcinoma compared with asymptomatic carriers. *Hepatology* 29: 946-953.

Bartholomeusz A, Schaefer S. (2004). Hepatitis B virus genotypes: comparison of genotyping methods. *Reveiews in Med Vir* 14: 3-16.

Baumert TF, Rogers SA, Hasegawa K, Liang TJ. (1996). Two core promoter mutations identified in a hepatitis B virus strain associated with fulminant hepatitis result in enhanced viral replication. *J Clin Invest* 98: 2268-2276.

Baumert TF, Barth H, Blum HE. (2005). Genetic variants of hepatitis B virus and their clinical relevance. *Minerva Gastroenterol Dietol* 51: 95-108.

Beckel-Mitchener A and Summers J. (1997). A novel transcriptional element in circular DNA monomers of the duck hepatitis B virus. *J Virol* 71: 7917-7922.

Beck J, Nassal M. (2007) Hepatitis B virus replication. *World Jour of Gastroenter* 13: 48-64.

Bell TG, Makondo E, Martinson NA, 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(10):e45750. doi: 10.1371/journal.pone.0045750.

Benn J, Schneider RJ. (1994). Hepatitis B virus HBx protein activates Ras-GTP complex formation and establishes a Ras, Raf, MAP kinase signalling cascade. *Proc Natl Acad Sci USA* 91: 10350-10354.

Berk PD, Popper H. (1978). Fulminant hepatic failure. *Am J Gastroenterol* 69: 349-400.

Biswas A, Banerjee A, Chandra PK, Datta S, Panigrahi R, Dutta D, De BK, Pal M, Guha SK, Chakrabarti S, Chakravarty R. (2011). Variations in the

Functional Domain of Basal Core Promoter of Hepatitis B Virus among Eastern Indian patients with prevalence of Genotypes A, C and D among the same ethnic group. *J Med Virol* 83: 253-260.

Biswas A, Panigraphi R, Banerjee A, Pal M, De Krishna B, Chakrabarti, Chakravarty R. (2012). Differential pattern of pre-S mutations/deletions and its association with hepatitis B virus genotypes in Eastern India. *Inf, Gen, and Evolution* 12: 384-391.

Blumberg BS, Alter HJ, Visnich S. (1965). A "NEW" ANTIGEN IN LEUKEMIA SERA. *JAMA* 191:541-546.

Bowyer SM, van Staden L, Kew MC, Sim JG. (1997). A unique segment of hepatitis B virus group A genotype identified in isolates from South Africa. *J Gen Virol* 78: 1719-1729.

Brechot C, Pourcel C, Louise A, Rain B, Tiollais P. (1980). Presence of integrated hepatitis B virus DNA sequences in cellular DNA of human hepatocellular carcinoma. *Nature* 286: 533-535.

Buckwold VE, Xu Z, Chen M, Yen TS, Ou JH.(1997). Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication. *J Virol* 70: 5845-5851.

Bulla GA and Siddiqui A. (1989). Negative regulation of the hepatitis B virus pre-S1 promoter by internal DNA sequences. *Virology* 170: 251-260.

Cabrerizo M, Bartolome J, Carreno V. (2002). In vitro infection of human peripheral blood mononuclear cells by a defective hepatitis B virus with a deletion in the PreS1 region of the viral genome. *J Viral Hepat* 9: 265-271.

Carman WF, Jacyna MR, Hadziyannis S, Karayiannis P, McGarcvey MJ, Markis A. (1989). Mutation preventing formation of hepatitis e antigen in patients with chronic hepatitis B infection. *Lancet* ii: 588-591.

Carman WF, Zanetti AR, Karayiannis P, Waters J, Manzillo G, Tanzi E, Zuckerman AJ, Thomas HC. (1990). Vaccine-induced escape mutant of hepatitis B virus. *Lancet* 336: 325-329.

Chan LYC. (2010). The significance of Hepatitis B virus genotypes and mutations in the development of Hepatocellular carcinoma in Asia. *J of Gasteroenter and Hepatol* 10: 1-13.

Chandra PK, Banerjee A, Datta S, Chakravarty R. (2007). G1862T mutation among hepatitis B virus-infected individuals: association with viral genotypes and disease outcome in Kolkata, Eastern India. *Intervirology* 50: 173-180.

Chandra PK, Biswas A, Datta S, Banerjee A, Panigrahi R, Chakrabarti S, De BK, Chakravarty R. (2009). Subgenotypes of hepatitis B virus genotype D (D1, D2, D3 and D5) in India: differential pattern of mutations, liver injury and occult HBV infection. *J Viral Hep* 16: 749-756.

Chang JJ, Lewin SR. (2007). Immunopathogenesis of hepatitis B virus infection. *Immunol Cell Biol* 85: 16-23.

Chattopadhyay S, Das BC, Hussain Z, Kar P. (2006a). Hepatitis B virus genotypes in acute and fulminant hepatitis patients from north India using two different molecular genotyping approaches. *Hepatol Res* 35: 79-82.

Chattopadhyay S, Das BC, Kar P. (2006b). Hepatitis B virus genotypes in chronic liver disease patients from New Dehli, India. *World J Gastroenterol* 12: 6702-6706.

Chaudhuri V, Tayal R, Nayak B, Acharya SK, Panda SK. (2004). Occult hepatitis B virus infection in chronic liver disease: full length genome analysis of mutant surface promoter. *Gastroenterol* 130: 1153-1168.

Chauhan R, Kazim SN, Bhattacharjee J, Sakhuja P, Sarin SK. (2006). Basal Core Promoter, Precore region mutations of HBV and their association with e

antigen, Genotype, and severity of liver disease in patients with Chronic Hepatitis B in India. *J Med Virol* 78: 1047-1054.

Chen MT, Billaud JN, Sallberg M, Guidotti LG, Chisari FV, Jones J, Hughes J, Milich DR. (2004). A function of the hepatitis B virus precore protein is to regulate the immune response to the core antigen. *Proc Natl Acad Sci* 101:14913-14918.

Chen BF, Liu CJ, Jow GM, Chen PJ, Kao JH, Chen DS. (2006). High prevalence and mapping of pre-S deletion in hepatitis B virus carriers with progressive liver diseases. *Gastroenterology* 130: 1153-1168.

Chen CY, Crowther C, Kew MC, 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: 580-592.

Choi MS, Kim DY, Lee DH, Lee JH, Koh KC, Paik SW, Rhee JC, Yoo BC. (2007). Clinical significance of pre-S mutations in patients with genotype C hepatitis B virus infection. *J Viral Hepat* 14: 161-168.

Chowdhurry A, Santra A, et al., (2005). Community-based epidemiology of hepatitis B virus infection in West Bengal, India: prevalence of hepatitis B e antigen-negative infection and associated viral variants. *J Gastroenterol Hepatol* 20: 1712-1720

Cooper A, Paran N, Shaul Y. (2003). The earliest steps in hepatitis B virus infection. *Biochim Biophys Acta* 1614: 89-96.

Dane DS, Cameron CH, Briggs M. (1970). Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *Lancet* 1: 695-698.

Datta S. (2008). An overview of molecular epidemiology of hepatitis B virus (HBV) in India. *Virology Journal* 5:156-183.

Datta S, Panigrahi R, Biswas A, Chandra PK, Banerjee A, Mahaptra PK, Panda CK, Chakrabati S, Bhattacharya SK, Biswas K, Chakravarty R. (2009). Genetic characterisation of hepatitis B virus in peripheral blood leukocytes: evidence for selection and compartmentalization of viral variants with the immune escape G145R mutation. *J Virol* 89: 9983-992.

Dienstag JL. (2008) Hepatits B Virus Infection. *NEJM* 359: 1486-1500.

Dimattia MA, Watts NR, Stahl SJ, Grimes JM, Steven AC, Stuart DI, Wingfield PT. (2013). Antigenic Switching of hepatitis B virus by alternative dimerization of the capsid protein. *Structure* 21:133–142.)

Ding X, Mizokami M, Yao G, Xu B, Orito E, Ueda R, Nakanishi M. (2001). Hepatitis B virus genotype distribution among chronic hepatitis B virus carriers in Shanghai, China. *Intervirology* 44: 43-47.

Dusheiko GM, Bowyer SM, Sjogren MH, Ritchie MJ, Santos AP, Kew MC. (1985). Replication of hepatitis B virus in adult carriers in an endemic area. *J Infect Dis* 152: 566-671.

Fan YF, Lu CC, Chen WC, Yao WJ, Wang HC, Chang TT, Lei HY, Shiau AL, Su IJ. (2001). Prevalence and significance of hepatitis B virus (HBV) pre-S mutants in serum and liver at different replicative stages of chronic HBV infection. *Hepatology* 33: 277-286.

Fattovich G, Bortolotti F, Donato F. Natural history of chronic hepatitis B: special emphasis on disease progression and prognostic factors. *J Hepatol* 48: 335-352.

Feitelson MA, Zhu M, Duan LX, London WT. (1993). Hepatitis B x antigen and p53 are associated in vitro and in liver tissues from patients with primary hepatocellular carcinoma. *Oncogene* 8: 1109-1117.

Fernholz D, Galle PR, Stemler M, Brunetto M, Bonino F, Will H. (1993). Infectious hepatitis B virus variant defective in pre-S2 protein expression in a chronic carrier. *Virology* 194: 137-148.

Flodell S, Cromsigt J, Schleucher j, Kidd-Ljunggren K, Wijmenga S. (2002). Structure elucidation of the hepatitis B virus encapsidation signal by NMR on selectively labelled RNAs. *J Biomol Struct Dyn* 19: 627-636.

François G, Kew M, Van Damme P, Mphahlele MJ, Meheus A. (2001). Mutant hepatitis B viruses: a matter of academic interest only or a problem with far-reaching implications? *Vaccine* 19: 28-29.

Galibert F, Mandart E, Fitoussi F, Tiollais P, Charnay P. (1979). Nucleotide sequence of hepatitis B virus genome (subtype ayw) cloned in E.coli. *Nature* 281: 646-650.

Gandhe SS, Chadha MS, Arankalle VA. Hepatitis B virus genotypes and serotypes in Western India: Lack of clinical significance. *J Med Virol* 63: 324-330.

Ganem D. (1991). Assembly of hepadnaviral virions and subviral particles. *Curr Top Microbiol Immunol* 168: 61-83.

Ganem D and Prince AM. (2004). Hepatitis B virus infection-natural history and clinical consequences. *NEJM* 350: 1118-1129.

Gerber MA, Hadziyannis S, Viisoulis C, Schaffner F, Paronetto F, Popper H. (1974). Electron microscopy and immunoelectronmicroscopy of cytoplasmic hepatitis B antigen in hepatocytes. *Am J Pathol* : 75: 489-502.

Gerlich WH, Robinson WS. (1980). Hepatitis B virus contains protein attached to the 5' terminus of its complete DNA strand. *Cell* 21: 801-809.

Gerner P, Schafer HM, Prange R, Pravitt D, Wirth S. (2003). Functional analysis of a rare HBV deletion mutant in chronically infected children. *Pediatr Res* 53: 891-897.

Glebe D, Urban S. (2007). Viral and cellular determinants involved in hepadnaviral entry. *World Jour of Gastroenter* 13:22-38.

Grethe S, Heckel JO, Rietschel W, Hufert FT. (2000). Molecular epidemiology of hepatitis B virus variants in nonhuman primates. *J Virol* 74: 5377-5381.

Gunther S, Li BC, Miska S, Kruger DH, Meisel H, Will H. (1995). A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. *J Virol* 69: 5437–5444.

Günther S, Piwon N, Iwanska A, Schilling R, Meisel H, Will H. (1996). Type, prevalence, and significance of core promoter/enhancer II mutations in hepatitis B viruses from immunosuppressed patients with severe liver disease. *J Virol* 70: 8318-8331.

Gunther S, Fisher L, Pult I, et al. (1999). Naturally occurring variants of hepatitis B virus. *Adv Virus Res* 52: 125-137.

Guo H, Mason WS, Aldrich CE, Saputelli JR, Miller DS, Jilbert AR, Newbold JE. (2005). Identification and characterisation of avihepadnaviruses isolated from exotic anseriformes maintained in captivity. *J Virol* 79: 2792-2742.

Guo X, Jin Y, Qian G, Tu H. (2008). Sequential accumulation of the mutations in the core promoter of hepatitis B virus is associated with the development of hepatocellular carcinoma in Qidong, China. J of Hepatol 49: 718-725.

Hadziyannis S, Gerber MA, Vissoulis C, Popper H. (1973). Cytoplasmic hepatitis B antigen in ‘ground-glass’ hepatocytes of carriers. *Arch Pathol* 96: 327-330.

Hannoun C, Norder H, Lindh M. (2000). An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietman. *J Gen Virol* 81: 2267-2272.

Heerman KM, Goldman U, Schwartz W, Seyyfarth T, Baumgarten H, Gerlich WH. (1994). Large surface proteins of hepatitis B virus containing the pre-s sequence. *J Virol* 52: 396-402.

Hiraga M, Nishizono A, Mifune K, Esumi M, Shikata T. (1994). Analysis of upstream region of hepatitis B virus core gene using in vitro transcription system. *J Med Virol* 43: 404-411.

Hoofnagle JH, Doo E, Liang TJ, Fleischer R, Lok AS. (2007). Management of hepatitis B: summary of a clinical research workshop. *Hepatology* 45: 1056-1075.

Huang M and Summers J. (1994). PET, a small sequence distal to the pregenome cap site, is required for expression of the duck hepatitis B virus pregenome. *J Virol* 68: 1564-1572.

Hussain SP, Schwank J, Staib F, Wang XW, Harris CC. (2007). TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer. *Oncogene* 26:2166-76.

Hutton T, Zhou S, Standring DA. (1992). RNA- and DNA-binding activities in hepatitis B virus capsid protein: a model for their roles in viral replication. *J Virol* 66: 5232-5241.

Huy TTT, Ushijima H, Sata T, Abe K. (2008). New Complex Recombinant Genotype of Hepatitis B Virus indentified in Vietnam. *J of Virol* 82: 5657-5663

Jean-Jean O, Levrero M, Will H, Perricaudet M, Rossignol JM. (1989). Expression mechanism of the hepatitis B virus (HBV) C gene and biosynthesis of HBe antigen. Virology 170: 99-106.

Huy TT, Ushijima H, Win KM, Luengrojanakul P, Shrestha PK, Zhong ZH, Smirnov AV, Taltavull TC, Sata T, Abe K. (2003). High prevalence of hepatitis B virus pre-s mutant in countries where it is endemic and its relationship with genotype and chronicity. *J Clin Microbiol* 41:5449-55.

Jiang S, Yang Z, Li W, Li X, Wang Y, Zhang J, Xu C, Chen PJ, Hou J, McCrae MA, Chen X, Zhuang H, Lu F. (2012). Re-evaluation of the carcinogenic significance of hepatitis B virus integration in hepatocarcinogenesis. *PLoS One.*7e40363.doi:10.1371/journal.pone.0040363

Kann M. (2002). Structural and Molecular Virology. In: Lai CL and Locarnini S. (eds.) *Hepatitis B Virus*. London: International medical press.

Kao JH, Wu NH, Chen PJ, Lai MY, Chen DS. (2000). Hepatitis B genotypes and the response to interferon therapy. *J Hepatol* 33: 998-1002.

Kao JH, Chen PJ, Lai MY, Chen DS. (2003). Basal core promoter mutations of hepatitis B virus increase the risk of hepatocellular carcinoma in hepatitis B carriers. *Gastroeneterology* 124: 327-334.

Kawamoto S, Ueda K, Mita E, Matsubara K. (1994). The packaging signal in hepatitis B virus pregenome functions only at the 5' end. *J Virol Methods* 49: 113-127.

Kay A, Zoulim F. (2007). Hepatitis B virus genetic variability and evolution. *Virus Res* 127: 164-176.

Kew MC. (1992). Chronic hepatitis B virus infection and hepatocellular carcinoma in Africa. S Afr J Sci 88: 524-528

Kew MC, Kramvis A, Yu MC, Arakawa K, Hodkinson J. (2005). Increased hepatocarcinogenic potential of hepatitis B virus genotype A in Bantu speaking sub-saharan Africans. *J of Med Virol* 75: 513-521.

Kidd-Ljunggren K, Myhre E, Bläckberg J. (2004). Clinical and serological variation between patients infected with different Hepatitis B virus genotypes. *J Clin Microbiol* 42: 5837-5841.

Kim Y, Hong YB, Jung G. (1999). Hepatitis B virus: DNA polymerase activity of deletion mutants. *Biochem Mol Biol Int* 47: 301-308.

Kim YJ, Kim HY, Kim JS, Lee JH, Yoon JH, Kim CY, Park BL, Cheong HS, Bae JS, Kim S, Shin HD, Lee HS. (2009). Putative association of transforming growth factor-alpha polymorphisms with clearance of hepatitis B virus and occurrence of hepatocellular carcinoma in patients with chronic hepatitis B virus infection. *J Viral Hepat* 17: 518:526.

Kimbi G, Kramvis A, Kew MC. (2004). Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa. *J Gen Virol* 85: 1211-1220.

Kimbi GC, Kew MC, Kramvis A. (2012). The effect of the G1888A mutation of subgenotype A1 of hepatitis B virus on the translation of the core protein. *Virus Res* 163: 334-340.

Kramvis A, Kew MC, Francois G. (2005). Hepatitis B virus genotypes. *Vaccine* 23: 2409-2423.

Kramvis A and Kew MC. (2007). Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes. *Hepatol Res* 37: S9-S19.

Kramvis A, Kew MC, Bukofzer S. (1998). Hepatitis B virus precore mutants in serum and liver of Southern African Blacks with hepatocellular carcinoma. *J Hepatol* 28: 132-141.

Kramvis A and Kew MC. (1999). The core promoter of hepatitis B virus. *J Viral Hepat* 6: 415-427.

Kramvis A, Bukofzer S, Kew MC, 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: 235-240.

Kramvis A, Weitzman L, Owiredu WK, Kew MC. (2002). Analysis of the complete genome of subgroup A' hepatitis B virus isolates from South Africa. *J Gen Virol* 83: 835-839.

Kramvis A and Kew MC. (2007). Molecular characterization of subgenotype A1 (subgroup Aa) of hepatitis B virus. *Hepatol Res* 37: S27-S32.

Kramvis A, Arakawa K, Yu MC, Nogueira R, Stram DO, Kew MC. (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 and Kew MC. (1998). Structure and function of the encapsidation signal of hepadnaviridae. *J Viral Hepat* 6: 357-367.

Kramvis A, Paraskevis D. (2012). Subgenotype A1 of Hepatitis B virus – Tracing Human Migrations in and out of Africa. *Antiviral Therapy, in press*

Kumar A, Kumar SI, Pandey R, Naik S, Agarwal R. Hepatitis B virus genotype A is more often associated with severe disease in northern India than is genotype D. *Indian J Gastroenterol* 24: 19-22.

Kurien P. (1994). Colonialism and ethnogenesis: A study of Kerala, India. *Theory and Society* 23: 385-417.

Kuroki K, Floerani M, Mimms LT, Ganem D. (1990). Epitope mapping of the preS1 domain of the hepatitis B virus large surface protein. Virology 176: 620-624.

Lai CL, Dienstag J, Schiff E, Leung NW, Atkins M, Hunt C, Brown N, Woessner M, Boehme R, Condreay L. (2003). Prevalence and clinical

correlates of YMDD variants during lamivudine therapy for patients with chronic hepatitis B. *Clin Infect Dis* 36: 687-696.

Lanford RE, Chavez D, Brasky KM, Burns RD 3rd, Rico-Hesse R. (1998). Isolation of a hepadnavirus from the wooly monkey, a New World primate. *Proc Natl Acad Sci USA* 95: 5757-5761.

Laperche S, Girault A, Beaulieu MJ, Bouchardeau F, Courouce AM. (2001). Determination of hepatitis B virus subtypes by an enzyme immunoassay method using monoclonal antibodies to type-specific epitopes of HBsAg. *J Viral Hep* 8: 447-453.

Lavanchy D. (2004). Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. *J Virol Hep* 11: 97-107

Le Bouvier GL. (1971). The heterogeneity of Australia antigen. *J Infect Dis* 123: 671-675.

Lee TH, Elledge SJ, Butel JS. (1995). Hepatitis B virus X protein interacts with a probable cellular DNA repair protein. *J Virol* 69: 1107-1114.

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: 581-592.

Li JS, Tong SP, Wen YM, Vitvitski L, Zhang Q, Trépo C. (1993). Hepatitis B virus genotype A rarely circulates as an HBe-minus mutant: possible contribution of a single nucleotide in the precore region. *J Virol* 67: 5402-5410.

Liang TJ. (2009). Hepatitis B: The virus and disease. *Hepatology* 49: S13-21.

Lin CM, Wang GM, Jow GM, Chen BF. (2012). Functional analysis of hepatitis B viruspre-s deletion variants associated with hepatocellular carcinoma. *J Biomed Sci* 19: 17.

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. *J Inf Dis* 175: 1285-1293.

Liu S, Zhang H, Gu C, Yin J, He Y, Xie J. (2009). Associations between hepatitis B virus mutations and the risk of hepatocellular carcinoma: a meta-analysis. J Natl Cancer Inst 10: 1066-1082.

Locarnini S, McMillan J, Bartholomuesz A. (2003). The Hepatitis B virus and Common Mutants. *Semin Liver Dis* 23: 005-020

Locarnini S. (2004). Molecular Virology of Hepatitis B Virus. *Sem In Liver Dis* 24: 3-10.

Lok AS, 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 USA* 91: 4077-4081.

Lok A, MacMahon BJ. (2007). Chronic hepatitis B. *Hepatol* 45: 507-539.

Lu X, Block T. (2004). Study of the early steps of the Hepatitis B Virus life cycle. *Int J Med Sci* 1:21-33

Madan Y, Batra Y, Sreenivas V, Mizokami M, Tanaka Y, Chalamalasetty SB, Panda S, Acharya SK. (2009). HBV genotypes in India: do they influence disease severity? *Hepatol Res* 39: 157-163.

Maeng CY, Ryu CJ, Gripon P, Guguen-Guillouzo C, Hong HJ. (2000).Fine mapping of virus-neutralizing epitopes on hepatitis B virus preS1. *Virology* 270: 9-16.

Magnius LO, Norder H.(1995). Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. *Intervirology* 38: 24-34.

Makondo E, Bell TG, Kramvis A. (2012). Genotyping and Molecular Characterization of Hepatitis B Virus from Human Immunodeficiency Virus-Infected Individuals in Southern Africa. *PLoS One*

Marion PL, Salazar FH, Alexander JJ, Robinson WS. (1980). State of hepatitis B viral DNA in a human hepatoma cell line. *J Virol* 33:795-806.

McCormack GP, Clewley JP. (2002). The application of molecular phylogenetics to the analysis of viral genome diversity and evolution. *Reviews in Med Virol* 12:221-238.

McIver CJ, Jacques CFH, Chow SSW, Munro SC, Scott GM, Roberts JA, Craig ME, Rawlinson WD. (2005). Development of Multiplex PCRs for Detection of Common Viral Pathogens and Agents of Congenital Infections. *J Clin Microbiol* 43: 5102-5110.

McLachlan A, Milich DR, Raney AK, Riggs MG, Hughes JL, Sorge J, Chisari FV. (1987). Expression of hepatitis B virus surface and core antigens: influences of pre-S and precore sequences. *J Virol* 61: 683-692.

Moolla N, Kew MC, Arbuthnot P.(2002). Regulatory elements of hepatitis B virus transcription. *J Viral Hepat* 9: 323-331.

Milich DR, Chen MK, Hughes JL, Jones JE. (1998). The secreted hepatitis B precore antigen can modulate the immune response to the nucleocapsid: a mechanism for persistence. *J Immunol* 160: 2013-2021.

Mun HS, Lee SA, Kim H, Hwang ES, Kook YH, Kim BJ. (2011). F141L pre-S2 mutation in hepatitis B virus increases the risk of hepatocellular carcinoma in patients with chronic genotype C infections. *J Virol* 85: 123-132.

Nassal M. (1992). The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. *J Virol* 66: 4107-4116.

Nassal M and Schaller H. (1993). Hepatitis B virus replication. *Trends Microbiol* 1: 221-228.

Naumann H, Schaefer S, Yoshida CF, Gaspar AM, Repp R, Gerlich WH. (1993). Identification of a new hepatitis B virus (HBV) genotype from Brazil that expresses HBV surface antigen subtype adw4. *J Gen Virol* 74: 1627-1632.

Neuveut C, Wei Y, Buendia MA. (2010). Mechanisms of HBV-related hepatocarcinogenesis. *J Hepatol* 52: 594-604.

Ni YH, Chang MH, Wang KJ, Hsu HY, Chen HL, Kao JH. (2004). Clinical relevance of hepatitis B virus genotype in children with chronic infection and hepatocellular carcinoma. *Gasteroenter* 127: 1733-1738.

Norder H, Hammas B, Lofdahl S, Courouce AM, Magnius LO. (1992). 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: 1201-1208.

Norder H, Couroucé AM, Magnius LO. (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 JW, Fields HA, Mushahwar IK, Magnius LO. (1996). Complete sequencing of a gibbon hepatitis B virus genome reveals a unique genotype distantly related to the chimpanzee hepatitis B virus. *Virology* 218: 214-223.

Okamoto H, Tsuda F, Sakugawa H, Sastrosoewignjo RI, Imai M, Miyakawa Y, Mayumi M. (1988). Typing hepatitis B virus by homology in nucleotide

sequence: comparison of surface antigen subtypes. *J Gen Virol* 69: 2575-2583.

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. *J Virol* 68: 8102-8110.

Orito E, Mizokami M. (2003). Hepatitis B virus genotypes and hepatocellular carcinoma in Japan. *Intervirology* 46: 408-412.

Orito E, Sugauchi F, Tanaka Y, Ichida T, Sata M, Tanaka E, Okanoue T, Sakugawa H, Watanabe H, Miyakawa H, Nishiguchi S, Kumada H, Ueda R, Mizokami M. (2005). Differences of hepatocellular carcinoma patients with hepatitis B virus genotypes of Ba, Bj or C in Japan. *Intervirology* 48: 239-245.

Ou JH, Laub O, Rutter WJ. (1986). Hepatitis B virus gene function: the precore region targets the core antigen to cellular membranes and causes the secretion of the e antigen. *Proc Natl Acad Sci USA* 83: 1578-1582.

Page, RDM. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**: 357-358.

Paran N, Geiger B, Shaul Y. (2001). HBV infection of cell culture: evidence for multivalent and cooperative attachment. EMBO J 20: 4443-4453.

Park JH, Cho EW, Lee YJ, Shin SY, Kim KL. (2000). Determination of the protective effects of neutralizing anti-hepatitis B virus (HBV) immunoglobulins be epitope mapping with recombinant HBV surface antigen protein. *Microbiol Immunol* 44: 703-710.

Pollack JR, Ganem D. (1993). An RNA stem-loop structure directs hepatitis B virus genomic RNA encapsidation. *J Virol* 67: 3254-3263.

Poustchi H, Mohamadkhani A, Bowden S, Montazeri G, Ayres A, Revill P, Farrell GC, Locarnini S, George J, Malekzadeh R. (2008). Clinical significance of precore and core promoter mutations in genotype D hepatitis B-related chronic liver disease. *J Viral Hepat* 15: 753-760.

Prevention of Hepatitis B in India-An Overview. (2002). World Health Organization South-East Asia Regional office, New Delhi.

Prince AM, Fuji H, Gershon RK. (1964). Immunohistochemical studies on the etiology of Anicteric hepatitis in Korea. *Am J Hyg* 79: 365-381.

Prince AM, Hargrove RL, Szmuness W, Cherubin CE, Fontana VJ, Jeffries GH. (1970). Immunologic distinction between infectious and serum hepatitis. *NEJM* 282: 987-991.

Radziwill G, Tucker W, Schaller H. (1990). Mutational analysis of the hepatitis B virus P gene product: domain structure and RNase H activity. J Virol 64: 613-620.

Raimondo G, Costantino L, Caccamo G, Pollicino T, Squadrito G, Cacciola I, Brancatelli S. (2004). Non-sequencing molecular approaches to identify preS2-defective hepatitis B virus variants proved to be associated with severe liver disease. *J Hepatol* 40: 515-519.

Raza SA, Clifford GM, Franceschi S. (2007). Worldwide variation in the relative importance of hepatitis B and hepatitis C viruses in hepatocellular carcinoma: a systematic review. *Brit J Cancer* 96: 1127-1134.

Rehermann B, Nascimbeni M. (2005). Immunology of Hepatitis B virus and Hepatitis C virus infection. *Nat Rev Immuno* 5: 215-229.

Roossinck MJ, Jameel S, Loukin SH, Siddiqui A. (1986). Expression of hepatitis B viral core region in mammalian cells. *Mol Cell Biol* 6: 1393-1400.

Roux KH. (2009). Optimization and Troubleshooting in PCR. *Cold Spring Harb Protoc*; doi: 10.1101/pdb.ip66

Russnak RH. (1991). Regulation of polyadenylation in hepatitis B viruses: stimulation by the upstream activating signal PS1 is orientation-dependent, distance-independent, and additive. *Nucleic Acids Res* 19: 6449-6456.

Sachse K. (2004). Specificity and performance of PCR detection assays for microbial pathogens. *Mol Biotechnol* 26: 61-80.

Sandesh K, Varghese T, Harikumar R, Beena P, Sasidharan VP, Bindu CS. (2006). Prevalence of Hepatitis B and C in the normal population and high risk groups in north Kerala. *Trop Gastroenterol* 27: 80-83.

Santos AO, Alvardo-Mora MV, Bothelho L, Vieira DS, Pinho JR, Carrilho FJ, Honda ER, Salcedo JM. (2010). Characterisation of hepatitis B virus (HBV) genotypes in patients from Rondonia, Brazil. *Virol J* 7: 315.

Schaefer S. (2007). Hepatitis B virus taxonomy and hepatitis B virus genotypes. *World J Gastroenterol* 13: 14-21.

Schaller H, Fischer M. (1991). Transcriptional control of hepadnavirus gene expression. Curr Top Microbiol Immunol 168: 21-3

Shi YH, Shi CH. (2009). Molecular characteristics and stages of chronic hepatitis B virus infection. *World J Gastroenterol* 15: 3099-3105.

Seddigh-Tonekaboni S, Waters JA, Jeffers S, Gehrke R, Ofenloch B, Horsh A, Hess G, Thomas HC, Karaviannis P. (2000). Effect of variation in the common ‘a’ determinant on the antigenicity of hepatitis B surface antigen. *J Med Virol* 60: 113-121.

Seeger C and Mason WS. (2000). Hepatitis B virus biology. *Microbiol Mol Biol Rev* 64: 51-68.

Sheldon J, Soriano V. (2008). Hepatitis B virus escape mutants induced by antiviral therapy. *J Antimicrob Chemo* 61: 766-768.

Siddiqui A, Jameel S, Mapoles J. (1986). Transcriptional control elements of hepatitis B surface antigen gene. *Proc Natl Acad Sci USA* 83: 566-570.

Simmonds P, Midgley S. (2005). Recombination in the genesis and evolution of hepatitis B virus genotypes. *J Virol* 79: 15467-15476.

Skelton M. (2008). Characterization of mutants and splice variants of Hepatitis B virus isolated from South African black Hepatocellular carcinoma patients. *Dissertation, University of the Witwatersrand*.

Sobotta D, Sominskaya I, Jansons J, Meisel H, Schmitt S, Heermann KH, Kaluza G, Pumpens P, Gerlich WH. (2008). Mapping of immunodominant B-cell epitopes and the human serum albumin-binding site in natural hepatitis B virus surface antigen of defined genosubtype. *J Gen Virol* 81:369-78.

Stuyver L, De Gendt S, Van Geyt C, Zoulim F, Fried M. (2000). A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *J Gen Virol* 81: 67-74.

Stuyver LJ, Locarnini SA, Lok A, Richman DD, Carman WF, Dienstag JL, Schinazi RF, and the HEP DART International Committee. (2000). Nomenclature for Antiviral-Resistant Human Hepatitis B virus Mutations in the Polymerase Region. Hepatology 33: 751-757.

Sugauchi F, Ohno T, Orito E, Sakugawa H, Ichida T, Komatsu M, Kuramitsu T, Ueda R, Miyakawa Y, Mizokami M. (2003). Influence of hepatitis B virus genotypes on the development of preS deletions and advanced liver disease. *J Med Virol* 70: 537-544.

Sugauchi F, Kumada H, Acharya SA, Shrestha SM, Gamutan MTA, Khan M, Gish RG, Tanaka Y, Kato T, Orito E, Ueda R, Miyakawa Y, Mizokami M.

(2004). Epidemiological and sequence differences between two subtypes (Aa and Ae) of hepatitis B virus genotype A. *J Gen Virol* 85: 811-820.

Sumi H, Yokosuka O, Seki N, Arai M, Imazeki F, Kurihara T, Kanda T, Fukai K, Kato M, Saisho H. (2003). Influence of hepatitis B virus genotypes on the progression of chronic type B liver disease. *Hepatology* 37: 19-26.

Summers J, O'Connell A, Millman I. (1975). Genome of hepatitis B virus: restriction enzyme cleavage and structure of DNA extracted from Dane particles. *Proc Natl Acad Sci USA*. 72: 4597-4601.

Summers J, Smolec JM, Snyder R. (1978). A virus similar to human hepatitis B virus associated with hepatitis and hepatoma in woodchuck. *Proc Natl Acad Sci USA* 75: 4533-4537.

Summers J, Mason WS. (1982). Replication of the genome of a hepatitis B--like virus by reverse transcription of an RNA intermediate. *Cell* 29: 403-415.

Takahashi K, Aoyama K, Ohno N, Iwata K, Akahane Y, Baba K, Yoshizawa H, Mishiro S. (1995). The precore/core promoter mutant T1762/A1764 of hepatitis B virus: clinical significance and an easy method for detection. *J Gen Virol* 76: 3259-3164.

Takahashi T, Akahane Y, Hino K, Ohta Y, Mishiro S. (1998). Hepatitis B virus genomic sequence in the circulation of hepatocellular carcinoma patients: comparative analysis of 40 full-length isolates. *Archives of Virol* 143: 2313-2326.

Takahashi K, Ohta Y, Kanai K, Akahane Y, Iwasa Y, Hina K, Ohno N, Yoshizawa H, Mishiro S. (1999). Clinical implications of mutations C-to-T1653 and T-to-C/A/G1753 of hepatitis B virus genotype C genome in chronic liver disease. *Arch Virol* 144: 1299-1308.

Takashima H, Araki K, Miyazaki J, Yamamura K, Kimoto M. (1992). Characterisation of T-cell tolerance to hepatitis B virus (HBV) antigen in transgenic mice. *Immunology* 75: 398-405.

Tanaka Y, Hasegawa I, Kato T, Orito E, Hirashima N, Acharya SK, Gish RG, Kramvis A, Kew MC, Yoshihara N, Shrestha SM, Khan M, Miyakawa Y, Mizokami M. (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.

Tanaka Y, Mukaide M, Orito E, Yuen MF, Ito K, Kurbanov F, Sugauchi F, Asahina Y, Izumi N, Kato M, Lai CL, Ueda R, Mizokami M. (2006). Specific mutations in enhancer II/core promoter of hepatitis B virus subgenotypes C1/C2 increase the risk of hepatocellular carcinoma. *J Hepatol* 45: 646-653.

Tatematsu K, Tanaka Y, Kurbanov F, Sugauchi F, Mano S, et al. (2009). A genetic variant of hepatitis B virus divergent fromknown human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. *J Virol* 83: 10538-10547.

Testut P, Reanrd CA, Terradillos O, Vityitski-Trepo L, Tekaia F, Degott C, Blake J, Boyer B, Buendia MA. (1996). A new hepadnavirus endemic in artic ground squirrels in Alsaka. *J Virol* 70: 4210-4219.

Thakur V, Guptan RC, Kazim SN, Malhotra V, Sarin SK. (2002). Profile, spectrum and significance of HBV genotypes in chronic liver disease patients in Indian subcontinent. *J Gastroeneterol Hepatol* 2: 165-170.

Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-80.

Thyagarajan SP, Murugavel KG. (2000). Prevalence of hepatitis B virus infection in the general population of India. *Ind J Gastroenterol* 7: 456-458

Toillais P, Pourcel C, Dejean A. (1985). The hepatitis B virus. *Nature* 317: 489-495.

Tong SP, Li JS, Vitvitski L, Trépo C. (1992). Replication capacities of natural and artificial precore stop codon mutants of hepatitis B virus: relevance of pregenome encapsidation signal. *Virology* 191: 237-245.

Tong S, Li J, Wands JR. (1995). Interaction between duck hepatitis B virus and a 170-kilodalton cellular protein is mediated through a neutralizing epitope of the pre-S region and occurs during viral infection. *J Virol* 69: 7106-7112.

Tong MJ, Blatt LM, Kao JH, Cheng JT, Corey WG. (2007). Basal core promoter T1762/A1764 and precore A1896 gene mutations in hepatitis B surface antigen-positive hepatocellular carcinoma: a comparison with chronic carriers. *Liver Int* 27: 1356-1363.

Torresi J. (2002). The virological and clinical significance of mutations in the overlapping envelope and polymerase genes of hepatitis B virus. *Journal of Clin Virol* 25: 97-106.

Tuttleman JS, Pourcel C, Summers J. (1986). Formation of the pool of covalently closed circular viral DNA in hepadnavirus-infected cells. *Cell* 47: 451-460.

Usuda S, Okamoto H, Iwanari H, Baba K, Tsuda F, Miyakawa Y, Mayumi M. (1999). Serological detection of hepatitis B virus genotypes by ELISA with monoclonal antibodies to type-specific epitopes in the preS2-region product. *J Virological Methods* 80: 97-112.

Vaudin M, Wolstenholme AJ, Tsiquaye KN, Zuckerman AJ, Harrison TJ. (1988). The complete nucleotide sequence of the genome of a hepatitis B virus isolated from a naturally infected chimpanzee. *J Gen Virol* 69: 1383-1389.

Vermeulen M, Dickens C, Lelie N, Walker E, Coleman C, Keyter M, Reddy R, Crookes R, Kramvis A. (2012). 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* 52:880-92.

Vivekanandan P, Abraham P, Sridharan G, Chandy G, Raghuraman S, Daniel H, Subramaniam T. (2004). Distribution of Hepatitis B Virus Genotypes in Blood Donors and Chronically Infected Patients in a Tertiary Care Hospital in Southern India. *Clin Infect Dis* 38: 81-86.

Wang GH, Seeger C. (1993). Novel mechanism for reverse transcription in hepatitis B viruses. *J Virol* 67: 6507-6512.

Wang HC, Huang W, Lai MD, Su IJ. (2006). Hepatitis B virus pre-S mutants, endoplasmic reticulum stress and hepatocarcinogenesis. *Cancer Sci* 97: 683-688.

Warren KS, Heeney JL, Swan RA, Herivanto Verschoor EJ. (1999). A new group of hepadnaviruses naturally infecting orang-utans (Pongo pygmaeus). *J Virol* 73: 7860-7865.

Weber M, Bronsema V, Bartos H, Bosserhoff A, Bartenschlager R, Schaller H. (1994). Hepadnavirus P protein utilizes a tyrosine residue in the TP domain to prime reverse transcription. *J Virol* 68: 6-13.

World Health Organisation. (2012). 'Fact Sheet N° 204 – Hepatitis B'. http://www.who.int/mediacentre/factsheets/fs204/en/

Wright TL, Lau JY. (1993). Clinical aspects of hepatitis B virus infection. Lancet 342: 1340-1344.

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* 58: 301-307.

Yaginuma K and Koike K. (1989). Identification of a promoter region for 3.6-kilobase mRNA of hepatitis B virus and specific cellular binding protein. *J Virol* 63: 2914-2920.

Yeh CT, So M, NG J, Yang HW, Chang ML, Lai MW, Chen TC, Lin CY, Yeh TS, Lee W. (2010). Hepatitis B virus-DNA level and basal core promoter A1762T/G1764A mutation in liver tissue independently predict postoperative survival in hepatocellular carcinoma. *Hepatol* 52: 1922-1933.

Yousif M, Kramvis A. (2012). Genotype D of hepatitis B virus and its subgenotypes: An update. *Hepatol Research* 10: 1-10.

Yu X and Mertz JE. (1996). Promoters for synthesis of the pre-C and pregenomic mRNAs of human hepatitis B virus are genetically distinct and differentially regulated. *J Virol* 70: 8719-8726.

Yu H, Yuan Q, Ge SX, Wang HY, Zhang YL, Chen QR, Zhang J, Chen PJ, Xia NS. (2010). Molecular and Phylogenetic Analyses Suggest an Additional Hepatitis B Virus Genotype "I". *PLoS ONE* 5: 92-97

Yuh CH and Ting LP. (1991). C/EBP-like proteins binding to the functional box-alpha and box-beta of the second enhancer of hepatitis B virus. *Mol Cell Biol* 11: 5044-5052.

Yuh CH, Chang YL, Ting LP. (1992). Transcriptional regulation of precore and pregenomic RNAs of hepatitis B virus. *J Virol* 66: 4073-4084.

Zhang D, Ma S, Zhang X, Zhao H, Ding H, Zeng C. (2010). Prevalent HBV point mutations and mutation combinations at BCP/PreC region and their association with liver disease progression. *BMC Inf Dis* 10: 271-276.

Zoulim F, Saputelli J, Seeger C. (1994). Woodchuck hepatitis virus X protein is required for viral infection in vivo. *J Virol* 68: 2026-2030.

Zoulim F and Locarnini S. (2009). Hepatitis B Virus Resistance to Nucleos(t)ide Analogues. *Gasteroenter* 137: 1593-1608