UNIVERSITY OF GHANA
COLLEGE OF HEALTH SCIENCES

ULTRA-DEEP SEQUENCING ANALYSIS OF HEPATITIS B VIRUS
PRECORE/CORE VARIANTS AMONG PRISON-INMATES

BY

FARRID BOADU
(10198945)

THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF MPHIL MEDICAL MICROBIOLOGY DEGREE

DEPARTMENT OF MICROBIOLOGY
JULY, 2017
DECLARATION

DECLARATION BY THE CANDIDATE

I hereby declare that this thesis is the result of my own research undertaken at the Virology unit of the School of Biomedical and Allied Health Sciences (SBAHS), under the supervision of Rev’d. Fr Dr Kwamena Sagoe of the Department of Medical Microbiology (SBAHS) and Dr Adwoa Agyei of the Department of Medicine, School of Medicine and Dentistry. Neither all nor part of this work has been presented for another degree elsewhere and that references to other people’s work have been duly acknowledged.

Signature………………………… Date……/……/………………

Farrid Boadu

DECLARATION BY SUPERVISORS

We hereby declare that the principal work and presentation of this thesis were supervised by us in accordance with guidelines on supervision of thesis laid down by the University of Ghana.

Principal supervisor

Signature………………………… Date……/……/………………

(Rev’d Fr Dr Kwamena William Coleman Sagoe)

Co-supervisor

Signature………………………… Date……/……/………………

(Dr. Adwoa Agyei)
DEDICATION

I dedicate this work to my wife, Tsoenee, my family and the World-Reinstatement Movement (WRM)
ACKNOWLEDGEMENT

Glory to God Almighty who always causes us to triumph in Christ and makes the savour of his knowledge known by us in every place (paraphrased). I am thankful to God Almighty for enabling me to undertake this thesis and to bring it to its successful completion.

I wish to also express my heartfelt gratitude to my supervisors Rev Dr Kwamena W.C Sagoe and Dr Adwoa Agyei for their invaluable role and support in directing and seeing to the completion of this work.

My sincere appreciation also goes to the lecturers and staff of the Department of Medical Microbiology for the opportunity to enrol in this programme and to be imparted with the requisite knowledge and training.

Many thanks also go to Dr Kyei who provided reagents for the laboratory analysis.

I cannot overlook the technical assistance of the staff of the Virology Unit of the School of Biomedical and Allied Health Sciences particularly the assistance of Mr Prince Pappoe-Ashong. I specially thank you for your contribution.

I am also highly indebted to my family for their love and support throughout the period.

Last but not least, I wish to express my profound gratitude to the training committee of the Kintampo Health Research Centre for granting me sponsorship to pursue this programme.
# TABLE OF CONTENTS

DECLARATION .............................................................................................................. i

DEDICATION ............................................................................................................. ii

ACKNOWLEDGEMENT ............................................................................................... iii

TABLE OF FIGURES .................................................................................................. vii

LIST OF TABLES ....................................................................................................... viii

LIST OF ABBREVIATIONS AND ACROYNMS ......................................................... ix

CHAPTER ONE ......................................................................................................... 1

1.0 INTRODUCTION ................................................................................................. 1

1.1 BACKGROUND .................................................................................................... 1

1.2 PROBLEM STATEMENT ..................................................................................... 4

1.3 JUSTIFICATION ................................................................................................. 5

1.4 AIM ..................................................................................................................... 5

1.5 SPECIFIC OBJECTIVES ..................................................................................... 5

CHAPTER TWO ......................................................................................................... 6

2.0 LITERATURE REVIEW ...................................................................................... 6

2.1 HEPATITIS B VIRUS AND ITS GENOME ....................................................... 6

2.1.1 GENOMIC CHARACTERISTICS OF THE PRECORE/CORE ORF ............. 8

2.1.2 GENOMIC CHARACTERISTICS OF THE CORE PROMOTER REGION .... 9

2.2 HBV REPLICATION ............................................................................................ 12

2.3 EPIDEMIOLOGY OF HBV .................................................................................. 15

2.3.1 PREVALENCE OF HBV AMONG PRISON IN-MATES .............................. 16
4.1.4 MUTATIONAL ANALYSIS ........................................................................................................ 50

CHAPTER FIVE .......................................................................................................................... 54

5.0 DISCUSSION AND CONCLUSION ..................................................................................... 54

5.1 DISCUSSION ........................................................................................................................ 54

5.1.1 PHYLOGENETIC ANALYSIS .......................................................................................... 54

5.1.2 MUTATIONAL ANALYSIS ............................................................................................ 56

5.2 LIMITATIONS ...................................................................................................................... 58

5.3 CONCLUSION ...................................................................................................................... 58

5.4 RECOMMENDATION .......................................................................................................... 58

REFERENCES ............................................................................................................................ 59

APPENDICES ............................................................................................................................... 75
TABLE OF FIGURES

Figure 2.1 Schematic structure of the Hepatitis B virion ................................................................................. 10
Figure 2.2 Genomic organization of the Hepatitis B virion .............................................................................. 11
Figure 2.3 Schematic illustrations of the relationship of the basic core promoter, precore, core regions and the X gene of the HBV genome ......................................................................................... 12
Figure 2.4 Schematic replication cycle of HBV ................................................................................................. 14
Figure 2.5 Geographic distribution of Hepatitis B genotypes ............................................................................ 21
Figure 2.6 Location and nature of the mutation stabilising the secondary structure of pregenomic encapsidation signal ........................................................................................................... 25
Figure 2.7 CD4+ and B-cell immunologic epitopes of the core protein .......................................................... 28
Figure 4.1 Representative Agarose Gel (2% in 1X TAE Buffer) electrophoregram of S-gene .................. 44
Figure 4.2 Representative Agarose Gel (2% 1X TAE Buffer) electrophoregram of S 1 genomic fragment .................................................................................................................................................. 45
Figure 4.3 Phylogenetic tree of 13 clones or contigs of sample 4S1 (precore/core ORF) ......................... 47
Figure 4.4 Phylogenetic tree of 12 clones or contigs of Sample 5S1 (precore/core ORF) ....................... 48
Figure 4.5 Phylogenetic tree of 14 clones or contigs of Sample 10S1 (precore/core ORF) ................... 49
LIST OF TABLES

Table 2.2 Core and precore mutations of clinical importance ........................................... 29

Table 3.1 A table showing the name, position on reference genome, sequence and purpose of all primers used .................................................................................................................................................................................. 38

Table 4.1 Summary of Demographic and serological data of study participants .................. 42

Table 4.2 Summary of PCR results showing number of successfully amplified targets and PCR positive rate ............................................................................................................................................................................................. 42

Table 4.3 Mutations detected in core and precore regions in the contigs/clones of 4S1 during analysis ............................................................................................................................................................................................................................................................. 51

Table 4.4 Mutations detected in core and precore regions in the contigs of 5S1 during analysis 52

Table 4.5 Mutations detected in core and precore regions in the contigs of 10 S1 during analysis ............................................................................................................................................................................................................................................................. 53
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3’</td>
<td>three prime</td>
</tr>
<tr>
<td>5’</td>
<td>five prime</td>
</tr>
<tr>
<td>Aa</td>
<td>amino acids</td>
</tr>
<tr>
<td>A (Ala)</td>
<td>Alanine</td>
</tr>
<tr>
<td>A</td>
<td>Adenine (nucleic acid – purine)</td>
</tr>
<tr>
<td>ALT</td>
<td>Alanine serum transaminase</td>
</tr>
<tr>
<td>anti-HBc</td>
<td>anti-HBcAg antibody</td>
</tr>
<tr>
<td>anti-HBe</td>
<td>anti-HBeAg antibody</td>
</tr>
<tr>
<td>anti-HBs</td>
<td>anti-HBsAg antibody</td>
</tr>
<tr>
<td>BCP</td>
<td>Basic Core Promoter</td>
</tr>
<tr>
<td>Bp (bp)</td>
<td>base pairs</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine (nucleic acid – pyrimidine)</td>
</tr>
<tr>
<td>CccDNA</td>
<td>covalently closed circular DNA</td>
</tr>
<tr>
<td>CHB</td>
<td>Chronic Hepatitis B</td>
</tr>
<tr>
<td>C-terminal</td>
<td>carboxyl terminal</td>
</tr>
<tr>
<td>D (Asp)</td>
<td>Aspartic acid</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
</tbody>
</table>
EBP
Enhancer Binding Protein

EcoR1
E. coli restriction site 1

ER
Endoplasmic Reticulum

ε
Encapsidation signal

F (Phe)
Phenylalanine

G (Gly)
Glycine

G
Guanine (nucleic acid – purine)

H (His)
Histidine

HBc
Hepatitis B core protein

HBeAb
anti-HBeAg antibody

HBeAg
Hepatitis B e-Antigen

HBsAb
anti-HBsAg antibody

HBsAg
Hepatitis B surface Antigen

HBV
Hepatitis B Virus

HBx
Hepatitis B x protein

HCC
Hepatocellular Carcinoma

HIV
Human Immunodeficiency Virus

I (Ile)
Isoleucine
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG</td>
<td>Immunoglobulin gamma</td>
</tr>
<tr>
<td>IgM</td>
<td>Immunoglobulin mu</td>
</tr>
<tr>
<td>INF</td>
<td>Interferon</td>
</tr>
<tr>
<td>IU</td>
<td>International Unit</td>
</tr>
<tr>
<td>K (Lys)</td>
<td>Lysine</td>
</tr>
<tr>
<td>Kbp</td>
<td>kilo base pairs</td>
</tr>
<tr>
<td>L (Leu)</td>
<td>Leucine</td>
</tr>
<tr>
<td>LHBs</td>
<td>Large Hepatitis B surface protein</td>
</tr>
<tr>
<td>M (Met)</td>
<td>Methionine</td>
</tr>
<tr>
<td>MHBS</td>
<td>Middle Hepatitis B surface protein</td>
</tr>
<tr>
<td>MHR</td>
<td>Major Hydrophilic Region</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>N (Asn)</td>
<td>Asparagine</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Centre for Biotechnology Information</td>
</tr>
<tr>
<td>NGS</td>
<td>Next Generation Sequencing</td>
</tr>
<tr>
<td>N-terminal</td>
<td>amino terminal</td>
</tr>
<tr>
<td>ORF</td>
<td>Open Reading Frame</td>
</tr>
<tr>
<td>P (Pro)</td>
<td>Proline</td>
</tr>
</tbody>
</table>
P1  forward primer
P2  reverse primer
PCR Polymerase Chain Reaction
PEG Polyethylene Glycol
pgRNA pre-genomic RNA
pre-C pre-Core
R (Arg) Arginine
rcDNA relaxed circular DNA
RNA Ribonucleic acid
RNAse Ribonuclease
RT Reverse transcriptase
S (Ser) Serine
sgRNA subgenomic RNA
SHBs Small Hepatitis B surface protein
SNP Single Nucleotide Polymorphism
SVPs Subviral Particles
T (Thr) Threonine
T Thymine (nucleic acid – pyrimidine)
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBE</td>
<td>Tris/Borate/EDTA</td>
</tr>
<tr>
<td>V (Val)</td>
<td>Valine</td>
</tr>
<tr>
<td>W (Trp)</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>Y (Tyr)</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>YMDD</td>
<td>tyrosine-methionine-aspartic acid-aspartic acid motif</td>
</tr>
<tr>
<td>α</td>
<td>alpha</td>
</tr>
<tr>
<td>µL</td>
<td>micro (mu) litre</td>
</tr>
<tr>
<td>ºC</td>
<td>degrees Celsius</td>
</tr>
</tbody>
</table>
ABSTRACT

Infection with hepatitis B virus constitutes a major public health problem globally. The lack of proofreading of the reverse transcriptase results in considerable mutations culminating into genotypes, sub-genotypes and variants. HBV quasispecies in an infected host has been linked to disease progression and therapeutic outcome. Limited data exist on the quasispecies diversity in other clinically relevant regions particularly the precore/core ORF which also is a determinant factor in the progression and treatment of this infection. Therefore, the need to explore this phenomenon is needed especially in populations with frequent exposure to HBV to help understand the pathogenesis of co-infections, recombination and evolution of HBV. The aim of this study was to examine the diversity and mutations at subpopulations level of HBV across the precore/core genomic region. Initially, viral DNA was extracted from twenty-six (26) archived plasma samples of prison-inmates who participated in a nationwide HBV/HIV surveillance study. Out of twenty-six samples, five (5) precore/core ORF (designated as S1) was successfully amplified and next-generation sequencing done with the Illumina MiSeq platform. Clones from 3 inmates were analysed for co-infection and clinically relevant mutations. A common pattern of mutations that was observed across the precore region was W28*. Additionally, in the core ORF mutations such as I97F, P130T and R181P were observed. Phylogenetic analysis of the individual clones of HBV revealed the predominance of genotype E at the subpopulation level. However, almost half (5 / 13) of the clones from 4S1 clustered with genotype G reference sequences. Majority of the 5S1 clones clustered with reference sequences of genotype E with the exception of one which clustered with genotype A. Clinically relevant mutations may be found in most quasispecies and will have implications for disease progression and e-antigenemia. HBV clones observed may give insight into the evolution of A/E and G/E recombinants.
CHAPTER ONE

1.0 INTRODUCTION

1.1 BACKGROUND

Hepatitis B viral infection constitutes a major public health problem worldwide. Globally two billion persons have been exposed to the virus and about 240 million are chronically infected. It is estimated that approximately 780000 chronic carriers die annually as a result of end-stage liver diseases including liver cirrhosis and hepatocellular carcinoma (HCC) (WHO 2015).

Globally, Sub-Saharan Africa and East Asia record the highest prevalence with 5-10% of infected adults progressing to the chronic carrier status while 85-90% of infants infected at birth become chronic carriers later in life (Gao et al., 2015). Ghana is recognized as one of the highly endemic countries in the sub-region with seroprevalence rates of 17% in HIV patients (Sagoe et al., 2012), 13% in pregnant women (Cho et al., 2012), 15% in blood donors (Mutocheluh et al., 2014) , and 25.5% in prison inmates (Adjei et al., 2008).

Prison-inmates appears to be the worst affected in terms of Hepatitis B virus (HBV) infection judging from the high seroprevalence rate among this population (Adjei et al., 2006; Adjei et al., 2008). Practices such as homosexuality, paid sex, injecting drug use, tattooing and needle sharing have been documented as some risk factors accounting for this high prevalence (Adjei et al., 2006; Moore et al., 2010). Thus prison-inmates pose a high risk and remain a portal source of transmission of the virus to the general population upon release (Jack 2011)
HBV belongs to the family *Hepdnaviridae* and has a partially-double stranded circular DNA genome of 3.2 kbp. The genome is organized into four overlapping open-reading frames (ORFs); polymerase (P), surface (S), precore/core (C), and the X ORFs (Miller *et al.*, 1989; Zhang *et al.*, 2016). The lack of proof-reading of the reverse transcriptase results in high mutation rate with the attendant effect being the generation of genetic variants. Factors such as high replication rate ($10^{11}$ virions per day) (Gong *et al.*, 2013) and host enzymatic factors (Homs *et al.*, 2014) have also been attributed to this phenomenon.

Thus at each point in the host circulation is a heterogeneous viral population made up of closely related but un-identical genetic variants termed quasispecies. Viral quasispecies, have been observed to have important clinical implications on viral transmission, viral persistence and liver damage (Domingo and Gomez 2007; Gong *et al.*, 2013). These outcomes are driven by viral mutants which emerge from the viral quasispecies under the selective pressure of host immune system or antiviral therapy. Thus in the absence of antiviral therapy, the immune system is the primary driver of natural evolution of the virus which results in increased viral quasispecies diversity (Homs *et al.*, 2014).

The precore/core ORF of the HBV genome has been described as the preferred site in the study of quasispecies variability (Rodriguez-Frias *et al.*, 2013; Homs *et al.*, 2014). This is against the background that the precore/core ORF harbours dominant epitopes to which host immune system recognizes and responds to in driving the natural evolution of the virus (Alexopoulou 2009). In addition, the precore/core ORF is the only non-overlapping segment of the HBV genome which regulates viral replication (Rodriguez-Frias *et al.*, 2013; Homs *et al.*, 2014).
Studies have reported the role of precore/core mutations in the natural progression of chronic HBV infection and their effect on antiviral response (Tseng et al., 2015; Yang et al., 2008; Chou et al., 2007; Yan et al., 2015). Mutations such as A1792T/G1794A in the basal core promoter have been associated with increased risk of liver cirrhosis and HCC development (Tseng et al., 2015; Yang et al., 2008; Chou et al., 2007). Some studies report a decreased risk in HCC development with the G1896A precore mutation whereas others report otherwise (Yang et al., 2008). It’s also been reported that the percentages of basal core and core variants prior to therapy influence the outcome of interferon therapy which makes the study of viral quasispecies at the precore/core ORF very crucial (Yang et al., 2013).

Direct sequencing by Sanger after cloning or PCR has been the mainstay in the investigation of viral quasispecies variability. (Yamani et al., 2015). However, this method is unable to detect viral mutants in less than 20% of the total viral population (Yamani et al., 2015; Solmone et al., 2009). Hence sequence-specific methods such as line-probe assay (LIPAs), restriction fragment length/mass polymorphism, oligonucleotide microarray and gene chip technology and pyrosequencing, have been designed to address the shortfall of the Sanger sequencing. However, these methods only detect specific mutations and minor variants to as low as 5% (Solmone et al., 2009).

The advent of next-generation sequencing (NGS) technologies has revolutionized the study and detection of viral mutations in a heterogeneous viral population. This method which involves the massive parallel pico-liter-scale amplification and sequencing of individual DNA molecules makes it possible to analyse thousands of clonally amplified regions of about 200 nucleotides simultaneously. Thus NGS allows for the detection of minority viral variants (Yamani et al., 2015; Gong et al., 2013; Solmone et al., 2009).
Application of NGS has been seen in HBV studies that sought to characterize HBV mutations at the RT region of the viral polymerase (Solmone et al., 2009) and even the major hydrophilic region (MHR) of the S-gene (Yamani et al., 2015). However, few has focused on detecting and determining the quasispecies variability in the precore/core segments of the HBV genome especially among unique populations such as prison inmates.

1.2 PROBLEM STATEMENT

HBV quasispecies composition and evolutionary changes that occur are critical factors linked to the control and treatment of chronic hepatitis B infection (CHB). Recent reports lend credence to the fact that quasispecies diversity is a clinically relevant determinant of the progression, prognosis and response to treatment (Homs et al., 2014). In view of this studies have explored quasispecies diversity in segments of the HBV genome like the reverse transcriptase and the major hydrophilic region of the S-gene using next-generation sequencing (Nishijima et al., 2012; Yamani et al., 2015; Gong et al., 2013). However, limited data exist on the quasispecies diversity in other clinically relevant regions particularly the precore/core ORF which also is a determinant factor in the progression and treatment of this infection. Therefore, the need to explore this phenomenon.
1.3 JUSTIFICATION

Most studies investigating quasispecies diversity have been done in the clinical setting. However, few or none has been conducted among prison-inmates who represent a unique group of people with high HBV transmission. Given the precore/core as the only non-overlapping segment of the HBV genome with dominant epitopes which drives the immune response (Alexopoulou 2009), it will be useful to investigate quasispecies diversity in this region in order to gain much understanding into the natural evolution of the virus among inmates who invariably are treatment-naive. Furthermore, the HBV A/E recombinant has been found in the minority in Ghana but it is unclear how the evolution of this recombinant occur.

1.4 AIM

The aim of this study is to evaluate HBV species diversity and mutations at subpopulation level in the precore/core ORF of the HBV genome.

1.5 SPECIFIC OBJECTIVES

I. To determine the phylogenetic relatedness of HBV clones across the precore/core region.

II. To determine the patterns of mutations in the core and precore regions at subpopulation level
CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 HEPATITIS B VIRUS AND ITS GENOME

HBV is a prototype member of the family Hepadnaviridae and genus Orthohepadnavirus. It is one of the smallest enveloped animal viruses and measures 42nm in diameter in the matured virion. The matured virion, a Dane particle, consists of a partially double-stranded circular DNA genome of 3.2 kbp enclosed by an icosahedral nucleocapsid. The nucleocapsid is in turn enclosed by the viral envelope containing the three viral encoded proteins, the small (SHBs), middle (MHBs) and large (LHBs) surface proteins (Seeger and Mason 2000; Coppola et al., 2015) as depicted in figure 2.1 below.

The HBV genome is organized into four overlapping open reading frames (ORF): pre-S/S (surface proteins), pre-C/C (pre-core/core), X (transcriptional co-activator) and P (DNA polymerase). The pre-S/S ORF is embedded in the P ORF but translated in a different reading frame. ORFs C and X, however, overlap the P ORF by ¼ and 1/3 of their respective sequence lengths (Miller et al., 1989; Zhang et al., 2016). Figure 2.2 below shows the overlapping ORF’s.

The Pre-S/S ORF encodes three structurally related envelope proteins which are synthesized from alternative initiation codons. These proteins namely the large (LHBs), middle (MHBs) and small (SHBs), are encoded specifically by the Pre-S1, pre-S2 and S genes respectively. These proteins although have the same carboxy terminus differ in terms of their amino-terminal extensions. As such the SHBs protein composed of 226 amino acids (aa), extends by an extra 55aa at its N-
terminal portion to form the M-protein whiles the L protein extends by an extra 108aa or 119aa depending on the genotype (Seeger and Mason 2000; Zhang et al., 2016).

The S protein forms the Hepatitis B surface antigen (HBsAg) and serves as a marker of infection and chronicity. Apart from it being the primary marker of diagnosis of infection, HBsAg is also the target of immunoprophylaxis. The dominant epitopes of this antigen to which neutralising antibodies bind to reside in the “α” determinant (124-147 aa) region which lies within the major hydrophilic region spanning the 99-169aa (Carman et al., 1990; Zuckerman and Zuckerman 2003; Coppola et al., 2015).

Another important aspect of the viral genome is the ORF P which encodes the viral polymerase. This enzyme apart from its polymerization activity also functions as a reverse transcriptase and an RNase. The unique feature of this polymerase similar to the retrovirus is its lack of the 3′-5′-exonuclease proofreading capacity. As a result, HBV shows an estimated mutation rate of 1.4 x 10^{-5} - 3.2 x 10^{-5} nucleotide per site per year (Girones and Miller 1989; Chotiyaputta and Lok 2009) higher than other DNA but less than RNA viruses. The implication of this high mutation rate is the generation of genetic variants, genotypes, subgenotypes and quasispecies during viral replication. These mutations are known to occur either spontaneously or under immune or antiviral pressure. Thus viral mutants can be generated in the absence of antiviral therapy and possibly be transmitted to a susceptible host (Lazarevic 2014).

One other crucial part of the viral genome is the ORF X. The ORF X encodes a viral transactivator protein (HBxAg) which modulates both host and viral gene expression. Studies have shown that the HBxAg encoded by the ORF X play a crucial role in the pathogenesis and development of HCC. By interacting with various signalling pathways, the HBxAg causes a dysregulation in the cell growth, cell cycling process and apoptosis (Zhang et al., 2016; Sukowati et al., 2016).
2.1.1 GENOMIC CHARACTERISTICS OF THE PRECORE/CORE ORF

The Pre-C/C ORF encodes the precore and the core proteins, two distinct proteins, derived from two in-frame initiation codons (ATG). The first initiation codon leads to the transcription of the precore mRNA and the second initiation codon to the transcription of the pregenomic RNA (Alexopoulou and Karayiannis 2014). Upon translation of the precore mRNA, the precore/core protein which results is directed to the endoplasmic reticulum by a 19aa signal peptide at its amino terminus. The signal peptide is later cleaved and the resultant protein released into endoplasmic reticulum lumen where further processing results in a removal of 34aa from the carboxy terminus end. The resulting protein is the soluble HBeAg which is released into circulation. HBeAg shares 149aa with HBcAg except that it differs from the latter by 10aa which it retains from the precore region. Despite this similarity, both proteins differ in antigenic epitopes due to conformational changes in HBeAg protein (Ou et al. 1986; Alexopoulou 2009).

HBeAg is thought not to have any biological function. However, it has been suggested that it induces tolerance in neonates born to infected mothers as it crosses the placenta and perhaps underlies the high number of neonates who progress to the chronic carrier stage. Similarly, HBeAg is thought to have an immunomodulatory function in infected adults by suppressing the T-helper cells thereby perpetuating the infection (Alexopoulou 2009).

The pregenomic RNA, on the other hand, is translated into both core and polymerase proteins. Assembly of the core particles into the viral nucleocapsid coincide with encapsidation of the pgRNA and the viral polymerase. Initiation of encapsidation occurs when viral polymerase binds to a stem-loop structure at the 5’ end of pgRNA termed as the encapsidation signal or ε. This signal is not only important in pgRNA packaging but also in the initiation of reverse transcription (Alexopoulou 2009).
2.1.2 GENOMIC CHARACTERISTICS OF THE CORE PROMOTER REGION

The core promoter (CP) is important in viral replication by directing the transcription of the precore and pgRNAs. It consists of the basic core promoter and its adjoining upstream regulatory sequences. The BCP overlaps the 3’ end of the X ORF and the 5’ end of the precore ORF and occurs between nucleotide 1742-1849 adjacent the pre-core region (Figure 2.3) (Alexopoulou 2009). The BCP by means of cis-acting elements independently directs the transcription of both the precore mRNA and pgRNA. Thus the role of BCP in regulating viral replication is critical and mutations in this region may impact on viral gene expressions, viral replication and the overall viral pathogenesis (Yu and Mertz 1996; Alexopoulou and Karayiannis 2014).

Regulating the activity of the core promoter is the function of the enhancer II element which overlaps partially with the CP and its upstream regulatory sequences. The region also harbours nucleotide motifs which make up the transcription factor binding sites (Alexopoulou 2009).
Figure 2.1 Schematic structure of the Hepatitis B virion.

Viral DNA genome with the polymerase protein is enclosed by the icosahedral capsid which is in turn encased by the viral envelope. Embedded within the envelope are the small, middle and large antigens (Perkins .2002).
Figure 2.2 Genomic organization of the Hepatitis B virion

The overlapping ORFs X, S, core, and the polymerase are shown as arrows (Jalali 2006).
Figure 2.3 Schematic illustrations of the relationship of the basic core promoter, precore, core regions and the X gene of the HBV genome.

Also indicated are transcriptional regulatory elements, the promoters, enhancers and direct repeat (Alexopoulou 2009).

2.2 HBV REPLICATION

Replication of the hepatitis B virus begins with the entry of the virus into hepatocytes, the site of predilection. Even though the precise mechanism of viral entry is not fully understood, two entry pathways have been proposed. The endocytic pathway in which viral nucleocapsid is released via endocytic vesicles into the cytoplasm and the fusion pathway in which viral envelope fuses directly with the plasma membrane (Grimm et al., 2011). Viral particle binds to an unknown hepatocyte-specific pre-S1 receptor after attachment of liver cells membranes to probably cell-associated
heparan sulfate proteoglycans (Urban et al., 2010). Following this, viral nucleocapsids containing relaxed circular partially double-stranded DNA (rcDNA) genome together with a viral polymerase is released into the cytoplasm. Upon entry, viral nucleocapsid is transported via microtubules to the nucleus where rcDNA is repaired to form the covalently closed circular DNA (cccDNA). This repair involves completion of the incomplete rcDNA plus-strand by viral polymerase using RNA-primers both of which are removed subsequently by cellular enzymes. The resulting DNA strands are then covalently ligated to form the cccDNA (Urban et al., 2010; Grimm et al., 2011; Seeger and Mason 2000). Subsequently, cccDNA is transcribed into sub-genomic RNA (sgRNA) and pregenomic RNA (pgRNA) using the host cell transcription machinery. Both host transcriptional factors such as enhancer binding protein (EPB) and hepatocyte nuclear factors (HNF) and viral proteins (core and regulatory X-protein) regulate this process thereby modulating the expression of viral genes as these interact with promoters of the four open reading frames (ORFs) (Levrero et al., 2009). The sgRNA is translated in the cytoplasm into the three enveloped proteins namely the L, M and S proteins coded by the L, M and S genes respectively of the S ORF and the regulatory X-protein of the X ORF. On the other hand, the pgRNA is translated into the core protein and the viral polymerase. Some pgRNAs form complexes with the core and viral polymerase via its epsilon stem-loop structure and auto assemble into RNA-containing viral nucleocapsids (Urban et al., 2010). RNA-containing nucleocapsid undergoes maturation by reverse transcription of pgRNA followed by a plus-strand DNA synthesis to form a DNA-containing nucleocapsid. The lack of proofreading of the reverse transcriptase results in frequent mutations in the viral genome leading to the production of genetically distinct viral species (quasispecies). The resulting DNA-containing nucleocapsids serve a dual purpose. Whereas some are re-imported into the nucleus to form additional cccDNA molecules the rest are enveloped and secreted via the endoplasmic reticulum
(ER) as shown in figure 2.4 below. After budding into the ER lumen, the envelope proteins are secreted by the cell either as small, non-infectious subviral spherical or filamentous particles (SVPs) of 22 nm diameter or as infectious virions of 42 nm (Dane particles). In most cases, the non-infectious SVPs outnumber the infectious virions in excess of a 1,000-to 1,000,000-fold (Grimm et al., 2011).

**Figure 2.4** Schematic replication cycle of HBV. Retrieved from [www.clinicaloptions.com](http://www.clinicaloptions.com)
2.3 EPIDEMIOLOGY OF HBV

Humans are the sole reservoir for HBV which is considered highly infectious, about 50-100 times compared to HIV (Horvat 2011). The infection presents either as acute or chronic. Acute infection represents the onset of infection characterized by the presence of HBsAg, immunoglobulin M antibody to the core antigen, HBcAg and HBeAg. The incubation period in the acute stage spans from one month to 6 months upon infection. Most people in the acute stage are asymptomatic with a few showing symptoms that last for a few weeks. Typical symptoms include loss of appetite, nausea, vomiting, fatigue, abdominal pain and jaundice in extreme cases. The infection is self-limiting in about 65% of cases, however, a small percentage develop acute liver failure which could lead to death (Zhang et al., 2016).

Chronic carrier status by convention is characterized by the persistence of the HBsAg in the blood of an infected person for a minimum of (6) months and the switch from IgM to IgG to the core antigen with or without HBeAg. (Crawford 1999). According to WHO report of 2015, about 780,000 of chronic carriers die each year as a result of liver cirrhosis and hepatocellular carcinoma. Transmission of the virus can either occur perinatally or by apparent or inapparent percutaneous or permucosal exposure to blood or other body fluids of an infected person. Risk factors for infection include transfusion with infected blood and blood products, sexual promiscuity, intravenous drug injection, tattooing and renal dialysis (Liaw and Chu 2009). Infants infected in-utero stand a high chance of becoming chronic carriers. Among this, about 90% of infected neonates become chronic carriers, and the risk of chronic infection is up to 30% in children infected at 1-4 years of age (Hyams 1995; Gao et al., 2015). Globally, it has been estimated that individuals infected during adulthood have a 5% risk of progressing to the chronic carrier status (Liaw and Chu 2009).
Infection rates vary from region to region with Southeast Asia and sub-Saharan Africa with the highest rates 8.0-20.0% (Lavanchy 2004).

In Ghana, however, different prevalence have been reported among different study populations. A seroprevalence of 17% have been reported among HIV patients (Sagoe et al., 2012), 13% among pregnant women (Cho et al., 2012), 15% among blood donors (Mutocheluh et al., 2014) and 25.5% among prison inmates (Adjei et al., 2008)

2.3.1 PREVALENCE OF HBV AMONG PRISON IN-MATES

Several studies point to the relationship that exists between incarceration and high transmission of blood-borne viruses and HBV is no exception. In fact, it has been documented that the prevalence of HBV among prison-inmates is higher than what pertains in the general population (Adjei et al., 2006; Adjei et al., 2008; Moore et al., 2010). In view of this, prison-inmates represents a high-risk group in terms HBV and serve as a reservoir of infection even among themselves as well as the general populace. Whereas some individuals were infected before incarceration there seems to be on-going of transmission within the prisons demonstrated by some studies (Angela Ayiku, student research work; Moore et al., 2010). In Ghana study by Adjei et al., 2006 puts the HBV prevalence among prison-inmates to be 17.4% and 25.5% respectively (Adjei et al., 2008).

Some risk factors attributed to this high prevalence among prison-inmates are intravenous drug injection, previous imprisonments, body piercing, tattooing and high-risk sexual behaviours such as sodomy, homosexuality and paid sex. Other studies report of the poor state of prisons, and inadequate health care as contributing factors to the high HBV prevalence (Adjei et al., 2006; Moore et al., 2010).
2.4 NATURAL HISTORY OF CHRONIC HEPATITIS B INFECTION

Hepatitis B virus is not cytopathogenic and therefore injury to the liver is mainly due to the host cell-mediated immune response to infected hepatocytes. Chronic hepatitis B is dynamic and the course of the disease is influenced by the interplay between the host immune system and the virus (Gao et al., 2015). Clinically, four phases of the disease have been identified: immune tolerant phase, immune clearance, inactive, and reactivation phase (Azmi et al., 2014). The immune tolerant phase is characterized by persistently normal serum alanine aminotransferase (ALT) levels, high HBV DNA levels and presence of Hepatitis B e antigen, but with no evidence of liver injury. The immune clearance phase is characterized by presence of HBeAg, persistently high ALT and HBV DNA levels with some degree of liver inflammation. However, HBeAg seroconversion may occur at the late stage of the immune clearance. Patients may progress from this stage to the immune inactive phase which is characterized by normal ALT level, low/undetectable HBV DNA (< 2000 IU/mL or < 10^4 virus copies/mL), absence of HBeAg and presence of anti-HBe, as well as no or minimal histological injury. The reactivation phase is characterized by rebound viraemia, presence of anti-HBe, elevated ALT levels, liver inflammation and the absence of the HBeAg. (Gao et al., 2015). It is noteworthy that not all chronically infected individuals go through the four phases and even if they do not sequentially. Certain factors such as age, route of infection affect the phases individuals go through. Individuals infected at an early age begin from the immune tolerant phase which lasts for about 20-30 years (Hadziyannis and Papatheodoridis 2006a; Alexopoulou and Karayiannis 2014).
2.5 TREATMENT OF HEPATITIS B

Over the last 15 years, treatment of hepatitis B has seen a marked improvement owing to the availability of antiviral drugs which has increased therapeutic options. At the moment, two classes of antiviral agent have been approved for treatment which includes; standard or pegylated interferon (Peg-IFN) and five nucleoside or nucleotide analogues (Santantonio and Fasano 2014).

The goal of antiviral therapy in the chronic hepatitis B is to achieve sustained suppression of viral replication so as to prevent or reduce the chances of progressing to liver cirrhosis, HCC and ultimately liver-related deaths. In fact, studies have demonstrated a reduction in risk of disease progression with a marked reduction in viral replication as a result of antiviral therapy (Mommeja et al., 2003; Chen et al., 2011; Santantonio and Fasano 2014).

Peg-IFN is similar but different from standard interferon and has come to replace the latter because of the convenience in administration (once weekly compared to thrice weekly subcutaneous injection), longer serum half-life and effective viral suppression. There are two forms of the pegylated interferon; Peg-IFN alpha 2a and Peg-IFN alpha 2b. Both are similar in terms of efficacy but Peg-IFN alpha 2a has been licensed for use globally. The interferons have immunomodulatory effect as well as antiviral activity (Santantonio and Fasano 2014).

Nucleoside or nucleotide analogues are taken orally and include lamivudine, telbivudine, entecavir (nucleoside analogues) and tenofovir, adefovir (nucleotide analogues) (Gish 2009). Lamivudine which used to be the first line treatment drug has been discouraged from use due to a high level of virological resistance. At the moment entecavir and tenofovir have become the first line treatment drugs because of their high barrier to resistance and their ability to achieve adequate viral suppression (Kim et al., 2014). The mode of action of virtually all nucleoside or nucleotide
analogues is the premature termination of viral DNA replication by incorporation of the enzymatically modified analogues into the growing DNA chain (Kim et al., 2014).

2.6 HBV GENOTYPES

HBV is classified into ten genotypes (A-J) based on at least 8% nucleotide divergence in their full-length genome (Okamoto et al., 1988; Norder et al., 1992a; Norder et al., 1992b; Yano et al., 2015). These genotypes show distinct geographical distribution (figure 2.5). Genotype A is found mainly in Africa, United States of America and Europe, HBV/B and HBV/C mainly in the Asia; genotype D is dominant in Africa, Europe, Mediterranean countries, and India; genotype G is observed in France, Germany, and the United States; genotype H is commonly encountered in Central and South America. In recent times genotypes, I have been reported in Vietnam and Laos and genotype J identified in the Ryukyu Islands in Japan (Sunbul 2014).

Genotype E, on the other hand, is known to be predominant in West and Central Africa, spanning over a wide geographical area from Senegal to Namibia and extending as far as to Central African Republic in the East (Botha et al., 2005; Hübschen et al., 2008). Although studies reports of a low genetic diversity in genotype E, strains from Ghana and Benin are reported to be highly diverse reflecting the relatively longer natural history of this genotype in these countries (Botha et al., 2005; Hübschen et al., 2008; Ampah et al., 2016). In Ghana just like countries in Sub-Saharan Africa, HBV/E is the predominant genotype with prevalence rates of (96.7%–99.1%), 95.1% and 100% based on reports by Huy et al., 2006, Garmiri et al., 2009 and Ampah et al., 2016 respectively.
Based on an intragenotypic diversity of 4% certain genotypes have been further classified into sub-
genotypes. These are A (A1-A7), genotype B (B1-B9), genotype C (C1-16), genotype D (D1-D8),
and genotype F (F1-F4) (Fang 2004; Cao 2009). Genotype E, however, is not classified into sub-
genotype because of the low intragenotypic diversity of 1.8% (Huy et al., 2006).

Inter-genotypic recombinants have also been observed to occur in HBV when genomes of different
genotypes co-exist in the same host. Recombination is an essential factor of HBV genetic
variability with probable clinical implications (Boyce et al., 2017; Kay and Zoulim 2007). Studies
have reported on recombinants such A/D, A/E, C/D, B/C and G/C reflecting the effect of migration
on the epidemiological landscape of the infection (Sugauchi et al., 2002). According to Garmiri et
al., 2009, 4.7% A/E recombinants were found among blood donors from Guinea and Ghana. The
identical point of A/E recombination among multiple blood donors strongly suggested that these
strains are in circulation and are infectious and transmissible (Garmiri et al., 2009). Boyce et al.,
2017, also reports of the detection and confirmation of D/E recombinant by next-generation
sequence (NGS) in an HIV patient co-infected with HBV in Ghana. In certain regions, these
recombinants have been observed to be gaining ascendancy over the parent genotype as is the case
with B/C recombinant in the Asia (Sugauchi et al., 2002).

Apart from genotype recombination, some studies have reported on mixed infection or dual
infection. Most importantly dual infection of genotype G with other genotypes, in particular, has
been observed. For instance, G-A, G-D, G-H and G-F dual infection have been reported (Kato et
al., 2002; Beggel et al., 2012; Kuyl et al., 2013). Genotype G has been noted as an aberrant
genotype which lacks replication fitness and therefore rarely exist as a mono-infection. However
in the presence of other genotypes replication fitness is restored and can outcompete the other
genotype (Kuyl et al., 2013).
Genotypes and sub-genotypes have been observed to influence disease progression, clinical outcome and response to antiviral therapy. For instance, genotype C has been observed to be highly associated with increased risk of hepatocellular carcinoma compared to genotype B. In the event of an acute onset of hepatitis B, sub-genotype C2 is more likely to cause chronic infection than sub-genotype B2 (Zhang et al., 2016).

**Figure 2.5** Geographic distribution of Hepatitis B genotypes (Hussain 2013).
2.7 HBV SEROTYPES AND SUBTYPES

HBV is classified into four serotypes and ten subtypes. Serotypes and subtypes arise as a result of amino acid substitutions in the major hydrophilic determinants of the HBsAg or the S protein. The S protein of the hepatitis B virus bears a pair of mutually exclusive determinants d or y and w or r which arise from single amino acid substitutions at positions 122 and 160 of the HBsAg respectively (Yano et al., 2015). These mutually exclusive determinants in conjunction with the major “a” determinant which is common among all subtypes results in four major serotypes designated adr, adw, ayr and ayw and ten identifiable subtypes namely adw2, adw4, adrq+, adrq−, ayw1, ayw2, ayw3, ayw4, ayr and adw3 (Norder et al., 1992; Magnus and Norder 1995; Yano et al., 2015) Studies have identified ayw4 as the predominant subtype in Ghana (Huy et al., 2006; Ampah et al., 2016; Garmiri et al., 2009).

Serotypes and subtypes like genotypes, show distinct geographic distributions and affect the antigenic characteristics of the hepatitis B virus (Le Bouvier et al., 1972; Norder et al., 1992). Serotypes play an important role in the production of HBV vaccines and in the development of HBV diagnostic assays. By far serotype adw of genotype A 2 has been used as the precursor for all genetically engineered HBV vaccines (Lin et al., 2013). All HBV vaccines thus contain the recombinant HBsAg derived from yeast cultures. In the development of diagnostic assays, antibodies raised against the immunodominant “a” determinant region are employed in these assays to detect the HBsAg. Mutations inducing a conformational change in the “a” determinant region can alter its antigenic epitopes allowing the escape of HBsAg from detection by diagnostic assays (Horvat 2011).
2.8 HBV MUTATIONS

The error-prone nature of the reverse transcriptase of HBV results in considerable genetic variation during replication. As such the HBV evolves constantly generating viral mutants in the form of genotypes, subgenotypes, and quasispecies (Zhang et al., 2016). Both endogenous factors such as the individual’s immunity, and exogenous factors such as vaccines, and antiviral agents have been documented to drive the selection of mutants (Malik et al., 2012). Mutations in the HBV genome not only impact on the replication fitness of the virus but also influence disease progression and response to antiviral agents (Kidd-Ljunggren et al., 2002; Malik et al., 2012). These mutations are not limited to the polymerase or reverse transcriptase region (P ORF) but have also been observed in the Pre-S/S ORF, Pre C/C ORF and the X ORF as well. Studies over the period have explored the effects of mutations in these ORFs on viral pathogenesis and have shown that these mutations are associated with disease outcomes (Kidd-Ljunggren et al., 2004; Malik et al., 2012)

2.8.1 PRECORE VARIANTS

It has been observed that most patients in the Mediterranean countries infected with HBV harbour the precore mutant virus (Alexopoulou 2009). A point mutation at position 1896 of the precore region which converts a tryptophan codon (codon 28), TGG, to a translational stop codon (TAG) occur more readily accounting for this observation. The effect of this mutation is the abrogation of HBeAg translation. As such this precore stop codon is predominantly seen in HBeAg negative CHB patients (Shi et al., 2012; Lazarevic 2014). This mutation is selected during seroconversion from an HBeAg positive immune clearance phase to an HBeAg negative inactive phase. The absence of the HBeAg results in the loss of the tolerogenic effect HBeAg has on the immune system. Thus the immune system becomes activated leading to seroconversion (Alexopoulou
Other mutations with similar phenotypic effect are loss of the precore/core translation initiation codon from (ATG to ACG or CUG), frameshift and deletions, as well as mutation of the second codon to a stop codon leading to the synthesis of nonsense proteins (Okamoto et al., 1990; Alexopoulou and Karayiannis 2014). It has been noted that certain genotypes are more predisposed to expressing the pre-stop codon than others. Genotypes such as B, D and E have a high predisposition to this mutation compared with genotypes A, C and F because of the presence of a T nucleotide at this position (Jalali 2006; Alexopoulou 2009). It’s been noted that at nucleotide 1858 codon 15(CCC) a C nucleotide base pairs with a G nucleotide at position 1896 codon 28(TAG) which give rise to a more stable encapsidation secondary structure. However, for genotypes such as B, C and D, the nucleotide at position 1858 is a T which favours the replacement of G to A at position 1896 of codon 28 (TAG) as illustrated in figure 2.6 below. The conversion of a TGG to a TAG stop codon ensures the stability of the secondary structure and proper functioning of the encapsidation signal. Conversely, a disruption of this C-G base pairing in genotype A because of the development of the stop codon destabilizes the stem-loop structure and affects the encapsidation signal. This explains why G1896A and for that matter, HBeAg negative mutant is rare in Genotype A and common in genotype D (Alexopoulou 2009).
Figure 2.6 Location and nature of the mutation stabilising the secondary structure of pregenomic encapsidation signal in genotype A and non-A genotypes (Alexopoulou 2009).

2.8.2 BASIC CORE PROMOTER VARIANTS

Mutations in the BCP region may impact on viral gene expressions, viral replication and the overall viral pathogenesis (Yu and Mertz 1996; Alexopoulou and Karayiannis 2014). Reports of basal core promoter variants almost invariably center on the double mutations G1762A and A1764T which has been described in varied states of HBV infection (Okamoto et al., 1994) They have been detected with increased frequency in patients with fulminant hepatitis (Hou et al., 2002), HBeAg and anti-HBe positive chronic hepatitis (Horikita et al., 1994), and hepatocellular carcinoma (Hsia et al., 1996) but less so in asymptomatic chronic carriers (Kurosaki et al., 1996). Although this mutation seems to be present in varied states of HBV disease in different geographical areas, some report the rarity of these double mutations particularly in patients with
fulminant hepatitis B living in the United States and Brazil. This observation may be related to the differences in the prevailing genotypes in these areas (Blackberg and Kidd-Ljunggren 2000).

This double mutation is known to reduce expression of the HBeAg as a result of decreased precore mRNA. However, by means of up-regulation of the pregenomic mRNA, there is increased viral replication (Hakami et al., 2013; Alexopoulou and Karayiannis 2014). BCP mutation has been observed to occur in HBV infected persons irrespective of their HBeAg status. Whereas a section of investigators have associated HBeAg negativity due to BCP double mutation to severe liver disease, others did not come to this conclusion (Alexopoulou 2009).

According to Chan et al., 1999, genotypes which harbour a C at nucleotide position 1858 are more predisposed to developing the A1762T/G1764A mutations in contrast to the precore mutant G1896A which has a T in this same nucleotide position. Moreover, this double mutation has been observed to be often accompanied by a change at position 1753, from T to C or G among patients who are anti-HBe positive (Nagasaki et al., 1998). This core promoter mutant T1753C, as well as C1766T, apart from the classical A1762T/G1764A in the core promoter have been shown to possibly impact strongly on viral replication and HBeAg expression (Parekh et al., 2003).

Apart from substitution mutations, deletions of 1-21 nucleotide base pairs have also been reported by other studies within the core promoter region in different clinical cases of HBV infection often characterized by low or high viraemia (Parekh et al., 2003).

2.8.3 CORE GENE VARIANTS

The core protein is known to have antigenic epitopes which are targets of B and T cells immune response. As such the capsid protein, from the core protein, has the propensity to elicit B-cell activation and proliferation producing immunoglobulins both in a T cell-dependent and a T cell-
independent fashion (Milich et al., 1997). Thus both B cell epitopes and T-cell epitopes are essential in the production of capsid-specific antibodies by activated B-cells. The immunodominant CD4+ T-cell epitopes span residues 1-20 and 50-69 or 1-25 and 61-85 and 117-131 in European and Fast Eastern patients respectively. On the other hand, the B-cell major antigenic determinant resides around amino acid position 80, namely 74-89, the second one at residues 107-118 and a third at amino acid 130-138 (see figure 2.7) (Alexopoulou 2009).

It has been established that infection with precore stop codon invariably results in a chronic liver disease with acute exacerbations and the reason is that this variant shows a high prevalence of mutations in the core region (Hadziyannis and Papatheodoridis 2006b). Mutations in the core region which results in amino acid changes may alter the antigenic epitopes in the core region and may have adverse clinical implications (see table 2.1). This is corroborated by studies in which treatment naïve patients with an active liver disease have majority of the core protein amino acid changes concentrated in both B- and T-cell epitopes (Alexopoulou et al., 1997; Alexopoulou et al., 1998).

Heterogeneity in the core protein occurs during infection and studies have linked this to viral persistence. At each point of HBV chronic infection mixed viral populations or quasispecies occur including mixed core variants (Alexopoulou et al., 1997). It has been shown that novel variants may emerge during the course of the infection which were absent at the initial stages but their extremely low numbers make them undetectable. Some minor variants subsequently become dominant as a result of selective pressure from the host immune system. Thus the emergence of a new mixed viral population of core variants and the selection of new mutations that predominated later at any one-time point may be due to the amino acids substitution in B- and CD4+ T- cell epitopes (Alexopoulou et al., 1997; Alexopoulou 2009).
Deletions of varying lengths occur within the core gene and this usually affects the central core region and the carboxy-terminal region (Akarca and Lok 1995; Ackrill et al., 1993; Garmiri et al., 2009). The effect of such deletions may result in truncated core species i.e. if it is an in-frame deletion or premature core synthesis termination that is if it’s out of frame deletions. Deletion core variants and their attendant effect have been reported in both immunocompetent (Marinos et al., 1996) and immunosuppressed (Gunther et al., 1996) individuals. Among immunocompetent, this variant was associated with low viraemia and early anti-HBe seroconversion whereas in long-term immunosuppressed individuals following kidney or liver transplants this variant was associated with liver cirrhosis and end-stage liver disease.

**Figure 2.7** CD4+ and B-cell immunologic epitopes of the core protein (Alexopoulou 2009).
Table 2.1 Core and precore mutations of clinical importance

<table>
<thead>
<tr>
<th>CORE MUTATIONS</th>
<th>PRECORE MUTATIONS</th>
<th>NUCLEOTIDE CHANGES</th>
<th>CLINICAL IMPLICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W28*Stop</td>
<td>G1896A</td>
<td>Change in HBeAg serostatus</td>
</tr>
<tr>
<td>P5R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D32N/H</td>
<td></td>
<td>G1994A/C</td>
<td></td>
</tr>
<tr>
<td>E43K</td>
<td></td>
<td>G2027A</td>
<td></td>
</tr>
<tr>
<td>P50Y/H/A</td>
<td></td>
<td>C2048T/G, C2049A</td>
<td></td>
</tr>
<tr>
<td>A131P/N/G</td>
<td></td>
<td>G2291C/A, C2292A/G, T2293C</td>
<td></td>
</tr>
<tr>
<td>P135Q</td>
<td></td>
<td>C2304A</td>
<td></td>
</tr>
<tr>
<td>S181P/H</td>
<td></td>
<td>T2441C, C2442A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W28*Stop</td>
<td>G1869A</td>
<td>Disease Severity (association with HCC)</td>
</tr>
<tr>
<td>P5T/L/H</td>
<td></td>
<td>C1913A, C1914A/T, G1915A/C</td>
<td></td>
</tr>
<tr>
<td>P5R</td>
<td></td>
<td>C1914G, ,</td>
<td></td>
</tr>
<tr>
<td>E83D</td>
<td></td>
<td>A2149T/C</td>
<td></td>
</tr>
<tr>
<td>I97F/L</td>
<td></td>
<td>A2189T/C, C2191T</td>
<td></td>
</tr>
<tr>
<td>L100I</td>
<td></td>
<td>C2198A</td>
<td></td>
</tr>
<tr>
<td>P130T</td>
<td></td>
<td>C2198A</td>
<td></td>
</tr>
<tr>
<td>Q182K/Stop</td>
<td></td>
<td>C2444A/T, A2445G</td>
<td></td>
</tr>
</tbody>
</table>

(Kim et al. 2016; Shen and Yan 2014)
2.9 NEXT GENERATION SEQUENCING

Classification of Hepatitis B virus of approximately 3.2 kbp into genotypes, sub-genotypes, and the identification of recombinants and mutations have become possible because of the availability of sequencing technologies. Sanger sequencing or first generation sequencing has been the main sequencing technology until the introduction of next-generation sequencing (Sanger and Coulson 1975; Tinhofer et al., 2015). This method uses fluorescently labelled chain-terminating nucleotides and capillary electrophoresis to separate varying lengths of fragments generated during the amplification of the original DNA strand by PCR (Radford et al., 2012; Tinhofer et al., 2015). Major limitations of this method are that it is time consuming, laborious, costly and is of low throughput when dealing with larger fragments/genomes (Radford et al., 2012). Moreover, the introduction of bias during cloning and the inability to detect viral variants in minor populations further limits its application (Chevaliez et al., 2012).

The advent of next-generation sequencing revolutionized genomic studies allowing up to millions of DNA fragments to be sequenced concurrently (Frese et al., 2013; Tinhofer et al., 2015). Thus NGS is also termed massive parallel sequencing and is of high throughput, lower cost per sequenced base compared to Sanger sequencing (Solmone et al. 2009).

Ever since the first NGS platform 454 FLX (Roche) was introduced unto the market in 2005, other platforms such as Illumina platforms (MiSeq, HiSeq etc.), SOLiD (Applied Biosystems), Heliscope (Helicos), Ion Torrent PGM (Life Technologies) and Pac Bio RS (Pacific Bioscience) have followed and quite recently Oxford Nano sequencing technology (Srinivasan and Batra 2014). Common to all of these platforms are the steps of template preparation, sequencing, imaging and data analysis. The differences, however, pertains to the sequencing chemistries and different
protocols. As a result, each platform comes with its own merits and demerits with regards to suitability to specific applications based on read-lengths, error rates etc. (Barzon et al., 2011).

In template preparation, template DNA is sheared into fragments either by nebulization, sonication or restriction enzymes. Fragments are ligated on both ends to adapters or linkers to generate short gun library. Each library fragment is amplified on a solid surface either on a bead or flat Silicon-derived surface with covalently attached adapters that hybridize the library adapters. Following this, each amplified library fragment set are sequenced by direct step-by-step detection of nucleotide base (Mardis et al., 2016). Signals generated during sequencing as bioluminescence, fluorescence, PH change to four-colour imaging, depending on the sequencing platform, are captured in real time. The massive data generated presents a significant challenge of data storage and data analysis.

NGS has proven to very instrumental in the field of virology. Full genome reconstruction of important pathogens such HIV (Vrancken, et al., 2010), human rhinovirus (Tapparel et al., 2011), and several enteric virus have become possible because of NGS. Detection and quantification of mutations in minor viral variants as well as detecting recombinants in chronic hepatitis B patients have been done by NGS (Gong et al., 2013; Boyce et al., 2017; Solmone et al., 2009; Yamani et al., 2015; Yan et al., 2015).

A study by Yan et al., 2015 explored the viral quasispecies among treatment naïve patients using deep sequencing focusing on mutations and quantification of these mutations both in the basal core and precore region of the hepatitis B virus in relation to both viral and host factors. In addition to the classical double mutations in the basal core A1762T/G1764A and precore mutant G1896A which were found in this study, other twelve hotspots with prevalence > 20% was also found. These hotspots were single nucleotide polymorphisms occurring at nucleotide positions (nt.1719,
Another study by Yamani et al., 2015 demonstrated the existence and relevance of viral quasispecies in the Major hydrophilic region (MHR) of the S-gene using ultra-deep sequencing. This study revealed that rate of MHR variation was strongly associated with disease progression and disease severity. Immune escape mutants P120Q/T, T123A, P127T, Q129H/R, M133L/T, and G145R were the most frequently encountered MHR variants in that study. This study also detected the major MHR variants G119R, Q129R, T140I, and G145R which impaired secretion of HBsAg among a cohort of the patients with advanced liver disease. These studies and others lend credence to the usefulness of NGS in the study of HBV viral quasispecies and their clinical relevance.
CHAPTER THREE

3.0 METHODOLOGY

3.1 STUDY DESIGN

The study was conducted using archival plasma samples of a cross-section of prison-inmates from a selected prison in Ghana. Samples were derived from prison inmates who participated in a previous nationwide surveillance study which focused on determining HIV and HBV prevalence from forty-two (42) prisons across the country (Ghana) from (February- March 2013).

3.2 STUDY POPULATION

Twenty-six (26) samples from the nationwide surveillance were examined in a separate study to ascertain HBV transmission among prison-inmates (Angela Ayiku student research work). These samples had been tested and confirmed as HBV positive and the presence of anti-HBc IgM and HBeAg determined by ELISA.

3.3 SAMPLE SIZE

HBV DNA was extracted from these twenty-six samples (26) and subjected to PCR to amplify the surface gene and precore/core gene designated as S1. Five samples (5) out of the 26 were successfully amplified for the precore/core gene (S1). These were subjected to next-generation sequencing (NGS).
3.3.1 SAMPLE COLLECTION

Archived plasma samples in a study bank were used. Prison-inmates who participated in the nationwide surveillance signed an informed consent as well as specimen banking forms during enrolment (Appendix I)

3.4 LABORATORY ANALYSIS

Plasma samples stored at -20 °C were removed and allowed to thaw at room temperature for at least 30 minutes.

3.4.1 EXTRACTION OF VIRAL DNA FROM PLASMA

Two hundred microlitres (200 μL) of plasma was added to a sterile microcentrifuge tube containing 25 μL of Proteinase K. This was followed by the addition of a 200 μL lysis buffer. The resultant mixture was pulse vortexed for 15 seconds and incubated at 56 °C for 15 minutes. Subsequently, 250 μL of absolute ethanol was added and vortexed for 15 min and the lysate incubated at room temperature for 5 min. The lysate was added to a labelled viral spin column in a collection tube and span at 6500 rpm for 1 minute. The collection tube together with filtrate was discarded and a new wash tube was fitted to the viral spiral column. A washing buffer of 500 μL was then added to the viral spiral column and centrifuged at 6500 rpm for a minute. This step was repeated and the viral spiral column within a new collection tube was then centrifuged at 15000 rpm to remove any residual wash buffer.
Following this, a clean recovery tube was fitted to the viral column and 35 μL RNase-free water added to the column. This was incubated at room temperature for 1 minute and finally spun at 15000 rpm for 1 minute to elute viral DNA of 35 μL. A total of 35 μL of Hepatitis B viral DNA was extracted from each sample using PureLink™ Viral RNA/DNA Mini Kit (Invitrogen, USA) based on the manufacturer’s instruction. (User Guide PureLink™ Viral RNA/DNA Mini Kit 2016). Viral DNA was stored at -20 °C until further analysis.

3.4.2 AMPLIFICATION OF HBV WHOLE GENOME BY NESTED PCR (1ST ROUND)

Extracted HBV DNA was amplified using the LongAmp® Taq 2X Master Mix according to the manufacturer's protocol (LongAmp® Taq 2X Master Mix New England Biolabs Inc, 2016). Each reaction mix contained 18 μL of the LongAmp® Taq 2X Master Mix, 5 μL of template DNA, 0.75 μL of 2.5 mM MgCl₂, 9.45 μL of nuclease-free water and 0.9 μL of 10 μM HBV forward and reverse primers each in a total volume of 36 μL. The primer pair P1 (5’-GTGGTGGACTTCTCTCAATTTC- 3’) and P2 (5’-CGGTATAAAGGGACTCAGAT- 3’) was applied in the 1st round PCR to generate a product of approximately 3.2 kb genome of the virus. Thermocycling conditions included an initial denaturation at 94 °C for 2 min, followed by 25 cycles at 94 °C for 30 s, 47 °C for 1 min, and 65 °C for 7 min and a final extension step, 65 °C for 15 min.
3.4.4 AMPLIFICATION OF HBV SURFACE-GENE

Following the first round amplification of the whole genome, the second round was set up with the intention of amplifying the hepatitis B surface gene to ascertain the presence of HBV and quality of DNA extracted. The second round was set up using a modified protocol by (Gous et al., 2010).

In the second round, a 15.45 μL of 1:5 dilutions of the 1st round PCR product was used as a template. The reaction mix consisted of 18 μL of LongAmp® Taq 2X Master Mix, 0.75 μL of 2.5 mM MgCl₂, 0.9 μL of 10 μM each of HBV forward (HBV-F) and reverse primers (HBV-R) (Table 3.1) all in a total volume of 36 μL, this time without nuclease-free water. Thermocycling conditions were as follows; 94 °C initial denaturation for 2 min, 42 cycles at 94 °C for 30 s, 52 °C for 1 min, and 65 °C for 2 min and a final extension step, 65 °C for 10 min. PCR was carried out in an MJ Research PTC- 200 Peltier Thermal cycler to generate a product of approximately 541 bp.

3.4.3 AMPLIFICATION OF PRECORE/CORE ORF (S1) GENOMIC FRAGMENT

Based on the successful amplification of the hepatitis B surface gene, a second round of PCR was set up to amplify precore/core ORF or S1 fragment based on a protocol by Zhang et al., 2007. From each first-round PCR amplicons, a 1:5 dilution was prepared which served as the template for the second-round PCR. The reaction mix consisted of 18 μL of the LongAmp® Taq 2X Master Mix, 6 μL of template DNA, 0.75 μL of 2.5 mM MgCl₂, 9.45 μL of nuclease-free water and 0.9 μL of 10 μM HBV forward (FA1-L) and reverse primers (FA1-R) (Table 3.1) each in a total volume of 36 μL. Thermocycling conditions included an initial denaturation at 94 °C for 5 min,
followed by 45 cycles at 94 °C for 30 s, 52 °C for 1 min, and 72 °C for 2 min and a final extension step, 72 °C for 10min.

3.4.5 AGAROSE GEL ELECTROPHORESIS

All PCR products were analysed on a 2% ethidium bromide-stained agarose gels in 1X TAE buffer. 1000 bp molecular marker, TrackIt DNA Plus ladder (Invitrogen, USA) was used to estimate the size of PCR products. The PCR products together with the marker were electrophoresed at 120V for 30 min and bands visualized under ultraviolet trans-illuminator. The image was captured by an inbuilt camera onto a computer.
Table 3.1 A table showing the name, position on the reference genome, sequence and purpose of all primers used

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Position</th>
<th>Sequence (5’-3’)</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1WRS</td>
<td></td>
<td>GTGGTGGAATTCTCTCTCAATTTC</td>
<td>Whole genome</td>
</tr>
<tr>
<td>P5W</td>
<td></td>
<td>CGGTATAAAGGGACTCAGAT</td>
<td></td>
</tr>
<tr>
<td>FA1-L</td>
<td></td>
<td>TTTTACCTCTGCTGCTATCATCTCT</td>
<td>S1 fragment (1014bp)</td>
</tr>
<tr>
<td>FA1-R</td>
<td></td>
<td>TCTTTTCCCAAGAATATGTTG</td>
<td></td>
</tr>
<tr>
<td>HBV-F</td>
<td></td>
<td>GTGGTGGAATTCTCTCTCAATTTC</td>
<td>S-gene (541 bp)</td>
</tr>
<tr>
<td>HBV-R</td>
<td></td>
<td>CGGTATAAAGGGACTCAGAT</td>
<td></td>
</tr>
</tbody>
</table>

(Zhang et al., 2007; Gous et al., 2010)
3.4.5 PCR AMPLIFICATION OF PRECORE/CORE FRAGMENT FOR SEQUENCING

To obtain higher volumes of amplicons for the purpose of sequencing, a repeat of the second round PCR for S1 genomic fragment was done. Each component of the reaction mix this time was increased by a factor of two. Thus reaction mix consisted of 36 μL of the LongAmp® Taq 2X Master Mix, 12 μL of template DNA, 1.5 μL of 2.5 mM MgCl₂, 18.90 μL of nuclease-free water and 1.8 μL of 10 μM HBV forward (FA1-L) and reverse primers (FA1-R) each in a total volume of 72 μL. Thermocycling conditions included an initial denaturation at 94 °C for 5 min, followed by 45 cycles at 94 °C for 30 s, 52 °C for 1 min, and 72 °C for 2 min and a final extension step, 72 °C for 10 min. PCR products were labelled, packaged and sent for commercial next-generation sequencing.

3.5 PHYLOGENETIC AND MUTATIONAL ANALYSIS

Quality control of the reads obtained from NGS in each sample was done. This was followed by trimming of the reads which were then assembled. Scaffolding was done which was used as the basis to generate contigs. All these procedures were done using Genesious software version R 10. Contigs generated were then subjected to NCBI BLAST to obtain the identity of sequences obtained and their relatedness to HBV genotypes in the genomic database NCBI. Some reference genomic sequences of HBV from genotypes A to H were downloaded and used for phylogenetic analysis to determine the relatedness of the clones. The molecular phylogenetic analysis was done by Maximum Likelihood method using the software MEGA version 6.0. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model. The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa
clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log-likelihood value. Construction of the phylogenetic trees was with a bootstrap value of 100. Evolutionary analyses were conducted in MEGA 6.0 (Tamura and Nei 1993; Tamura et al., 2013)

3.6 ETHICAL CONSIDERATION

 Archived plasma samples which were investigated in this study were obtained from prison-inmates who had earlier on consented to the banking of their specimens and usage of these specimens in further studies from the parent study. Approval was obtained from the College of Health Science Protocol and Ethical Review Committee.
4.0 RESULTS

Of the twenty-six archival plasma samples from which viral DNA was successfully extracted, all were successfully amplified for the S-gene whereas only 5 (19.2%) were amplified for the precore/core genomic fragment (Table 4.1).

NGS data were generated for the precore/core region of 5 inmates and 3 successfully analysed for phylogenetic relationships and mutations. Succinct demographic and serological data for these individuals are shown in Table 4.2.
Table 4.1 Summary of PCR results showing the number of successfully amplified targets and PCR positive rate

<table>
<thead>
<tr>
<th></th>
<th>S-GENE</th>
<th>PRECORE/CORE GENE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of DNA Extracts</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Number Successfully Amplified</td>
<td>26</td>
<td>5</td>
</tr>
<tr>
<td>Positivity Rate (%)</td>
<td>100.0</td>
<td>19.2</td>
</tr>
</tbody>
</table>

Table 4.2 Summary of Demographic and serological data of study samples

<table>
<thead>
<tr>
<th>ID</th>
<th>AGE</th>
<th>SEX</th>
<th>ANTI-HBc IgM</th>
<th>HBeAg</th>
</tr>
</thead>
<tbody>
<tr>
<td>4S</td>
<td>25</td>
<td>M</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>5S1</td>
<td>18</td>
<td>M</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>6S1</td>
<td>37</td>
<td>M</td>
<td>R</td>
<td>NR</td>
</tr>
<tr>
<td>9S1</td>
<td>39</td>
<td>M</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>10S1</td>
<td>24</td>
<td>M</td>
<td>R</td>
<td>R</td>
</tr>
</tbody>
</table>

HBeAg and Anti-HBc IgM reaction by ELISA; M- Male; R- Reactive; NR- Non-reactive
4.1.2 GEL ELECTROPHORESIS

Amplicons were run on 2% agarose gel stained with ethidium bromide. Two fragments of the HBV genome both S-gene and precore/core ORF after amplification were run on the gel. A band size of approximately 541bp was expected for the S-gene and 1014bp for the precore/core fragment. Gel electrophoregrams indicating successfully amplified S-gene of expected band size and that of the precore/core are shown in Figure 4.2 and Figure 4.3 respectively.
**Figure 4.1** Representative Agarose Gel (2% in 1X TAE Buffer) electrophoregram of S-gene amplification showing expected band sizes. The last lane represents the 100bp (NEB) Marker labelled M. Lanes one has the Positive control with lane two to five being the amplicons from study participants. Lane six represents the Negative Control; M- Molecular Marker; PC- Positive Control; NC- Negative Control
**Figure 4.2** Representative Agarose Gel (2% 1X TAE Buffer) electrophoregram of S1 genomic fragment of HBV encompassing the precore/core ORF. Lane one contains the 100 bp molecular Marker labelled M. Lanes four to eight contain successfully amplified S1 amplicons of expected band size of 1014 bp as indicated above.

M- Molecular Marker; NC- Negative Control; PC- Positive Control
4.1.3 PHYLOGENETIC ANALYSIS

Three phylogenetic trees were constructed by the neighbour-joining method using the clones of each sample together with 41 full-length reference sequences of genotypes A-H retrieved from the NCBI nucleotide database. Accession numbers of each reference sequence are indicated together with the HBV genotype.

For 13 clones of sample 4S1, two distinct clusters were obtained. Five (5) of these clones clustered closely with genotype G sequences whereas eight (8) clustered closely with reference sequences of genotype E (Figure 4.3). For sample 5S1 eleven (11) out of twelve (12) clones clustered closely with the genotype E reference sequences. The remaining one clustered closely with genotype A reference sequences (Figure 4.4). All 14 clones of sample 10S1 clustered closely with genotype E reference sequences (Figure 4.5)
Figure 4.3 Phylogenetic tree for 13 clones or contigs of the HBV precore/core ORF for sample 4S1
Figure 4.4 Phylogenetic tree for 12 clones or contigs for the HBV precore/core ORF of Sample

Reference sequences include genotypes A-H from the NCBI database and bootstrap values are indicated at the nodes.
Figure 4.5 Phylogenetic tree for 14 clones or contigs for the HBV precore/core ORF of Sample 10S1.

Reference sequences include genotypes A-H from the NCBI database and bootstrap values are indicated at the nodes.
4.1.4 MUTATIONAL ANALYSIS

Several clinically relevant mutations were found across majority of the clones generated for the precore and core regions. For all eleven (11) contigs or clones for sample 4S1, majority harboured the precore stop codon W28* which abrogates HBeAg synthesis. Three mutations I97F, P130T, R181P, were commonly detected in the core region of virtually all the clones (Table 4.3).

No mutations of clinical relevance were seen in the precore region of the clones generated for sample 5S1. Core mutations, however, were found in only six (6) clones. The mutations I97F, P130T, R181P were found in only clone. Two clones harboured I97F, R181P, mutations whereas one clone harboured I97F (Table 4.4).

Precore and core mutations were detected in fourteen (14) clones of sample 10S1. In all the fourteen clones, only two had the precore stop codon W28*. The rest did not show mutations of clinical importance in this region. Of all the fourteen clones, 12 clones had the mutations I97F, P130T and R181P. (Table 4.5)
Table 4.1 Mutations detected in core and precore regions in the contigs/clones of 4S1 during analysis

<table>
<thead>
<tr>
<th>CLONES</th>
<th>CORE MUTATIONS</th>
<th>PRECORE MUTATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>4S1_1</td>
<td>I97F, P130T, R181P,</td>
<td>W28*</td>
</tr>
<tr>
<td>4S1_2</td>
<td>I97F, P130T, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>4S1_3</td>
<td>I97F, P130T, R181P,</td>
<td>W28*</td>
</tr>
<tr>
<td>4S1_4</td>
<td>-</td>
<td>W28*</td>
</tr>
<tr>
<td>4S1_5</td>
<td>I97F, P130T, R181P,</td>
<td>W28*</td>
</tr>
<tr>
<td>4S1_9</td>
<td>-</td>
<td>W28*</td>
</tr>
<tr>
<td>4S1_363</td>
<td>I97F, P130T, R181P,</td>
<td>W28*</td>
</tr>
<tr>
<td>4S1_331</td>
<td>I97F, P130T, R181P,</td>
<td>W28*</td>
</tr>
<tr>
<td>4S1_262</td>
<td>I97F, P130T, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>4S1_254</td>
<td>I97F, P130T, R181P,</td>
<td>W28*</td>
</tr>
<tr>
<td>4S1_125</td>
<td>I97F, P130T, R181P,</td>
<td>W28*</td>
</tr>
</tbody>
</table>
Table 4.2 Mutations detected in core and precore regions in the contigs of 5S1 during analysis

<table>
<thead>
<tr>
<th>CLONES</th>
<th>CORE MUTATIONS</th>
<th>PRECORE MUTATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5S1_14</td>
<td>I97F, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>5S1_25</td>
<td>I97F, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>5S1_8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5S1_13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5S1_12</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5S1_21</td>
<td>I97F, P130T, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>5S1_23</td>
<td>I97F, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>5S1_39</td>
<td>I97F,</td>
<td>-</td>
</tr>
<tr>
<td>5S1_41</td>
<td>I97F, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>5S1_5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5S1_4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CLONES</td>
<td>CORE MUTATIONS</td>
<td>PRECORE MUTATIONS</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>10S1_1</td>
<td>I97F, P130T, R181P,</td>
<td>W28*</td>
</tr>
<tr>
<td>10S1_3</td>
<td>I97F, P130T, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>10S1_7</td>
<td>I97F, P130T, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>10S1_154</td>
<td>I97F, P130T, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>10S1_113</td>
<td>I97F, P130T, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>10S1_43</td>
<td>I97F, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>10S1_12</td>
<td>I97F, R181P,</td>
<td>W28*</td>
</tr>
<tr>
<td>10S1_70</td>
<td>I97F, P130T, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>10S1_26</td>
<td>I97F, P130T, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>10S1_193</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10S1_69</td>
<td>I97F, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>10S1_189</td>
<td>I97F, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>10S1_50</td>
<td>I97F, R181P,</td>
<td>-</td>
</tr>
<tr>
<td>10S1_134</td>
<td>I97F, P130T, R181P,</td>
<td>-</td>
</tr>
</tbody>
</table>
CHAPTER FIVE

5.0 DISCUSSION AND CONCLUSION

5.1 DISCUSSION

Hepatitis B virus exists as heterogeneous viral quasispecies circulating within a single host owing to lack of proofreading of the reverse transcriptase as well as high replication rate. Next-generation sequence, unlike first generation sequencing, affords the opportunity to explore all genetic variants including minor variants. Therefore, this study was aimed at evaluating the diversity of HBV at the sub-population level in the precore and core region.

5.1.1 PHYLOGENETIC ANALYSIS

Based on phylogenetic analysis most of the clones for samples analysed clustered around genotype E. This is consistent with earlier studies which have established the predominance of genotype E in Ghana. (Garmiri et al., 2009; Ampah et al., 2016). All of the 14 clones of sample 10S1 as seen in figure 12 clustered around genotype E, meaning these are closely related. However, in the case of samples 4S1 and 5S1, an interesting observation was made. Almost half of the clones of 4S1 clustered around genotype E reference sequences whereas the other half clustered around genotype G reference sequences. This is possibly an indication of a dual infection. Genotype G is rarely seen in this part of the world and the presence of this tells of the contact history of the inmate from which this was found. Genotype G has been described as an aberrant genotype of unknown origin with poor replication fitness and little genetic variation (Kuyl et al. 2013). As a result of these characteristics, genotype G rarely exist as a mono-infection and requires the presence of other
genotypes to be replication efficient. In view of this, studies have reported of dual infection with this genotype such as G-A, G-H, G-C and even G-F (Kato et al. 2002; Osiowy et al. 2008; Beggel et al. 2017; Kuyl et al. 2013) Co-infection of genotype G with other genotypes has been associated with certain risk groups such as intravenous drug users and men who have sex with men (MSM) (Kuyl et al. 2013). This association that has been made may underlie this incident of possible dual infection of genotype E and G in this study. This is because the behavioural pattern seen among prison-inmates is similar to the study population described by Kuyl et al., 2013.

Additionally, it cannot be established whether or not this case of dual infection is one of co-infection or superinfection. In any case, the chances of genotype recombination occurring over time cannot be ruled out. Another likely event is genotype G outcompeting the already established genotype E to become the predominant species as reported by (Kato et al. 2002).

Another unique observation in this study was that, one clone out of the rest in sample 5S1 clustered closely with genotype A reference sequences. This perhaps could be another situation of a dual infection or possible recombination. Even though genotype E is seen as the predominant genotype in Ghana, certain studies have reported the presence of genotype A (Garmiri et al. 2009). There have also been reports of A/E recombination in Ghana (Garmiri et al. 2009; Ayiku 2015), unpublished). In the case of Angela Ayiku, this recombinant was seen in two prison-inmates who form part of my study populations.
5.1.2 MUTATIONAL ANALYSIS

Mutational analysis was done using Geno2pheno software on the clones generated for the three samples namely 4S1, 5S1 and 10S1 which were analysed. It was observed that the clones exhibited common patterns of mutations both in the core and precore region which are of clinical relevance. The classical mutation W 28* (G1896A) was observed as the common mutation in the precore region. Despite the fact that this mutation generates a stop codon which prematurely abrogates the translation of the HBeAg, all of the subjects in this study tested positive for this antigen by ELISA. This can be expected as studies (Yan et al., 2015) have reported the presence of this mutation at the late stage of HBeAg chronic stage prior to seroconversion to anti-HBeAg negative phase of the infection. Therefore, the presence of this mutation is a reflection of the initiation of the gradual build-up towards seroconversion although the infection among these subjects is recent or of an acute onset. Moreover, the prevailing genotype in this case predominantly genotype E also determine the predisposition to this mutation.

For non-genotype A, the presence of a T nucleotide at position 1858 favours the nucleotide substitution of G to A at position 1896 to ensure the stability of the encapsidation signal (Jalali 2006; Alexopoulou 2009) which perhaps accounts for the presence of W28* mutation.

Apart from the abrogation of the HBeAg translation, W28* has also been associated with disease severity (hepatocellular carcinoma) as reported by some studies (Kim et al., 2016; Shen and Yan 2014). However some other studies reports of no such association with HCC (Yan et al., 2015).

In the case of the core region, a common pattern of mutation was also observed. Mutations such as R181P, P130T and I97F/L were seen to be common at the subpopulation level of the samples analysed. According to Kim et al., 2016; Shen and Yan 2014), the presence of these mutations has
been linked both to disease progression in terms of severity and HBeAg serostatus. R181P mutation within the core region has been associated with HBeAg negative status whereas I97F and P130T have been associated with hepatocellular carcinoma. As explained earlier with respect to the presence of W28*, the presence of R181P has been associated with serostatus reveals a background initiation of seroconversion from an HBeAg status to anti-HBeAg status even though all the study subjects were HBeAg positive. I97F and P130T mutations, on the other hand, have been associated with HCC. Therefore, the presence of these mutations with their possible concomitant effect is predictive of the future outcome of the infection.
5.2 LIMITATIONS

- Lack of genotyping data among prison-inmates to ascertain the extent of the phenomenon of dual infection and possible recombination
- Limited clinical data such as HBV DNA level, HBcAg level as well as the lack of data on contact history or sexual habits of study subjects
- Poor amplification success

5.3 CONCLUSION

This study has shown that common pattern of mutation exists across the precore and core region which are of clinical relevance particularly the precore- W28* stop codon and I97F, P130T, R181P core mutations. It also evidently shown that the predominant genotype among the viral subpopulation is genotype E with dual genotypic infection with genotypes G and A respectively.

5.4 RECOMMENDATION

Further studies in this area are highly recommended with increased sample size to ascertain the true picture of the extent of genotype dual infection and recombination.
REFERENCES


Alexopoulou, A. & Karayiannis, P., 2014. HBeAg negative variants and their role in the natural history of chronic hepatitis B virus infection. , 20(24), pp.7644–7652.


Boyce, C.L. et al., 2017. Identification and comparative analysis of hepatitis B virus genotype D
/E recombinants in Africa. Virus Genes. Available at: "http://dx.doi.org/10.1007/s11262-017-1469-4.


Chan, H.L., Hussain, M. & Lok, A.S., 1999. Different hepatitis B virus genotypes are associated with different mutations in the core promoter and precore regions during hepatitis B e antigen seroconversion. Hepatology (Baltimore, Md.), 29(3), pp.976–84. Available at:


Cho, Y. et al., 2012. The prevalence and risk factors for hepatitis B surface Ag positivity in pregnant women in eastern region of Ghana. Gut and Liver, 6(2), pp.235–240. Available at:


Kuyl, A.C. Van Der et al., 2013. High prevalence of hepatitis B virus dual infection with genotypes A and G in HIV-1 infected men in Amsterdam , the Netherlands , during 2000-2011.


Moore, O. et al., 2010. Hepatitis B transmission event in an English prison and the importance of


Solmone, M. et al., 2009. Use of Massively Parallel Ultradeep Pyrosequencing To Characterize the Genetic Diversity of Hepatitis B Virus in Drug-Resistant and Drug-Naive Patients and
To Detect Minor Variants in Reverse Transcriptase and Hepatitis B S Antigen, 83(4), pp.1718–1726.


Tapparel, L., Cordey, S., Junier, T., Farinelli, L., Van Belle, S., Soccal, P.M., Aubert, J.,
Zdobnov, E. and Kaiser, L., 2011. Rhinovirus genome variation during chronic upper and

Tinhofer, I. et al., 2015. Next-generation sequencing: hype and hope for development of
personalized radiation therapy? *Radiation Oncology*, pp.1–8. Available at:
http://dx.doi.org/10.1186/s13014-015-0481-x.

Tran Thien Tuan Huy, Koichi Ishikawa, William Ampofo, 4 Taku Izumi, 1 Akira Nakajima,
length Genome sequences indicate the endemicity of genotype E in West Africa
Characteristics of Hepatitis B Virus in Ghana: Full Length Genome Sequences Indicate the

Tseng, T.-C. et al., 2015. Higher proportion of viral basal core promoter mutant increases the
risk of liver cirrhosis in hepatitis B carriers. *Gut*, 64(2), pp.292–302. Available at:

Urban, S. et al., 2010. Hepatology Snapshot The replication cycle of hepatitis B virus. *Journal of

Vrancken, B., Lequime, S., Theys, K. and Lemey, P., 2010. Covering all bases in HIV research:

WHO, 2015. WHO | Hepatitis B. *WHO*. Available at:


APPENDICES

APPENDIX 1 SPECIMEN BANKING FORM

AGREEMENT TO PARTICIPATE IN SPECIMEN BANKING

Study title: National Health and HIV Survey of Prison Inmates and Prison Officers in Ghana

You have agreed to participate in a study known as National Health and HIV Survey of Prison Inmates and Prison Officers in Ghana, sponsored by the Ghana AIDS Commission and partners.

You are being asked to sign this second consent form to indicate if you are willing to allow the samples collected solely for research purposes, which will be referred to as your 'Specimen', to be saved or 'banked' for use in future research studies. At this time, we do not know what future research studies may be done using your Specimen. The specimen bank will be maintained by the Clinical Virology Laboratory, University of Ghana Medical School.

Your signature below will allow your Specimen to be stored in the specimen bank, with the possibility that it will be used in future research studies. It is very unlikely that any future research performed using your Specimen would benefit you directly, but it may provide important medical knowledge that could be helpful in understanding the transmission of viral pathogens in prisons.

Information about how your Specimen will be used and protected is as follows.

A. Confidentiality and Privacy of Medical Record. If you sign this form, you give the Clinical Virology Laboratory permission to store your Specimen in a controlled specimen bank, along with portions of your personal health information collected related to the Main Study. This information could be used to link the specimen back to you. Clinical Virology Laboratory will protect your confidentiality by making sure that no information that could be used to identify you will be used.
or disclosed by Clinical Virology Laboratory without your authorization or without legally required protections in place. The following rules will apply:

The Clinical Virology Laboratory in charge of the specimen bank will determine for which research studies to release all or part of your Specimen. 1) In most cases, your Specimen will be 'de-identified'; that is, the researcher who is given your Specimen will not be given enough information to identify you. In these cases, you will not be contacted prior to your Specimen being released to the researcher. 2) Your authorization would not be required for researchers to use partially de-identified (in accordance with legal standards) Specimens for future research studies; however, such researchers would be required to sign a Data Use Agreement, which would protect your privacy by limiting how they could use your Specimen. 3) If an approved researcher wants to use your Specimen and feels that is important that he or she be able to identify you to a) Collect information about you that was not collected as part of your participation in the Main Study; or, b) Collect additional samples of your blood then this would be considered using an identifiable Specimen meaning the Specimen could be linked back to you). In that case, someone associated with Clinical Virology Laboratory will contact you to provide further information about the proposed study so you can decide whether you will agree to participate. If you decide to participate, you will then be contacted by a researcher for the new study and asked to sign a separate consent form for that study. Clinical Virology Laboratory will apply national legal standards in making decisions about who can review your records in preparing for the study and about who can contact you to provide information about the new study. 4) It is possible that your Specimen could be used for future research purposes without your consent or authorization if a committee of people who know about research, privacy and medical ethics (such as the Noguchi Memorial Institute for Medical Research Institutional
Review Board or the Ghana Health Service Ethical Review Committee) decided that use of your information is necessary and that use of it would be of low risk to you and your privacy. Clinical Virology Laboratory will ensure that all specimens stored at the Clinical Virology Laboratory will be kept confidential and only shared by the Clinical Virology Laboratory in accordance with the above rules; no other people, including relatives or Ghana Prisons Officials will have access to the stored samples or information about them without your written consent. Appropriate physical and computer security measures will be maintained to limit access to Specimens. Papers or articles written by Clinical Virology Laboratory researchers which are based on studies involving your Specimen will not identify you by name.

B. Control and Ownership of the Specimens(s)

By consenting to participate in the specimen bank, you authorize the banking of your Specimen for research conducted in accordance with the rules described above. If you should have a need for the Specimen at some later date for a medical purpose, it usually can be removed from the specimen bank for that purpose unless there is no specimen left.

C. Withdrawal of Your Consent If you decide at some time in the future that you no longer wish your stored Specimen to be used in future studies, you have the right to request that the Specimen be withdrawn from the specimen bank. However, withdrawal cannot be guaranteed and may be impossible. For example, it is possible that the Specimen might no longer be identifiable as belonging to you, or that it might already have been released for research studies and used up. To request withdrawal of the Specimen from the specimen bank, please write to Dr. Kwamena Sagoe, Clinical Virology Laboratory, Department of Microbiology, University of Ghana Medical School, P. O. Box 4236, Accra (Mobile: 0277408528).
D. **Length of Storage** Specimens in the bank will be stored for an indefinite period of time, until research funding is exhausted or the Specimen is no longer usable.

E. **Signature**

You have the right to refuse to sign this form. Refusing to sign this form will not affect your participation in the Main Study or your access to health services in the prison. If you chose to sign, your signature below indicates that you have read this form and discussed it with researchers associated with the Main Study, and that you wish to participate in the specimen bank in accordance with the terms described in this form.

……………………………  ……………………………
Participant’s Signature  Date

……………………………  ……………………………
Witness (only if consent presented orally)  Date

……………………………  ……………………………
Participant’s legal representative  Date

(If patient unable to sign)

**2% Agarose Gel (Ethidium Bromide stained)**

- 3g Molecular Grade Agarose

- 150 ml 10X TAE Buffer

- Heat in microwave for approximately 3 minutes or until agarose is completely dissolved;
• Cool to ~ 50°C

• Add 8μl ethidium bromide and mix;

• Pour into gel tray with inserted comb and allow to set

Ethidium Bromide Stock (10 mg/ml)

• 0.1 g Ethidium Bromide

• 10 ml Best quality water

• Mix well;

• Store in a dark bottle at 4°C.

TAE Buffer (10X)

• 24.2 g Tris-base

• 5.68 ml Glacial acetic acid

• 10 ml 0.5 M EDTA (pH 8)

• Make up to 1L distilled water;

• Autoclave;

• Store at room temperature.