CLINICAL AND MOLECULAR CHARACTERIZATION OF HEPATITIS B VIRUS INFECTION IN HUMAN IMMUNODEFICIENCY VIRUS-POSITIVE SOUTHERN AFRICAN ADULTS, FACILITATED BY NEWLY DEVELOPED BIOINFORMATIC TOOLS

Trevor Graham BELL

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

Johannesburg, March 2013

Frontispiece

The "Three Red Dots", which the research nurse requested, to direct new participants to the enrollment office at Shongwe Hospital, Mpumalanga Province, South Africa. Blood samples from HIV-positive, treatment naïve, adults were screened at the University of the Witwatersrand for the presence of hepatitis B virus. These data were used to characterize hepatitis B virus in the Shongwe region. A database and bioinformatic tools were developed to assist in the analyses.

Declaration

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

_____

Signature

_____

Date

Papers and Presentations

The following published papers describe aspects of this work:

Bell, T. G., Makondo, E., Martinson, N. A. and Kramvis, A. (2012). Hepatitis B virus infection in human immunodeficiency virus infected Southern African adults: occult or overt - that is the question. *PLOS ONE* **7(10)**: e45750

[Reproduced in Chapter 4]

Makondo, E., Bell, T. G. and Kramvis, A. (2012). Genotyping and molecular characterization of hepatitis B virus from human immunodeficiency virus-infected individuals in Southern Africa. *PLOS ONE* **7(9)**: e46345

[Reproduced in Appendix C]

Bell, T. G. and Kramvis, A. (2013). Mutation Reporter Tool: An online tool to interrogate loci of interest, with its utility demonstrated using hepatitis B virus. *Virology Journal* **10** (62). doi: 10.1186/1743-422X-10-62

[Reproduced in Chapter 6]

Bell, T. G. and Kramvis, A. (2013). Fragment Merger Tool: An online tool to merge long overlapping sequence fragments. *Viruses* **5**: 824–833

[Reproduced in Chapter 7]

Aspects of this work have been presented in the following posters:

Bell, T. G., Makondo, E., Martinson, N. and Kramvis, A. (2010). Hepatitis B virus infection in HIV-positive adults from rural South Africa: occult or overt - that is the question. *Poster presentation by Bell, T. G. at the 25th Annual Meeting on the Molecular Biology of Hepatitis B Viruses, October 9–13, 2010, Taipei, Taiwan*

Makondo, E., Bell, T. G. and Kramvis, A. (2010). Genotyping and molecular characterization of hepatitis B virus (HBV) from human immunodeficiency virus infected South African patients. *Poster presentation by Makondo, E. at the 25th Annual Meeting on the Molecular Biology of Hepatitis B Viruses, October 9–13, 2010, Taipei, Taiwan*

Chauhan, R., Bell, T. G., Gopalakrishnan, D. and Kramvis, A. (2011). Molecular characterization of the surface overlapping polymerase region of hepatitis B virus (HBV) isolated from HIV infected southern Africans experiencing virological breakthrough (VBTH) during lamivudine based antiretroviral treatment. *Poster presentation by Chauhan, R. at the 26th Annual Meeting on the Molecular Biology of Hepatitis B Viruses, October 9–12, 2011, Orlando, Florida, USA*

Gopalakrishnan, D., Bell, T. G. and Kramvis, A. (2011). Quasispecies evolution in the basic core promoter/precore region of hepatitis B virus (HBV) isolated from HBV/HIV co-infected South Africans following initiation of anti-retroviral treatment. *Oral presentation by Gopalakrishnan, D. at the 26th Annual Meeting on the Molecular Biology of Hepatitis B Viruses, October 9–12, 2011, Orlando, Florida, USA*

Abstract

Hepatitis B virus (HBV) and human immunodeficiency virus (HIV) share transmission routes and are endemic in sub-Saharan Africa. Mpumalanga is the second-smallest of the nine South Africa provinces, but has the second-highest HIV prevalence of 15.4%. No in-depth study on HBV/HIV co-infection in this province has been undertaken. Routinely, the prevalence of HBV infection is determined serologically by detecting HBsAg. However, HBV infection can occur in the absence of HBsAg, generally termed occult HBV infection (OBI). HIV infection has been shown to be a risk factor for the development of OBI. The "Taormina" definition of OBI, which requires stringent polymerase chain reaction (PCR) amplification of at least two different regions of the HBV genome to determine HBV-positivity, was used to determine the prevalence and characteristics of HBV infection in antiretroviral treatment (ART)-naïve HIV-positive adults in a new rural adult cohort.

The study group consisted of 298 adults (114 men and 184 women) with median age, CD4 count and BMI of 34 years, 147 cells mm^{-3} and 22 kg m^{-2}, respectively. The presence of HBV serological markers was determined by enzyme linked immunoassay (ELISA) tests. HBV DNA-positivity was determined by PCR of at least two of three different regions of the HBV genome and real-time PCR was used to determine the viral loads. Liver fibrosis was determined using the aspartate aminotransferase-to-platelet ratio index. Bioinformatic tools and a database to store clinical and molecular data were developed to facilitate processing and analysis of clinical and molecular data.

An online web interface provided access to data from the database and the R programming language was used to analyze clinical data extracted from the database. Of the 298 participants, 231 (77.5%) showed at least one HBV marker, with 53.7% HBV DNA-negative (resolved) and 23.8% HBV DNA-positive (current infection) [8.7% HBsAg-positive;15.1% HBsAg-negative].

Only the total number of lifetime sexual partners distinguished HBV DNA-positive and HBV DNA-negative participants, implicating sexual transmission of HBV and/or HIV. It is plausible that sexual transmission of HBV and/or HIV may result in a new HBV infection, superinfection and re-activation as a consequence of immunosuppression. Three HBsAg-negative HBV DNA-positive participants had HBV viral loads <200 IU/ml and were therefore true occult HBV infections. The majority of HBsAg-negative HBV DNA-positive participants did not differ from HBsAg-positive HBV DNA-positive (overt) participants in terms of HBV viral loads, amino transaminase (ALT) levels or frequency of liver fibrosis.

Subgenotype A1 was the prevalent strain of HBV circulating in this cohort. The mutations found in this subgenotype differ from those found in other genotypes. Identifying these mutations required careful analysis and a thorough knowledge of sequence data. Since processing of sequence data is labour-intensive and time-consuming, tools, which can assist or automate some of the processes and can reduce the time required to analyze data, were developed. A series of online analysis tools executed via a CGI script, using a Python library was developed, to process and analyze FASTA and chromatogram sequence data. Chromatogram quality was assessed via density plots and a visual representation produced by the "Quality Score Analyzer" tool. Then the "Automatic Contig Generator Tool" was used to generate contigs from forward and reverse chromatogram files. The "Fragment Merger" tool, which can process both FASTA and chromatogram data, was then used to merge three overlapping fragments to yield a single HBV S region sequence. The HBV serotype was determined using the "Serotyper" tool and phylogenetic analysis was automated via the "Pipeline: TreeMail" tool. The "Mutation Reporter Tool", which extracts and summarizes nucleotide and amino acid variation at specified loci of interest, was used to analyze both the S and the basic core promoter/precore (BCP/PC) regions. Subgenomic fragments from the BCP/PC and S regions were placed into their correct orientation within a full-length backbone template for submission to GenBank by the "PadSeq" tool. Additional tools were developed to streamline the process. The "Automatic Alignment Clean-up Tool" selectively and conservatively cleans and disambiguates sequence data. The "Sequence Fetcher" tool was used to fetch sequences from GenBank, by accession number, and provide a single FASTA file for download. Protein sequence data from any of the overlapping reading frames were translated and extracted by the "Babylon" tool, and the "Sequence Divergence Calculator" determined pairwise sequence divergence. The "Wildtype 2x2" tool was used to statistically analyze the frequency of mutant to wild-type polymorphisms.

This is the first southern African study which strictly applied the “Taormina” definition of OBI. Close to a quarter of HIV-positive participants were HBV DNA-positive, of which the majority were HBsAg-negative and were only detected using nucleic acid testing. Of the HBsAg-negative individuals, only three had viral loads of <200 IU/ml and were therefore true occult according to the “Taormina” definition. The remaining HBsAg-negative individuals had higher HBV viral loads and the term *HBsAg-covert* (for cryptic overt) was coined to describe these, because this group of patients was clinically indistinguishable from patients with HBsAg-positive HBV DNA-positive overt infection. The HBsAg-negativity in the HBsAg-covert individuals may be the result of the presence of mutations that alter the S region in such a way that the HBsAg is not detected by commercial assays, or impair either virion or S antigen secretion.

The presence of HBsAg-negative infection and the circulation of mutant strains in the community has important public health implications. As HBsAg-covert infection and OBI are transmissible, either sexually, or via blood or organ donations, or can be reactivated as a result of immunosuppresion, they are public health risk factors. Whilst the clinical implications of HBsAg-negative infection are unclear, they can cause liver disease and hepatocellular carcinoma (HCC) in the same way as HBsAg-positive infections. In fact, in the present study, HBsAg-positive and HBsAg-negative groups were clinically indistinguishable. Therefore, it is important that HBsAg-negative infection, either HBsAg-covert or true occult, be taken into consideration when determining the prevalence of HBV infection in HIV-infected individuals.

The unique combination of viral genotypes, human hosts and socio-economic factors of HBV/HIV co-infected individuals in southern Africa requires further in-depth studies and targeted solutions, which consider the overall well-being of the people involved. The development of bioinformatic tools is a positive step in this direction.

Preface

The aims and objectives of this study were to:

1. Establish a rural cohort in Mpumalanga in order to characterize hepatitis B virus (HBV) in human immunodeficiency virus (HIV) co-infected individuals. No HBV prevalence data are available for this province, which has an HIV prevalence of 15.4%.
2. Examine occult HBV infection in HBV/HIV co-infected individuals, using the "Taormina" definition.
3. Establish a custom database to store clinical and molecular data from the cohort.
4. Develop new bioinformatic tools to facilitate the clinical analyses and molecular characterization of HBV infection.

This work was undertaken in the Hepatitis Virus Diversity Research Programme (HVDRP), Department of Internal Medicine, School of Clinical Medicine, Faculty of Health Sciences at the University of the Witwatersrand (Wits) in Johannesburg, between June 2008 and March 2013. The study was designed to contain both wet and dry laboratory components, with a focus on developing bioinformatic tools to process and analyze data generated in the wet laboratory. This thesis is presented as a set of published academic papers, with supporting chapters providing both the necessary background and additional material.

The study involved establishing a new, rural cohort site at Shongwe Hospital, in Mpumalanga province, which is approximately 500 km from Johannesburg. This cohort was established with the assistance of my supervisor, and leader of the HVDRP, Professor Anna Kramvis, and Dr Neil Martinson, from the Wits Perinatal HIV Research Unit (PHRU). Sister Agatha Nkosi at Shongwe Hospital enrolled participants and managed the cohort site. Dr Precious Gabashane was the clinical officer at Shongwe Hospital.

As this document consists of clinical and bioinformatics components, background for each of these areas is included, and the thesis is in a style that will be understandable to readers with expertise in either of these fields.

The structure of this thesis is as follows:

- Chapter 1: Clinical and Virological Introduction and Literature Review.
- Chapter 2: Bioinformatics Introduction and Literature Review.
- Chapter 3: Establishment of Cohort and Database.
- Chapter 4: **Paper I:** *Hepatitis B virus infection in human immunodeficiency virus infected Southern African adults: occult or overt - that is the question*.
- Chapter 5: Development of Bioinformatic Tools. This chapter discusses the structure of the bioinformatics tools, which were developed in this study, with a detailed description of each tool.
- Chapter 6: **Paper II:** *Mutation Reporter Tool: An online tool to interrogate loci of interest, with its utility demonstrated using hepatitis B virus*.
- Chapter 7: **Paper III:** *Fragment Merger: An online tool to merge overlapping long sequence fragments*.
- Chapter 8: Concluding Discussion

Paper IV: *Genotyping and molecular characterization of hepatitis B virus from human immunodeficiency virus-infected individuals in Southern Africa*, reproduced in Appendix C. The tools described in this thesis were developed and tested extensively while analyzing the data, which are presented in this paper.

References for literature cited in the published papers appear within the references section of the papers themselves. References for all other citations appear in the References section at the end of the thesis. Some references may appear in papers and in the thesis. For clarity and convenience, Internet sites have been cited in footnotes, rather than in the References section.

Acknowledgements

I extend my thanks to Professor Anna Kramvis, the Leader of the Hepatitis Virus Diversity Research Programme (HVDRP) and my supervisor, for her constant support, encouragement, advice and guidance. I feel privileged to have had the opportunity to be part of the HVDRP team and to benefit from Professor Kramvis' mentorship and expertise. The support I enjoyed from the HVDRP team is also gratefully acknowledged. Members of the HVDRP tested the various online tools developed during this work and provided valuable comments, suggestions and ideas. The quality and stability of these tools was only enhanced by their feedback. I would especially like to thank Ms Euphodia Makondo, an MSc(Med) student and collaborator on the study.

I would like to thank Sisters Agatha Nkosi and Rosalina Candlovu, who were responsible for recruiting and enrolling approximately 300 participants in the cohort study, and for following-up selected participants for a period of almost two years.

The Wits Bioinformatics Node, under Professor Scott Hazelhurst, kindly allowed me to install a computer server within their precinct. The assistance provided by Shaun Aron is also acknowleged. Wits CNS arranged a domain name and provided technical support.

Funding for this work was received from the National Bioinformatics Node (NBN), the Medical Research Council (MRC), the Poliomyelitus Foundation (PRF), the National Research Foundation (NRF), the Univeristy of the Witwatersrand and the Cancer Association of South Africa (CANSA). I thank all of these organizations for making this funding available. This work could not have been undertaken without their financial support.

To my parents, Lesley and Keith, for their constant support and generosity, unconditionally given.

To Karl, who never stopped believing in me. Still waters run deep.

– T.

Contents

List of Tables

List of Figures

Abbreviations

ε Hepatitis B virus encapsidation signal

A (Nucleotide) Adenine

ALT Alanine transaminase

Anti-HBc Antibodies to HBV core antigen

Anti-HBe Antibodies to HBV e antigen

Anti-HBs Antibodies to HBV surface antigen

ART Antiretroviral therapy

AST Aspartate transaminase

BCP Hepatitis B virus basic/basal core promoter

BMI Body mass index

BQW Best quality water

C (Nucleotide) Cytosine

cccDNA Covalently closed circular DNA

CHB Chronic hepatitis B

DNA Deoxyribonucleic acid

dNTP Deoxyribonucleotide

ELISA Enzyme-linked immunosorbent assay

G (Nucleotide) Guanine

HAART Highly active antiretroviral therapy (multiple drugs acting on different viral targets)

HBcAg Hepatitis B virus core antigen

HBeAg Hepatitis B virus e antigen

HBsAg Hepatitis B virus surface antigen

HBV Hepatitis B virus

HBx Hepatitis B virus X protein

HCC Hepatocellular carcinoma

HIV Human immunodeficiency virus

km kilometer

LHBs Hepatitis B virus large surface protein

MHBs Hepatitis B virus medium/middle surface protein

nm Nanometer

OBI Occult hepatitis B virus infection

OHB Occult hepatitis B

ORF Open reading frame

PCR Polymerase chain reaction

pgRNA Pregenomic RNA

RFLP Restriction fragment length polymorphism

RNA Ribonucleic acid

SHBs Hepatitis B virus small surface protein

T (Nucleotide) Thymine

WHO World Health Organization

A glossary of technical terms is provided in Section 3.3.1.

For Karl

— for everything

“All our science, measured against reality, is primitive and childlike

– and yet it is the most precious thing we have.”

— Albert Einstein (1879–1955)

“The true delight is in the finding out rather than in the knowing.”

— Isaac Asimov (1920–1992)

Chapter 1

Clinical and Virological Introduction and Literature Review

1.1 Hepatitis B Virus

1.1.1 Classification: The Hepadnaviruses

According to the "Baltimore" classification system (Baltimore, 1971), in which viruses are divided into seven groups[1] based on the viral genome type (single- or double-stranded DNA or RNA) and the method of replication, hepatitis B virus (HBV) belongs to Group VII, designated "dsDNA-RT". This group contains double-stranded DNA viruses that replicate via a single-stranded RNA intermediate (pararetrovirus). The only members of the group known to date are the *Hepadnaviridae*, infecting animals, and the *Caulimoviridae*, which are the only double-stranded DNA viruses infecting plants.

The family *Hepadnaviridae* consists of enveloped, partially double-stranded DNA viruses with small, circular genomes, which replicate via an RNA intermediate. All members of the family are hepatotropic and can cause liver disease, including hepatitis, fulminant hepatitis, fibrosis, cirrhosis and hepatocellular carcinoma (liver cancer). The name "hepadnavirus" derives from "hepar" (Greek for "liver") and "DNA". Analysis of a "fossil" hepadnavirus sequence recently discovered in zebra finches suggests that the virus family may be between 19 and 40 million

[1] Baltimore's original system did not include Group VII, as "gapped" double-stranded DNA viruses were not yet known.

years old (Gilbert and Feschotte, 2010). Additionally, epidemics of HBV, which infects humans, coincide with human dispersal patterns over the last 40,000 years (Paraskevis *et al.*, 2012).

Two genera within the family are recognized: the genus *Orthohepadnavirus*, which infects mammals, and the genus *Avihepadnavirus*, which infects birds. Orthohepadnavirus infections have been reported in humans, captive and wild gibbons and orangutans (Norder *et al.*, 1996; Warren *et al.*, 1999; Noppornpanth *et al.*, 2003; Sa-nguanmoo *et al.*, 2008), gorillas (Grethe *et al.*, 2000; Starkman *et al.*, 2003; Makuwa *et al.*, 2006), chimpanzees (Zuckerman *et al.*, 1978; MacDonald *et al.*, 2000; Takahashi *et al.*, 2000), baboons (Dickens *et al.*, 2013), woolly monkeys (Lanford *et al.*, 1998), ground squirrels (Marion *et al.*, 1980), arctic ground squirrels (Testut *et al.*, 1996), Richardson ground squirrels (Minuk *et al.*, 1986) and woodchucks (groundhogs) (Summers *et al.*, 1978), with Avihepadnavirus infections reported in ducks (Mason *et al.*, 1980), herons (Sprengel *et al.*, 1988), snow geese (Schettler, 1971; Chang *et al.*, 1999), Ross's geese, storks (Pult *et al.*, 2001), cranes (Prassolov *et al.*, 2003) and five exotic captive Anseriforms (Guo *et al.*, 2005). HBV, the orthohepadnavirus type species (Gust *et al.*, 1986), with a genome of approximately 3,200 nucleotides, is the smallest virus known to infect humans. A complete HBV DNA sequence has recently been obtained from a liver biopsy of a mummified Korean child from the sixteenth century (Bar-Gal *et al.*, 2012) and this 500 year old genome is currently the oldest full viral genome sequenced.

1.1.2 Historical Perspective

Hippocrates [c. 460 BCE[2] to c. 370 BCE] correctly recorded the symptoms of hepatitis and described it as a disease "produced by black bile, when it flows into the liver" (Lai, 2002). Almost two thousand years later, Lürman (1885) reported an outbreak of hepatitis (as "jaundice") among staff at the A. G. *Weser* "shipbuild[ers], engineering works and iron foundry" in Bremen, Germany. Lürman's original German report[3] is clearly a description of "long-incubation-period" hepatitis (MacCallum, 1972a), also referred to as "serum hepatitis" (Barker *et al.*, 1970) or "homologous serum hepatitis". Lürman concluded that a smallpox vaccine, containing human lymph, was the source of the infection (Lai, 2002). By the late 1930s, cases of jaundice in children (who had received adult serum as a prophylactic against measles) and adults (who had received a yellow fever vaccine containing small amounts of human serum) were reported in Great Britain

[2] "Before Common Era"; neutral equivalent to "BC"

[3] A translation prepared for this thesis, thought to be the first full English translation published, is available in Appendix A.

(Findlay and MacCallum, 1937; MacNalty, 1938; Propert, 1938; MacNalty, 1939). An outbreak of jaundice amongst soldiers in western United States occurred in 1942 (Sawyer *et al.*, 1944) and "a few cases of homologous serum jaundice" were reported in British and Allied troops in Britain from 1941 to 1944 (Marshall, 1949). Reports from this period distinguish between "epidemic, infective" and "serum" hepatitis or jaundice. MacCallum proposed that the former be termed "hepatitis virus A" and the latter "hepatitis virus B" (MacCallum, 1947, 1972b). Clinical features, epidemiology, pathogenesis and immunity relating to "virus A" and "virus B" were reported as early as 1953 (MacCallum, 1953). The first edition of *Clinical Practice in Infectious Diseases* makes no mention of "serum hepatitis" (Harries *et al.*, 1940). The fourth edition, released a decade later, provides a detailed discussion of the disease (Harries *et al.*, 1951). The first reported case of "virus B" in South Africa is difficult to determine. A meeting in 1968 reports that it had not yet been proven that "infective" and "serum" hepatitis were caused by two different viruses (Bickersteth Medical Society, 1968). A book review in Afrikaans from 1970, reports that "serum hepatitis" is "rare" (H. P. W., 1970). "Serum hepatitis" was not included in the proceedings of the annual meeting of the *Medical Association of South Africa* in any of the years[4] which were accessible.

The description of "Australia" antigen (Blumberg *et al.*, 1965; Alter and Blumerg, 1966; Blumberg *et al.*, 1967; Thomas London *et al.*, 1969), which occurred only in blood from individuals with "serum hepatitis" (Prince, 1968), and the description of the infectious ("Dane") particles (Dane *et al.*, 1970) and its association with "Australia antigen" (MacCallum, 1971), eventually led to the discovery of HBV itself. Work describing "SH" (serum hepatitis) antigen and later "HBsAg" (HBV surface antigen) began in the 1970s. By 1979, the genome of the virus had been sequenced (Galibert *et al.*, 1979a). Baruch Blumberg, who discovered HBsAg and developed one of the first HBV vaccines, is pictured in Figure 1.1.

The HBsAg protein is the principle component of the HBV vaccine. A vaccine prepared by two Americans, microbiologist Maurice Hilleman [1919–2005] and researcher Saul Krugman [1911–1995], from human blood serum, was released in 1981. This was followed in 1986, by a recombinant vaccine developed by Chilean biochemist Pablo D. T. Valenzuela [1941–]. This recombinant vaccine, the first against a major human cancer, is 95% effective in preventing infection (World Health Organization, 2012a) and is still used today. At present, approximately 2 billion people worldwide have been exposed to the virus, between 240 and 350 million are

[4] 1928, 1929, 1930, 1946, 1948, 1949, 1952, 1954, 1955, 1959, 1961, 1961, 1963.

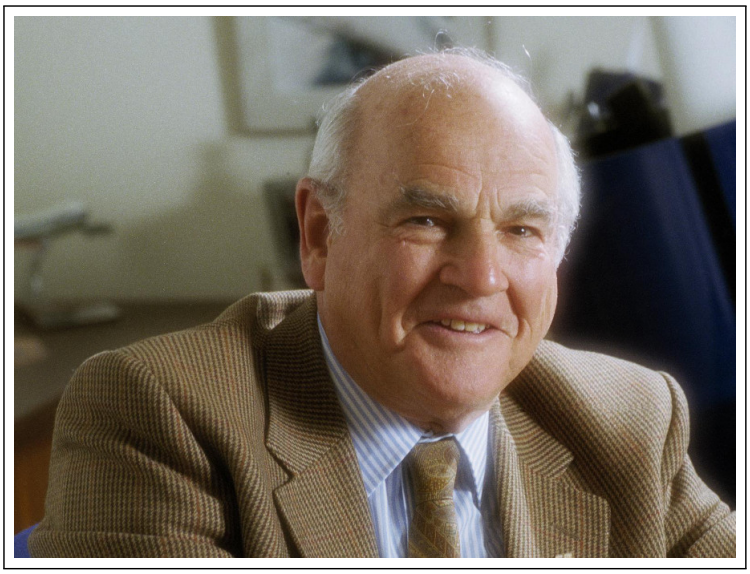


Figure 1.1. *Baruch Blumberg, in 1999. Image from Wikipedia.*

Table 1.1. Proteins produced from the various reading frames of HBV

ORF (Gene)	Protein	Components	Size[a]	Notes	Type[b]
Surface (S)	HBsAg	S	226	Small surface (SHBs)	S
Surface (S)		S and PreS2	281	Medium surface (MHBs)	S
Surface (S)		PreS1, PreS2 and S	389 to 400	Large surface (LHBs)	S
Polymerase (Pol)	Polymerase		832 to 845		S
Core (C)	HBcAg	Core	183 to 185		S
Core (C)	HBeAg	PreCore and Core	212 to 224		NS
X	HBx		154		NS

[a] Size in amino acids, which varies by genotype; for further details, see Table 1.2; [b] "S" = structural/particulate; "NS" = non-structural/secretory

chronically infected, and 600,000 die annually as a result of clinical outcomes of the infection (World Health Organization, 2009, 2012a).

1.1.3 Genome

The HBV genome is composed of circular, partially double-stranded DNA (Figure 1.2). A viral DNA polymerase protein, with retrovirus-like reverse transcriptase activity (Summers and Mason, 1982), is covalently bound to the 5' end of the negative (long) DNA strand, which ranges in length from 3020 to 3320 nucleotides. The positive (short) DNA strand ranges in length from 1700 to 2800 nucleotides. The presence of the viral polymerase means that the negative strand is not covalently closed. Watson-Crick pairing between the negative and positive strands keeps the genome circular by bridging the gap in the negative strand. The genome codes for seven proteins, in four overlapping open reading frames (ORF), as shown in Figure 1.2 and Table 1.1. A summary of the sizes of all HBV proteins is provided in Table 1.2.

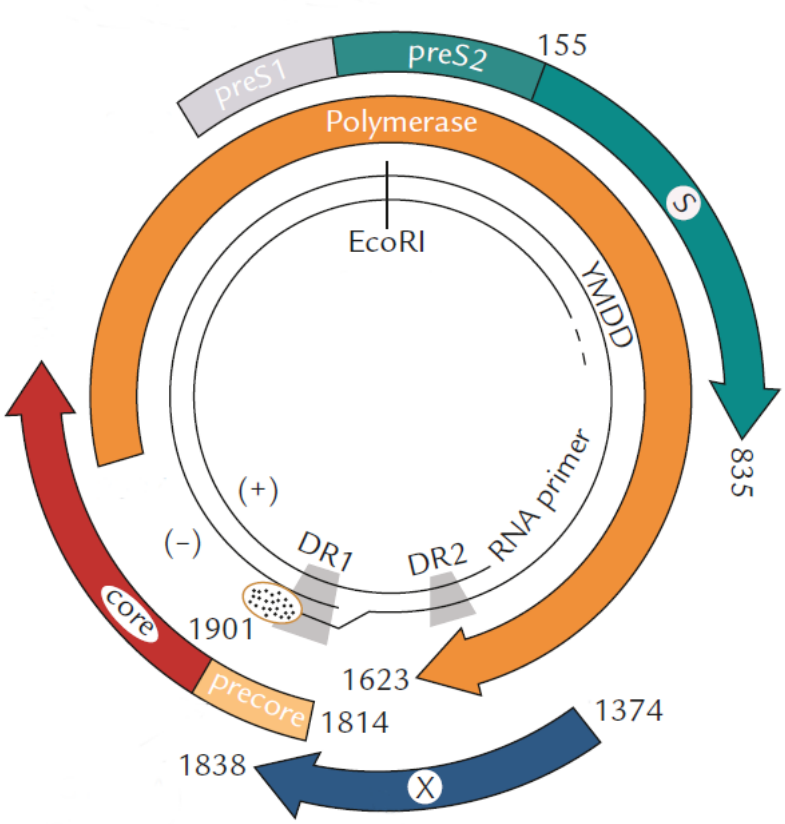


Figure 1.2. *The circular genome of HBV, showing the two partially overlapping strands, coding for seven proteins from four overlapping reading frames. The length of the genome varies by genotype. Reproduced from Lin and Kao (2010) with permission. Co-ordinates in the original image, which are not conserved across all genotypes, have been removed.*

Table 1.2. Summary of protein lengths for all HBV genotypes

Gt.	Length	PC	HBcAg	HBeAg	PreS1	PreS2	HBsAg	MHBs	LHBs	Pol.	HBx
A	3221	29	185	214	119	55	226	281	400	845	154
B	3215	29	183	212	119	55	226	281	400	843	154
C	3215	29	183	212	119	55	226	281	400	843	154
D	3182	29	183	212	108	55	226	281	389	832	154
E	3212	29	183	212	118	55	226	281	399	842	154
F	3215	29	183	212	119	55	226	281	400	843	154
G	3248	29	195	224	118	55	226	281	399	842	154
H	3215	29	183	212	119	55	226	281	400	843	154

Adapted from Kramvis *et al.* (2005); all values are number of amino acids, except for "Length" which is number of nucleotides; "Gt." is HBV genotype; "PC" is PreCore; "Pol." is polymerase.

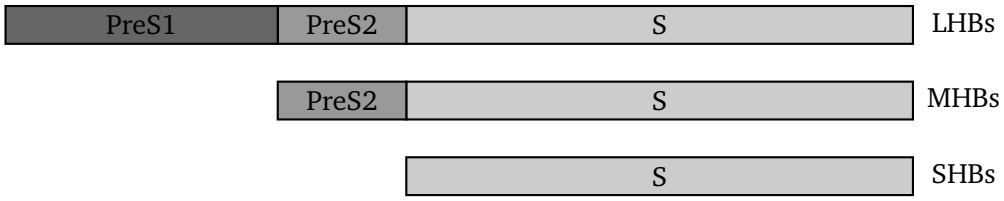


Figure 1.3. *The three domains of HBsAg (proportionally to scale; differences in the length of the PreS1 domain between genotypes is not indicated) showing the small, medium and large proteins. Redrawn from Locarnini* et al. *(2003).*

1.1.4 Viral Proteins

1.1.4.1 Surface Protein

The surface gene (S gene), which codes for the HBV surface antigen (HBsAg), contains three in-frame start codons, designated as "PreS1", "PreS2" and "S". This arrangement results in the production of proteins of three sizes, termed "small" HBs (S region only), "medium" HBs (PreS2 and S regions) and "large" HBs (PreS1, PreS2 and S regions). These are designated as SHBs, MHBs and LHBs, respectively (Figure 1.3).

Codon positions 124 to 147 of the surface gene code for the *a* determinant epitope, which is targeted by anti-HBs antibodies during immune response. Mutations in this region may change the structure of the determinant, rendering antibodies unable to bind to it (Paulij *et al.*, 1999; Salisse and Sureau, 2009). The amino acids found at several positions in the surface gene are used to determine the HBV serological subtype.

In addition to the *a* determinant, there are several allelic variations in three antigenic determinants of HBsAg (*d/y*, *w/r* and *q*), which have been recognized over the years, and have led to the classification of HBV into ten serological subtypes (Blumberg *et al.*, 1965; Magnius and Norder, 1995; Arauz-Ruiz *et al.*, 2002; Purdy *et al.*, 2007), as shown in Table 1.3. An algorithm to deduce serological subtype from the amino acid variants at key loci has been published (Purdy *et al.*, 2007). These subtypes are not distributed geographically, but do correlate loosely with HBV genotype (Kramvis *et al.*, 2008) (Section 1.1.5).

1.1.4.2 Polymerase Protein

The largest protein, the 90 kDa viral DNA polymerase (Pol), is between 834 and 845 amino acids in length and is coded for by the polymerase (P) gene, which completely overlaps the surface gene. The P gene consists of four domains: the N-terminal protein, the spacer, the rt (reverse transcription) domain, and the RNaseH domain (Figure 1.4).

Figure 1.4. *The domains of HBV polymerase (proportionally to scale). Redrawn from Zoulim and Locarnini (2009).*

1.1.4.3 Core Protein and HBeAg

The core (C gene) codes for the 21.5 kDa HBV core antigen (HBcAg), which is 183 to 195 amino acids in length. Dimers of HBcAg spontaneously assemble into the icosahedral viral capsid (nucleocapsid) in the cytoplasm of hepatocytes (Kann, 2002).

Translation from an upstream (5' end) start codon (designated "PreCore" or "Pre-C"), produces a precursor protein with 29 additional amino acids at the 5' end. This precursor protein is post-translationally modified by cleaving 10 amino acids from the amino terminal, and a variable number from the carboxyl end (Figure 1.5) (Kramvis *et al.*, 1997; Ou, 1997). The resulting 17 kDa protein, termed "e" antigen (HBeAg), is targeted to the endoplasmic reticulum by a signal peptide on the amino terminal. HBeAg is a water-soluble, non-particulate protein, expressed on the hepatocyte surface or secreted in the serum, and is not required for HBV replication (Cabrerizo *et al.*, 2000; Visvanathan *et al.*, 2007). The crystal structure of HBeAg, which has recently been determined, reveals that the protein has the ability to flip its structure between one which is not assembled into capsids and one which is (DiMattia *et al.*, 2013; Zlotnick *et al.*, 2013). The biological effects of HBeAg are not fully understood. However, it may modulate the response of the host immune system (Kann, 2002) and is thought to be a T-cell tolerogen (Chen *et al.*, 2004). HBeAg-positivity has been shown to inhibit viral replication (Lamberts *et al.*, 1993; Guidotti *et al.*, 1996; Scaglioni *et al.*, 1997) and therefore reduced HBeAg expression can result in increased viral replication (Kramvis and Kew, 1999).

HBeAg-positivity is an indicator of an active HBV infection and ongoing viral replication. HBeAg persistence indicates a chronic infection, as seroconversion of HBeAg to anti-HBe is associated with a decrease in both viral replication and liver damage (Farrell and Denstag, 2002). However, HBeAg-negativity may also result from one or more mutations in the HBV genome (Table 1.8). Most Black South Africans who are HBV-positive are infected during early childhood (Botha *et al.*, 1984) and all but 5% are HBeAg-negative by early adulthood (Song *et al.*, 1984; Ahn *et al.*, 2003). This high prevalence of HBeAg-negativity is not found elsewhere in the world (Song *et al.*, 1984; Ahn *et al.*, 2003).

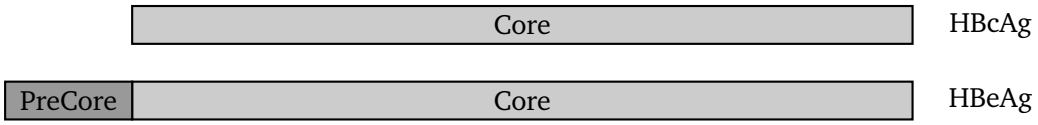


Figure 1.5. *HBcAg and HBeAg proteins (proportionally to scale; differences in the length of the Core domain between genotypes is not indicated).*

1.1.4.4 X ORF

The X gene codes for the smallest HBV protein, termed "HBx". This non-structural protein of 154 amino acids is found only in the *Orthohepadnaviridae*. HBx has been shown to act as a transcriptional transactivator (reviewed in Rossner (1992)), inhibit apoptosis and deregulate cell cycle checkpoints (Benn and Schneider, 1995; Wang *et al.*, 1995; Yang and Cho, 2012), and stimulate HBV transcription and replication (Tang *et al.*, 2005). It is also thought to play a role in the development of hepatocellular carcinoma (HCC) by binding to, and inactivating, the p53 transcription factor and tumor suppressor (Wang *et al.*, 1994; Truant *et al.*, 1995).

1.1.5 Genotypes and subgenotypes

HBV was initially classified according to serological subtype (serotype) (Table 1.3). However, after the analysis of sequence data from 18 full HBV genomes of various serotypes, four genotypes, denoted as "groups" A to D, were recognized (Okamoto *et al.*, 1988). These groups were distinguished by intergroup nucleotide divergence of at least 8%. An additional five genotypes, E and F (Norder *et al.*, 1992; Naumann *et al.*, 1993; Norder *et al.*, 1994), G (Stuyver *et al.*, 2000), H (Arauz-Ruiz *et al.*, 2002) and I (Olinger *et al.*, 2008; Tran *et al.*, 2008; Arankalle *et al.*, 2010; Osiowy *et al.*, 2010; Yu *et al.*, 2010), have subsequently been recognized. A tenth genotype, J, isolated from a single individual, has been proposed by Tatematsu *et al.* (2009). Subgenotypes have been recognized in genotypes A to D, F and I, and these are named numerically, for example "A1" or "D3" (Kramvis *et al.*, 2005). Isolates are considered subgenotypes if they show sequence divergence between 4% and less than 8% (Kramvis *et al.*, 2005).

The definition of "genotype" and "subgenotype" has been formalized as follows. Complete HBV sequences which diverge by more than 7.5% are considered to be different genotypes, whereas sequences which diverge between 4% and 7.5%, with strong bootstrap support, are subgenotypes (Kramvis *et al.*, 2008). Sequences showing divergence of less than 4% are classified as the same subgenotype. Genome length and protein size vary by genotype, as indicated in Table 1.2.

Table 1.3. HBV serotypes

Gt.	*ayw1*	*ayw2*	*ayw3*	*ayw4*	*adw2*	*adw3*	*adw4*	*ayr*	*adrq+*	*adrq-*
A	•	•			•		•			
B	•				•	•				•
C		•	•		•			•	•	•
D		•	•	•		•				
E		•		•						
F			•	•	•		•			
G					•					
H							•			

Adapted from Kramvis *et al.* (2005) and Purdy *et al.* (2007); "Gt." is HBV genotype; subgenotype I1 reported as *adw*, I2 as *ayw* (Olinger *et al.*, 2008) and genotype J as *ayw* (Tatematsu *et al.*, 2009).

Table 1.4. Location of HBV genotypes

Gt.	**Location**
A	Africa; India; Northwestern Europe; North America
B	Asia; Oceania
C	Asia; Oceania; Korea
D	Predominantly Mediterranean and India, but worldwide
E	West African Coast; Madagascar
F	Aboriginal population of South America
G	Carriers in France, Germany, United Kingdom, Italy, USA
H	Amerindians of Central America; California; Mexico
I	*Laos; Vietnam*
J	*Borneo or Japan*

Adapted from Kramvis *et al.* (2005). The existence of genotypes I and J is still considered controversial and require confirmation.

HBV genotypes correlate loosely with serotype (Kramvis *et al.*, 2005), as shown in Table 1.3, and strongly with geographic location (reviewed in Kramvis *et al.* (2005)); see Table 1.4. Both genotypes A and D are found in southern Africa, with genotype A being more prevalent (Bowyer *et al.*, 1997).

Disease progression, clinical manifestation of illness, treatment response and distribution of mutations may differ between HBV genotypes and subgenotypes (Rodriguez-Frias *et al.*, 1995; Mayerat *et al.*, 1999; Sanchez-Tapias *et al.*, 2002; Sumi *et al.*, 2003; Westland *et al.*, 2003); reviewed in Kramvis and Kew (2005). Genotype A is strongly associated with chronic infection when compared with genotype D (Mayerat *et al.*, 1999). In HBeAg-positive chronic infections, subgenotype A2 is more prevalent, with anti-HBe-positive infections showing a greater prevalence in genotype D (Kramvis and Kew, 2005). HBV DNA viral load levels are significantly lower in subgenotype A1 compared to subgenotype A2 and genotype D (Tanaka *et al.*, 2004).

Recombination between HBV strains is possible. *In vivo* recombination between wild-type duck HBV genomes was the first to be reported (Sprengel *et al.*, 1987). Since then, recombination has been reported *in vitro* (Hino *et al.*, 1991), in HBV DNA from HCC samples (Georgi-Geisberger *et al.*, 1992) and between wild-type and mutant genomes (Tran *et al.*, 1991). Genotype A to D

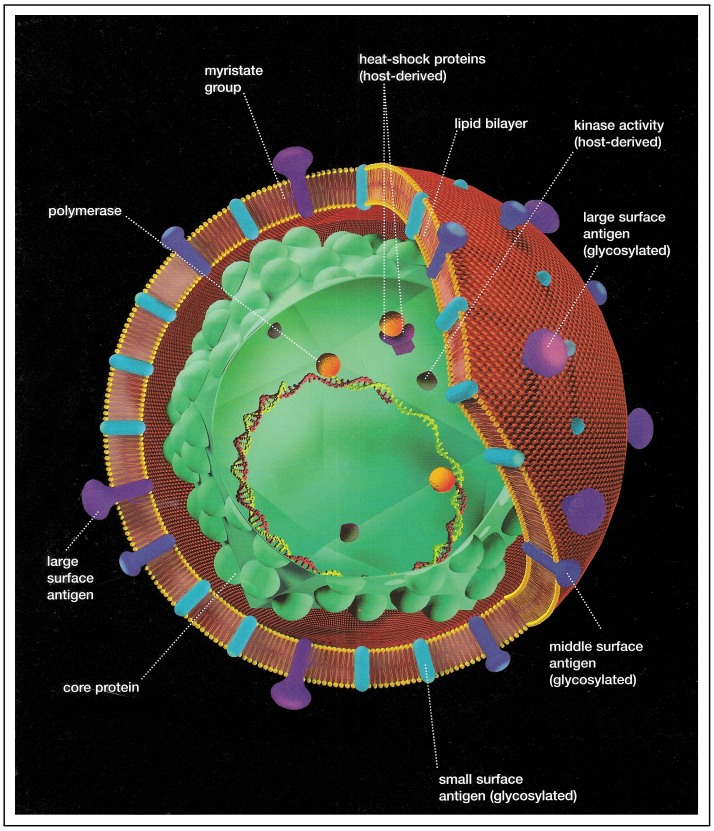


Figure 1.6. *The structure of the HBV virion, from Lai and Locarnini (2002). © 2002, International Medical Press. Used with permission.*

recombination has been found in 11.5% of isolates from Black South African adults (Owiredu *et al.*, 2001), with D to A recombination also being reported (Bollyky *et al.*, 1996; Bowyer and Sim, 2000).

1.1.6 Structure

The HBV virion (virus particle) is ∼42 nm in diameter and is one of the smallest enveloped viruses infecting animals (Figure 1.6). Five of the seven HBV proteins are present in the structure of the virion. The icosahedral nucleocapsid, consisting of HBcAg, contains both the viral DNA and a DNA polymerase. This capsid is enclosed in an outer lipid membrane (envelope), which is derived from the host hepatocyte and into which all three surface antigens (small, medium and large) are embedded, to form the infectious, or Dane, particle. In addition to this infectious virion, non-infectious filamentous and spherical sub-viral particles (22 nm in diameter) lacking a protein core and viral nucleic acid, also occur. Neither HBeAg, which is secretory (extra-particular or extra-cellular), nor HBx, are present in the structure of HBV.

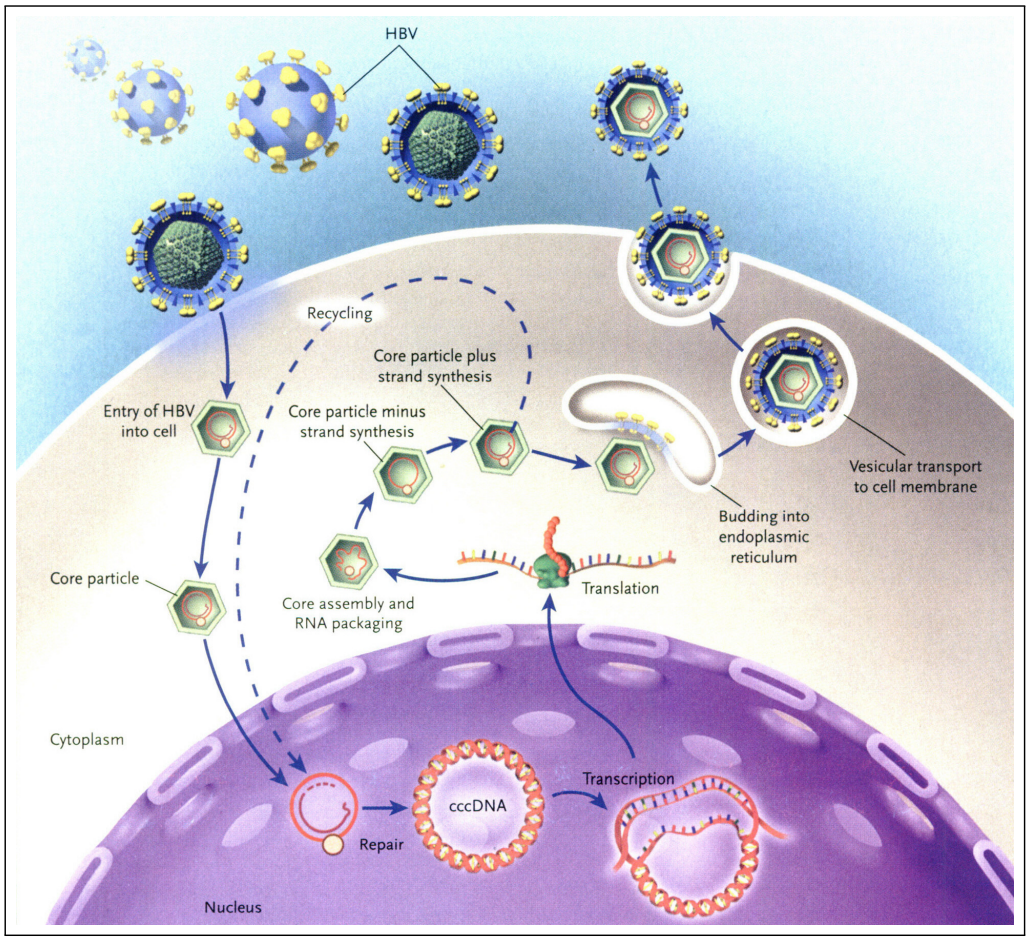


Figure 1.7. *The HBV life cycle. Reproduced, with permission, from Ganem and Prince (2004). Copyright Massachusetts Medical Society.*

1.1.7 Life cycle and Replication

HBV replication is a complex process (Figure 1.7), which involves several stages: binding of the virion to the host hepatocyte, transport of the virus within the cell, conversion of relaxed circular DNA (rcDNA) to covalently closed circular DNA (cccDNA), transcription, expression of core and polymerase proteins, encapsidation, reverse transcription, synthesis of positive strand DNA, circularization, progeny capsid trafficking and formation of a closed circular DNA pool (Jilbert *et al.*, 2002).

The specific cellular receptor or receptors, which the virus uses to gain entry to the hepatocyte, have not yet been identified. Evidence suggests that the HBV envelope attaches irreversibly to a receptor, or receptors, and is taken into the cell via endocytosis (Jilbert *et al.*, 2002; Urban *et al.*, 2010). As all three surface proteins (SHBs, MHBs and LHBs) are present in the viral envelope, any of the HBsAg domains (preS1, preS2 or S) may be the ligand for the unknown receptor (Jilbert *et al.*, 2002), although evidence suggests a hepatocyte-specific preS1 receptor (Urban

et al., 2010). These findings are supported by a recent study reporting that a receptor-binding region of the preS1 region interacts with a multiple transmembrane transporter, which is expressed predominantly in the liver (Yan *et al.*, 2012).

After entering the cell, the envelope is removed and the viral core is transported along microtubules, passes through the nucleopore, and enters the nucleus, where the rcDNA is converted to cccDNA by host enzymes (Tuttleman *et al.*, 1986; Jilbert *et al.*, 2002; Urban *et al.*, 2010). During this process, the positive viral DNA strand is completed and the 5' covalently-linked viral polymerase is removed. The negative strand of the cccDNA in the nucleus serves as the template for the transcription of viral RNA by host RNA polymerase II. The mRNA synthesized for the LHBs, MHBs, SHBs and X proteins is subgenomic in length, whereas that synthesized for the HBeAg, core and polymerase proteins, are longer than the genome (Will *et al.*, 1987; Jilbert *et al.*, 2002).

Both the 5' and 3' ends of the pregenomic RNA contain an encapsidation signal (ϵ), which takes the form of a stem-loop structure with a bulge (Junker-Niepmann *et al.*, 1990). The binding of the polymerase protein to the 5' copy of this bulge initiates encapsidation of the RNA pregenome and reverse transcription (Summers and Mason, 1982; Pollack and Ganem, 1994). Core protein dimers then encapsidate the polymerase-pregenome complex. Negative strand DNA synthesis, by reverse transcription (Summers and Mason, 1982), is then initiated by the polymerase protein. The RNaseH activity of the polymerase degrades the RNA template. A short 18-nucleotide sequence at the 5' end of the RNA pregenome, which is not degraded, initiates positive strand DNA synthesis by binding to the 5' end of the negative strand. Positive strand synthesis proceeds from this primer towards the 5' end of the negative strand. A terminal redundancy, *r*, at the 3' end of the negative strand then causes the 3' end of the positive strand to "jump" to *r* (circularizing the genome), from where positive strand synthesis continues (Summers and Mason, 1982; Will *et al.*, 1987; Wang and Seeger, 1993; Jilbert *et al.*, 2002).

Nucleocapsids containing double-stranded DNA and the polymerase protein either return to the nucleus or are secreted from the cell as infectious virions (Jilbert *et al.*, 2002; Urban *et al.*, 2010). Those which are secreted from the cell first move into the lumen of the endoplasmic reticulum, where they are enveloped and then move through the Golgi body into a secretory vesicle, which is released from the hepatocyte (Ganem and Prince, 2004). Both infectious virions and non-infectious sub-viral particles may be released by the hepatocyte (Urban *et al.*, 2010).

Although HBV DNA is known to integrate into the host at a variable frequency, this integration is not required for HBV replication (Dejean *et al.*, 1983; Nakamura *et al.*, 1988).

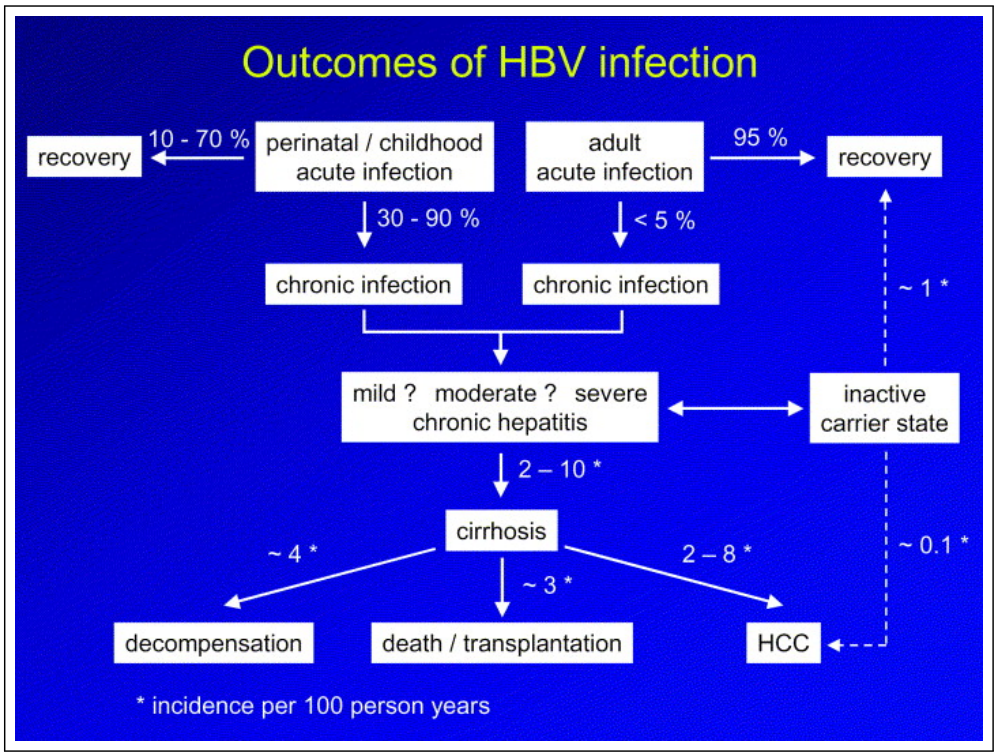


Figure 1.8. *Outcomes of HBV infection. Reproduced from de Franchis* et al. *(2003), with permission from Elsevier.*

Table 1.5. Stages of chronic HBV infection

Characteristic	Immunotolerant	Immunoactive	Inactive Carrier
HBsAg	Yes		
Anti-HBs	No	No	Yes
HBeAg	Yes	Possible	No
Anti-HBe	No	No	Yes
HBV DNA	High	Decrease	Very low
Aminotransferases	Normal/Slightly elevated	Increase	Normal
Symptoms	None	Possible	
HBV Replication	Yes	Yes	Very low

Adapted from de Franchis *et al.* (2003).

1.1.8 Natural History of HBV Infection

The outcomes of HBV infection are indicated in Figure 1.8. Both adult and childhood acute infection can either resolve or progress to chronic infection. Chronic infection can either develop into an inactive carrier state, or progress to chronic hepatitis and liver cirrhosis and hepatocellular carcinoma (liver cancer) (de Franchis *et al.*, 2003). Various stages of chronic HBV infection are recognized, as detailed in Table 1.5. Two variants of chronic hepatitis are possible: HBeAg-positive and HBeAg-negative chronic hepatitis. Serological patterns of acute and chronic infections are shown in Table 1.6.

Table 1.6. Serological patterns in HBV infection

Serology	Incub.	Acute Infection	Recovery	Chronic >6 months	Chronic <6 months	Imm. Carrier	Inactive
HBsAg	+	+	-	+	+	-	+
Anti-HBs	-	-	+	-	-	+	-
Anti-HBc IgG	-	+/-	+	+	+	-	+
Anti-HBc IgM	-	+	+/-	+/-	+/-	-	*
HBeAg	+	+	-	+	-	-	-
Anti-HBe	-	-	+	-	+	-	+
HBV DNA	+	+	-	+	+	-	+

"IgG" is "Total"; "*" means "not applicable"; Adapted from Kao and Chen (2002) and Torbenson and Thomas (2002).

1.1.8.1 Acute Infection

Infections, which develop within 6 months of HBV exposure, and then resolve within 6 months, are considered acute (de Franchis *et al.*, 2003). These are self-limiting and are indicated by elevated alanine transaminase (ALT) levels and the presence of anti-HBc in the serum (Goodman, 2002; de Franchis *et al.*, 2003). Liver injury or disease is a result of the host immune system initiating apoptosis (programmed cell death) and/or necrosis of infected hepatocytes (Goodman, 2002). Fulminant hepatitis is a severe manifestation of an acute infection.

1.1.8.2 Chronic Infection

If the host immune system is unable to clear the virus from the hepatocytes, the infection will become chronic, which is formally defined as the persistence of HBsAg in the serum for longer than six months (Farrell and Denstag, 2002). The extent of liver damage, which results from the continuing inflammatory response, varies between individuals (Goodman, 2002). Several factors may increase the chances that an infection will develop to chronicity, including the age and immune status of the individual, as well as the persistence of HBeAg. Individuals infected at birth have a 90% chance of developing chronic infection and this decreases to less than 5% for those infected during adulthood (Farrell and Denstag, 2002). Acute HBV infection in individuals with compromised immunity, such as those infected with HIV, is more likely to develop into chronic HBV (Farrell and Denstag, 2002). Persistence of HBeAg for more than three months is an indication that the infection is likely to become chronic (Farrell and Denstag, 2002). Seroconversion of HBeAg to anti-HBe is associated with a decrease in both viral replication and liver damage (Farrell and Denstag, 2002). However, the loss of HBeAg may also be associated with the development of mutant HBV strains.

Four phases of the natural history of chronic HBV are recognized, although all infected individuals do not necessarily experience each phase: (1) immune tolerance, (2) immune clearance

(HBeAg-positive) or immune active phase, (3) inactive carrier (HBeAg-negative) and (4) HBsAg clearance and/or reactivation (Yim and Lok, 2006; McMahon, 2009). The immune tolerance phase, which lasts from 1 to 4 decades, is characterized by elevated levels of HBV DNA in the serum, HBeAg-positivity, normal ALT levels and reduced immunological response to HBV infection (Yim and Lok, 2006; McMahon, 2009). The liver may, however, show some symptoms of infection (Kao and Chen, 2002). During the second phase, HBV DNA levels and ALT levels may fluctuate, the liver may show active inflammation or fibrosis and HBeAg seroconversion may occur (Yim and Lok, 2006; McMahon, 2009). In the third phase, ALT levels return to normal, anti-HBe is present, liver inflammation and fibrosis decrease and HBV DNA serum levels become undetectable (Yim and Lok, 2006; McMahon, 2009). This phase may persist indefinitely in some individuals, resulting in a positive prognosis (Yim and Lok, 2006). However, in some individuals, a fourth phase occurs, characterized by the reactivation of HBV replication, elevated HBV DNA serum levels, elevated ALT levels and further liver damage (Yim and Lok, 2006).

Fibrosis Fibrosis, defined as "the presence of excess collagen due to new fibre formation" (Anthony *et al.*, 1977), is a fibrous scarring, which varies between individuals, but is almost inevitable in chronic HBV infection (Goodman, 2002). It results from abnormal fibrogenesis during the healing of liver tissue (Schuppan and Afdhal, 2008).

Cirrhosis Cirrhosis is defined as "a diffuse process characterized by fibrosis and the conversion of normal liver architecture into structurally abnormal nodules" (Anthony *et al.*, 1977). It is considered to be an advanced stage of liver fibrosis in which the hepatic vasculature and architecture are distorted such that blood flow in the liver is compromised (Schuppan and Afdhal, 2008).

HCC Hepatocellular carcinoma (HCC, liver cancer) is one of the leading causes of death worldwide, particularly in southern China and sub-Saharan Africa, where it is responsible for 10% of all deaths (Bunz, 2008). Important risk factors for HCC include infection with HBV, infection with hepatitis C virus, exposure to aflatoxin and alcohol-induced liver disease (Bunz, 2008). Fibrosis disrupts interactions between hepatocytes, which are normally quiescent, resulting in uncontrolled growth (Bréchot *et al.*, 2000). Hepatocytes, which are not eliminated via apoptosis or the immune response, can therefore become fully transformed (Bréchot *et al.*, 2000).

1.1.8.3 Occult HBV Infection

Presence of HBV in a sample is routinely determined by an enzyme-linked immunosorbent assay (ELISA) test for HBsAg. This test is relatively inexpensive, particularly when samples are processed in bulk, and is easy to administer. A positive result for HBsAg indicates the presence of HBV. However, a variant form of HBV infection, in which HBsAg test results are negative, has been described (Bréchot *et al.*, 1981; Hoofnagle *et al.*, 1978; Hu, 2002; Raimondo *et al.*, 2007, 2013). Such infections are termed "occult HBV infections" (OBI). The clinical and virological relevance of OBI are considered particularly challenging to understand, with the issues being debated for the last three decades (Raimondo *et al.*, 2008). OBI has been referred to as "inapparent HBV infection", "serologically silent hepatitis B", "silent hepatitis B", "surface antigen negative carriers" and "unrecognized hepatitis B infection" (reviewed in Torbenson and Thomas (2002)).

OBI is defined by the *Taormina* expert panel as "the presence of HBV DNA in liver (with detectable or undetectable HBV DNA in the serum) of individuals testing HBsAg-negative by currently available assays. When detectable, the amount of HBV DNA in the serum is usually very low (<200 IU/ml)" (Raimondo *et al.*, 2008). As liver biopsies are not commonly available, particularly in resource-limited environments, OBI is usually detected by the analysis of serum samples (Raimondo *et al.*, 2008). Samples, which are HBsAg-negative, with serum HBV DNA levels < 200 IU/ml, are considered "true" OBI. Samples in which serum HBV DNA levels are similar to HBsAg-positive ("overt") infection, but are nevertheless HBsAg-negative, have been termed "false" OBI by the *Taormina* panel. "False" OBI may be explained by the inability of commercial detection assays to detect HBsAg variants produced by HBV surface (S) gene mutants (Raimondo *et al.*, 2008). Seronegative OBI (negative for all serological marks) accounts for approximately 20% of all OBI cases, with the remainder being seropositive OBI (50% anti-HBc-positive and 35% anti-HBc-positive) (Torbenson and Thomas, 2002; Raimondo *et al.*, 2008; Hollinger and Sood, 2010).

The *Taormina* panel proposed a "gold standard" for detecting OBI (Raimondo *et al.*, 2008). This involves amplifying HBV DNA from liver and/or blood samples, using nested PCR, or real-time PCR, with a detection limit of less than 10 copies of HBV DNA per reaction. Oligonucleotide primers for three different, highly conserved, regions of the HBV genome should be used (Raimondo *et al.*, 2007). Samples, which yield amplicons from at least two different genomic regions, are considered to be occult infections (reviewed in Raimondo *et al.* (2007)).

The clinical implications of OBI in immunocompromised individuals, such as those infected

with HIV, remain unclear (Mphahlele *et al.*, 2006; Raimondo *et al.*, 2007). Furthermore, HIV has been identified as a risk factor for the development of OBI (Mphahlele *et al.*, 2006).

1.1.9 Epidemiology

Prevalence of HBV infection, as measured by HBsAg-positivity, varies greatly around the world (Table 1.7). Areas of high endemicity include sub-Saharan Africa, China, India and other parts of south-east Asia. It is estimated that 65% to 98% of populations in sub-Saharan Africa have been exposed to HBV and 8% to 20% are chronic carriers of HBV (Kramvis and Kew, 2007), far exceeding the 4% to 6% lifetime exposure rates and 0.2% to 0.5% carrier rates in regions of low endemicity.

The virus is transmitted between individuals via bodily fluids, either by direct blood contact, or by unprotected sexual contact (World Health Organization, 2012a). In sub-Saharan Africa, this transmission is mainly horizontal during the first few years of childhood (Barin *et al.*, 1981; Whittle *et al.*, 1983; Kew, 1996; World Health Organization, 2012a). Horizontal transmission results from direct contact with blood, such as via open wounds, or contact with saliva (Davis *et al.*, 1989; Heiberg *et al.*, 2010). In many African countries, horizontal transmission routes may also include traditional tattooing, scarification and circumcision (Robson *et al.*, 1994). Although transmission by insect bites remains unproven, in West African countries transmission has been associated with bed bugs and skin diseases, rather than with traditional scarification and circumcision (Mayans *et al.*, 1990; Robson *et al.*, 1994).

In developed countries, transmission is typically via unprotected sexual contact, from mother to child at birth (perinatal or vertical transmission) or via intravenous drug use (IDU) (World Health Organization, 2012a). Therefore, in developed countries, HBV infection often occurs later in life, at the time of sexual debut, rather than during childhood, as in sub-Saharan African countries (Botha *et al.*, 1984). HBV infection is a hazard for health-care workers, but cannot be transmitted by food or water (World Health Organization, 2012a).

HBV infection can be prevented by vaccination prior to exposure to the virus. In 1992, when the World Health Assembly passed a resolution to recommend global hepatitis B vaccination, only 31 countries worldwide were administering the vaccine (World Health Organization, 2012c). By 2010, the vaccine had been introduced into the national infant immunization programme in 179 countries (World Health Organization, 2012c), with 87% of countries in Africa administering the vaccine in 2008 (François *et al.*, 2008). Over one billion doses of the vaccine, which is 95% ef-

Table 1.7. Worldwide prevalence of HBV; Reproduced from Lok *et al.* (2007).

Prevalence	High	Intermediate	Low
Carrier rate	8% to 20%	3% to 7%	0.1% to 2%
Geographic distribution	Southeast Asia China Pacific Islands Sub-Saharan Africa Alaska (Eskimos)	Mediterranean basin Eastern Europe Central Asia Japan Latin/South America Middle East	United States Canada Western Europe Australia New Zealand
Predominant Age at Infection	Perinatal and Early childhood	Early childhood	Adult
Predominant Mode of Transmission	Maternal-Infant Percutaneous	Percutaneous Sexual	Sexual Percutaneous

fective in preventing infection, have been administered worldwide (World Health Organization, 2012a,c). In South Africa, universal hepatitis B vaccination, given at 6, 10 and 14 weeks, was introduced into the Expanded Programme on Immunization (EPI) in April 1995 (Aspinall, 1995; Tsebe *et al.*, 2001). Prior to this, more than 70% of Black South Africans had been exposed to HBV and almost 10% were HBsAg carriers (Kiire, 1996; Tsebe *et al.*, 2001). Five years after the introduction of the vaccine, the HBsAg carrier rate in children younger than five years had decreased favourably (Tsebe *et al.*, 2001). Despite no catch-up vaccination programme for adolescents and no school-based vaccination programme, the implementation of HBV vaccination in South Africa is considered a success (Burnett *et al.*, 2012). To date, no national HBV surveillance programme has been conducted in South Africa.

1.1.10 BCP/Pre-Core/Core Mutations

The region of the HBV genome from position 1591 to 1822 (numbered according to the *Eco*RI site (Galibert *et al.*, 1979b)) is the core promoter (CP), which consists of the BCP (basic core promoter) and the URR (upper regulatory region) (Kramvis and Kew, 1999). Initiation of pre-core mRNA and pgRNA transcription *in vivo* requires only the BCP region. Several upstream regulatory sequences are found in the URR. The BCP region, from 1742 to 1849, contains the direct repeat 1 (DR1), which is essential for reverse transcription (Ganem and Varmus, 1987). The pre-core region starts at position 1814 and extends to position 1900, with the core region starting at position 1901.

HBeAg expression may be down-regulated, or even abolished entirely, as a result of mutations in four locations, effective at the three different stages of protein synthesis (Kramvis and Kew, 2007), as follows (Figure 1.9 and Figure 1.10).

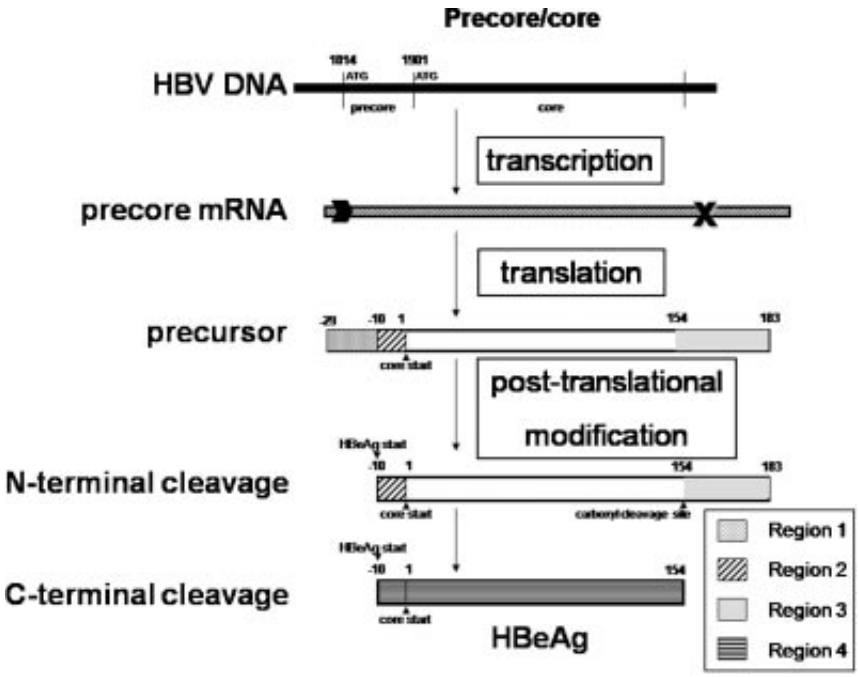


Figure 1.9. *HBeAg expression. Reproduced from Revill* et al. *(2010) with permission.*

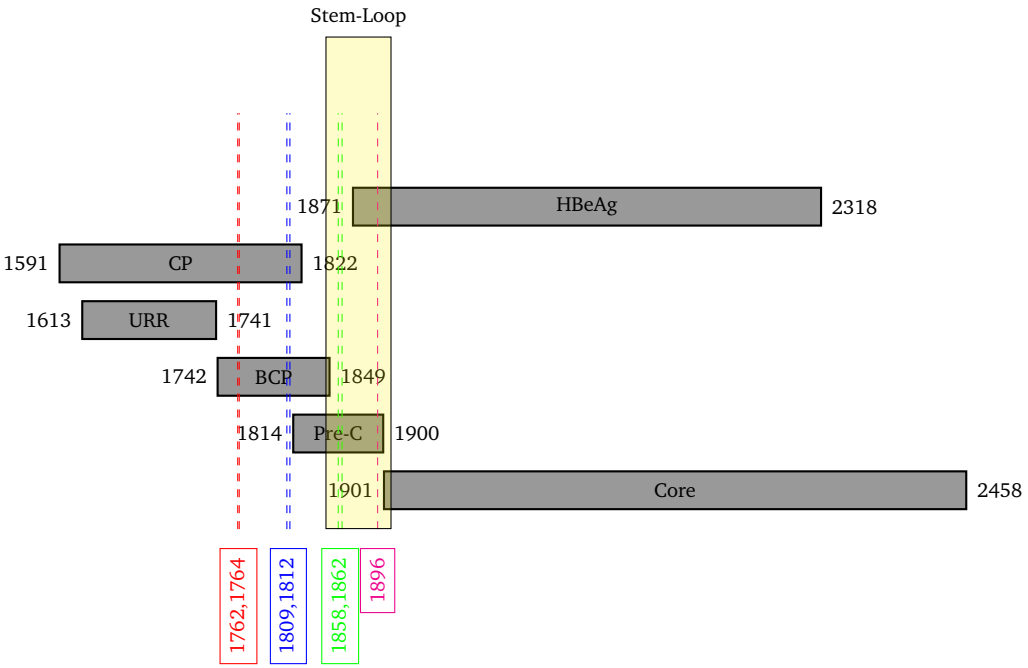


Figure 1.10. *Schematic representation of the BCP, PC and Core regions, showing the positions of key mutations.*

A1762T/G1764A This double mutation, located in the BCP region and consisting of an A → T transversion at 1762 and a G → A transition at 1764, results in the down-regulation of transcription of pre-core mRNA, thereby reducing HBeAg expression and increasing pre-genomic RNA packaging and viral progeny production (Buckwold *et al.*, 1997). The down-regulation of transcription is a result of a liver-enriched transcription factor being unable to bind to the mutant strains at positions 1759 to 1769 (Buckwold *et al.*, 1996). The mutations are present in chronic and fulminant infections, and sera, tumors and liver tissue of hepatocellular carcinoma infections, but are less prevalent in asymptomatic infections, immunocompromised individuals, and carriers with no HBV markers (Kramvis and Kew (1999) and references therein). The mutations occur more frequently in genotype C than in genotypes B or D (Kramvis and Kew, 1999; Kao *et al.*, 2003). HBV isolates, which have C1858, are more likely to develop these mutations, since C1858 precludes the development of the 1896 mutation (Li *et al.*, 1993; Lok *et al.*, 1994; Kramvis and Kew, 1999). These mutations develop before or during HBeAg seroconversion, and may or may not be associated with HBeAg-negativity (Kramvis and Kew, 1999). HBeAg-positive individuals with these mutations have lower viral loads and lower DNA polymerase levels (Takahashi *et al.*, 1995; Baptista *et al.*, 1999). Those with these mutations have lower HBeAg expression (Baptista *et al.*, 1999), meaning more HBcAg is targeted, resulting in increased liver damage. The mutations are rarely detected alone: A1764 alone slightly reduces viral replication, and T1762 alone occurs only very rarely. The double mutation is non-synonymous, resulting in a lysine to methionine change at amino acid 130 of the X protein, and a valine to isoleucine at position 131. These changes could therefore affect the structure of the X protein (Kramvis and Kew, 1999; Baptista *et al.*, 1999). These mutations predominate in South African subgenotype A1 HCC individuals, compared to asymptomatic ones (Kramvis and Kew, 2007), and have been implicated as a risk factor for HCC (Baptista *et al.*, 1999).

1809–1812 The region immediately preceding the pre-core start codon, termed the "Kozak" sequence, is critical for initiating translation. Variations in the Kozak sequence may result in sub-optimal translation of the downstream protein. In HBV, the pre-core start codon is located within the BCP region in all genotypes, at co-ordinate 1814, with the Kozak sequence occurring at positions 1809-1812. Position 1813 is conserved as a C. Of the HBV strains circulating in South Africa, 80% have a double or triple mutation at 1809, 1811 and/or 1812 (Baptista *et al.*, 1999; Ahn *et al.*, 2003). Mutations in the Kozak sequence in

subgenotype A1 cause leaking scanning, which results in a reduction of HBeAg expression (Ahn *et al.*, 2003). Mutations in this region affect HBeAg expression at the translational level. Together with the 1762/1764 mutations described above, these mutations in the "Kozak" sequence can reduce HBeAg expression to very low levels.

1862 A transversion from G to T at 1862 is common in South African HBV strains and occurs more frequently in genotype A, and particularly in subgenotype A1, than in other genotypes (Kramvis *et al.*, 1997; Kramvis and Kew, 1998, 2007). This mutation effects a valine to phenylalanine missense mutation in the pre-core region, which may interfere with signal peptide cleavage. The mutant HBeAg produced by this variant is not secreted, but rather accumulates in the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) (Chen *et al.*, 2008). Position 1862 is located in the bulge of ε, a region which is critical for the initiation of reverse transcription of the pregenomic RNA. The effect of this mutation is at the post-translational level.

1896 A transition from G to A at 1896 creates a "STOP" codon (nonsense mutation) in the pre-core/core protein, thereby truncating it to a polypeptide of only 28 amino acids in length (Carman *et al.*, 1989, 1991). The expression of HBeAg is therefore terminated by this mutation. Position 1896 is located in the stem loop of ε, opposite position 1858. A mutation from G to A at 1896 requires a T at position 1858 for stable Watson-Crick base-pairing. Therefore, genotypes with C1858 (genotypes A and H, and some genotype C and F sequences) are less likely to develop the G1896A mutation than those with T1858 (genotypes B, D and E) (Li *et al.*, 1993; Lok *et al.*, 1994; Kramvis and Kew, 2005). The high prevalence of HBeAg-negativity in South Africa, where subgenotype A1 predominates, is therefore not as a result of the G1896A mutation (Kramvis *et al.*, 1997). Mutations at this position reduce viral replication in genotype A, but not in genotype D (Chen *et al.*, 2008). This is because the G1896A mutation destabilizes ε in genotype A, which has C1858, but stabilizes it in genotype D, which has T1858. The effect of this mutation is at the translational level.

A summary of the mutations affecting HBeAg expression is provided in Table 1.8.

1.1.11 Drug Resistance Mutations (S)

Sequence heterogeneity resulting from low fidelity replication is considered a feature of the HBV and HIV genomes, owing to the lack of proofreading ability of their viral polymerases (Steinhauer and Holland, 1986; Yamamoto *et al.*, 1994). The nucleotide substitution rate, per site per year,

Table 1.8. BCP/PC/C Mutations

Co-ord.	Mut.	Pos.	Type	HBeAg	Notes	In SA	Ref.
1762 1764	A → T G → A	BCP	Transcription	Reduced	Down-regulates transcription of pre-core mRNA	Yes	A
1809-1812	Various	Kozak	Translation	Reduced	Leaky scanning	Yes	B
1862	G → T	ε (Bulge)	Post-translation	Reduced	Intracellular retention and impaired secretion of HBeAg	Yes	C
1896	G → A	ε	Translation	Abolished	STOP codon truncates pre-core/core protein	No; requires C1858	D

A: Buckwold *et al.* (1997); B: Ahn *et al.* (2003); C: Chen *et al.* (2008); D: Carman *et al.* (1989)

for HBV has been estimated to be 1.4×10^{-5} to 8×10^{-5} (Okamoto *et al.*, 1987; Orito *et al.*, 1989; Mimms, 1995; Fares and Holmes, 2002; Osiowy *et al.*, 2006), approximately the same as retroviruses (10^{-5}), but 10^4 times higher than DNA genomes (Holland *et al.*, 1982; Orito *et al.*, 1989).

Mutations conferring resistance to a range of drugs have been described in HBV (Lindström *et al.*, 2004; Suzuki *et al.*, 2006; Ijaz *et al.*, 2008; Suzuki *et al.*, 2008) and HIV (Pillay *et al.*, 2000; O'Meara *et al.*, 2001; Hoffmann *et al.*, 2007; McMahon *et al.*, 2007; Wang *et al.*, 2007). A comprehensive resistance mutation database is currently available for HIV (Hoffmann *et al.*, 2007). Although many studies on HBV/HIV co-infections in areas of low HBV endemicity (such as North America and Europe) have been performed, further investigation is needed in Africa, where both HBV and HIV are highly endemic, unique combinations of genotypes of both viruses circulate, and there is a paucity of such studies (Burnett *et al.*, 2005).

The synthetic dideoxynucleoside analogue lamivudine has shown strong *in vitro* antiviral action (by inhibiting reverse transcriptase) against both HBV and HIV (Doong *et al.*, 1991; Coates *et al.*, 1992) and is widely used in the treatment regimes of both diseases. However, long-term use of lamivudine can induce viral resistance in HBV, followed by treatment failure, in both immunocompetent and immunosuppressed patients (Honkoop *et al.*, 1997; Lindström *et al.*, 2004). Resistance to lamivudine is a result of one of several possible mutations in the YMDD motif of the C domain of the HBV DNA polymerase gene (Bozdayi *et al.*, 2003). Lamivudine is incorporated less efficiently into the proviral HBV strand as a result of the mutation (Severini *et al.*, 1995). Between one quarter and one third of patients will develop YMDD mutations after one year of lamivudine treatment, with another third showing the mutation after a second year of treat-

ment (Tassopoulos *et al.*, 1999; Leung *et al.*, 2001; Wolters *et al.*, 2002). This can be interpreted as an approximately 29% risk of a lamivudine-treated patient developing drug resistance after 1 year of treatment. The most common YMDD mutation changes the methionine (M) residue at position rt204 to either valine (V), isoleucine (I), or, more rarely, serine (S) (Allen *et al.*, 1998; Buti *et al.*, 1998; Fu and Cheng, 1998; Niesters *et al.*, 1998; Bozdayi *et al.*, 2003). Double or triple lamivudine-resistant mutations have also been reported (Thibault *et al.*, 1999; Santos *et al.*, 2004). Mutations are also known to occur downstream of the YMDD motif, and in other domains of the polymerase gene (Delaney *et al.*, 2001). Mutations in the precore region of the HBV genome have been reported in lamivudine-treated individuals in which YMDD mutations were found (Suzuki *et al.*, 2002). No relationship exists between the appearance of HBV and HIV lamivudine-resistant mutations, nor is there any evidence of co-evolution of resistance between the two viruses (Pillay *et al.*, 2000).

Mutations in HBV and/or HIV which confer resistance to interferon-α, lamivudine, emtricitabine, adefovir and entecavir have previously been reported (Thio *et al.*, 2005). HBV/HIV co-infection increases the chronicity of HBV, prolongs viraemia and increases liver-related morbidity (Levy and Grant, 2006). Response to treatment may be influenced by viral mutations. For example, mutations in the precore and core regions may reduce the effectiveness of interferon-α and YMDD mutations result in reduced suppression of HBV by way of reduced sensitivity to nucleoside analogues (Tisdale *et al.*, 1993; Wainberg *et al.*, 1995; Hunt *et al.*, 2000).

1.2 HIV

The human immunodeficiency virus (HIV) is a retrovirus, belonging to the *Retroviridae* family. These enveloped RNA viruses use reverse transcription to produce viral DNA, which is subsequently integrated into the host genome. As a member of the *Lentivirus* genus, HIV is characterized by a long incubation period and the ability to infect non-dividing host cells. CD4+ T cells, macrophages and dendritic cells of the immune system of infected individuals are specifically targeted and killed by HIV, resulting in host immunodeficiency. Individuals with such compromised immunity will be increasingly susceptible to infections and disease. After 10 to 15 years, HIV infection can develop into Acquired Immunodeficiency Syndrome (AIDS). Individuals at this advanced stage of infection will present with severe illness, including tuberculosis and cryptococcal meningitis, and cancers such as lymphomas and Kaposi's sarcoma. The World Health Organiza-

tion recognizes four clinical stages of HIV infection, progressing from asymptomatic to advanced infection. Individuals classified as "clinical staging IV" are considered to have developed AIDS.

In 1981, AIDS was formally recognized as an illness by medical professionals in the USA (Sharp and Hahn, 2010), with the name "AIDS" formally being adopted the following year. At the time, AIDS was seen almost exclusively in homosexual men and injecting drug users, and was associated with Kaposi's sarcoma, *Cryptococcus* and *Pneumocystis* pneumonia. However, the long incubation period of HIV and the lack of a formal description of AIDS prior to 1981, meant that prior cases of infection may not have been correctly diagnosed. Retrospective analyses have since attributed several deaths between 1959 and 1980 to AIDS. Analysis of preserved blood samples from an individual who died in the Congo in 1959 suggest that the first human HIV infection took place in, or prior to, 1959 (Zhu *et al.*, 1998).

In the early 1980s, two independent groups (researchers Luc Montagnier and Francoise Barré-Sinoussi in France, and Robert Gallo in the USA[5]) identified a novel retrovirus associated with AIDS. Montagnier's group named the virus "LAV" (lymphadenopathy associated virus) and Gallo's group named it "HTLV-III" (human T-lymphotropic virus type III). In 1986, the virus was renamed "human immunodeficiency virus".

Contact with blood or other bodily fluids of an HIV-infected individual can result in transmission of the virus. Common routes of infection include sexual intercourse, mother-to-child transmission, blood transfusion, or contaminated equipment, such as that used for injections, tattoos, body-piercing or surgery. Prevention strategies include the use of male and female condoms, testing and counselling, pre-exposure prophylaxis (PrEP) for the HIV-negative partner, post-exposure prophylaxis (PEP), male circumcision, elimination of mother-to-child transmission, anti-retroviral therapy and the use of sterile equipment by injecting drug users. HIV infection is determined by the presence of antibodies in the blood of an individual after a window period of 3 to 12 weeks.

Over the last three decades, 25 million people worldwide have died as a result of HIV infection. Of the 34 million individuals presently infected with HIV, 22.5 million (∼60%) reside in sub-Saharan Africa (World Health Organization, 2012b).

In 2009, the ten countries comprising southern Africa (Angola, Botswana, Lesotho, Malawi, Mozambique, Namibia, South Africa, Swaziland, Zambia, Zimbabwe) accounted for 34% of all HIV infections worldwide (UNAIDS, 2010a), with the largest epidemic in South Africa. The highest worldwide adult HIV prevalence of almost 26% occurs in Swaziland (UNAIDS, 2010a),

[5]Montagnier and Barré-Sinoussi were awarded the 2008 Nobel Prize in Physiology or Medicine for their discovery of HIV; Gallo was not included as a recipient.

Table 1.9. HIV prevalence in South Africa in 2008, by sex, age, race and province; reproduced from Shisana *et al.* (2009).

Variable	n	HIV+ (%)	95% CI
Sex			
Male	5938	7.9	6.8-9.2
Female	8284	13.6	12.5-14.8
Total	14222	10.9	10.0-11.9
Age Group			
2-14	3414	2.5	1.9-3.5
15-24	3617	8.7	7.2-10.4
25+	7191	16.8	15.3-18.4
Population Group			
African	8702	13.6	12.6-14.8
White	1327	0.3	0.1-0.9
Coloured	3067	1.7	1.3-2.4
Indian	1102	0.3	0.1-1.2
Province			
Western Cape	2098	3.8	2.7-5.3
Eastern Cape	1984	9.0	7.2-11.2
Northern Cape	1227	5.9	4.5-7.8
Free State	960	12.6	10.5-15.1
KwaZulu-Natal	2464	15.8	13.4-18.6
North West	1156	11.3	9.1-14.0
Gauteng	2093	10.3	8.3-12.7
Mpumalanga	988	15.4	11.9-19.7
Limpopo	1252	8.8	6.5-11.9

which shares a border with South Africa. Between 2001 and 2009, in four of the five sub-Saharan countries with the highest HIV prevalence (Ethiopia, South Africa, Zambia and Zimbabwe), new HIV infections were reduced by more than 25% (UNAIDS, 2010a).

HIV is endemic in South Africa. A summary of the findings of the most recent comprehensive HIV prevalence survey in South Africa, undertaken during 2008, is shown in Table 1.9. According to this study, 5.6 million people (10.6%) of the population are HIV positive (Shisana *et al.*, 2009). HIV prevalence among antenatal attendees in 2010 ranged from 18.5% to 39.5% by province, with a national prevalence of 30.2% (National Department of Health, 2011). The national prevalence was highest, at 42.6%, in the group aged 30 to 34 years. South Africa has the largest ART roll-out program of any country, with almost one million of the 2.6 million people estimated to need treatment receiving antiretroviral treatment in 2009 (UNAIDS, 2010b).

As there is no cure nor vaccine for HIV at present, treatment involves reducing viral replication to very low levels. Initially this was achieved by using a single antiretroviral drug. As HIV evolved resistance to a single drug, treatment options then recommended switching to a second drug or using more than one drug in combination. This strategy of combining three or more antiretroviral drugs into a single treatment regime, termed “HAART” (highly-active antiretroviral

Table 1.10. Eligibility criteria for HAART initiation in South Africa

Criterion (Either)	Before 2009	2009-2011	From August 2011
CD4 cell count/μl	< 200	< 200, or < 350 for co-infection with tuberculosis and/or pregnant, and children	< 350
WHO Staging		IV	

Table 1.11. HAART regiments in South Africa

	Lamivudine (3TC)	Stavudine (d4T)	Efavirenz (EFV)	Nevirapine (NVP)	Tenofovir (TDF)	Emtricitabine (FTC)
Prior to April 2010						
Drug 1	•					
Drug 2		•				
Drug 3			•	•		
From April 2010 (new patients only)						
Drug 1					•	
Drug 2	•					•
Drug 3			•	•		

When more than one drug is shown, either may be prescribed, depending on the needs of the patient.

treatment), is now used as the standard treatment regime worldwide. Many different combinations of drugs have been developed and patients will be switched between the regimens as the virus evolves resistance to one or more of the drugs. As at the end of 2011, slightly more than half of the eligible individuals worldwide were receiving treatment, with 8 million people in low- and middle-income countries on treatment (World Health Organization, 2012b). Three-quarters of a million people in high-income countries were receiving treatment by the end of 2010 (UNAIDS, 2011). By the end of 2010, seventeen low- to middle-income countries (excluding South Africa) were providing treatment to at least 70% of those in need (UNAIDS, 2011). As at the end of 2010, the treatment need in low- to middle-income countries was highest in sub-Saharan Africa, at 73%, with 76% of those receiving treatment residing in this region (UNAIDS, 2011). However, the largest increase in the number of people receiving treatment, from 3.9 million in 2009 to 5.1 in 2010, occurred in sub-Saharan Africa (UNAIDS, 2011).

The eligibility criteria for HAART initiation in South Africa are detailed in Table 1.10. The government-approved HAART regimens available in South Africa are detailed in Table 1.11.

1.3 HBV and HIV Co-Infection in Africa

Individuals co-infected with HBV and HIV may suffer an increased disease burden. HIV-positivity negatively impacts on the progression of HBV (Chung, 2006). Although previously it was thought that HBV did not have an effect on HIV disease progression, a recent study reported that HBV-

positivity in HIV-positive patients has a significant impact on HIV outcomes by increasing the risk of AIDS events or mortality (Chun *et al.*, 2012). Co-infected individuals may experience HBV reactivation (Peters, 2007; Burnett *et al.*, 2005) and have an increased risk of mortality from liver-related disease (Thio, 2009). Additionally, they may suffer from a lower response to HBV vaccination and experience a higher chance of developing chronic disease (Kottilil *et al.*, 2005). Individuals infected with both HBV and HIV exhibit the following features: higher levels of serum HBV DNA, lower ALT levels, lower HBeAg seroconversion rates, increased risk of liver cirrhosis, HBV reactivation and decreased response to HBV vaccination (Farrell and Denstag, 2002). Accurate identification of HBV infection, particularly in HIV-infected individuals, is therefore important.

HBV and HIV are two of the most important blood-borne pathogens in sub-Saharan Africa, where both viruses are endemic. An extensive literature review of HBV and HIV co-infection in Africa was undertaken. The NCBI "PubMed" database[6] was searched using the following search phrase: **("hiv"[title] or "human immunodeficiency"[title]) and ("hbv"[title] or "hepatitis b"[title]) and ("country"[title] or "demonym"[title])**, where "country" and "demonym" were replaced with the name and demonym[7] of each country on the African continent. Articles, which were easily accessible online, were included in this review. HBV, HIV and co-infection prevalence data were extracted and are presented in Table 1.12. Each study is numbered for reference.

Most studies determined HBsAg-positivity by serological (ELISA) testing only. A wide variety of different test kits were used, from laboratory-based systems to point-of-care testing. Using the data from Table 1.12, the prevalence of HBsAg-positivity in HIV-positive patients in Africa ranges from 0.4% in a South African adult cohort, to 100% in a small Ugandan cohort of "hepatocarcinoma" patients. The median prevalence is 11.6% and the average prevalence is 14.3%$\pm$14.4%. Testing for the presence of HBV DNA by PCR and determination of HBV viral loads is not routinely done in African studies. Even in cases when patients were screened for both HBV and HIV infection, the prevalence of co-infection was not always reported. In studies in which both HIV-positive and HIV-negative patients were examined, HBsAg-positivity was higher in HIV-positive patients in all but three studies. This difference showed a borderline statistical significance ($p=0.056$).

Testing samples for the presence of occult HBV infection (Section 1.1.8.3), which is expensive and requires specialist skills and equipment, is not undertaken routinely in resource-limited settings. With the exception of South Africa, the extent of occult HBV infection in Africa is not

[6] *http://www.ncbi.nlm.nih.gov/pubmed/*

[7] A demonym is the name for a resident of a country; for example, a resident of Malawi is a *Malawian*.

known, as only a few studies outside South Africa investigated occult HBV infection: studies 19 and 21 each reported 10% occult HBV infection in HIV-positive patients, and study 47 reported possible occult HBV infection in two of 45 patients. Likewise, the determination of HBV genotype and subgenotype was not determined in most African studies. Studies that do report genotype, however, indicate that genotype E is present in West Africa (studies 18 and 19) and genotype A is present in East Africa (study 25).

Table 1.12. Prevalence of HBV, HIV and HBV/HIV co-infection in African countries (Page 1 of 2)

No.	Country	Cohort	Year	HBsAg+ Only		HIV+		HBsAg+ and HIV+		Reference
				Count	%	Count	%	Count	%	
01	Botswana	Adults	2005-2009			266/266	100.0	14/266	5.3	Patel *et al.* (2011)
02	Burkina Faso	Blood donors	2002	96/500	19.2	49/500	9.8	12/49	24.5	Kania *et al.* (2009)
03	Burkina Faso	Blood donors	2009	673/4520	14.9	100/4520	2.2	17/100	17.0	Nagalo *et al.* (2011)
04	Burkina Faso	Pregnant women	2003-2005	40/429	9.3	108/429	25.2	14/108	13.0	Simpore *et al.* (2004)
05	Burkina Faso	Pregnant women	2004-2005	33/336	9.8	207/336	61.6	24/207	11.6	Simpore *et al.* (2006)
06	Burkina Faso	Pregnant women	2005-2006	30/379	7.9	48/379	12.7	5/48	10.4	Ilboudo *et al.* (2007)
07	Burkina Faso	Pregnant women	2007-2009			115/115	100.0	14/115	12.2	Ilboudo *et al.* (2010)
08	Burkina Faso	Pregnant women	2009	8/276	2.9	138/276	50.0	8/138	5.8	Ouermi *et al.* (2009)
09	Cameroon	Pregnant women	2000-2003	51/650	7.8	301/650	46.3	28/301	9.3	Kfutwah *et al.* (2012)
10	Cameroon	Treatment naïve HIV+ patients	2001-2003			169/169	100.0	14/169	8.3	Laurent *et al.* (2010)
11	Cameroon	Treatment naïve HIV+ patients	2005-2010			690/690	100.0	87/690	12.6	Zoufaly *et al.* (2012)
12	Cameroon	Blood donors	2008	564/4644	12.1	206/4644	4.4	36/206	17.5	Ymele *et al.* (2012)
13	Cameroon	Blood donors	2011-2012	55/543	10.1	22/543	4.1	6/22	27.3	Noubiap *et al.* (2013)
14	Ethiopia	Blood donors	2010	126/6063	2.1	129/6063	2.1	2/129	1.6	Yami *et al.* (2011)
15	Ethiopia	Blood donors	2003-2007	298/6361	4.7	239/6361	3.8	17/239	7.1	Tessema *et al.* (2010)
16	Ethiopia	Street dwellers	2004	33/302	10.9	27/302	8.9	9/27	33.3	Moges *et al.* (2006)
17	Ethiopia	Pregnant women	2008	10/165	6.1	3/165	1.8	1/3	33.3	Ramos *et al.* (2011)
18	Gambia	HIV+ patients				570/570	100.0	70/570	12.3	Stewart *et al.* (2011)
19	Ghana	HIV+ patients	2007			838/838	100.0	140/838	16.7[a]	Chadwick *et al.* (2013)
20	Ghana	HIV+ patients	2007			138/138	100.0	18/138	13.0	Sagoe *et al.* (2012)
21	Ivory Coast	Treatment naïve HIV+ patients	2006			495/495	100.0	63/495	12.7 [b]	N'Dri-Yoman *et al.* (2010)
22	Ivory Coast	Women	1995-1996	43/428	10.0	222/428	51.9	23/222	10.4	Combe *et al.* (2001)
23	Ivory Coast	HIV+ children	2000-2003			280/280	100.0	34/280	12.1	Rouet *et al.* (2009)
24	Ivory Coast	Pregnant women		85/997	8.5	499/499	100.0	45/499	9.0	Rouet *et al.* (2004)
25	Kenya	HIV+ adults	2006-2008			379/379	100.0	27/379	7.1	Kim *et al.* (2011)
26	Kenya	HIV+ patients				378/378	100.0	23/378	6.1	Harania *et al.* (2008)

[a] 9.9% occult HBV infection
[b] 10% occult HBV infection

Table 1.12. Prevalence of HBV, HIV and HBV/HIV co-infection in African countries (Page 2 of 2)

No.	Country	Cohort	Year	HBsAg+ Only		HIV+		HBsAg+ and HIV+		Reference
				Count	%	Count	%	Count	%	
27	Malawi	Male agricultural workers	1998	72/469	15.4	189/469	40.3	32/189	16.9	Sutcliffe *et al.* (2002)
28	Malawi	Adult inpatients	2004	34/194	17.5	172/226	76.1	31/172	18.0	Nyirenda *et al.* (2008)
29	Niger	Pregnant women	2008	80/495	16.2	10/495	2.0	3/10	30.0	Mamadou *et al.* (2012)
30	Nigeria	HIV+ patients and blood donors	2006	22/362	6.1	102/362	28.2	29/102	28.4	Balogun *et al.* (2012)
31	Nigeria	HIV+ patients	2009			273/273	100.0	18/273	6.6	Adekunle *et al.* (2011)
32	Nigeria	HIV+ patients and HIV- blood donors	1999-2002	152/665	22.9	490/665	73.7	127/490	25.9	Uneke *et al.* (2005)
33	Nigeria	HIV+ patients	2004-2006			1564/1564	100.0	262/1564	16.8	Idoko *et al.* (2009)
34	Nigeria	HIV+ patients				260/260	100.0	30/260	11.5	Adewole *et al.* (2009)
35	Rwanda	HIV+ children	2010			88/88	100.0	6/88	6.8	Mutwa *et al.* (2012)
36	Rwanda	Pregnant women	2001-2004			82/82	100.0	2/82	2.4	Pirillo *et al.* (2007)
37	Senegal	HIV+ patients	1998-2002			363/363	100.0	61/363	16.8	Diop-Ndiaye *et al.* (2008)
38	South Africa	Hospital inpatients	1998-2001	72/295	24.4	167/295	56.6	27/167	16.2[c]	Mphahlele *et al.* (2006)
39	South Africa	HIV+ and HIV- pregnant women	1999-2001	41/710	5.8	710/710	100.0	44/710	6.2	Burnett *et al.* (2007)
40	South Africa	HIV+ miners	2002-2006			537/537	100.0	106/537	19.7	Hoffmann *et al.* (2008)
41	South Africa	HIV+ urban/rural patients	2004-2007			192/192	100.0	44/192	22.9[d]	Lukhwareni *et al.* (2009)
42	South Africa	HIV+ adults	2008			242/242	100.0	1/242	0.4[e]	Barth *et al.* (2011)
43	South Africa	HIV+ rural patients	2008-2009			1765/1765	100.0	126/1765	7.1	Boyles and Cohen (2011)
44	South Africa	HIV+ urban outpatients				502/502	100.0	24/502	4.8[f]	Firnhaber *et al.* (2008, 2009, 2012)
45	South Africa	HIV+ military personnel				1771/1771	100.0	106/1771	6.0	Matthews *et al.* (2011)
46	Tanzania	HIV+ children	2006			167/167	100.0	2/167	1.2	Telatela *et al.* (2007)
47	Tanzania	HIV+ treatment naïve patients	2006			260/260	100.0	45/260	17.3[g]	Nagu *et al.* (2008)
48	Tanzania	Blood donors	2004-2005	141/1598	8.8	59/1557	3.8	5/59	8.5	Matee *et al.* (2006)
49	Tanzania	HIV+ adults	2004-2011			17539/17539	100.0	1079/17539	6.2	Hawkins *et al.* (2012)
50	Tanzania	HIV+ children	2005			104/104	100.0	9/104	8.7	Pippi *et al.* (2008)
51	Uganda	Pregnant women	2001-2004			164/164	100.0	8/164	4.9	Pirillo *et al.* (2007)
52	Uganda	Primary hepatocarcinoma patients	2008	13/15	86.7	3/15	20.0	3/3	100.0	Ocama *et al.* (2011)
53	Zambia	HIV+ adults	2008			323/323	100.0	32/323	9.9	Kapembwa *et al.* (2011)
54	Zimbabwe	Pregnant women	2003-2005	13/341	3.8	83/341	24.3	4/83	4.8	Mavenyengwa *et al.* (2010)

[c] 22% (31/140) occult HBV infection
[d] 44% (34/78) occult HBV infection
[e] 10% (6/60) occult HBV infection in anti-HBc-positive samples
[f] 88.4% (38/43) occult HBV infection in anti-HBc-positive alone samples
[g] 4.4% (2/45) possible occult HBV infection

1.4 Aims and Objectives: Clinical and Virological

1. Establish a rural cohort in Mpumalanga in order to characterize hepatitis B virus (HBV) in human immunodeficiency virus (HIV) co-infected individuals. No HBV prevalence data are available for this province, which has an HIV prevalence of 15.4%.
2. Examine occult HBV infection in HBV/HIV co-infected individuals, using the *Taormina* definition.

Chapter 2

Bioinformatics Introduction and Literature Review

2.1 History and Context

Bioinformatics is a field of study concerned with computational analysis and storage of biological data. The field is broad, ranging from the study of DNA and proteins, to structural biology, drug design and comparative genomics. An early pioneer in the field was the American physical chemist Dr Mary Dayhoff [1925–1983], who used mathematical and computational approaches to perform biochemical, medical and macromolecular evolution analyses, including protein sequence alignment and comparison (Hunt and Dayhoff, 1983).

One of the earliest projects, which aimed to manage the storage and retrieval of biological data, was the GenBank sequence database (Benson *et al.*, 2012), established thirty years ago. The growth in the number of nucleotides stored in GenBank has followed Moore's Law, by doubling approximately every 18 months (GenBank, 2013). The latest GenBank release (version 194.0 of 20 February 2013) contains 150,141,354,858 bases from 162,886,727 reported sequences (GenBank, 2013). GenBank, the DNA Data Bank of Japan and the European Nucleotide Archive (ENA) in the United Kingdom form the International Nucleotide Sequence Database Collaboration. Data are synchronized between all three databases daily.

In addition to these databases, many others exist, including genome databases, protein sequence, structure and interaction databases, microarray databases and meta-databases. A list of biological databases on Wikipedia[1] includes almost 100 entries.

[1] *http://en.wikipedia.org/wiki/List_of_biological_databases*, accessed 02 March 2013.

2.2 Free Software

In addition to biological databases, a large variety of biological analysis software is available. As with software in any field, the licensing terms and commercial costs of these packages vary widely. Packages, which may be free of cost, may not necessarily be open-source, for example.

The Free Software Foundation[2] (FSF) defines free software as software which "respects the users' freedom" in the sense that "users have the freedom to run, copy, distribute, study, change and improve the software". As such, "free" is "a matter of liberty, not price". Free software, therefore, does not necessarily have to be made available at no cost or be a non-commercial project.

The term "open-source" is often used when referring to "free" software. However, the two terms are not synonymous, although there is some overlap. Open-source software may, or may not, be free software, depending on the restrictions placed on users by the software. If the user is not free to distribute, change and improve the software, even if it is open-source, then it cannot be considered free software. Most software for which a license is purchased, is not free or open-source. The user does not have the freedom to distribute the software, or to use it on any computer he or she chooses.

2.3 Bioinformatics Software, Services and Resources

A glossary of technical terms is provided in Section 3.3.1. Software available for bioinformatic analysis ranges from complex, integrated analysis suites, which run on all major operating system platforms, to highly-specialized command-line tools, which require compilation from source-code on selected operating systems only, to online services available on the Internet. Analysis using several, separate stand-alone tools can be automated via scientific workflow tools. Large-scale analyses can be undertaken via cloud services offering virtual servers, or via large computing clusters, or server farms. Custom GNU/Linux distributions with specialized bioinformatics software pre-installed are available for download. Analysis tools and services are available on the Internet as integrated web-services. Modules and code libraries to facilitate bioinformatics software development are available for many major programming languages (Open Bioinformatics Foundation, 2012). The Open Bioinformatics Foundation (Open Bioinformatics Foundation, 2012) supports a wide range of bioinformatics projects. These include bioinformatics projects for the open-source

[2] *http://www.fsf.org* and *http://www.gnu.org/philosophy/free-sw.html*

programming languages Java, Perl, Python and Ruby, and projects such as BioSQL and EMBOSS. BioPython[3] is a suite of Python libraries and applications for "biological computation". The project is mature and actively developed, with approximately two new versions released annually.

The range of bioinformatics programs available is illustrated by the number of entries in lists on Wikipedia and other sites: a page listing open-source bioinformatics software contains 45 entries[4], a list of sequence alignment software contains almost 200 entries in various categories[5], and 29 sequence alignment visualization packages are available[6]. A directory of links provided by another site[7] lists many hundreds of programs, databases and resources in various categories. A list of phylogeny programs includes 392 packages and 54 free web servers[8]. An attempt has been made to establish a web-based database, called *MetaBasis*, which lists bioinformatics software tools and databases (Atlamazoglou *et al.*, 2006), but this does not appear to be maintained. The URL provided in the paper is no longer available, but a version of the resource, available at a different address[9], indicates that 3229 published bioinformatics tools and databases are indexed. Furthermore, lists of online websites, which themselves provide lists of resources, are also available (Gilbert, 2004).

Maintaining software programs, databases and websites requires an ongoing commitment of time, effort and resources. Such maintenance may be undertaken by the researcher who developed the resource. Over time, this person may no longer be able to maintain the resource because of other commitments, or may leave the institution. One fifth of URLs published in MEDLINE disappear within 10 years (Wren, 2004). Tools may last longer than their authors anticipated, as they may address a very specific need, but attract no further development (Gilbert, 2004). Attempts to archive all bioinformatics resources have not been entirely successful (Gilbert, 2004). This may be due, in part, to the number of archiving sites available, and the disparate, *ad hoc* nature of a new and developing field. At the 2010 International Conference on Bioinformatics, several journals agreed to comply with the Minimum Information about Bioinformatics investigation (MIABi) when publishing conference papers (Tan *et al.*, 2010). This initiative aims to "ensure data and software persistence and perpetuity, database and resource re-instantiability and reproducibility of results, author and contributor identity disambiguation" (Tan *et al.*, 2010).

[3] *http://www.biopython.org*
[4] *http://en.wikipedia.org/wiki/List_of_open_source_bioinformatics_software*, accessed 02 March 2013,
[5] *http://en.wikipedia.org/wiki/List_of_sequence_alignment_software*, accessed 02 March 2013.
[6] *http://en.wikipedia.org/wiki/List_of_alignment_visualization_software*, accessed 02 March 2013.
[7] *http://bioinformatics.ca/links_directory/*, accessed 02 March 2013
[8] *http://evolution.gs.washington.edu/phylip/software.html*, accessed 02 March 2013.
[9] *http://bioserver-1.bioacademy.gr/Metabasis/*, accessed 02 March 2013.

Available programs and resources may be generic, such as integrated bioinformatics suites, which can process sequence data from any source, but may not offer specialized analyses, to specific programs, such as the specialized, command-line components of the EMBOSS suite (Rice *et al.*, 2000). Resources may be aimed at general users, or at technical researchers with specialist skills. Despite the range of resources available, much time is needed to find and download programs of interest (Gilbert, 2004). After successfully installing a program on the appropriate operating system, it must be carefully tested and assessed. Large, complex programs can be difficult to use without training or assistance. In many instances, biologists may prefer to continue using a program with which they are comfortable and familiar, rather than committing to learning a new one. However, analysis of large datasets often requires the use of specialist, command-line tools and *ad hoc* scripting, neither of which a biologist may be familiar with (Kumar and Dudley, 2007).

Analysis needs of researchers are so diverse and so specific, that existing solutions may not always be suitable. Whilst free, open-source solutions allow for modification of the program, most biologists are not computer programmers and are unable to implement any such changes themselves. Solutions to this would include embedding bioinformaticists into research laboratories, or possibly producing even more flexible and generic programs. However, overly-generic programs may be more complex and unintuitive for the end-user to operate, as they attempt to fit all possible needs. As such, solutions developed by researchers with an intimate knowledge of the data being analyzed have the advantage of being tailored to nuances and subtleties, which general solutions may fail to address. Solutions, which have been specifically targeted to a known, understood and quantifiable problem, may be better. Accommodating the different backgrounds and skills, which biologists using bioinformatic tools have, is a challenge (Schneider *et al.*, 2010). The technical aspects of biological data analysis may be challenging for biologists, who may not be sufficiently computer literate, and thus resources, such as the book by Gibas and Jambeck (2001), are available to address this need.

2.3.1 HBV Resources

More specifically, several HBV-related resources are available on the Internet. These include genotyping tools and resistance mutation databases. Table 2.1 lists some of these resources with their respective URLs.

Identifying the HBV (sub)genotype of a sample is important, but may often be difficult be-

Table 2.1. A selection of HBV Resources Available on the Internet

Name	URL
HBV STAR	*http://www.vgb.ucl.ac.uk/starn.shtml*
Oxford HBV Subtyping Tool	*http://www.bioafrica.net/rega-genotype/html/subtypinghbv.html*
HBV Blast Search	*http://www.bioafrica.net/blast/hbvblast.html*
HBVseq	*http://hivdb.stanford.edu/HBV/HBVseq/development/HBVseq.html*
HBVdb	*http://hbvdb.ibcp.fr/*
SeqHepB	*http://www.seqhepb.com/*
HBVRegDB	*http://lancelot.otago.ac.nz/HBVRegDB/*
Hepatitis Virus Database	*http://s2as02.genes.nig.ac.jp/*
Hepatitis Virus Database	*http://www.ibibiobase.com/projects/hepatitis/index.htm*
RegaDB	*http://regaweb.med.kuleuven.be/software/regadb*

cause of differences in genome length, recombination and general sequence variability. "HBV STAR", by Myers *et al.* (2006), is an adaptation of a tool previously used to genotype HIV samples (Myers *et al.*, 2005). The tool uses a position-specific scoring matrix (PSSM) to genotype HBV samples from whole- and sub-genomic fragments, without generating phylogenetic trees. Recombinant genotypes can also be detected. PSSM values, from genotype-specific alignments, are transformed into a Z score, which is used to predict the genotype of the query sequence.

The "Oxford HBV Subtyping Tool" uses phylogenetic methods to identify the subtype of a sequence (de Oliveira *et al.*, 2005; Alcantara *et al.*, 2009). A BLAST search (Altschul *et al.*, 1990) is performed to identify a specific region of a reference HBV sequence, after which an alignment against selected genotypes is undertaken. A phylogenetic tree is then constructed and recombination analysis performed. The phylogenetic results are interpreted by scripts written in the JAVA and PHP languages, and the results are plotted using the R statistical programming language (R Core Team, 2012). The "HBV Blast Search" tool performs a BLAST search on the query sequences, using a variety of reference datasets, such as the complete GenBank HBV database, or the HBV genotype reference database.

The "HBVseq" site determines the HBV genotype of submitted samples and compares them to consensus reference sequences. Any differences noted are used to query a local HBV drug resistance database. Results returned include the prevalence of each mutation by genotype. The drug resistance database was created by annotating publicly available HBV sequences.

"HBVdb" (Hayer *et al.*, 2013) is a specialized database containing computer-annotated HBV sequences, providing genotyping, drug resistance profiling and bioinformatic analyses. The database consists of almost 40,000 sequences, approximately 10% of which are full genomes. All HBV sequences from the ENA were extracted and annotated, using a set of 16 manually-annotated, non-recombinant, full-length reference sequences, covering genotypes A to H.

"SeqHepB" is a sequence analysis program and relational database system for chronic HBV, which is targeted at clinicians, laboratories, clinical trials and drug development. The service operates primarily as a commercial entity requiring registration (Yuen *et al.*, 2007). The company has registered several patents relating to the identification of HBV drug resistance mutants in diagnostic, therapeutic and clinical testing. The service determines HBV genotype and key mutations of clinical importance associated with antiviral resistance. Three-dimensional models of sections of HBV reverse transcriptase are included.

An online database and comparative genomic analysis tool called "HBVRegDB" provides comparison, detection and visualization of regulatory elements in HBV sequence data (Panjaworayan *et al.*, 2007). The curated database contains genomic sequence data, annotations, alignments and information about conserved regions.

The "Hepatitis Virus Database" provides various services such as genotyping, sequence alignment and map viewing, for all hepatitis viruses. The site, which requires the Java Runtime Environment (JRE) in the client web-browser, was last updated in September 2010. Another site also called the "Hepatitis Virus Database" provides access to reference strains of all hepatitis viruses (A to "G", excluding "F") and an overview of each virus. "RegaDB" (Libin *et al.*, 2007) is a viral data and analysis management environment. It has been designed to store clinical data related to HIV and HCV treatment. It is aimed at offering tools for sequence analysis of these viruses, and to facilitate collaboration between researchers at different institutions. However, use of this system requires an investment of time and infrastructure.

Often, solutions are developed by biologists, who are not necessarily programmers or bioinformaticists, to address a specific, localized need at the time. There may not be an initial commitment to make the program or code available, or to maintain it over an extended period. Software may be released as a courtesy or service to the community, without the expectation at the time that it will be widely adopted, or maintained over an extended period of time. Whilst many of these tools are a valuable resource for the research community, not all of them provide a command-line and/or programmatic interface for their reuse, or to facilitate easy integration into other systems. However, these tools were not designed to provide such functionality. The value of some of these tools could be leveraged by automatically extracting content from their web-sites and processing this content independently. As bioinformatics is a new and evolving field, there is still space for the development of additional analysis tools. This development should be kept up to date by incorporating new results generated from the wet laboratory.

2.4 Aims and Objectives: Bioinformatics

1. Establish a custom database to store clinical and molecular data from the cohort.
2. Develop new bioinformatic tools to facilitate the clinical analyses and molecular characterization of HBV infection.

Chapter 3

Establishment of Cohort and Database

3.1 Introduction and Context

Whilst HIV surveillance is undertaken annually in South Africa, no corresponding HBV surveillance has ever been performed. As such, HBV prevalence in South Africa is determined from localized studies in urban and rural areas (Section 1.3 and Table 1.12). As these are not formalized national studies, they represent only a small sample population, and their findings cannot easily be compared or extrapolated to the entire country. Therefore, additional studies in other regions of the country can provide greater insight into the prevalence and characteristics of HBV infection in South Africa. In the province of Mpumalanga, large parts of which are rural, the HIV infection rate is particularly high, at 15.4% (Shisana *et al.*, 2009). This high prevalence of HIV and the high endemicity of HBV in South Africa, required that a study examining HBV/HIV co-infection, particularly in a rural setting, be undertaken. Additionally, as HIV-positivity has been shown to be a risk factor for occult HBV infection (OBI), it was important to determine the prevalence OBI within Mpumalanga province. Although HBV subgenotypes A1 and D3 are known to predominate in South Africa (Kimbi *et al.*, 2004), there are no data for Mpumalanga province, which borders the countries of Swaziland and Mozambique. Moreover, travel between these three countries occurs frequently, and HBV has not been characterized in either Swaziland or Mozambique. Therefore, a study in this province provided us with the opportunity to examine the distribution of HBV genotypes circulating within this region.

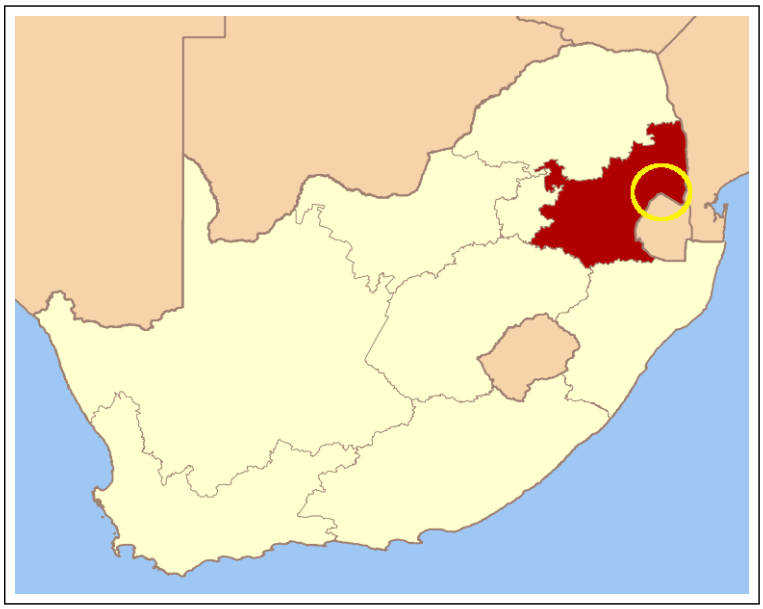


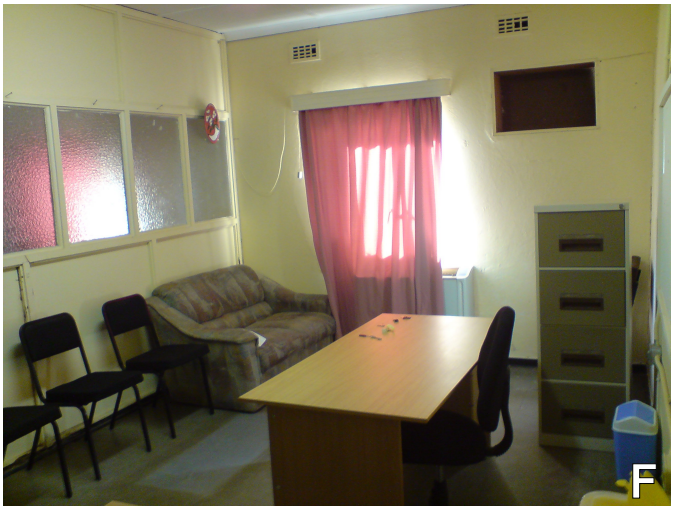
Figure 3.1. *Location of Shongwe Hospital (the centre of the yellow circle) within Mpumalanga Province (shaded red). South Africa is shown in cream.*

3.2 Shongwe Hospital

Study Site Shongwe Hospital, in Mpumalanga province, South Africa, was identified as a suitable rural cohort study site (Figure 3.1). Mpumalanga, one of nine South African provinces, accounts for 6.4% of the land area of the country (Bornman and Whittall, 1997). The town of Malalane (formerly Malelane), on the southern border of the Kruger National Park, lies 30 km north of the hospital. The border town of Jeppe's Reef, providing access to the country of Swaziland, is located 10 km south of the hospital. The South African border with Mozambique is 60 km east of Malalane, at the border town of Komatiespoort. Principal economic activities in the Malalane region include tourism, sugar production and agriculture (Bornman and Whittall, 1997). Inhabitants of the region include the Swazi, Shangaan and Pedi tribes (Bornman and Whittall, 1997). Shongwe Hospital is surrounded by informal settlements, with schools, informal traders and shops lining the main road through the area. A Right-to-Care- and USAID-funded (Voluntary Counselling and Training, VCT) clinic is located at the hospital. Photographs of Shongwe Hospital and surrounding areas are presented in the Colour Plate I (overleaf).

Ethics and Informed Consent Ethical clearance for this study was obtained from the Mpumalanga Department of Health and from the Human Ethics Research Committee of the University of the Witwatersrand, as amended. Documentation is included for reference in Appendix B. The Informed Consent document was approved by the Human Ethics Research Committee of the

Colour Plate I



[A] Fields near Shongwe Hospital.
[B] Rural dwellings near Shongwe Hospital.
[C] The main entrance to Shongwe Hospital.
[D] Buildings surrounding the VCT clinic.
[E] The main entrance to the VCT clinic.
[F] Nurse's office and interview room.

University of the Witwatersrand and is included in Appendix B. The document was translated from English into isiZulu, Sesotho, Siswati and Xitsonga.

Research Nurse and Documentation A retired nurse, Sister Agatha Nkosi, who had previously worked at Shongwe Hospital, attended a Good Clinical Practice (GCP) training course before commencing recruitment of study participants. A locum nurse, Sister Rosalina Candlovu, who also attended GCP, continued with enrollments when Sr Agatha was unavailable. The Clinical Report Form (CRF) and other documentation are provided in Appendix B.

Participant Recruitment Based on results from South African urban and rural studies, it was anticipated that 10% of HIV-positive participants would also be positive for HBV (Burnett *et al.*, 2005; Firnhaber *et al.*, 2008). Thus, a sample size of 30 co-infected individuals was considered to be statistically sufficient for our study (Hogg and Tanis, 2005), and therefore the objective was to recruit 300 participants. Ultimately, 298 participants, all of whom were treatment-naïve, but were about to start anti-retroviral therapy, were enrolled in the study. After participants signed the Informed Consent, the research nurse completed the CRF and drew blood samples. Enrollment of participants continued for a period of four months. A list of participants, who were positive for HBV (Chapter 4 and Appendix C), was provided to the hospital's chief medical officer, and to the research nurse, who contacted these participants to notify them and to ask them to return for follow-up visits at 3, 6, 12 and 18 months after commencement of ART. Not all eligible participants returned for follow-up blood-draws at all time-points.

Cohort Demographics The cohort of 298 adults consisted of 114 men (38%) and 184 women (62%). The median age, CD4 cell count and BMI were 34 years, 147 cells mm^{-3} and 22 kg m^{-2}, respectively. Men were older than women and had lower CD4 counts. Box-and-whisker plots of age, BMI, ALT and CD4 cell count for men and women in the cohort are provided in Figure 3.2. Medians are reported here, as these data are skewed. Sixty-two percent of the participants reported never using a condom, with 29% reporting occasional usage. One third of the cohort had at most a grade 5 education. Almost three-quarters had at most a grade 11 education, with the majority of the remainder having a grade 12 education. Median individual income was R950 (US$100) per month, with a median household income of R1400 (US$150) per month.

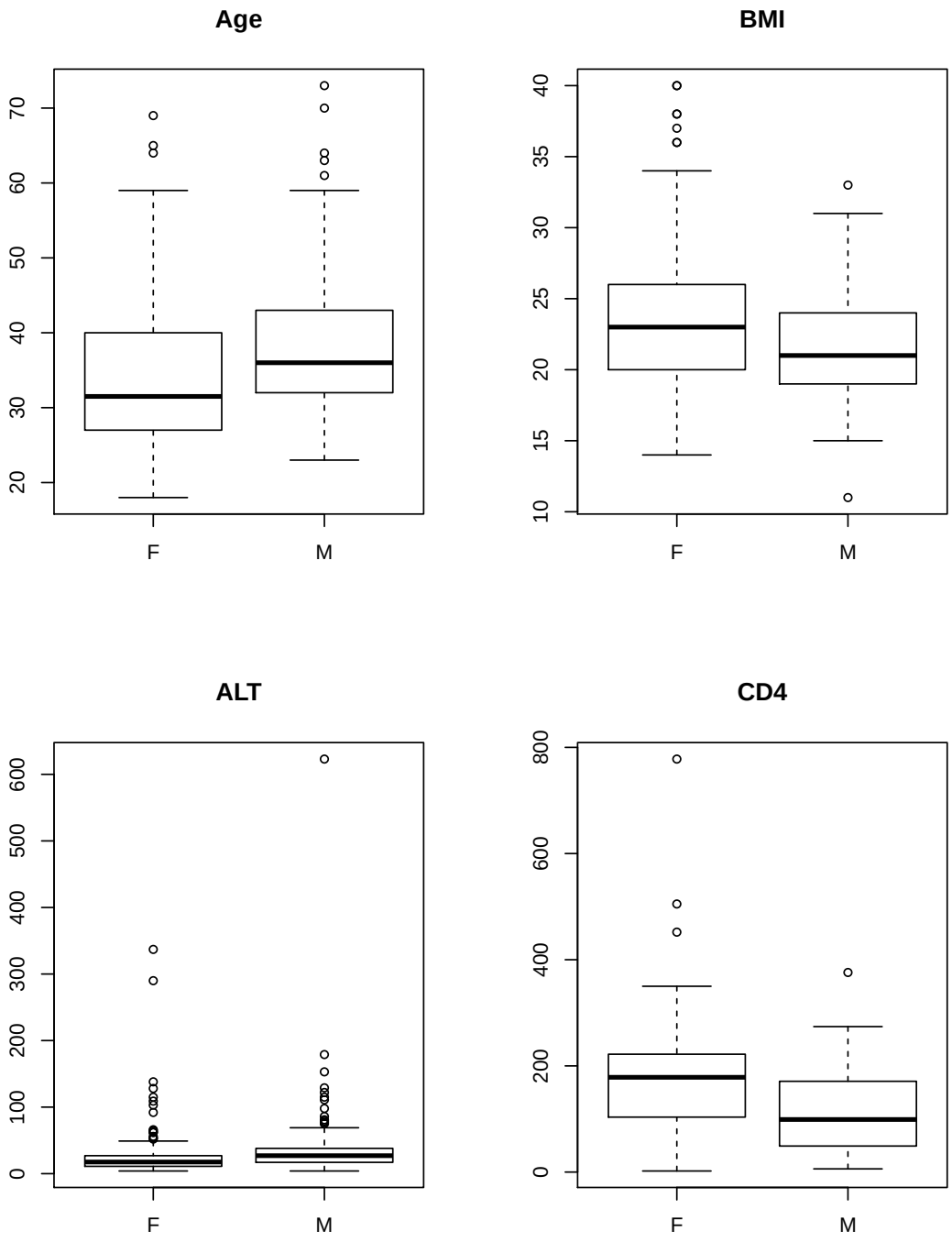


Figure 3.2. *Box-and-whisker plots of the age, BMI, ALT and CD4 cell counts of the cohort.*

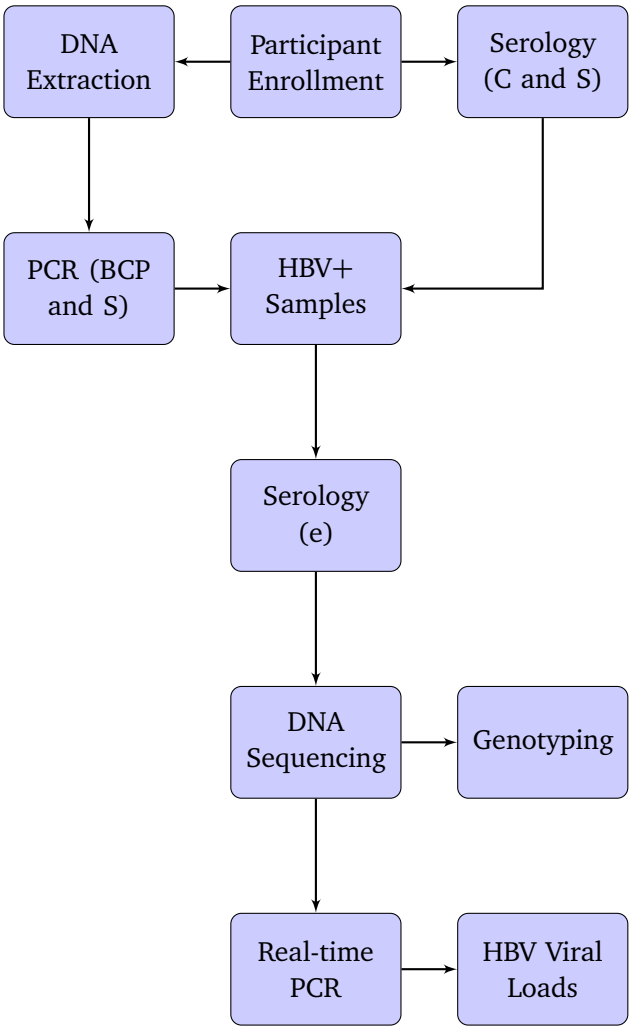


Figure 3.3. *Flowchart of wet laboratory procedures undertaken on blood samples.*

Data Collection Participant data from the CRF were entered into a database (Section 3.3). Additional clinical data were sourced from the National Health Laboratory Services (NHLS) and the *TherapyEdge-HIV (TE)*^{TM} patient information system at Shongwe Hospital. Obtaining a full set of clinical data for all participants was not possible. In some instances, data were simply not available (missing or never collected). In other cases, duplicate hospital numbers meant that it was not possible to find the data for the participants in question. More complete data were available for the enrollment (baseline) time-point than for the various follow-up time-points.

Wet Laboratory A summary flowchart of the wet laboratory procedures undertaken is shown in Figure 3.3. Further details of laboratory materials and methods, including DNA extraction, ELISA serological tests, PCR and real-time PCR protocols and primers, are provided in the two published papers (Chapter 4 and Appendix C).

3.3 Database Establishment and Usage

3.3.1 Glossary

The following technical terms are used in this section and in Chapter 5.

Adaptor A software routine, which facilitates communication between two different systems.

Apache The Apache **HTTP** web server is one of several web server software programs, which is responsible for delivering web content. A web server does this by responding to incoming requests for web pages from client computers, processing and preparing the output pages, and sending these pages to the requesting client computer. *http://httpd.apache.org/*

CGI A common Gateway Interface is a method in which web pages are generated on the server by another program, such as a programming language.

Class A programming construct, which represents a generalized concept or object. A class may contain variables and/or procedures (methods). Instances of the class, which are created in a program, store data and/or execute the instructions in the methods.

Django A free and open source web application framework, written in Python. *https://www.djangoproject.com/*

Framework A software framework is a set of reusable libraries or routines. A web application framework is such a framework designed to assist the development of dynamic websites.

GNU/Linux Any of a large number of free and open-source, Unix-like operating systems, built on the kernel written by Linus Torvalds. As all variants of "Linux" use GNU ("GNU's Not Unix") libraries and routines, it is considered preferable to refer to "GNU/Linux", rather than "Linux".

HTTP Hypertext Transfer Protocol is the name of the common transfer protocol, which is used to deliver web page on the Internet (World Wide Web).

Library A collection of reusable, modular, programming routines, which may be accessed used by other, independent, programs. Synonymous with *Module*.

Module See *Library*.

PGSQL An abbreviation for "PostgreSQL".

PHP The "PHP: Hypertext Preprocessor" is an open-source scripting language, which runs on a web server. PHP code, embedded into web pages, is processed by PHP on the web server, which then creates and generates the required web page. *http://php.net/*

PostgreSQL The PostgreSQL object-relational database management system is a "database server", which is responsible for storing and accessing data from a database. *http://www.postgresql.org/*

Psycopg A **PGSQL**-Python adaptor. *http://initd.org/psycopg/*

Python A free and open-source scripted, high-level, general programming language. *http://www.python.org/*

Query A request, written in **SQL**, to a database server to retrieve records from the database, which match the query terms.

RPostgreSQL A **PGSQL**-R adaptor. *https://code.google.com/p/rpostgresql/*

R A free and open-source scripted programming language, designed for statistical calculations and the production of graphics. *http://www.r-project.org/*

SELECT An **SQL** instruction to retrieve records from a database.

SQL Structured Query Language is the programming language used to manage and retrieve data stored in a database server.

URL A Uniform Resource Locator is a "web address" or "web-site address", which typically starts with the sequence, *http://*.

Ubuntu Ubuntu is a **GNU/Linux** operating system, which is available in both desktop and server versions. *http://www.ubuntu.com/*

View A database View is a predefined database query. It is often easier and quicker to **SELECT** a View, rather than specifying a whole query.

Wrapper A software routine, which takes responsibility for executing another routine independently.

3.3.2 Structure

The PostgreSQL (PGSQL) database management system was installed onto a remote dedicated Ubuntu GNU/Linux server. The Apache HTTP Server was also installed on the server to provide access to the database via an Internet web-browser. The Python, R and PHP programming languages were installed, as well as the *Psycopg* and *RPostgreSQL* adaptors. Creation of the database and tables, importing of data, querying and database management were typically undertaken from the PGSQL command shell (*psql*) directly on the server. Some database tasks were undertaken via the *phpPgAdmin*[1] web-based front-end.

PGSQL was selected, in part, because custom functions, written in the Python Procedural Language (*PL/PythonU*), can be stored in the database. The database was constructed to store clinical, demographic and molecular data, at multiple time-points, for participants from multiple different studies. To ensure data accuracy and consistency, several house-keeping tables were created. These typically stored codes and descriptions, which were used with foreign key lookups and constraints. This meant that only valid codes from house-keeping tables were permitted for most of the data fields. Data were normalized to provide as much flexibility as possible for current and future use. Data were imported into the database from plain text files, which were prepared directly by hand, or exported from a spreadsheet.

A simplified entity-relationship (ER) model of the database is shown in Figure 3.4. This model excludes all house-keeping and some other tables for clarity. Examples of the content stored in some of these tables are illustrated in Section 3.3.3.4.

Participant Table The "participant" table stores invariant (static) participant data, including a participant code as the primary unique key, height and date of birth. A study code links the participant to a study, details of which are stored in the *study* table (not shown).

[1] *http://phppgadmin.sourceforge.net/doku.php*

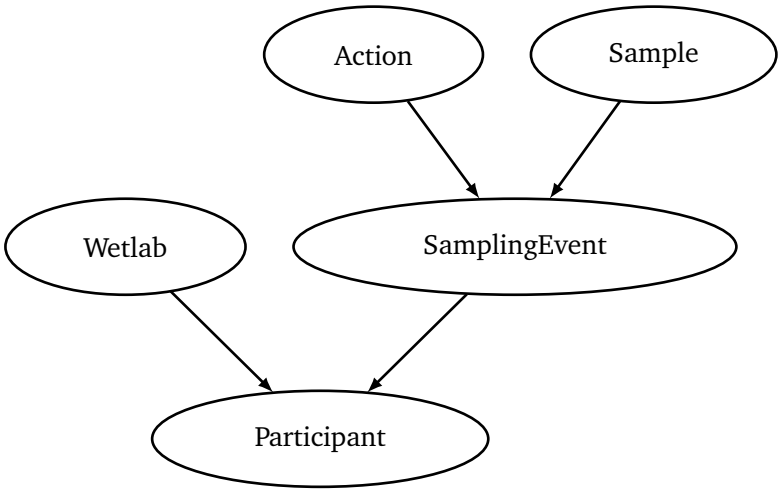


Figure 3.4. *A simplified ER model of the database, showing the most important foreign key relationships and tables.*

Sampling Event Table As participants may present themselves for one or more visits, a "samplingevent" table stores data relevant to each visit (variant data), with a sampling event code uniquely identifying each record. An entry is created in this table for each visit for which the participant was present. The sampling event code was created by appending a single upper-case letter, representing the visit number, to the participant code. For example, when participant "SHH001" enrolls in the study, a sampling event code of "SHH001A" is used. If that participant returns at a later time for a follow-up visit, a sampling event code of "SHH001B" would be used. In this study, this table held 431 records.

Action Table An "action" table stored actions associated with a sampling event, and the date on which these occurred. Examples of actions are sample collection, delivery and receipt. Certain wet laboratory procedures were also added as actions, such as DNA extraction and serological tests. Data in this table were used to check, for example, dates on which DNA was extracted from samples or whether specific serology tests had been undertaken on samples of interest. In this study, this table held 4100 records.

Wetlab Table A "wetlab" table stored wet laboratory results and clinical data for each participant. This table was not linked to a sampling event, as certain data points, such as clinical data, may have been generated at other times. Instead, a "time-point" value was assigned to each result. Fields in this table included the test performed (such as "ALT" or "HBsAg"), the date and the result or outcome of the test. These data were normalized, in that each record in the table stored one test and one result for one time-point for one participant only, rather than a single

record holding results for all possible tests for one participant. This was by design, as the tests and results available for each participant were not the same, and some tests may be repeated. Moreover, this arrangement allowed additional data, such as the date on which the test was performed and the name of the technician or person performing the test, to be stored. Storing these additional data points for each test in one row, for all tests, is not practical. However, this normalized approach meant that extracting a full set of test results for each participant required a complex set of multiple nested queries. This was achieved via a Python script, which processed the data and prepared a consolidated table (Section 3.3.3). In this study, this table held 6545 records.

Sample A "sample" table stored details such as the location and remaining volume of plasma and serum samples for each samplingevent. In this study, this table held 431 records.

3.3.3 Data Mining and Processing

3.3.3.1 Query Interface

An SQL query web interface was created to view extracted data from the remote database. This interface required an SQL query starting with the "SELECT" keyword as input and returned an HTML table of the output results. The interface was used for routine database maintenance and to extract results, which were required for analyses. Database results could easily be made available directly to other users and collaborators by providing them with the URL of the interface and a query string. In cases where the queries were complex, this query string selected from a predefined database view.

For security reasons, the script connected to the database as a user with limited ("SELECT" only) privileges. Additionally, to prevent the execution of SQL statements used for routine database maintenance (such as "UPDATE", "INSERT" or "DROP"), the script only processes SQL statements starting with the "SELECT" keyword. The output from a sample query is shown in Figure 3.5.

The "actiontype" and "person" fields hold codes, which are stored, with descriptions, in separate house-keeping tables. The action code "EN" means "Enrollment" and "DE" means "Delivery", for example. In this example, the query does extract the description for each code, as stored in the house-keeping table.

Results

```
select samplingevent, actiontype, actiondate, actionnotes, actionperson from
action where samplingevent = 'SHH001A' order by actiondate;
```

samplingevent	actiontype	actiondate	actionnotes	actionperson
SHH001A	EN	2009-07-22	None	R011
SHH001A	DE	2009-07-28	None	R011
SHH001A	AR	2009-07-29	None	R004
SHH001A	CE	2009-07-31	None	R001
SHH001A	EX	2009-08-05	None	R001
SHH001A	S1	2009-10-14	7th Floor Labs	R001
SHH001A	S2	2010-02-16	HVDRP Labs	R007
SHH001A	S3	2010-02-17	HVDRP Labs	R007
SHH001A	S4	2010-02-19	HVDRP Labs	R001
SHH001A	S5	2010-02-19	HVDRP Labs	R007
SHH001A	QP	2010-06-14	Internal Medicine BioRad	R001
SHH001A	EX	2010-07-26	New extraction for QPCR rerun	R001

12 records

Figure 3.5. *The output HTML table produced by the query interface.*

3.3.3.2 Online Web Interface

The number of participants enrolled and returning for follow-up visits from the start of the study to the end of the follow-up period (a period of almost two years) were made available for viewing via a web interface. A Python CGI script on the server called an R script, which connected to the database via an adaptor to run a query, created a graph of the results, and made this available for viewing online. Graphs, as output by this script, are shown in Figure 3.6. The number of days between the enrollment date for each participant and the various expected follow-up dates for all participants was also made available as a graph via the web interface, as shown in Figure 3.7. These graphs, which are generated each time the web interface is viewed, were accessed regularly during the course of the study. The web interface presented additional data from the database, such as the number of samples, which had tested positive for HBV, and the date on which the hospital had been notified of this. The output from a number of predefined queries was also available. These included lists of samples from which DNA had been extracted and for which serology tests had been completed.

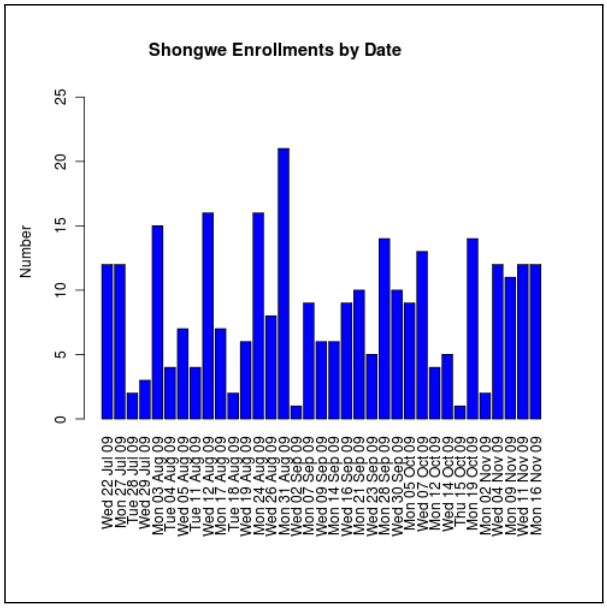


Figure 3.6. *The number of enrollments and the dates on which they enrolled in the study.*

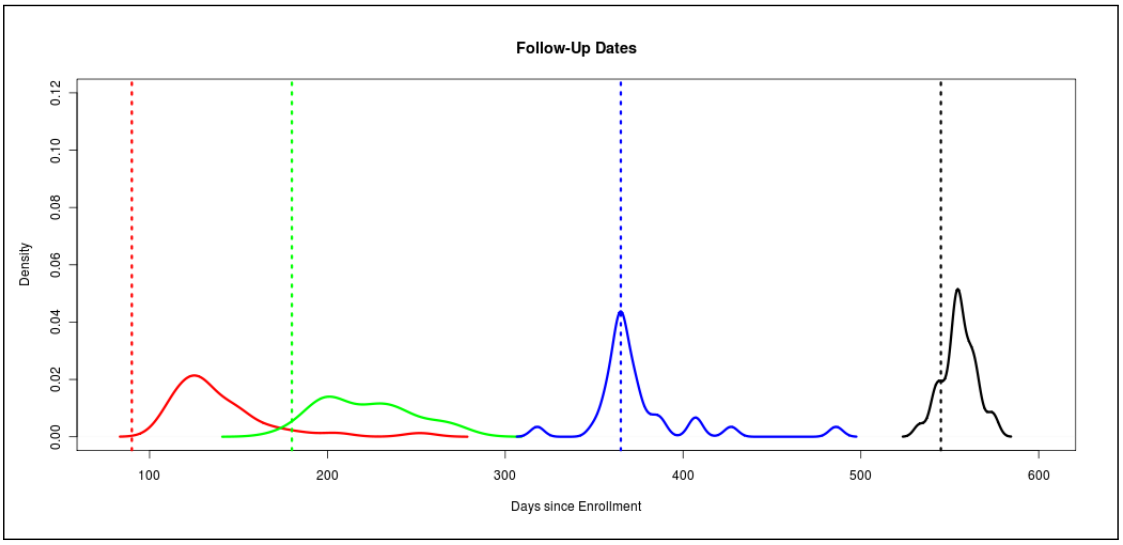
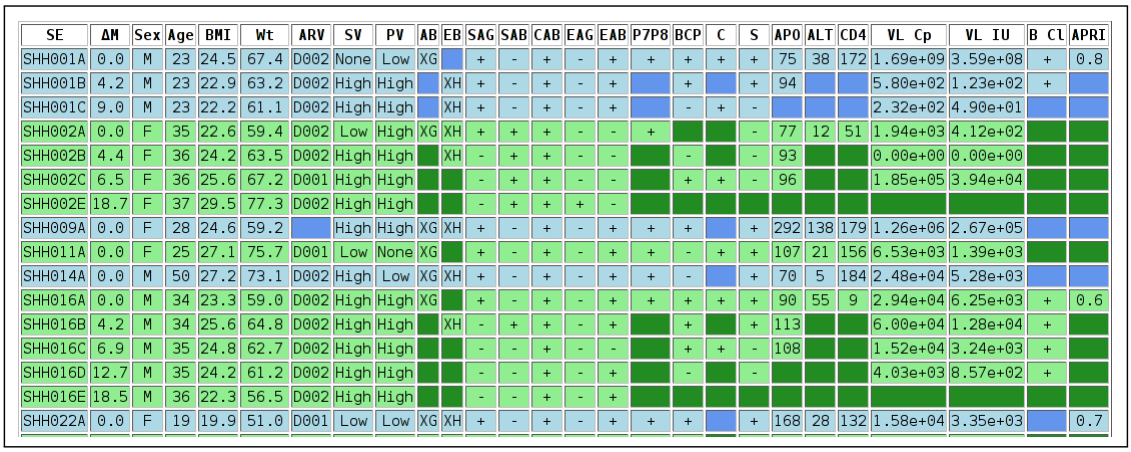


Figure 3.7. *The number of days since enrollment for all participants, plotted as a density function. The dotted lines represent the time-points of 3, 6, 12 and 18 months for reference.*



SE	ΔM	Sex	Age	BMI	Wt	ARV	SV	PV	AB	EB	SAG	SAB	CAB	EAG	EAB	P7P8	BCP	C	S	APO	ALT	CD4	VL Cp	VL IU	B Cl	APRI
SHH001A	0.0	M	23	24.5	67.4	D002	None	Low	XG		+	-	+	-	+	+	+	+	+	75	38	172	1.69e+09	3.59e+08	+	0.8
SHH001B	4.2	M	23	22.9	63.2	D002	High	High		XH	+	-	+	-	+		+		+	94			5.80e+02	1.23e+02	+	
SHH001C	9.0	M	23	22.2	61.1	D002	High	High		XH	+	-	+	-	+		-	+	-				2.32e+02	4.90e+01		
SHH002A	0.0	F	35	22.6	59.4	D002	Low	High	XG	XH	+	+	+	-	-	+			-	77	12	51	1.94e+03	4.12e+02		
SHH002B	4.4	F	36	24.2	63.5	D002	High	High		XH	-	+	+	-	-		-		-	93			0.00e+00	0.00e+00		
SHH002C	6.5	F	36	25.6	67.2	D001	High	High			-	+	+	-	-		+	+	-	96			1.85e+05	3.94e+04		
SHH002E	18.7	F	37	29.5	77.3	D002	High	High			-	+	+	+	-											
SHH009A	0.0	F	28	24.6	59.2		High	High	XG	XH	+	-	+	-	+	+	+		+	292	138	179	1.26e+06	2.67e+05		
SHH011A	0.0	F	25	27.1	75.7	D001	Low	None	XG		+	-	+	-	+	+	-	+	+	107	21	156	6.53e+03	1.39e+03		
SHH014A	0.0	M	50	27.2	73.1	D002	High	Low	XG	XH	+	-	+	-	+	+	-		+	70	5	184	2.48e+04	5.28e+03		
SHH016A	0.0	M	34	23.3	59.0	D002	High	High	XG		+	-	+	-	+	+	+	+	+	90	55	9	2.94e+04	6.25e+03	+	0.6
SHH016B	4.2	M	34	25.6	64.8	D002	High	High		XH	-	+	+	-	+		+		+	113			6.00e+04	1.28e+04	+	
SHH016C	6.9	M	35	24.8	62.7	D002	High	High			-	-	+	-	-		+	+	-	108			1.52e+04	3.24e+03	+	
SHH016D	12.7	M	35	24.2	61.2	D002	High	High			-	-	+	-	+		-		-				4.03e+03	8.57e+02	+	
SHH016E	18.5	M	36	22.3	56.5	D002	High	High			-	-	+	-	+											
SHH022A	0.0	F	19	19.9	51.0	D001	Low	Low	XG	XH	+	-	+	-	+	+	+		+	168	28	132	1.58e+04	3.35e+03		0.7

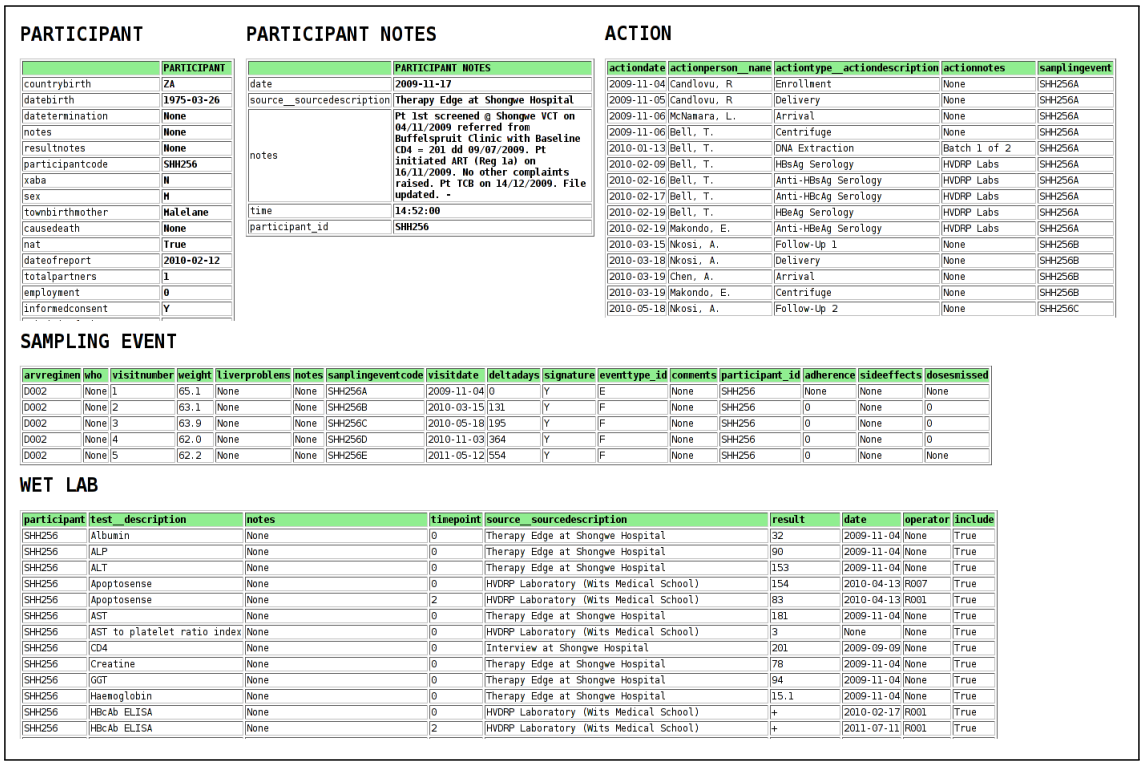
Figure 3.8. *The consolidated data table, which shows all results for all participants.*

3.3.3.3 Consolidated Data Table

The consolidated data table (Figure 3.8) shows multiple test results for each participant in one table. This table is output by a Python script (Appendix D), which queries data from the database and processes it. The query, which is run by the script, extracts data from the "wetlab" table for all participants. As these data are normalized, they need to be manipulated into a "horizontal" table. The Python script searches the query results for all tests, which match a list of those tests required for inclusion in the table, and places these results in a single row of the table for each time-point, for each participant. Tests, which are missing for a given participant (that is, for which there is no record in the database table), are shown in the output table as an empty shaded cell. Some columns, such as "ΔM", "Age" and "BMI", contain calculated data, for, respectively, the number of months between enrollment and the time-point, the age of the participant at the time-point, and the BMI of the participant. The name of the person responsible for generating a test result is shown as a tooltip when the mouse is held over a value in a table cell.

3.3.3.4 Individual Reports

The Django web application framework was used to create a pilot implementation of an interface to display details from all relevant database tables for a single participant. The web interface allows the user to select a study from a list of studies in the database and then presents a list of all participants in the selected study. After the user selects a participant, a web page shows data for the selected participant from the following database tables: participant table, participant notes table, sampling event table, action table and wetlab table. Example output is shown in Figure 3.9



PARTICIPANT

	PARTICIPANT
countrybirth	ZA
datebirth	1975-03-26
datetermination	None
notes	None
resultnotes	None
participantcode	SHH256
xaba	N
sex	M
townbirthmother	Malelane
causedeath	None
nat	True
dateofreport	2010-02-12
totalpartners	1
employment	0
informedconsent	Y

PARTICIPANT NOTES

	PARTICIPANT NOTES
date	2009-11-17
source__sourcedescription	Therapy Edge at Shongwe Hospital
notes	Pt 1st screened @ Shongwe VCT on 04/11/2009 referred from Buffelspruit Clinic with Baseline CD4 = 201 dd 09/07/2009. Pt initiated ART (Reg 1a) on 16/11/2009. No other complaints raised. Pt TCB on 14/12/2009. File updated. -
time	14:52:00
participant_id	SHH256

ACTION

actiondate	actionperson__name	actiontype__actiondescription	actionnotes	samplingevent
2009-11-04	Candlovu, R	Enrollment	None	SHH256A
2009-11-05	Candlovu, R	Delivery	None	SHH256A
2009-11-06	McNamara, L.	Arrival	None	SHH256A
2009-11-06	Bell, T.	Centrifuge	None	SHH256A
2010-01-13	Bell, T.	DNA Extraction	Batch 1 of 2	SHH256A
2010-02-09	Bell, T.	HBsAg Serology	HVDRP Labs	SHH256A
2010-02-16	Bell, T.	Anti-HBsAg Serology	HVDRP Labs	SHH256A
2010-02-17	Bell, T.	Anti-HBcAg Serology	HVDRP Labs	SHH256A
2010-02-19	Bell, T.	HBeAg Serology	HVDRP Labs	SHH256A
2010-02-19	Makondo, E.	Anti-HBeAg Serology	HVDRP Labs	SHH256A
2010-03-15	Nkosi, A.	Follow-Up 1	None	SHH256B
2010-03-18	Nkosi, A.	Delivery	None	SHH256B
2010-03-19	Chen, A.	Arrival	None	SHH256B
2010-03-19	Makondo, E.	Centrifuge	None	SHH256B
2010-05-18	Nkosi, A.	Follow-Up 2	None	SHH256C

SAMPLING EVENT

arvregimen	who	visitnumber	weight	liverproblems	notes	samplingeventcode	visitdate	deltadays	signature	eventtype_id	comments	participant_id	adherence	sideeffects	dosesmissed
D002	None	1	65.1	None	None	SHH256A	2009-11-04	0	Y	E	None	SHH256	None	None	None
D002	None	2	63.1	None	None	SHH256B	2010-03-15	131	Y	F	None	SHH256	0	None	0
D002	None	3	63.9	None	None	SHH256C	2010-05-18	195	Y	F	None	SHH256	0	None	0
D002	None	4	62.0	None	None	SHH256D	2010-11-03	364	Y	F	None	SHH256	0	None	0
D002	None	5	62.2	None	None	SHH256E	2011-05-12	554	Y	F	None	SHH256	0	None	None

WET LAB

participant	test__description	notes	timepoint	source__sourcedescription	result	date	operator	include
SHH256	Albumin	None	0	Therapy Edge at Shongwe Hospital	32	2009-11-04	None	True
SHH256	ALP	None	0	Therapy Edge at Shongwe Hospital	90	2009-11-04	None	True
SHH256	ALT	None	0	Therapy Edge at Shongwe Hospital	153	2009-11-04	None	True
SHH256	Apoptosense	None	0	HVDRP Laboratory (Wits Medical School)	154	2010-04-13	R007	True
SHH256	Apoptosense	None	2	HVDRP Laboratory (Wits Medical School)	83	2010-04-13	R001	True
SHH256	AST	None	0	Therapy Edge at Shongwe Hospital	181	2009-11-04	None	True
SHH256	AST to platelet ratio index	None	0	HVDRP Laboratory (Wits Medical School)	3	None	None	True
SHH256	CD4	None	0	Interview at Shongwe Hospital	201	2009-09-09	None	True
SHH256	Creatine	None	0	Therapy Edge at Shongwe Hospital	78	2009-11-04	None	True
SHH256	GGT	None	0	Therapy Edge at Shongwe Hospital	94	2009-11-04	None	True
SHH256	Haemoglobin	None	0	Therapy Edge at Shongwe Hospital	15.1	2009-11-04	None	True
SHH256	HBcAb ELISA	None	0	HVDRP Laboratory (Wits Medical School)	+	2010-02-17	R001	True
SHH256	HBcAb ELISA	None	2	HVDRP Laboratory (Wits Medical School)	+	2011-07-11	R001	True

Figure 3.9. *Extracts of tables shown in the Individual Reports. Some tables have been truncated for space reasons and the position of individual tables has been adjusted.*

3.3.3.5 R

The statistical programming language, R, was used to analyze clinical and demographic data. An R script was written (Figure 3.10), which included several queries to extract data from the database. This script then processed and analyzed the data, and generated plots as required. The results of these analyses were used in the two published papers (Chapter 4 and Appendix C).

3.4 Conclusion

The database proved invaluable in tracking the progress of the study during the enrollment and follow-up phases, and in managing the samples and analyses undertaken thereafter. The web-based interfaces allowed members of the research team and collaborators to query and view data from any computer or mobile device with a compatible web-browser, which is connected to the Internet. Scripts, which accessed the database, facilitated the analysis of data for preparation and presentation of the results. Improvements were made and additional functionality was implemented in response to difficulties encountered while the database was used during the course of this study.

```
# Assumes required libraries have been installed in R:
# Rdbi (pgUtils and RdbiPgSQL)

# Define an OddsRatio function
OR = function(q) { return (exp(coef(summary(q)))) }

# Define a Confidence Interval function
CI = function(q) { return (exp(confint(q))) }

# Define the function to query the database
getQuery = function(q)
{
  library('RdbiPgSQL')
  PG = PgSQL()
  DBconn = dbConnect(PG, host='HOST', user='USER', dbname='DATABASE',password='PASSWORD')
  Q = dbGetQuery(DBconn, q)
  dbDisconnect(DBconn)
  return (Q)
}

# Define a function to output the glm details
glmShowAll = function(q)
{
   cat ("\n\n---␣GLM␣Model␣---\n")
   print (q)
   cat ("\n\n---␣GLM␣Model␣Summary␣---\n")
   print (summary(q))
   cat ("\n\n---␣Odds␣Ratios␣---\n")
   print (OR(q))
   cat ("\n\n---␣Confidence␣Intervals␣---\n")
   print (CI(q))
   return (NULL)
}

# DATA SET 1: ALL PARTICIPANTS: OUTCOME: HBV POSITIVE (followup TRUE) or NEGATIVE (failure)
# DATA SET 2: ALL HBV POSITIVE PARTICIPANTS: OUTCOME OCCULT (occult TRUE) or OVERT (failure)

########################
### DATA SET 1: ALL ###
########################

# Query a view from the database
All = getQuery('select␣*␣from␣clinical_all2;')

# Cast all columns to the correct data type (just in case)
All$age = as.numeric(All$age)
All$bmi = as.numeric(All$bmi)
All$hbsab = as.factor(All$hbsab)
All$participant = as.character(All$participant)
All$agesex = as.numeric(All$agesex)
All$cd4 = as.numeric(All$cd4)
All$hbsag = as.factor(All$hbsag)
All$sex = as.factor(All$sex)
All$alt = as.numeric(All$alt)
All$followup = as.logical(All$followup)
All$height = as.numeric(All$height)
All$totalpartners = as.numeric(All$totalpartners)
All$apopto = as.numeric(All$apopto)
All$hbcab = as.factor(All$hbcab)
All$occult = as.logical(All$occult)
All$weight = as.numeric(All$weight)

# Run a multivariate logistic regression (removing all serology variables)
temp = glm(All$followup ~ All$age + All$bmi + All$agesex + All$cd4 + All$sex + All$alt + All$totalpartners
     + All$apopto, family=binomial(link="logit")) # this shows that alt is significant
glmShowAll(temp)

# Bin data
All$ageBin = cut(All$age, c(0, 19, 29, 39, 49, 59, 69, 79))
All$cd4Bin = cut(All$cd4, c(0,200,350, 999))
All$altBin = cut(All$alt, c(0, 30, 999))

# Run a multivariate logistic regression with binned data
temp = glm(All$followup ~ All$ageBin + All$bmi + All$agesex + All$cd4Bin + All$sex + All$altBin + All$
     totalpartners + All$apopto, family=binomial(link="logit")) # this shows that alt is significant
glmShowAll(temp)
```

Figure 3.10. *The first section of the R script, which analyzes clinical data.*

Future Work: Sequence Data and Binary Objects The database structure includes tables for storing molecular sequence data (direct, Sanger sequencing and ultra-deep pyrosequencing data), as well as binary objects, such as chromatogram trace files. However, as these features are under development and were not used in the present study, the details are not shown here. Future improvements include creating tools to analyze sequence data from the database, and incorporating clinical and demographic data into these analyses.

Chapter 4

Paper I

Hepatitis B virus infection in human immunodeficiency virus infected Southern African adults: occult or overt - that is the question

Bell, T. G., Makondo, E., Martinson, N. A. and Kramvis, A. (2012)

PLOS One **7**(10): e45750.

References References for literature cited in the following paper appear within the references section of the paper itself, and not necessarily in the references section of this thesis.

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

OPEN ACCESS Freely available online

PLOS ONE

Hepatitis B Virus Infection in Human Immunodeficiency Virus Infected Southern African Adults: Occult or Overt – That Is the Question

Trevor G. Bell[1], Euphodia Makondo[1], Neil A. Martinson[2,3], Anna Kramvis[1]*

1 Hepatitis Virus Diversity Research Programme, Department of Internal Medicine, University of the Witwatersrand, Johannesburg, South Africa, **2** Perinatal HIV Research Unit, University of the Witwatersrand, Johannesburg, South Africa, **3** Johns Hopkins University School of Medicine, Baltmore, Maryland, United States of America

Abstract

Hepatitis B virus (HBV) and human immunodeficiency virus (HIV) share transmission routes and are endemic in sub-Saharan Africa. The objective of the present study was to use the *Taormina* definition of occult HBV infection, together with stringent amplification conditions, to determine the prevalence and characteristics of HBV infection in antiretroviral treatment (ART)-naïve HIV^{+ve} adults in a rural cohort in South Africa. The presence of HBV serological markers was determined by enzyme linked immunoassay (ELISA) tests. HBV DNA-positivity was determined by polymerase chain reaction (PCR) of at least two of three different regions of the HBV genome. HBV viral loads were determined by real-time PCR. Liver fibrosis was determined using the aspartate aminotransferase-to-platelet ratio index. Of the 298 participants, 231 (77.5%) showed at least one HBV marker, with 53.7% HBV DNA^{-ve} (resolved) and 23.8% HBV DNA^{+ve} (current) [8.7% $HBsAg^{+ve}$: 15.1% $HBsAg^{-ve}$]. Only the total number of sexual partners distinguished HBV DNA^{+ve} and HBV DNA^{-ve} participants, implicating sexual transmission of HBV and/or HIV. It is plausible that sexual transmission of HBV and/or HIV may result in a new HBV infection, superinfection and re-activation as a consequence of immunesuppression. Three $HBsAg^{-ve}$ HBV DNA^{+ve} participants had HBV viral loads $<$200 IU/ml and were therefore true occult HBV infections. The majority of $HBsAg^{-ve}$ HBV DNA^{+ve} participants did not differ from $HBsAg^{+ve}$ HBV DNA^{+ve} (overt) participants in terms of HBV viral loads, ALT levels or frequency of liver fibrosis. Close to a quarter of HIV^{+ve} participants were HBV DNA^{+ve}, of which the majority were $HBsAg^{-ve}$ and were only detected using nucleic acid testing. Detection of $HBsAg^{-ve}$ HBV DNA^{+ve} subjects is advisable considering they were clinically indistinguishable from $HBsAg^{+ve}$ HBV DNA^{+ve} individuals and should not be overlooked, especially if lamivudine is included in the ART.

Citation: Bell TG, Makondo E, Martinson NA, Kramvis A (2012) Hepatitis B Virus Infection in Human Immunodeficiency Virus Infected Southern African Adults: Occult or Overt – That Is the Question. PLoS ONE 7(10): e45750. doi:10.1371/journal.pone.0045750

Editor: Sheila Mary Bowyer, University of Pretoria/NHLS TAD, South Africa

Received May 29, 2012; **Accepted** August 24, 2012; **Published** October 1, 2012

Funding: This study was supported by grants received from the South African Medical Research Council and the National Research Foundation (NRF; GUN#65530) awarded to AK. TGB received bursaries from the National Bioinformatics Network, Poliomyelitis Research Foundation (PRF) and NRF. EM received a Belgium Technical Cooperation Fellowship and bursaries from the NRF and PRF. The funders had no role in study design, data collection and analysis, decision to publish.

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

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

Introduction

Hepatitis B virus (HBV) and human immunodeficiency virus (HIV) share transmission routes and represent the two most important blood-borne pathogens in terms of prevalence, morbidity and mortality in sub-Saharan Africa, where both viruses are endemic. Of the 33.3 million adults and children living with HIV globally, 22.5 million reside in sub-Saharan Africa [1]. Moreover, it is estimated that 65% to 98% of populations in sub-Saharan Africa have been exposed to HBV and 8% to 20% are chronic carriers of HBV [2], far exceeding the 4% to 6% lifetime exposure rates and 0.2% to 0.5% carrier rates in regions of low endemicity. Thus, widespread co-infections are likely to occur, with 16% to 98% of HIV^{+ve} individuals in sub-Saharan Africa being carriers of HBV or showing exposure to HBV [3].

The progression of chronic HBV to cirrhosis, end-stage liver disease (ESLD), and hepatocellular carcinoma (HCC) is more rapid in HIV^{+ve} individuals than those with HBV alone [4], with a significant increase in hepatic-related mortality rates [5]. Furthermore, HBV co-infection negatively impacts on HIV outcomes [6]. Before the introduction of antiretroviral therapy (ART), the majority of HBV/HIV co-infected individuals were more likely to die from the clinical consequences of HIV than those of HBV [3]. However, since the introduction of ART, the disease profile has changed, with increases in the proportion of mortality attributed to HBV-associated ESLD [7]. Thus, HBV/HIV co-infection can potentially impact on the safety and effectiveness of ART, requiring an integrated approach for the appropriate management of co-infected individuals [8].

There is a paucity of comprehensive and standardized data describing HBV/HIV co-infection from southern African countries, where HIV prevalence is extremely high. Existing data show large discrepancies, with exposure rate to HBV in HIV^{+ve} South Africans varying from 28% to 99.8% and HBsAg prevalence ranging from 0.4% to 23% [9–18]. Differences can be attributed to different locations, study designs, laboratory measures and/or the composition of the study populations.

HIV infection has been implicated as a risk factor for the development of occult HBV infection (OBI) [12], defined by the *Taormina* expert panel as the "*Presence of HBV DNA in liver (with*

detectable or undetectable HBV DNA in the serum) of individuals testing HBsAg negative by currently available assays. When detectable, the amount of HBV DNA in the serum is usually very low (<200 IU/ml)" [19]. Because liver biopsies are not commonly available, especially in resource-limited environments, OBI is usually detected by the analysis of sera [19]. Furthermore, the experts differentiate between true occult (HBV viral load <200 IU ml^{-1}) and false occult where HBV DNA levels are comparable to those detected in $HBsAg^{+ve}$ infection (overt) and are usually as a result of infection by HBV variants with S gene escape mutants, producing HBsAg that is not recognized by detection assays [19]. The clinical implications of OBI are unclear.

The prevalence of OBI in HIV infected individuals varies depending on the definition used, the sensitivity of the assay and the HBV viral loads [11–13,16,17]. Furthermore, studies performed outside Africa, in areas of low HBV and HIV endemicity, cannot necessarily be extrapolated to Africa because of differences in host factors, epidemiology, transmission patterns and genotypes of the viruses between the two regions.

The objective of the present study was to use the *Taormina* definition of OBI [19], together with stringent amplification conditions, to determine the prevalence and characteristics of HBV infection in ART-naïve HIV^{+ve} adults entering a rural cohort in Mpumalanga Province, which has a HIV prevalence of 15.4% [20]. No in-depth studies have been undertaken to determine the prevalence and characteristics of HBV/HIV co-infection in this province.

Materials and Methods

Subjects

A new rural cohort was established at Shongwe Hospital in Mpumalanga Province in South Africa and 298 ART-naïve, HIV^{+ve} adults were enrolled from July to November 2009. All had qualified for ART according to the then-current South African ART guidelines (CD4 counts <200 cells mm^{-3}) [21] and were recruited while undergoing treatment-readiness counselling. Universal HBV vaccination at 6, 10, and 14 weeks of age was introduced into the South African Expanded Programme on Immunization (EPI) in 1995 and therefore none of the participants were likely to have received this vaccination and self-reported as unvaccinated. Clinical and demographic data (including ALT levels, CD4 T-cell count, age, sex, height and weight) were obtained from hospital records, the National Health Laboratory Services (NHLS) databases and the TherapyEdge-HIV (TE) ^{TM} electronic patient record. All participants signed informed consent. The study was approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand and Mpumalanga Department of Health Research Ethics Committee.

Serology

The presence of HBsAg, anti-HBsAg and anti-HBcAg was determined for 298 sera using the Monolisa^{TM} HBsAg ULTRA, HBsAb ULTRA and HBcAb PLUS ELISA kits (Bio-Rad, Hercules, CA), respectively. HBeAg and anti-HBe tests were performed on HBV DNA^{+ve} sera using the Monolisa^{TM} HBeAg-Ab PLUS kit. Anti-HBcAg IgM was determined for 17 anti-HBc^{+ve} HBV DNA^{+ve} samples for which serum was available using the ARCHITECT® kit (Abbott Diagnostics, Wiesbaden, Germany). The M30-Apoptosense® ELISA (Peviva AB, Stockholm, Sweden) was used on all sera to quantify the apoptosis-associated cytokeratin 18Asp396 neo-epitope as a measure of hepatocyte apoptosis [22].

Measurement of liver fibrosis

The aspartate aminotransferase (AST)-to-platelet ratio index (APRI) = (AST[/ULN]*100)/platelet count [10^9 L^{-1}], a noninvasive measure of liver fibrosis in patients with chronic HBV [23], was calculated for 163 subjects for whom AST levels and platelet counts were available. APRI indicates liver fibrosis only when liver disease has reached a severely advanced stage, with significant fibrosis defined as APRI$\geq$1.5, and no fibrosis as APRI$\leq$0.5 [24].

Polymerase chain reaction (PCR)

DNA was extracted from 200 µl blood plasma with the QIAamp DNA Blood Mini Kit (QIAGEN Gmbh, Hilden, Germany) and eluted into 75 µl of best-quality water (BQW). Known positive and negative sera and BQW were used as controls for the extraction. Three regions of the HBV genome were amplified in a MyCycler^{TM} thermocycler (Bio-Rad, Hercules, Ca, USA) using Promega Taq DNA polymerase (Promega, Madison, WI) (Table 1). To avoid cross-contamination and false positives, the precautions and procedures of Kwok and Higuchi [25] were strictly adhered to. DNA extraction, PCR, and electrophoresis were performed in physically separated venues.

Real-time PCR quantification of HBV DNA

PCR primers, HBV-Taq1 and HBV-Taq2 covering a region of the S gene (321 to 401 from the *EcoR*I site) with a FAM/TAMRA labelled TaqMan BS-1 probe [26] were used to quantify HBV DNA in an ABI 7500 Real Time PCR System (Applied Biosystems, Foster City, Ca, USA). A serial dilution of cloned plasmid DNA containing a single genome of HBV DNA, with concentrations ranging from 2×10^1 to 2×10^{11} IU ml^{-1}, was used as template to generate the standard curve. The second WHO International Standard for HBV Nucleic Acid Amplification Techniques (product code 97/750 National Institute for Biological Standards and Controls (NIBSC); Hertfordshire, UK), which has a final concentration of 10^6 IU ml^{-1} was used as the internal standard. The standard curve, blank, positive and negative controls, and samples were all tested in duplicate. The measured IU/ml for each reaction was calculated using the Ct (cycle threshold) value of each PCR interpolated against the linear regression of the standard curve. The lower detection limit of our assay is $\sim$20 IU ml^{-1}. The conversion formula of IU = copies/4.7 was used [11,27].

Statistical analysis

Clinical data were inspected visually. As all continuous variables showed a skewed distribution, the Mann-Whitney U test (Wilcoxon rank-sum test) was used to compare samples. Chi-squared and Fisher's exact test were used to compare categorical variables. Exhaustive multivariate logistic regression analyses were performed. The R statistical language was used throughout [28].

Results

Serological and nucleic acid testing for HBV

The study group consisted of 298 adults (114 men and 184 women) with median age, CD4 count and BMI of 34 years, 147 cells mm^{-3} and 22 kg m^{-2}, respectively. Men were older than women and had lower CD4 counts (Table 2).

The 298 participants were classified into five serogroups: 28 (9.4%) $HBsAg^{+ve}$, 57 (19.1%) isolated anti-HBc^{+ve}, 123(41.3%) anti-HBc^{+ve}anti-HBs^{+ve}, 11 (3.7%) anti-HBs^{+ve} alone and 79 (26.5%) serologically^{-ve} for HBV. Six percent of men (7/114) were anti-HBs^{+ve} alone compared to 2% (4/184) of women ($p<0.05$). The HBV serologically^{-ve} participants were significant-

Table 1. PCR primers and cycling parameters used for amplification of the three regions of the HBV genome.

Genome Region	Primer	Position[a]	Sequence	Cycles	Denaturation	Annealing	Extension	Size	Reference
Complete S PCR1F	2410(+)	2410-2439	5′-TCAATCGCCGCGTCGCAGAAGATCTCAATC-3′	40	94°C for 1 min	65°C for 1 min	72°C for 3 min	2126	[48]
Complete S PCR1R	1314(−)	1314-1291	5-′TCCAGACCXGCTGCGAGCAAAACA-3′						
Complete S PCR2F	2451(+)	2451-2482	5′-AATGTTAGTATTCCTTGGACTCATAAGGTGGG-3′	40	94°C for 1 min	66°C for 1 min	72°C for 3 min	2051	[48]
Complete S PCR2R	1280(−)	1280-1254	5′-AGTTCCGCAGTATGGATCGGCAGAGGA-3′						
Partial S PCR1F	231(+)	231-249	5′-TCACAATACCGCAGAGTCT-3′	40	94°C for 1 min	55°C for 1 min	72°C for 2 min	571	
Partial S PCR1R	801(−)	801-782	5′-AACAGCGGTATAAAGGGACT-3′						
Partial S PCR2F	256(+)	256-278	5′-GTGGTGGACTTCTCTCAATTTTC-3′	40	94°C for 1 min	55°C for 1 min	72°C for 2 min	541	[49]
Partial S PCR2R	796(−)	796-776	5′-CGGTATAAAGGGACTCACGAT-3′						
BCP PCR1F	1606(+)	1606-1625	5′-GCATGGAGACCACCGTGAAC-3′	40	94°C for 1 min	55°C for 1 min	72°C for 2 min	369	[48]; [50][b]
BCP PCR1R	1974(−)	1974-1955	5′-GGAAAGAAGTCAGAAGGCCAA-3′						
BCP PCR2F	1653(+)	1653-1672	5′-CATAAGAGGACTCTTGGACT-3′	40	94°C for 1 min	55°C for 1 min	72°C for 2 min	307	[48]; [50][b]
BCP PCR2R	1959(−)	1959-1941	5′-GGCAAAAAACAGAGTAACTCA-3′						

[a]Nucleotide position of HBV *adw* genome (GenBank accession number (AY233276), where position 1 is the *EcoR*I cleavage site.
[b]Modifications underlined.
(+) and "F" indicate forward (sense) direction; (−) and "R" indicate reverse (anti-sense) direction; PCR1 indicates first round PCR; PCR2 indicates second round PCR.
doi:10.1371/journal.pone.0045750.t001

ly younger than most HBV serologically^{+ve} groups and had significantly fewer lifetime sexual partners than those with isolated anti-HBs^{+ve} ($p<0.05$). There was no significant difference in serologically-negative and -positive individuals in terms of CD4 counts, age of sexual debut, BMI, ALT and Apoptosense® levels. Only five participants were HBeAg^{+ve} and they did not differ from HBeAg^{-ve} individuals in either demographic or clinical features.

Screening for HBV DNA was carried out using primers targeting three non-overlapping regions of the HBV genome (Table 1). A sample was considered to be HBV DNA^{+ve}, only if at least two regions amplified. Sixty-seven of 298 participants (22.5%) lacked HBV DNA and all HBV serological markers, ruling out HBV exposure and/or infection and with no antibodies against HBV would be susceptible to acquiring HBV infection. The remaining 231/298 (77.5%) showed at least one marker for HBV, with 160/298 (53.7%) HBV DNA^{-ve} (resolved) and 71/298 (23.8%) HBV DNA^{+ve} (current) [26/298 (8.7%) HBsAg^{+ve} (overt): 45/298 (15.1%) HBsAg^{-ve} ("occult")] (Figure 1).

Of the entire group of 298, 26 (8.7%)/28 (9.4%) HBsAg^{+ve} participants were HBV DNA^{+ve} and together with the 45 (15.1%) HBsAg^{-ve} HBV DNA^{+ve} participants were classified into 6 serogroups (Figure 2). Within the HBsAg^{-ve} groups, the frequency of HBV DNA was significantly higher in anti-HBc^{+ve} alone individuals (16/57; 28.1%) compared to those anti-HBc^{+ve}anti-HBs^{+ve} (17/123; 13.8%) ($p<0.05$). The relative risk of an HBsAg^{-ve} individual, who was anti-HBc^{+ve} alone, being HBV DNA^{+ve} was twice as high as that of one with anti-HBc^{+ve}anti-HBs^{+ve}. The frequency of HBV DNA in the serologically^{-ve} group was not significantly different to that in the anti-HBc^{+ve} alone or anti-HBc^{+ve}anti-HBs^{+ve}. Moreover, HBV DNA was not detected in any of the 11 isolated anti-HBs^{+ve} individuals. Sufficient serum was available to test for anti-HBc IgM in 17 of 57 anti-HBc^{+ve} HBV DNA^{+ve} participants and all tested negative. Only three HBsAg^{-ve} HBV DNA^{+ve} participants had viral loads <200 IU ml^{-1}, thus meeting the *Taormina* criterion for true OBI [19]. These participants had serological patterns of groups A, D and E, respectively (Figure 2). All other HBsAg^{-ve} HBV DNA^{+ve} individuals had HBV viral loads >200 IU ml^{-1}.

Comparison of demographic and clinical characteristics between HBV DNA^{+ve} and HBV DNA^{-ve} groups

Visual inspection of plots and linear regression models of each of the continuous variables in Tables 2 and 3 (age, age at sexual debut, lifetime sexual partners, BMI (body mass index), ALT, Apoptosense, CD4 cell count, HBV viral load) against each other, for HBV DNA^{+ve} versus HBV DNA^{-ve}, and HBsAg^{+ve} versus HBsAg^{-ve} groups, revealed no significant correlation.

A multiple logistic regression model was used to determine predictors of HBV DNA positivity. In this model, only ALT levels were significant when all variables were included ($p<0.05$; OR = 1.01; 95% CI: 1.002–1.020). When the data were split according to gender, number of lifetime sexual partners was the only predictor in the females ($p<0.05$; OR = 1.16; 95% CI: 1.01–1.36) and ALT in the males ($p<0.05$; OR = 1.02; 95% CI: 1.004–1.030).

As shown in Table 2, the only variable that differentiated the HBV DNA^{+ve} and HBV DNA^{-ve} groups was number of lifetime sexual partners ($p<0.05$). Regardless of whether they were HBV DNA^{+ve} or HBV DNA^{-ve}, males were older than females, had a higher ALT and lower CD4 count. In the whole cohort and the HBV DNA^{-ve} group, females had a higher BMI ($p<0.05$) and fewer sexual partners than males ($p<0.05$). These differences were not seen in the HBV DNA^{+ve} group. The age of sexual debut was

Table 2. Characteristics of treatment-naïve HIV-infected adults: comparison of HBV DNA^{+ve} *versus* HBV DNA^{-ve} individuals[a].

Characteristic	All participants (n = 292)			HBV DNA^{-ve} participants (n = 71)			HBV DNA^{-ve} participants (n = 221)			Significance – p-values[b]			
	All	Male[A]	Female[A]	All[D]	Male[B]	Female[B]	All[D]	Male[C]	Female[C]	A	B	C	D
Numbers (%)	292[c]	113(39%)	179(61%)	71	30(42%)	41(58%)	221	83(38%)	138(62%)	n/a	0.54	0.72	n/a
Age in years	34(28–41)	36(32–43)	32(27–40)	35(28–41)	37(34–47)	31(25–39)	33(28–41)	35(32–41)	32(27–40)	<0.0001	0.01	0.002	0.42
Age at sexual debut in years[d]	17(16–19)	17(16–20)	17(16–18)	17(16–18)	17(16–19)	16(15–18)	17(16–19)	17(16–20)	17(16–18)	0.035	0.12	0.11	0.36
Lifetime sexual partners[e]	3(1–5)	4(2–8)	2(1–4)	3(2–5)	4(2–8)	3(2–5)	3(1–5)	4(2–8)	2(1–3)	<0.0001	0.34	<0.0001	0.025
BMI in kg m^{-2}	22(20–25)	21(19–24)	23(20–26)	22(20–24)	21(20–23)	22(20–25)	22(20–26)	21(19–24)	23(20–26)	0.0051	0.37	0.007	0.43
ALT in U L^{-1}	21(12–32)	27(17–38)	18(11–27)	23(15–37)	31(19–59)	20(12–28)	20(12–31)	26(17–37)	17(11–27)	<0.0001	0.027	<0.0001	0.12
Apoptosense in U L^{-1}	110(89–150)	112(91–150)	109(89–150)	114(91–158)	117(91–160)	114(91–152)	109(89–142)	111(92–141)	109(88–147)	0.63	0.96	0.68	0.34
CD4 cells mm^{-3}	147(76–196)	99(49–171)	179(104–222)	148(74–199)	104(60–171)	180(103–245)	144(76–194)	98(46–169)	177(104–215)	<0.0001	0.006	<0.0001	0.83

[a]all values expressed as "Median (Interquartile Range)".
[b]comparing the columns marked by UPPER case letter; statistical significance is indicated by underlining.
[c]6 of the 298 participants were not included in the analyses because they lacked a complete data set.
[d]9 samples without data points omitted.
[e]2 samples without data points omitted.
doi:10.1371/journal.pone.0045750.t002

significantly different only when comparing males and females in the whole cohort.

Comparison of demographic and clinical characteristics between $HBsAg^{+ve}$ HBV DNA^{+ve} and $HBsAg^{-ve}$ HBV DNA^{+ve} groups

Data from the HBV DNA^{+ve} participants were examined by logistic regression for predictors of HBV DNA-positivity in the absence of HBsAg. Only increasing age was weakly significant. The female subset showed that age was a significant predictor ($p<0.05$; OR = 1.28; 95% CI: 1.06–1.72). No predictors in the male subset were significant.

In the $HBsAg^{-ve}$ HBV DNA^{+ve} group, men were older and had significantly lower CD4 cell counts compared to females ($p<0.05$). Although the difference in ALT levels between the $HBsAg^{+ve}$HBV DNA^{+ve} and $HBsAg^{-ve}$HBV DNA^{+ve} groups did not reach statistical significance (Table 3), individuals who were $HBsAg^{+ve}$anti-HBc^{+ve} HBV DNA^{+ve} [group C] had significantly higher ALT levels compared to individuals who were either serologically^{-ve} HBV DNA^{+ve} [group A] ($p<0.05$) or anti-HBc^{+ve}HBV DNA^{+ve} [group E] ($p<0.05$) (Figure 2). There was no significant difference between the $HBsAg^{+ve}$ and $HBsAg^{-ve}$ DNA^{+ve} groups when ALT levels were coded into binary groups: >29 U/L for males and >19 U/L for females. HBV viral loads did not differ significantly between $HBsAg^{+ve}$ and $HBsAg^{-ve}$ groups (Table 3).

Measurement of liver fibrosis using APRI score

Ten percent of 163 individuals, for which data were available, had elevated APRI scores (≥ 1.5), representing advanced fibrosis: 7.94% (10/126) HBV DNA^{-ve} [5.3% (2/38) seronegative and 9.1% (8/88) seropositive] and 16.2% (6/37) HBV DNA^{+ve} [26.7% (4/15) $HBsAg^{+ve}$HBV DNA^{+ve} and 9.1% (2/22) $HBsAg^{-ve}$ HBV DNA^{+ve}]. The frequency of liver fibrosis was significantly higher in $HBsAg^{+ve}$HBV DNA^{+ve} individuals compared to seronegative HBV DNA^{-ve} ones ($p<0.05$), but not to seropositive HBV DNA^{-ve} ones ($p = 0.07$). There was no significant difference between the $HBsAg^{+ve}$ and $HBsAg^{-ve}$ HBV DNA^{+ve} groups.

Discussion

In this group of 298 southern African ART-naïve HIV^{+ve} individuals, 231 participants had at least one HBV marker, giving an overall exposure to HBV of 77.5%, comparable to that in HBV monoinfected individuals [2]. In addition, almost one quarter of the group was HBV DNA^{+ve} (Figure 1) of whom almost two thirds were $HBsAg^{-ve}$. Direct comparison with other South African ART-naïve HIV^{+ve} cohorts is difficult because of the different markers were used to measure exposure. In Limpopo Province, exposure to HBV, measured by anti-HBc and/or anti-HBs positivity, was 28.2% in a rural cohort [17] and 39.2% in anti-natal HIV^{+ve} women [16]. This differs from the 63% HBV exposure rate (measured by at least one marker: HBsAg, anti-HBs or anti-HBc) found in a rural-urban HIV^{+ve} cohort in Limpopo [13] and the much higher exposure rate of 99.8% in hospital-admitted HIV^{+ve} patients [12]. In Gauteng Province, a 47% exposure was seen in an urban HIV^{+ve} cohort where ~15% were HBV-positive as follows: 4.8% $HBsAg^{+ve}$ [10], 7.6% anti-HBc^{+ve}HBV DNA^{+ve} [11] and 2.4% serologically^{-ve} HBV DNA^{+ve} [27].

The 9.4% HBsAg prevalence was comparable to that reported for some HIV^{+ve} South African cohorts: 6.2% in anti-natal women in Limpopo Province [16]; 7.1% in rural Eastern Cape (6.6% in ART-treated *versus* 8.8% in ART-naïve, $p>0.05$) [9]; and 6% in a

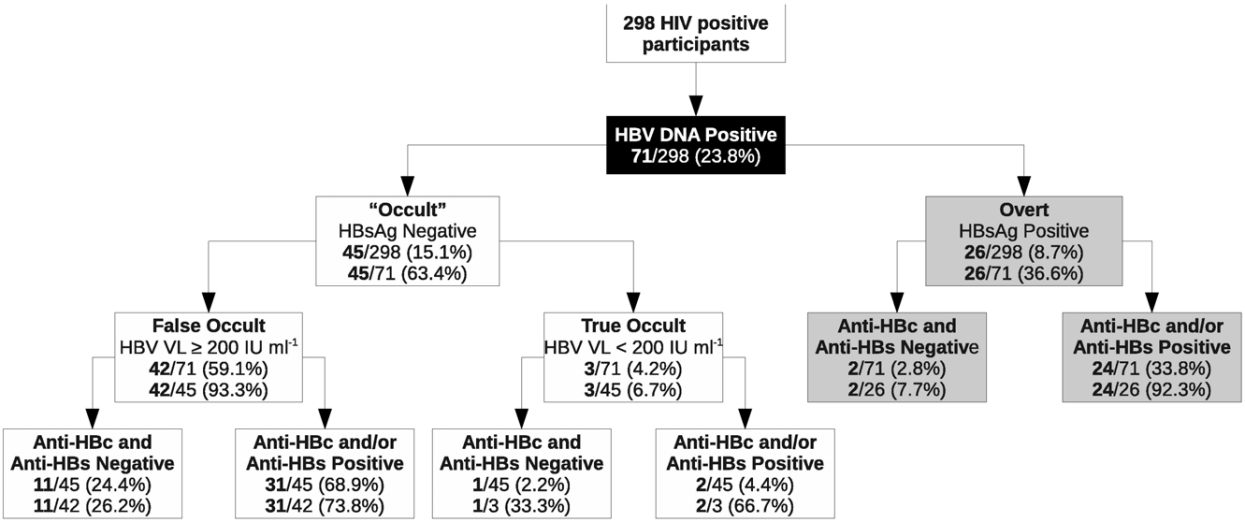


Figure 1. Serological and DNA markers for HBV detected in 71 of 298 HIV^{+ve} participants. Overt refers to HBsAg^{+ve} and ''occult'' to HBsAg^{-ve}. According to the *Taormina* definition, false occult infections are HBsAg^{-ve} with HBV viral load (VL) $\geq$200 IU ml^{-1} and true occult infections are HBsAg^{-ve} with HBV VL $<$200 IU ml^{-1} [19].
doi:10.1371/journal.pone.0045750.g001

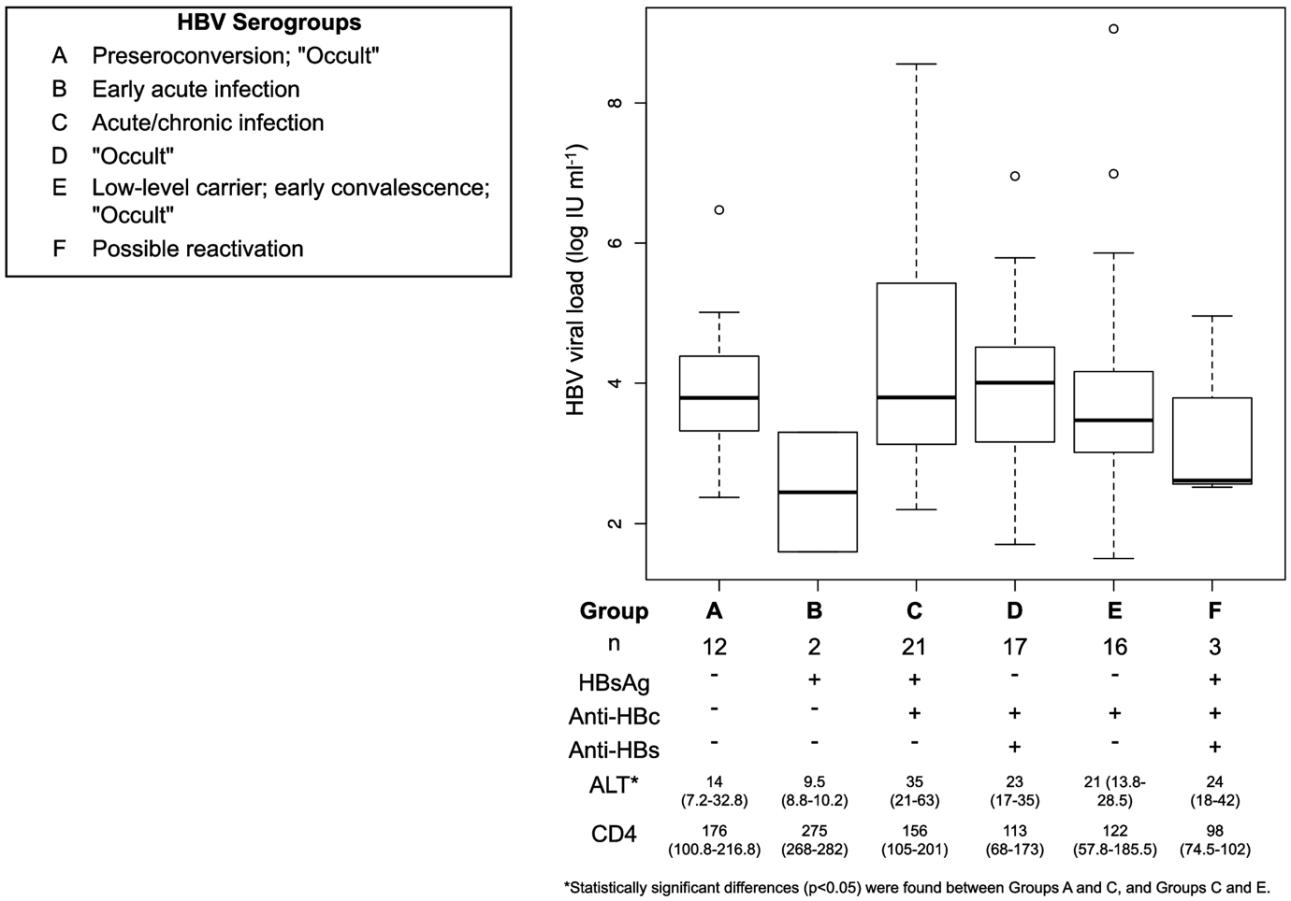


Figure 2. Box and whisker plot of HBV viral loads of the 71 HBV DNA^{+ve} participants separated into the six serological groups (A to F), interpreted according to Hollinger (2008) with modifications [**35**]**.** ''n'' indicates the number of participants in each group. ALT and CD4 cell counts for each group are indicated in the table below the plot as ''Median (Interquartile Range)''. Viral loads and CD4 cell counts did not differ significantly between the six serological groups. The five HBeAg^{+ve} participants belonged to serological group C.
doi:10.1371/journal.pone.0045750.g002

Table 3. Characteristics of treatment-naïve HIV-infected adults: comparison of $HBsAg^{+ve}HBV\ DNA^{+ve}$ *versus* HBsAg-ve HBV DNA^{+ve} individuals[a].

Characteristic	HBV DNA^{+ve} participants (n = 71)			$HBsAg^{+ve}HBV\ DNA^{+ve}$ (n = 26)			HBsAg-ve HBV DNA^{+ve} (n = 45)			Significance–p values[b]			
	All	Male[A]	Female[A]	All[D]	Male[B]	Female[B]	All[D]	Male[C]	Female[C]	A	B	C	D
Numbers (%)	71	30(42%)	41(58%)	26	13(50%)	13(50%)	45	17(38%)	28(62%)	n/a	0.42	0.54	n/a
Age in years	35(28–41)	37(34–47)	31(25–39)	34(27–28)	35(33–39)	29(25–35)	36(29–45)	38 (35–47)	34(28–40)	0.01	0.07	0.035	0.14
Age at sexual debut in years[c]	17(16–18)	17(16–19)	16(15–18)	17(16–20)	18(16–22)	17(15–18)	16(15–18)	17(16–18)	16(15–18)	0.12	0.11	0.66	0.23
Lifetime sexual partners[d]	3(2–5)	4(2–8)	3(2–5)	3(1–6)	2(1–6)	3(1–5)	4(2–5)	4(2–10)	3(2–5)	0.34	0.98	0.17	0.38
BMI in kg m^{-2}	22(20–24)	21(20–23)	22(20–25)	23(20–25)	22(20–23)	25(20–27)	21(20–23)	21(20–22)	22(20–24)	0.37	0.29	0.64	0.30
ALT in U L^{-1}	23(15–37)	31(18–59)	20(12–28)	30(19–59)	35(22–60)	21(12–39)	20(13–32)	27(15–37)	19(13–25)	0.027	0.29	0.12	0.07
Apoptosense U L^{-1}	114(91–158)	117(91–160)	114(91–152)	113(84–156)	110(75–150)	114(91–161)	114(94–158)	121(96–188)	114(91–142)	0.96	0.27	0.37	0.62
CD4 cells mm^{-3}	148(74–199)	104(60–171)	180(103–245)	152 (101–202)	125 (98–184)	179 (132–255)	147(68–196)	87(57–165)	181(91–229)	0.006	0.13	0.016	0.52
LogHBVVL[e]	3.72 (3.04–4.41)	3.76 (2.95–4.47)	3.61 (3.08–4.35)	3.64 (2.73–4.80)	4.14 (2.74–5.44)	3.22 (2.54–4.01)	3.78 (3.15–4.35)	3.74 (2.99–4.28)	3.81 (3.19–4.44)	0.66	0.14	0.59	0.74

[a]all values expressed as "Median (Interquartile Range)".
[b]comparing the columns marked by UPPER case letter; statistical significance is indicated by underlining.
[c]9 samples without data points omitted.
[d]2 samples without data points omitted.
[e]"LogHBVVL" indicates the log of the HBV viral load rounded to two decimal places [A Mann-Whitney U test was used on the untransformed viral loads (IU ml^{-1})].
doi:10.1371/journal.pone.0045750.t003

country-wide study of treatment-naïve HIV^{+ve} military personnel and their family members [14]. On the other hand, the HBsAg prevalence was higher than the 0.4% in another rural cohort in Limpopo Province [17], double the 4.8% in a Gauteng urban cohort [10], but lower than the 11.3% in hospital-admitted Limpopo Province patients [12], the 19.7% in miners [15] and the 22.9% from a rural-urban cohort in Limpopo Province [13]. This difference in HBsAg prevalence correlates with the variations reported in HBV monoinfected individuals from different locales [16,29,30].

Regardless of whether they were HBV DNA^{+ve} or HBV DNA^{-ve}, males were older, had higher ALT levels and lower CD4 counts than females (Table 1). These differences are because males tend to come for treatment later than females [31]. In the cohort as a whole and in the HBV DNA^{-ve} group, males had significantly more partners than females, with BMI significantly lower. In the HBV DNA^{+ve} group, these factors did not differ between the genders. The only factor differentiating the HBV DNA^{+ve} versus HBV DNA^{-ve} participants was the number of lifetime sexual partners (Table 2), suggesting sexual transmission of HBV and/or HIV. It is plausible this mode of transmission may result in a new HBV infection, superinfection and re-activation as a consequence of immunesuppression. Twenty percent of the 71 HBV DNA+ve participants had possible markers of recent infection: 12 $serologically^{-ve}$ HBV DNA^{+ve} and 2 $HBsAg^{+ve}$ HBV DNA^{+ve} (Figure 2). Of the 17 anti-HBc^{+ve} HBV DNA^{+ve} sera tested for anti-HBc IgM, none were positive.

The HBV serology of the cohort, the high frequency of HBeAg-negativity, the absence of anti-HBc IgM and the relatively low HBV viral loads (Figure 2) reflect the natural history of HBV infection in sub-Saharan Africa, where most individuals are infected at childhood by horizontal transmission [2]. This means that most individuals have been exposed to HBV, and are protected by anti-HBV antibodies, by the time they become sexually active and acquire HIV. All isolated anti-HBs^{+ve} participants were HBV DNA^{-ve} and the presence of anti-HBs with anti-HBc reduced the risk of being HBV DNA^{+ve}. None of the participants had received HBV vaccination.

The HBsAg prevalence in this HIV^{+ve} cohort was not different to HIV^{-ve} cohorts [2,3]. This differs from observations in areas of low HBV and HIV endemicity, where HBV and HIV are acquired simultaneously and therefore HBsAg prevalence in HIV^{+ve} individuals is significantly higher than in HIV^{-ve} individuals [3]. Only four participants in the present study were $HBsAg^{+ve}$ alone: two were HBV DNA^{+ve}, whereas the other two were HBV DNA^{-ve}, even after repeated attempts to amplify HBV DNA, possibly indicating low viral loads undetectable by PCR. This might reflect the process of natural HBsAg clearance [32]. Although immune suppression by HIV may lead to the $HBsAg^{+ve}$anti-HBc^{-ve} profile [33], this is unlikely in these two cases, considering that ~59% of the participants were anti-HBc^{+ve}, with a third of these having isolated anti-HBc. Moreover, HIV^{+ve} patients with CD4<100 cells mm^3 are more likely to have isolated anti-HBc [34].

HBV DNA without HBsAg was detected in 15.1% of the participants (Figure 1). This is within the 8% to 18% range for South African HIV^{+ve} cohorts but again direct comparison is complicated by differences in study design [11–13,16,17]. Twelve participants were $serologically^{-ve}$ HBV DNA^{+ve}, which can occur before the appearance of HBsAg, in the preseroconversion phase (indicating a recent infection), or at the tail end of the infection [35]. Anti-HBsAg seroconversion, in the presence or absence of anti-HBc, decreased the relative risk of being HBV DNA^{+ve} in the

$HBsAg^{-ve}$ group. This agrees with findings in HBV monoinfected [36] and in HBV/HIV coinfected individuals [37].

There was no difference in the demographics of the HBV DNA^{+ve} subjects, with and without HBsAg (Table 2). In the presence of HBsAg, there was no difference between males and females, whereas in the absence of HBsAg, males were older and had lower CD4 counts than females. Thus older males with lower CD4 counts are more likely to be $HBsAg^{-ve}$ HBV DNA^{+ve}. Lower CD4 counts have been associated with $HBsAg^{-ve}$ viremia regardless of gender [37], however the median CD4 counts in that study were relatively higher (316 cells mm^{-3} versus 147cells mm^{-3} in the present study) [37].

In agreement with other studies [38,39], there were similar ALT levels in $HBsAg^{+ve}$ and $HBsAg^{-ve}$ HBV DNA^{+ve} participants and between HBV DNA^{+ve} and HBV DNA^{-ve} participants. The absence of transaminitis is as a result of the immunosuppressed state of the HIV^{+ve} subjects. Immunosuppression causes HBV reactivation and can lead to high viremia without clinical manifestation [40]. The APRI score was used to compare the frequency of liver fibrosis in the HBV^{+ve} *versus* HBV^{-ve} participants. The frequency of liver fibrosis was significantly higher in $HBsAg^{+ve}$ HBV DNA^{+ve} individuals compared to seronegative HBV DNA^{-ve} ones, but not relative to seropositive HBV DNA^{-ve} ones. It is intriguing that there was no difference in the frequency of liver fibrosis between HBV DNA^{+ve} individuals, with and without HBsAg.

The reactivation of an infection, which originated in childhood, can explain why no significant difference was seen in the HBV viral loads between the $HBsAg^{+ve}$ and $HBsAg^{-ve}$ participants (Table 2), nor between the different serological groups (Figure 2). Following HIV infection, HBV can reactivate in anti-HBs^{+ve} only individuals, with and without the reappearance of HBsAg [41]. Group F, which had the lowest CD4 count of <100 cells mm^{-3}, and by inference was the most immunosuppressed, was $HBsAg^{+ve}$ anti-HBc^{+ve} anti-HBs^{+ve} HBV DNA^{+ve} with a viral load $>10^2$ IU ml^{-1} (Figure 2). Spontaneous reverse seroconversion, where anti-HBs disappears and HBsAg reappears can also occur in the presence of CD4 counts <200 cells mm^{-3} [42]. Although HBV viral loads have been shown to be higher in HBV^{+ve} HIV^{+ve} individuals compared to HBV^{+ve} ones [43], the HBV viral loads detected in the present study were comparable to those detected in HBV mono-infected individuals [44]. This is probably because the majority of individuals were infected with subgenotype A1 [45], which is characterized by relatively low viral loads in mono-infected individuals compared to other genotypes or subgenotypes [44].

Only three $HBsAg^{-ve}$ HBV DNA^{+ve} patients had HBV loads <200 IU ml^{-1}, meeting the *Taormina* criterion for OBI. Thus the majority of $HBsAg^{-ve}$ HBV DNA^{+ve} would be classified as false "occult" [19]. It is possible that immunosuppression precludes true occult HBV infection. Because the majority of $HBsAg^{-ve}$ HBV DNA^{+ve} ("occult") participants did not differ from $HBsAg^{+ve}$ HBV DNA^{+ve} (overt) participants in terms of viral loads, CD4 counts, ALT levels and frequency of liver fibrosis, it may be more accurate to refer to these HBV infections as *HBsAg-covert* (HBsAg-***c***ryptic ***o***vert) instead of false "occult" [19].

HIV infection was demonstrated to be a risk factor for $HBsAg^{-ve}$ HBV infection [12], and pre-S mutations preventing HBsAg secretion [46], 'a' determinant mutations leading to detection escape and overlapping polymerase mutations affecting replication, may be responsible for this. This possibility was investigated and is presented in a follow-up paper, where 12 of 13 HBV S region sequences, from $HBsAg^{-ve}$ participants, had pre-S and/or S mutations [45]. Another possible explanation for HBsAg-negativity may be that HIV co-infection prevents HBsAg secretion, as shown in co-infected hepatic cell lines [47].

Despite the possible limitations of this study, including its cross-sectional nature, the absence of HIV viral loads, no HBV mono-infected patients and patients with higher CD4 counts for comparison, a number of important conclusions can be reached. The number of lifetime sexual partners was the only factor differentiating HBV DNA^{+ve} and HBV DNA^{-ve} infections, suggesting sexual transmission of HBV and/or HIV. HBV^{+ve} HIV^{+ve} individuals were found to have significantly higher lifetime sexual partners than HBV-monoinfected individuals [18]. HBV infection in HIV^{+ve} individuals was predominantly $HBsAg^{-ve}$, which did not differ significantly from $HBsAg^{+ve}$ infections in terms of viral loads, CD4 counts, ALT levels and frequency of liver fibrosis.

The detection of HBV DNA in the absence of HBsAg in this and other South African studies [11–13,17] has important implications for the clinical management of HIV in sub-Saharan Africa, where the burden of HBV/HIV co-infection is disproportionately high (24% in this study). Although the World Health Organization recommends that ART be initiated in HBV/HIV co-infected individuals irrespective of CD4 count, in South Africa we face a number of challenges. The most recent South African guidelines recommend initiation of treatment of patients with CD4 counts <350 cells mm^{-3} and HBsAg testing if ALT levels exceed 100 U L^{-1}. Considering that the highest median ALT levels (IQR) of 30 (19–59) U L^{-1} were found in the $HBsAg^{+ve}$ HBV DNA^{+ve} group (Table 3), which also had the highest frequency of advanced fibrosis, this cut-off value is inappropriate. Moreover, 65% of the 71 participants, who were HBV^{+ve} HIV^{+ve}, lacked HBsAg and HBV could only be detected by nucleic acid testing, which is unaffordable in resource-limited environments. Although the clinical significance of $HBsAg^{-ve}$ infection is under debate [32], it is imperative that HBV/HIV co-infection is detected before ART initiation, especially because lamivudine remains in two of the three drug regimens currently provided by the South African government and HBV can develop resistance to lamivudine. To determine the clinical relevance of *HBsAg-covert* HBV infection in our setting, prospective studies following ART initiation are in progress.

Acknowledgments

The authors thank Sisters Ntombikayise Elizabeth Agatha Nkosi and Rosalina Candlovu for enrolling participants and drawing blood samples, Dr Precious Gabashane, chief medical officer, Shongwe Hospital and all the study participants. Professor Jacky Galpin and Ms Elsabe Smit of the School of Statistics and Actuarial Science of the University of the Witwatersrand provided guidance with the statistical analyses. Some clinical data were provided by the National Health Laboratory Services (NHLS) Corporate Data Warehouse Team and by Dr Mhairi Maskew (Right-to-Care/TherapyEdge).

Author Contributions

Conceived and designed the experiments: TGB EM NM AK. Performed the experiments: TGB EM. Analyzed the data: TGB EM AK. Contributed reagents/materials/analysis tools: TGB AK. Wrote the paper: TGB AK. Set up rural cohort site: TGB NM AK.

References

1. UNAIDS (2010) UNAIDS report on the global AIDS epidemic 2010. Available: http://www.unaids.org/en/dataanalysis/. Accessed 2011 November 16.
2. Kramvis A, Kew MC (2007) Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes. Hepatol Res 37: S9–S19.
3. Burnett RJ, Francois G, Kew MC, Leroux-Roels G, Meheus A, et al. (2005) Hepatitis B virus and human immunodeficiency virus co-infection in sub-Saharan Africa: a call for further investigation. Liver Int 25: 201–213.
4. Chung RT (2006) Hepatitis C and B viruses: the new opportunists in HIV infection. Top HIV Med 14: 78–83.
5. Thio CL, Seaberg EC, Skolasky R, Phair J, Visscher B, et al. (2002) HIV-1, hepatitis B virus, and risk of liver-related mortality in the Multicenter Cohort Study (MACS). Lancet 360: 1921–1926.
6. Chun HM, Roediger MP, Hullsiek KH, Thio CL, Agan BK, et al. (2011) Hepatitis B Virus Coinfection Negatively Impacts HIV Outcomes in HIV Seroconverters. J Infect Dis 205: 185–193.
7. Thomas DL (2006) Growing importance of liver disease in HIV-infected persons. Hepatology 43: S221–229.
8. Soriano V, Sheldon J, Ramos B, Nunez M (2006) Confronting chronic hepatitis B virus infection in HIV: new diagnostic tools and more weapons. Aids 20: 451–453.
9. Boyles TH, Cohen K (2011) The prevalence of hepatitis B infection in a rural South African HIV clinic. S Afr Med J 101: 470–471.
10. Firnhaber C, Reyneke A, Schulze D, Malope B, Maskew M, et al. (2008) The prevalence of hepatitis B co-infection in a South African urban government HIV clinic. S Afr Med J 98: 541–544.
11. Firnhaber C, Viana R, Reyneke A, Schultze D, Malope B, et al. (2009) Occult hepatitis B virus infection in patients with isolated core antibody and HIV co-infection in an urban clinic in Johannesburg, South Africa. Int J Infect Dis 13: 488–492.
12. Mphahlele MJ, Lukhwareni A, Burnett RJ, Moropeng LM, Ngobeni JM (2006) High risk of occult hepatitis B virus infection in HIV-positive patients from South Africa. J Clin Virol 35: 14–20.
13. Lukhwareni A, Burnett RJ, Selabe SG, Mzileni MO, Mphahlele MJ (2009) Increased detection of HBV DNA in HBsAg-positive and HBsAg-negative South African HIV/AIDS patients enrolling for highly active antiretroviral therapy at a Tertiary Hospital. J Med Virol 81: 406–412.
14. Matthews GV, Manzini P, Hu Z, Khabo P, Maja P, et al. (2011) Impact of lamivudine on HIV and hepatitis B virus-related outcomes in HIV/hepatitis B virus individuals in a randomized clinical trial of antiretroviral therapy in southern Africa. AIDS 25: 1727–1735.
15. Hoffmann CJ, Charalambous S, Martin DJ, Innes C, Churchyard GJ, et al. (2008) Hepatitis B virus infection and response to antiretroviral therapy (ART) in a South African ART program. Clin Infect Dis 47: 1479–1485.
16. Burnett RJ, Ngobeni JM, Francois G, Hoosen AA, Leroux-Roels G, et al. (2007) Increased exposure to hepatitis B virus infection in HIV-positive South African antenatal women. Int J STD AIDS 18: 152–156.
17. Barth RE, Huijgen Q, Tempelman HA, Mudrikova T, Wensing AM, et al. (2011) Presence of occult HBV, but near absence of active HBV and HCV infections in people infected with HIV in rural South Africa. J Med Virol 83: 929–934.
18. Mayaphi SH, Roussow TM, Masemola DP, Olorunju SA, Mphahlele MJ, et al. (2012) HBV/HIV co-infection: the dynamics of HBV in South African patients with AIDS. S Afr Med J 102: 157–162.
19. Raimondo G, Allain JP, Brunetto MR, Buendia MA, Chen DS, et al. (2008) Statements from the Taormina expert meeting on occult hepatitis B virus infection. J Hepatol 49: 652–657.
20. Shisana O, Rehle T., Simyabi L. C., Zuma K., Jooste S., Pillay-van-Wyk V., Mbelle N., Van Zyl J Parker W., Zungu N. P., Pezi S. and the SABSSM III Implementation Team (2009) South African national HIV prevalence, incidence, behavior and communication survey 2008: A turning tide among teenagers? Cape Town: HSRC Press.
21. National Department of Health SA (2004) National Antiretroviral Treatment Guidelines: Jacana, Minuteman Press.
22. Leers MP, Kolgen W, Bjorklund V, Bergman T, Tribbick G, et al. (1999) Immunocytochemical detection and mapping of a cytokeratin 18 neo-epitope exposed during early apoptosis. J Pathol 187: 567–572.
23. Lebensztejn DM, Skiba E, Sobaniec-Lotowska M, Kaczmarski M (2005) A simple noninvasive index (APRI) predicts advanced liver fibrosis in children with chronic hepatitis B. Hepatology 41: 1434–1435.
24. Lo Re V, 3rd, Frank I, Gross R, Dockter J, Linnen JM, et al. (2007) Prevalence, risk factors, and outcomes for occult hepatitis B virus infection among HIV-infected patients. J Acquir Immune Defic Syndr 44: 315–320.
25. Kwok S, Higuchi R (1989) Avoiding false positives with PCR. Nature 339: 237–238.
26. Weinberger KM, Wiedenmann E, Bohm S, Jilg W (2000) Sensitive and accurate quantitation of hepatitis B virus DNA using a kinetic fluorescence detection system (TaqMan PCR). J Virol Methods 85: 75–82.
27. Firnhaber C, Chen CY, Evans D, Maskew M, Schulz D, et al. (2012) Prevalence of hepatitis B virus (HBV) co-infection in HBV serologically-negative South African HIV patients and retrospective evaluation of the clinical course of mono- and co-infection. Int J Infect Dis 16: e268–272.
28. R Development Core Team (2011) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. Available: http://www.R-project.org. Accessed 2010 June.
29. Kew MC (1996) Progress towards the comprehensive control of hepatitis B in Africa: a view from South Africa. Gut 38 Suppl 2: S31–36.
30. Vardas E, Mathai M, Blaauw D, McAnerney J, Coppin A, et al. (1999) Preimmunization epidemiology of hepatitis B virus infection in South African children. J Med Virol 58: 111–115.
31. Grinsztejn B, Smeaton L, Barnett R, Klingman K, Hakim J, et al. (2011) Sex-associated differences in pre-antiretroviral therapy plasma HIV-1 RNA in diverse areas of the world vary by CD4(+) T-cell count. Antivir Ther 16: 1057–1062.
32. Togashi H, Hashimoto C, Yokozawa J, Suzuki A, Sugahara K, et al. (2008) What can be revealed by extending the sensitivity of HBsAg detection to below the present limit? J Hepatol 49: 17–24.
33. Avettand-Fenoel V, Thabut D, Katlama C, Poynard T, Thibault V (2006) Immune suppression as the etiology of failure to detect anti-HBc antibodies in patients with chronic hepatitis B virus infection. J Clin Microbiol 44: 2250–2253.
34. Sun HY, Lee HC, Liu CE, Yang CL, Su SC, et al. (2009) Factors associated with isolated anti-hepatitis B core antibody in HIV-positive patients: impact of compromised immunity. J Viral Hepat 17: 578–587.
35. Hollinger FB (2008) Hepatitis B virus infection and transfusion medicine: science and the occult. Transfusion 48: 1001–1026.
36. Chu CM, Liaw YF (2012) Prevalence of and Risk Factors for Hepatitis B Viremia After Spontaneous Hepatitis B Surface Antigen Seroclearance in Hepatitis B Carriers. Clin Infect Dis 54: 88–90.
37. Cohen Stuart JW, Velema M, Schuurman R, Boucher CA, Hoepelman AI (2009) Occult hepatitis B in persons infected with HIV is associated with low CD4 counts and resolves during antiretroviral therapy. J Med Virol 81: 441–445.
38. Lo Re V, 3rd, Kostman JR, Gross R, Reddy KR, Mounzer K, et al. (2007) Incidence and risk factors for weight loss during dual HIV/hepatitis C virus therapy. J Acquir Immune Defic Syndr 44: 344–350.
39. Shire NJ, Rouster SD, Rajicic N, Sherman KE (2004) Occult hepatitis B in HIV-infected patients. J Acquir Immune Defic Syndr 36: 869–875.
40. Gerlich WH, Bremer C, Saniewski M, Schuttler CG, Wend UC, et al. (2010) Occult hepatitis B virus infection: detection and significance. Dig Dis 28: 116–125.
41. Vento S, di Perri G, Luzzati R, Cruciani M, Garofano T, et al. (1989) Clinical reactivation of hepatitis B in anti-HBs-positive patients with AIDS. Lancet 1: 332–333.
42. Thio CL (2009) Hepatitis B and human immunodeficiency virus coinfection. Hepatology 49: S138–145.
43. Gilson RJ, Hawkins AE, Beecham MR, Ross E, Waite J, et al. (1997) Interactions between HIV and hepatitis B virus in homosexual men: effects on the natural history of infection. AIDS 11: 597–606.
44. Tanaka Y, Hasegawa I, Kato T, Orito E, Hirashima N, et al. (2004) A case-control study for differences among hepatitis B virus infections of genotypes A (subtypes Aa and Ae) and D. Hepatology 40: 747–755.
45. Makondo E, Bell TG, Kramvis A (2012) Genotyping and molecular characterization of hepatitis B virus (HBV) from antiretroviral treatment naive human immunodeficiency virus (HIV)-infected southern African adults. PlosOne (in press): e46345. doi:10.1371/journal.pone.0046345.
46. Melegari M, Bruno S, Wands JR (1994) Properties of hepatitis B virus pre-S1 deletion mutants. Virology 199: 292–300.
47. Iser DM, Avihingsanon A, Wisedopas N, Thompson AJ, Boyd A, et al. (2010) Increased intrahepatic apoptosis but reduced immune activation in HIV-HBV co-infected patients with advanced immunosuppression. AIDS 25: 197–205.
48. Vermeulen M, Dickens C, Lelie N, Walker E, Coleman C, et al. (2012) Hepatitis B virus transmission by blood transfusion during 4 years of individual-donation nucleic acid testing in South Africa: estimated and observed window period risk. Transfusion 52: 880–892.
49. Lindh M, Andersson AS, Gusdal A (1997) Genotypes, nt 1858 variants, and geographic origin of hepatitis B virus–large-scale analysis using a new genotyping method. J Infect Dis 175: 1285–1293.
50. Takahashi K, Aoyama K, Ohno N, Iwata K, Akahane Y, et al. (1995) The precore/core promoter mutant (T1762A1764) of hepatitis B virus: clinical significance and an easy method for detection. J Gen Virol 76 (Pt 12): 3159–3164.

Chapter 5

Development of Bioinformatic Tools

Terminology A glossary of technical terms is provided in Section 3.3.1.

5.1 CLIMB

A set of Python methods to programmatically manipulate and process biological sequence data, named "Command Line Interface for Molecular Biology" (CLIMB), was developed in this study. The HBV genome (Section 1.1.3), which is circular, is 3200 nucleotides in length. Subgenomic fragments from several samples are routinely sequenced and examined for mutations at known loci (positions), or translated into amino acids. The CLIMB routines were designed to allow many of these tasks to be performed programmatically, either via a script or interactively, as in the following example. Python commands are shown on the left, and a descriptive comment explaining the command is shown in green on the right of each line.

```
1 import climb
2 S = climb.Sequence()                    # create a new object
3 S.load('aligned.fasta')                 # load a FASTA file
4 S.seqLength()                           # output length of each sequence
5 S.seqCase()                             # convert sequence to uppercase
6 S.extract('1809-1812', mapping=1751)    # display the "Kozak" sequence
```

Additional methods were added as the library was developed. Although, where appropriate, BioPython (Cock *et al.*, 2009) routines were used, the CLIMB library was developed independently, rather than extending BioPython. The reasons for this include greater flexibility and control,

the extension of skills obtained when developing solutions *de novo* and a lower risk of incompatibilities introduced by dependencies on external libraries, which are updated frequently.

The CLIMB module consists of a "Sequence" class, with methods, functions and variables. An "__init__" class allocates default values to a number of variables. The "load" method reads data from either a FASTA file or a chromatogram file in the standard Applied Biosystems "ABIF" file format[1]. The FASTA file may contain one or more sequences, which would typically be aligned. A list of class methods is provided in Table 5.1. In addition, the module contains variables and functions, which are not part of the "Sequence" class.

5.1.1 Common Gateway Interface

To facilitate development and testing of the CLIMB module, a front-end for users was required. Executing Python source code, with a graphical interface, on different operating system platforms can be difficult, particularly for non-computer-literate end-users. Updating the code, as improvements and corrections are made, would also be difficult, as the end-user would have to check that the latest version has been installed. For these reasons, it was decided to develop a web-based front-end for the module, as this could easily be used from any operating system platform and did not require installation of software. Furthermore, updates could be applied quickly and the front-end could easily be accessed from a wide variety of locations.

A Python Common Gateway Interface (CGI) was implemented on a GNU/Linux server (Section 3.3.2). Implementation of several features of the module, in the form of online tools, were in response to feedback received from members of the research group, who were using the front-end. As such, aspects of the tools were developed to address the specific needs of the users, who were testing and using them regularly.

Two of the tools have been described in papers. The paper discussing the "Mutation Reporter Tool", published in *Virology Journal*, is reproduced as Chapter 6, with the paper describing the "Fragment Merger Tool", published in *Viruses*, reproduced as Chapter 7. Other tools, which were developed during this study, are described below.

[1]The Python ABIF file reader, which was used, is available online at *http://www.interactive-biosoftware.com/software/resources/abif-reader*

Table 5.1. List of methods in the "Sequence" class.

Method	Description
baseDistribution	Return base distribution values (not percentages) at a given position
basePercentage	Return count or percentage of base(s) or fragment (no regular expressions)
blunt	Remove columns from multiple sequence alignment such that no sequence starts or ends with a gap
countMotif	Count mutations – that is, where mutSeq == sequence data
disambiguateColumns	Replace/correct/disambiguate ambiguous bases in mono-base gap-free columns
eliminateGapColumns	Eliminate columns containing >= threshold proportion of GAPs
extract	Return proteins from a nucleotide sequence
find	Return all occurrences of a regular expression in a nucleotide sequence
gvt	Return variation at each locus for two groups
load	Load sequence data from a file
mutationFinder	Return mutations found in self using given object reference as reference sequence
nucCopy	Crop sequences
save	Save sequence data to a file
seqAdd	Add sequences from another object (.seq)
seqCase	Change case of sequence data
seqCount	Return the number of sequences
seqDegap	Remove gaps by calling seqSR
seqGenotype	Return serotype according to Kramvis 2008
seqLength	Return sequence ID and length
seqRemoveByID	Remove sequence with regex match against sequence ID
seqRemoveByIndex	Remove sequence (indexed from zero)
seqRevComp	Reverse and/or complement sequences
seqSR	Replace "search" with "replace" in sequence
seqSerotype	Return serotype according to Purdy 2007
seqSlide	Place SStr starting at position SPos
status	Return status
translateLoci	Returns amino acids from all three reading frames for each position
unload	Unload file
wt2x2	Return position, wild-type, Fisher's and Chi for significant positions in 2x2 contingency table

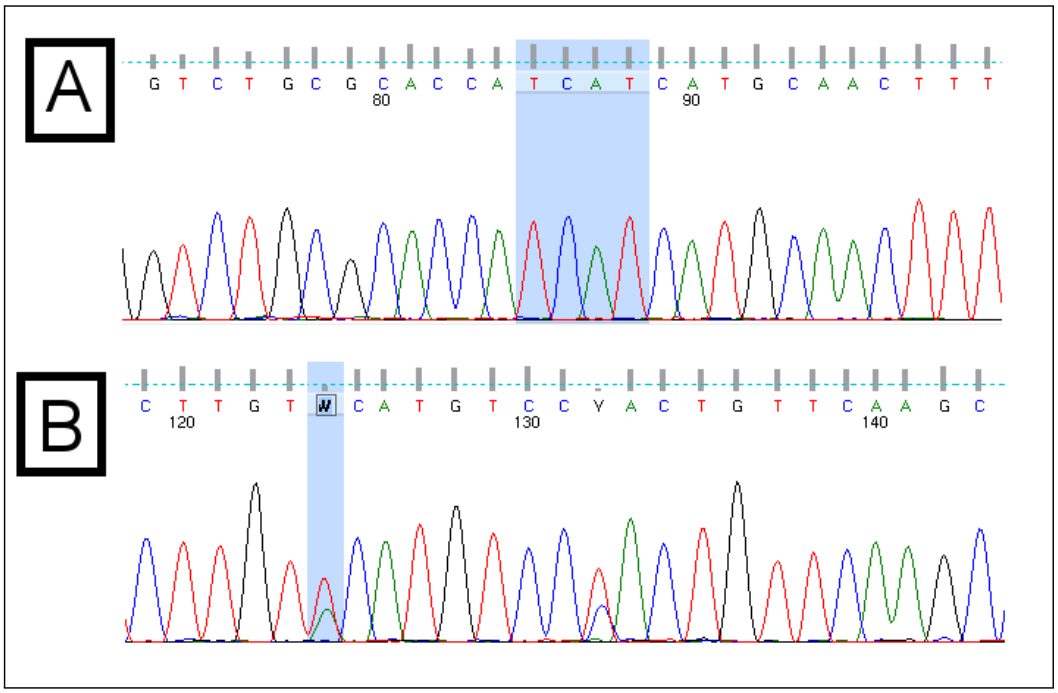


Figure 5.1. *Example chromatograms of the BCP/PC region of Shongwe samples. Panel **A** shows the "Kozak" region (TCAT) of subgenotype A1, followed by the "ATG" pre-core start codon. Panel **B** shows ambiguous (wobble) bases, which result from double peaks. These indicate a mixed population or the presence of quasispecies. Quality scores are indicated by the grey bars above each base call. The quality scores associated with the ambiguous "W" and "Y" bases in panel **B** are 15 and 10, respectively, compared with scores above 50 for each base of the "TCAT" motif in panel **A**.*

5.2 Bioinformatic Tools

Direct DNA sequencing of PCR amplicons is a routine laboratory procedure. The result of this sequencing reaction is a chromatogram file, which is also known as a trace file or an electropheretogram (Figure 5.1). A common file format for chromatograms is the Applied Biosystems format (ABIF[2]), which has a filename extension of "ab1". In a process known as "base calling", which is part of the sequencing service, each nucleotide in the sequence is automatically identified by a software program. A "quality score", implemented originally by the "phred" base-calling program, is assigned to each base call (Ewing *et al.*, 1998; Ewing and Green, 1998). This score, which is logarithmic, indicates the reliability of the base call. A value of 10 indicates a 1-in-10 (90%) probability that the base call is incorrect. A value of 20 indicates a 1-in-100 (99%) probability of an incorrect base call. Generally, quality scores greater or equal to 20 are considered reliable. Due to the nature of the sequencing reactions, the quality scores at the beginning and the end of chromatograms are generally too low to be considered reliable, and are therefore routinely removed before any downstream processing is done.

[2]The ABIF file format specifications are available online at *http://www.appliedbiosystems.com/support/software_community/ABIF_File_Format.pdf*

Table 5.2. List of the online tools developed and the workflow process at which each would be used.

Workflow	Tool Name and Description
Chromatograms	**Quality Score Analyzer** *Plot Chromatogram Quality Scores* **Automatic Contig Generator Tool** *Generate a contig from a forward and reverse chromatogram* **Fragment Merger Tool*** *Merge overlapping fragments into one sequence*
Alignment	**Automatic Alignment Clean-up Tool** *Eliminate "gap-columns" and disambiguate ambiguous bases*
Analysis	**Sequence Fetcher** *Fetch GenBank sequences in a batch* **Mutation Reporter Tool**# *Extract and summarize variation at specified loci* **Babylon** *Extract HBV protein sequences (ORFs)* **Sequence Divergence Calculator** *Determine pairwise divergence between sequences* **Wild-type 2x2** *Calculate 2x2 wild-type/mutant contingency tables*
Serotyping	**HBV Serotyper Tool** *Determine HBV Serotype*
Phylogenetics	**Pipeline: TreeMail** *Generate a phylogenetic tree*
GenBank Submission	**PadSeq** *Place two HBV sequence fragments on a backbone template*

*Discussed in Chapter 7
#Discussed in Chapter 6

A standard workflow in the HVDRP, as shown in our published paper (Makondo *et al.*, 2012) (Appendix C) and Figure 3.3, includes DNA extraction, PCR amplification, direct DNA sequencing by a third party service, viewing and checking of chromatograms, preparation of curated sequences, multiple sequence alignment, sequence analysis, serotyping, genotyping, phylogenetic analysis and preparation of sequences for submission to GenBank. During these processes, challenges were encountered in the analysis of HBV sequence data, and bioinformatic tools were developed to address these (Table 5.2).

5.2.1 Quality Score Analyzer

The quality scores of the base calls in a chromatogram are important. In some cases, the overall quality of an entire chromatogram is so poor that it cannot be used. In other cases, regions of the chromatogram are of poor quality. Submitting a poor quality chromatogram to an online tool will typically result in poor quality results or no results. An online tool was developed to assist users to determine the overall quality of a chromatogram file.

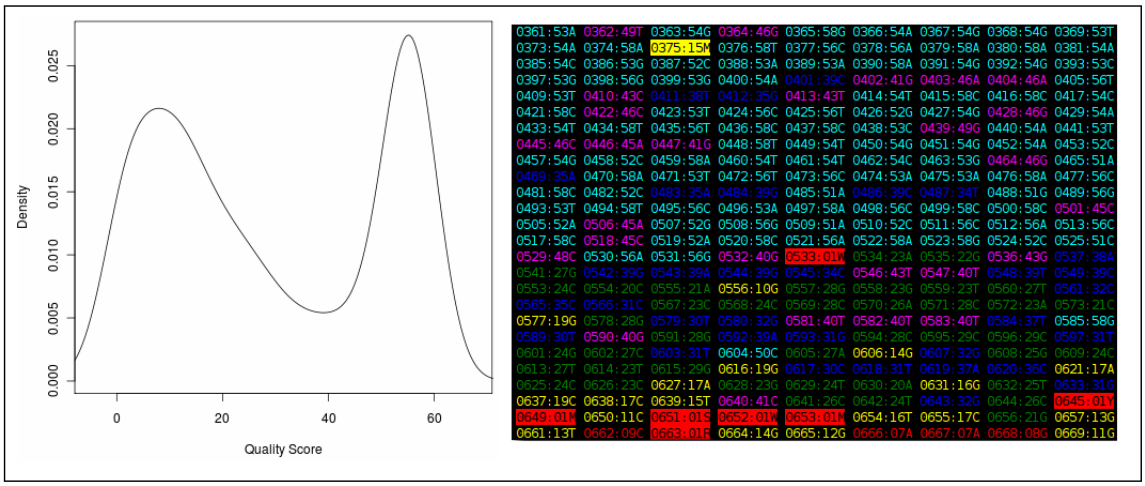


Figure 5.2. *The output of the "Quality Score Analyzer" tool, showing the density plot on the left and a section of the "heat map" on the right. Each entry in this map is in the format "XXXX:YYZ", where "XXXX" is the base position number in the sequence, increasing from "0001" for the first position in the file and "YY" is the quality score from the chromatogram. The "Z" is the base called at the position. The colour of each entry represents the quality score. Values in the range 0 to 9 (considered very poor) are shown in red, between 10 and 19 (poor) in yellow, between 20 and 29 (acceptable) in green, between 30 and 39 (good) in blue, between 40 and 49 (very good) in magenta, between 50 and 59 (excellent) in cyan. Quality scores higher or equal to 60, which are theoretical only, are shown in white. Ambiguous bases are shown in reverse colours (black text on a coloured background).*

The online "Quality Score Analyzer" requires an "ab1" chromatogram file as input and displays a box-and-whisker plot (not shown) and density plot of the quality scores, and a "heat map" (Figure 5.2). The "heat map" provides a visual impression of the overall quality of an entire chromatogram. Areas of interest, such as regions of low quality, can be examined in more detail. No trimming of the input chromatogram is performed.

5.2.2 Automatic Contig Generator Tool (ACGT)

To ensure accurate coverage of a DNA region, a PCR amplicon may be sequenced in both the forward and the reverse direction. A consensus sequence of the forward and reverse reads can be produced by a number of proprietary software programs, or by checking both chromatograms and editing the sequence data manually, which is time-consuming and error-prone.

The "Automatic Contig Generator Tool" (ACGT) generates a consensus sequence (contig) from a forward and reverse chromatogram file. Each chromatogram is trimmed[3] and the sequence designed as "reverse" is reversed and complemented. A pairwise alignment, using the BioPython wrapper for the "needle" alignment algorithm (Needleman and Wunsch, 1970), is then executed.

[3]In this implementation, the front-end uses default trimming parameters, which the user cannot adjust. This is a feature, which could be added in future. The "load" method of the "Sequence" class provides adjustable trimming parameters. Details of these, and the method used to trim chromatograms, are discussed in the paper describing the "Mutation Reporter Tool", reproduced in Chapter 6.

The tool then constructs a consensus sequence by examining the pairs of bases at each position in turn and determining which base to include. The tool generates a consensus from a forward and reverse sequence of the same amplicon region – it is not used to assemble overlapping regions. The quality of each residue and the average quality over a user-defined window of bases are used to assist in determining which residue should be retained. The default value for the window is to include five bases downstream and five bases upstream of the current residue. The tool determines which of the two residues to include by applying the following rules:

```
if the residues are equal then
    use residue 1
else
    if one residue is a gap then
        use the non-gap residue
    else
        if neither residue is ambiguous then
            if one residue has higher quality and above user-defined threshold then
                use this residue
            else
                if one residue has higher quality then
                    use this residue
                else
                    if one residue in higher quality window then
                        use this residue
                    else
                        use residue 2
                    end
                end
            end
        else
            if one residue is ambiguous then
                use non-ambiguous, regardless of quality
            else
                if one ambiguous base has higher quality then
                    use this residue
                else
                    use residue 1
                end
            end
        end
    end
end
```

Algorithm 1: Algorithm for ACGT.

The output of the tool is shown in Figure 5.3. A “consensus level” indicator line is included above the alignment. This line shows different characters depending on the level of consensus: both bases equal, bases mismatched, or one base is a gap. The characters used can be specified

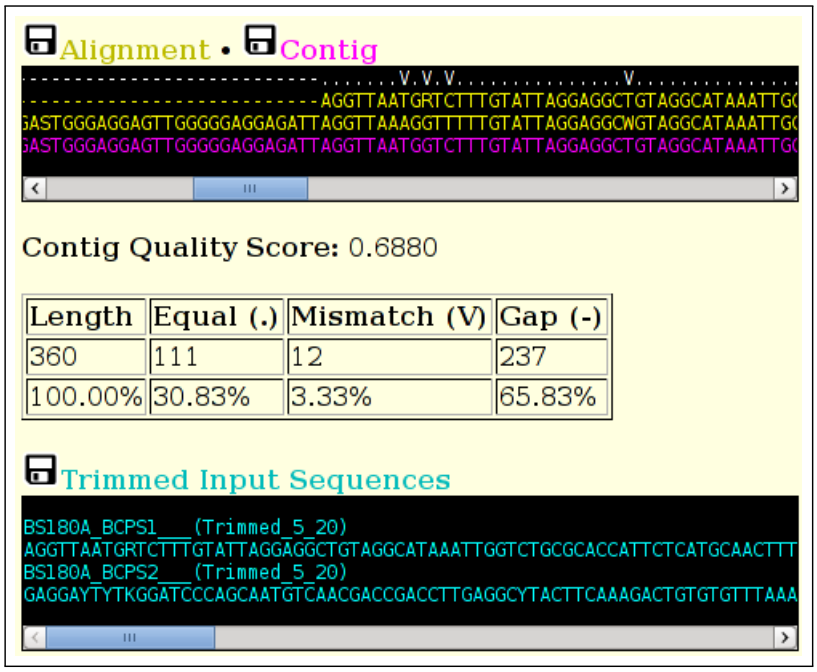


Figure 5.3. *A section of the output of the ACGT tool, showing the indicator line (top) in white, the two trimmed and aligned sequences (yellow) and the computed consensus sequence (magenta). The table displays totals of the various consensus groups. The trimmed sequences are displayed in cyan. Sequences can be downloaded in FASTA format by selecting the small black disk icon.*

on the main input screen. The font size of the sequence output can also be specified. A consensus "quality score" is calculated by adding to a running total a value of 1 when both residues are equal, or a value of 0.5 if one position is a gap. The final total is divided by the length of the consensus to give the consensus quality score. A table below the contig box summarizes the various consensus levels in the alignment. The alignment, contig and trimmed sequences can be downloaded in FASTA format.

5.2.3 Automatic Alignment Cleanup Tool (AACT)

When working with sequence data from several samples, it is routine to prepare a multiple sequence alignment. The data in this alignment will often require careful checking and curating. Whilst automated curation is difficult and may incorrectly manipulate the sequence, there are instances where some automatic correction may be possible. The "Automatic Alignment Clean-Up Tool" (AACT) performs one or both of the following two actions on sequences in an aligned FASTA file.

Eliminate "gap-columns": Columns (positions) in the multiple sequence alignment, which contain at least the specified threshold (expressed as a percentage) of gaps will be eliminated from the sequence. This is intended to remove "misreads", which occur often, but not exclusively,

in homopolymeric regions, and result in one or two bases in a column, consisting otherwise only of gaps.

Disambiguate gap-free "monobase" columns: Ambiguous bases in a column, which consists otherwise of only one base type and is free of gaps, are disambiguated to the base in the column if the ambiguous base represents the base in the column. For example, if a column contains base "A", except for one sequence which contains an "M", this "M" will be disambiguated to "A". If the sequence contained a "Y" instead of an "M", this position would not be disambiguated. All positions are converted to uppercase characters.

The user can specify which of the two components (or both) should be executed on the input file. A report is displayed showing details of how the alignment was changed. A link is provided to download the "cleaned" FASTA file.

Additionally, the user may specify that the start and/or end of the sequences be "blunted", such that all columns containing running gaps from the start of the sequence, and/or to the end of the sequence, are removed. Every sequence in the alignment, therefore, starts, and/or ends with a base, rather than a gap. Effectively, the alignment is trimmed to the length of the shortest sequence. Either or both of these two options can be selected in addition to the functionality discussed above. Alignments, using the "JalView" program[4], before and after processing by the tool are shown in Figure 5.4. The tool also outputs various numerical results, including the number of gap-columns, which were eliminated.

It is recommended that this tool only be used after the alignment has been examined carefully, as the tool may potentially remove mutations from the sequence. However, only ambiguous bases or residues in "gap-columns" would be removed in this way. The suggested approach would be to use the tool to clean an alignment instead of doing such cleaning manually. Submitting uncurated and unchecked data to the tool may result in deletions or changes to the sequence which are undesirable or incorrect.

5.2.4 Sequence Fetcher

This tool is not a replacement for general searching on GenBank. Instead, it fetches a batch of up to 25 sequences, specified as GenBank Accession Numbers, and makes the sequences available for download as a single file in FASTA format. This is useful when fetching standard sets of known sequences, or sequences referenced in the literature. A series of Accession Numbers can be copied

[4] *http://www.jalview.org/*

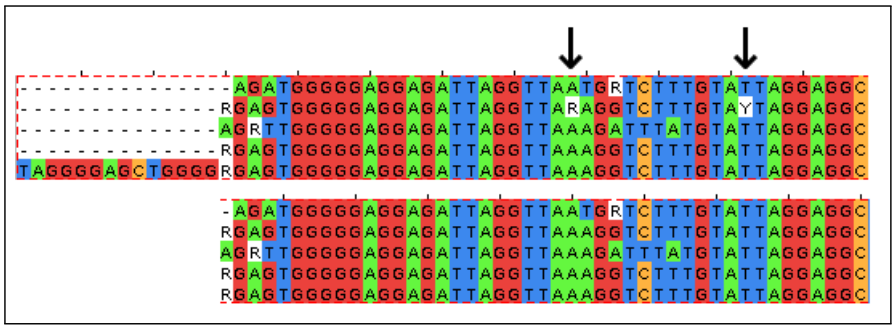


Figure 5.4. *The sequence alignment, which was provided as input to the 'AACT' tool (top) and the alignment produced by the tool (bottom). Columns containing at least 80% gaps (top, left of image) have been removed. Columns containing only one base type, with an ambiguous base, have been disambiguated (arrows). Columns containing a variety of bases are not altered. The alignments were visualized using the "JalView" program.*

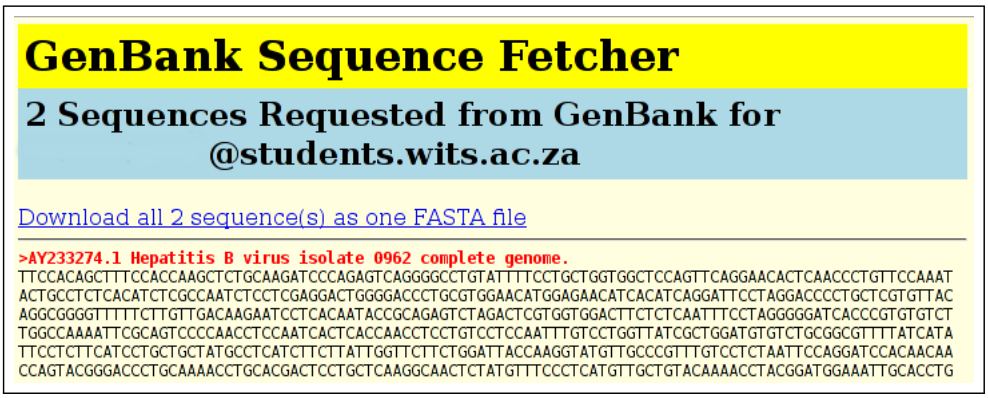


Figure 5.5. *Part of the output page of the "Sequence Fetcher". Sequence data, in FASTA format, can be downloaded via the link provided. The data is also displayed on the output screen for reference.*

from a source document and pasted into the tool, or entered manually. Accession Numbers, which are incorrect or are not found, are reported on the output page for reference. The tool uses the *Entrez.efetch* method of the BioPython framework to query GenBank and fetch the sequence data in FASTA format. The user is required to provide an email address on the input page, as this is required by GenBank. The sequence data are downloaded from GenBank, written to a file and made available to the user via a download link on the output page. The user may specify if the sequence data itself should also be included on the output page. An example output page is shown in Figure 5.5.

5.2.5 Babylon

The HBV genome codes for seven proteins, in four overlapping open reading frames (ORF), as shown in Figure 1.2 and Table 1.1. A summary of the sizes of all HBV proteins is provided in Table 1.2. The "Babylon" tool extracts (splits) HBV sequence data, from a single input file, into multiple files, with each output file containing either nucleotide or translated amino acid data for

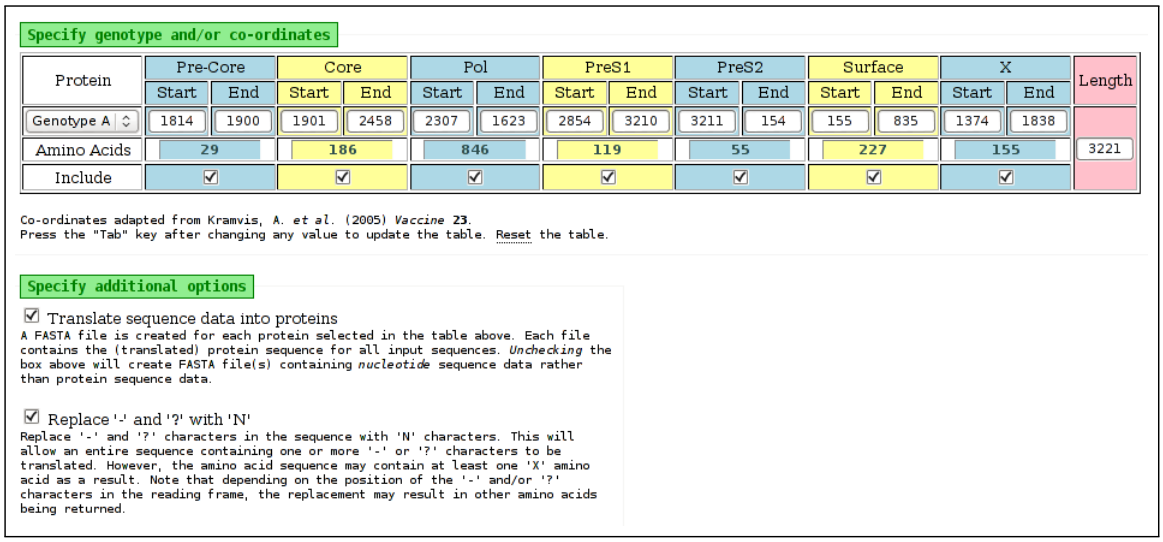


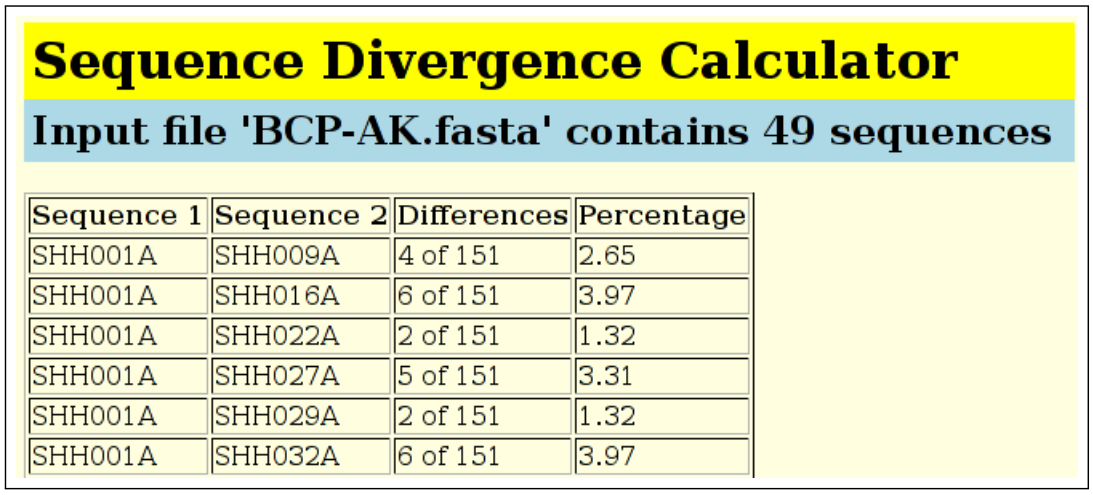
Figure 5.6. *Part of the input page of the "Babylon" Tool. Selecting a genotype from the list on the left will populate the nucleotide positions for each protein with default values from Kramvis* et al. *(2005). However, each of these positions can be edited, as necessary. The number of amino acids for each protein is determined automatically from the nucleotide values. An "Include" field for each protein specifies if it should be included in the output. Amino acid output is obtained by selecting the appropriate check-box on the input page. The "-" and "?" characters, which may be present in input sequence data, will be processed by the tool as an "N" character if the appropriate check-box is selected. It may be possible to translate nucleotides to amino acids when "-" and "?" characters are replaced with "N" characters.*

one HBV protein, as specified on the input page. The tool does not require full-length sequences and the co-ordinates used to extract the protein/s can be specified manually, if required. The tool processes a FASTA file, which should contain aligned nucleotide sequence data from samples belonging to a single HBV (sub)genotype. The input page of the "Babylon" tool is shown in Figure 5.6.

The tool extracts sequence data for each of the selected proteins from the FASTA file, optionally translating the data into amino acids, if specified. A separate output file (in FASTA format) is created for each selected protein, containing the nucleotide or amino acid data for all samples, for that protein only. The files can be downloaded individually or all together in one compressed archive ("ZIP") file.

5.2.6 Sequence Divergence Calculator

The extent of sequence divergence between HBV isolates is used to classify the virus into various genotypes and subgenotypes, as discussed in Section 1.1.5. The "Sequence Divergence Calculator" determines the pairwise sequence divergence between all sequences in an input file. The file, in FASTA format, should contain between two and 100 sequences, which do need to be aligned. Typically, these would be full-genome sequences, but this is not mandatory. All gaps in



Sequence Divergence Calculator

Input file 'BCP-AK.fasta' contains 49 sequences

Sequence 1	Sequence 2	Differences	Percentage
SHH001A	SHH009A	4 of 151	2.65
SHH001A	SHH016A	6 of 151	3.97
SHH001A	SHH022A	2 of 151	1.32
SHH001A	SHH027A	5 of 151	3.31
SHH001A	SHH029A	2 of 151	1.32
SHH001A	SHH032A	6 of 151	3.97

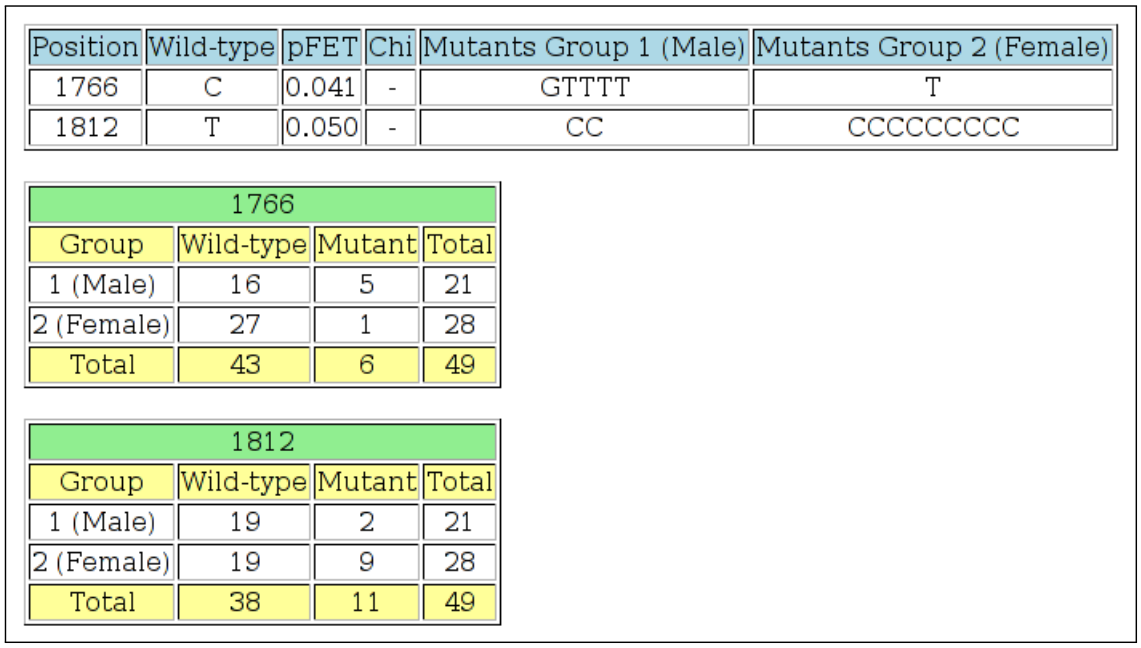
Figure 5.7. *Part of the output page of the "Sequence Divergence Calculator", showing the FASTA ID of each of the pairs of sequences, which have been compared. The "Differences" column shows the number of mismatches found and the total sequence length. This is displayed as a percentage in the "Percentage" column. In this extract, all samples vary by less than 4%.*

the sequence data are removed (the original file will not be changed) and each pair of sequences are aligned in turn. *ClustalW* is used to align the sequences, even though only two sequences are being aligned. The tool determines the divergence between the two sequences by totalling the number of matches and mismatches in each pairwise alignment. The results are displayed in a table on the output page, as shown in Figure 5.7.

5.2.7 Wild-type 2x2

When analyzing a set of HBV sequences, it is often desirable to compare the number of wild-type residues at a locus, with the number of mutant (non-wild-type) residues as a locus. In this case, "wild-type" refers to the residue, which occurs in the majority of the isolates. The "Wild-type 2x2" tool requires a FASTA file of (aligned) nucleotide or amino acid data as input. It calculates wild-type/mutant 2x2 contingency tables for sequences in the two specified groups, for all loci. Detailed output for loci, which are statistically significant at the specified threshold, is provided.

The input sequence data must be allocated into two groups using the number (numerical position) of sequences in the FASTA file. For example, if a file contains 20 sequences, with the first 5 representing "Group 1" and the remaining 15 representing "Group 2", this would be specified as "1-5" and "6-20", without the quotation marks. Groups may also be specified as individual numbers, such as "1,3,6,7,10", or as a mixture of both notations, such as "2,5,6-12". No spaces or other characters are permitted. If one of the groups is omitted entirely (left blank), all sequences, which are not allocated to the other group, will automatically be allocated.



Position	Wild-type	pFET	Chi	Mutants Group 1 (Male)	Mutants Group 2 (Female)
1766	C	0.041	-	GTTTT	T
1812	T	0.050	-	CC	CCCCCCCCC

1766			
Group	Wild-type	Mutant	Total
1 (Male)	16	5	21
2 (Female)	27	1	28
Total	43	6	49

1812			
Group	Wild-type	Mutant	Total
1 (Male)	19	2	21
2 (Female)	19	9	28
Total	38	11	49

Figure 5.8. *The output page of the "Wild-type 2x2 Tool". The first table shows positions at which the p-value is less than or equal to the threshold specified, the wild-type residue, the Fisher's Exact Test p-value ("pFET"), the mutant residues found in Group 1 and the mutant residues found in Group 2. At least one value in the 2x2 contingency table for each row in the table was less than or equal to 5, so Chi-squared tests were not performed in these cases. An offset value was specified on the input page, so the positions on the output page are adjusted accordingly. The next section of output shows the 2x2 contingency tables for each position listed in the first table. This output indicates that there is a statistically significant difference at the 5% level, between HBV isolates from males and females at positions 1766 and 1812 of the BCP region.*

For each position/locus in the sequence data, the majority residue (nucleotide or amino acid) is determined, and this is considered the "wild-type" residue for that locus. The number of mutant residues, at each position, is then determined. A 2x2 contingency table is constructed, for each position, using wild-type and mutant counts, for each of the two groups. If at least one cell in the table contains a value less than or equal to 5, a Fisher's Exact Test is performed; otherwise a Chi-Squared test is performed on the table data. If the resulting p-value is less than or equal to the threshold value specified on the input page, that position is considered as statistically significant, and the details of that position are included on the output page. The value of the optional "offset", as entered on the input page, is added to the position in the output. This can be used to obtain output positions, which correspond exactly with genome co-ordinates. The two groups can be allocated names by entering text into the appropriate box on the input page. Example output is shown in Figure 5.8.

Table 5.3. The decision tree, from Purdy *et al.* (2007), represented as a table. For each serotype, the required amino acids are indicated in the appropriate column. The numbered columns indicate the amino acid position within HBsAg.

Serotype	122	160	127	159	140
adr	K	R			
adw2	K	K	P		
adw3	K	K	T		
adw4	K	K	I or L		
ayr	R	R			
ayw3	R	K	T		
ayw4	R	K	I or L		
ayw1	R	K	P	A	
ayw2	R	K	P	Not A	Not S
ayw4	R	K	P	Not A	S

5.2.8 HBV Serotyper Tool

In addition to genotypic classification, HBV samples can be classified into one of nine serological subtypes (serotypes) (Magnius and Norder, 1995). This classification is determined by the amino acids present at either three or five known positions within the HBV surface antigen (HBsAg) (Wands *et al.*, 1984; Mimms *et al.*, 1990; Swenson *et al.*, 1991; Purdy *et al.*, 2007). HBV serotype is loosely correlated with genotype (Kramvis *et al.*, 2005). A published decision tree (tabulated in Table 5.3) summarizes the interpretation of the amino acid positions to determine the HBV serotype (Purdy *et al.*, 2007).

The "HBV Serotyper Tool", based on this decision tree, requires a FASTA file of nucleotide data as input. This file should contain one or more aligned sequences, which must include the start of the HBsAg sequence and the amino acid positions required for serotyping. The input page of the tool presents a default nucleotide motif for the start of the S gene, which may be edited by the user, if necessary. The serotype, amino acid motif and nucleotide sequence are output (Figure 5.9).

5.2.9 Pipeline: TreeMail

Phylogenetic analyses undertaken by members of our research group typically involves several programs from the "Phylip" suite (Felsenstein, 1989). These command-line tools are interactive and menu-driven, requiring the user to undertake several steps to complete an analyses. An output data file from one component of the suite must be renamed manually to prepare it for use as an input file for another component of the suite. This process is repetitive and time-consuming, especially when running several analyses.

Sequence ID	Serotype	Motif1	Motif3	Motif Sequence
SHH011A	ayw1	RKPAT	ArgLysProAlaThr	AGGAAACCTGCAACA
SHH014A	adw2	KKPAT	LysLysProAlaThr	AAAAAACCTGCAACA
SHH022A	adw2	KKPAT	LysLysProAlaThr	AAAAAACCTGCAACA
SHH032A	adw2	KKPAT	LysLysProAlaThr	AAAAAACCTGCAACA
SHH037A	ayw1	RKPAT	ArgLysProAlaThr	AGAAAACCTGCAACA
SHH039A	adr	KRPAT	LysArgProAlaThr	AAAAGACCTGCAACA
SHH043A	adw2	KKPRT	LysLysProArgThr	AAAAAACCTCGCACA

Figure 5.9. *Output of the HBV Serotyping Tool, showing the sequence ID, the serotype, the amino acid motif for all five positions using both the one-letter amino acid abbreviations ("Motif1") and the three-letter abbreviations ("Motif3"), and the nucleotides present at all five amino acid positions. All five amino acids are not required in order to deduce some serotypes, but all five positions are included for all samples for reference.*

Table 5.4. Programs from the "Phylip" suite, which are used by the "Pipeline: TreeMail" tool.

Program	Description
consense	Computes consensus trees using the majority-rule method
dnadist	Computes distances between samples from sequence data
neighbor	Computes an unrooted by neighbor-joining or UPGMA
seqboot	Generates multiple data sets by bootstrap resampling

The "Pipeline: TreeMail" tool runs the "Phylip" *dnadist* and *neighbor* programs on the input file, with parameters automatically set as required by the research group. The input page of the tool is shown in Figure 5.10. The input file must be in Phylip (".phy") format. The Kimura and lower-triangular settings are specified for *dnadist*, and the lower-triangular setting is specified for *neighbor*. The pipeline tool emails the resulting tree file (".tre") to the email address provided. The tool reports progress once the input file has been uploaded. A description of each of the "Phylip" programs used by the "Pipeline: TreeMail" tool is provided in Table 5.4.

If the "Bootstrap" mode is specified on the input page, the tool runs the Phylip *seqboot* program before *dnadist*, and *consense* after *neighbor*, as described in the online documentation for *seqboot*[5]. When in "Bootstrap" mode, the pipeline tool will create 1000 data sets with *seqboot* and will email the final consensus tree to the email address provided.

The tool also emails the final Phylip "outfile" from either *neighbor* (normal mode) or *consense* (bootstrap mode) as a second attachment, called "result.txt". When running in bootstrap mode, the actual bootstrap values will appear on the tree in this file ("result.txt") and are present in the consensus tree file. These are shown as values out of 1000, not percentages.

Running in bootstrap mode may increase the time required for the tool to complete. A bootstrap test run of 41 sequences of approximately 1000 nucleotides each took 15 minutes to process

[5] *http://evolution.genetics.washington.edu/phylip/doc/seqboot.html*

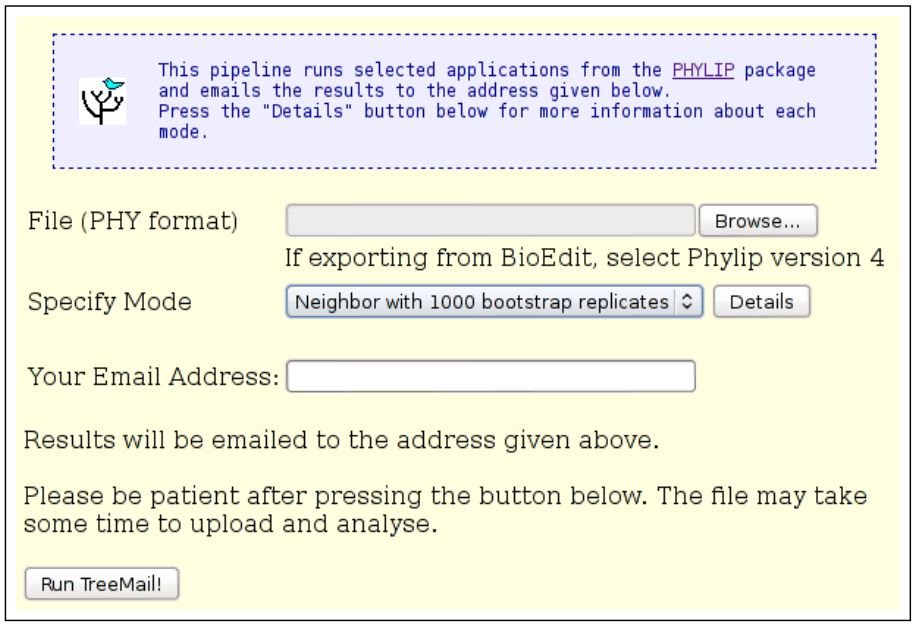


Figure 5.10. *The "Pipeline: TreeMail" input page. An input file in "Phylip" format is required. The user may select either "Neighbor" mode, which does not include bootstrap replicates, or "Bootstrap" mode, which includes 1000 bootstrap replicates. Results of the analysis, including the tree file, are emailed to the address provided.*

and email. The web-page will time-out if the analysis takes longer than 60 minutes.

5.2.10 PadSeq

Subgenomic fragments, particularly those less than 200 nucleotides in length, may not be accepted by GenBank. Such fragments can be submitted, in FASTA format, to the "PadSeq Tool", which places each of the two sequence fragments at the specified co-ordinates on a sequence backbone (template/scaffold) of the specified length. The tool can be used to generate artificial "full-length" sequences from two fragments, such as BCP/PC and S region fragments. This may be useful when attempting to genotype samples automatically.

The tool requires two input files (Figure 5.11). One file should contain all the sequence data for the first fragment, with each sequence in the file starting at the same co-ordinate. For example, this would be a file containing curated/checked BCP/PC sequences all starting at 1750. The second file should contain data for the second fragment, also all starting at the same position for the second fragment. For example, this would be a file containing curated/checked S region sequences all starting at 2854. The sequences do not all have to be the same length, but all the sequences in one file must start at the same position. The order of the sequence data in both files must be the same. For example, if the BCP/PC file contains sequence data for sample 1 then sample 2 then sample 3, the S region file must contain sequence data in the same order. The

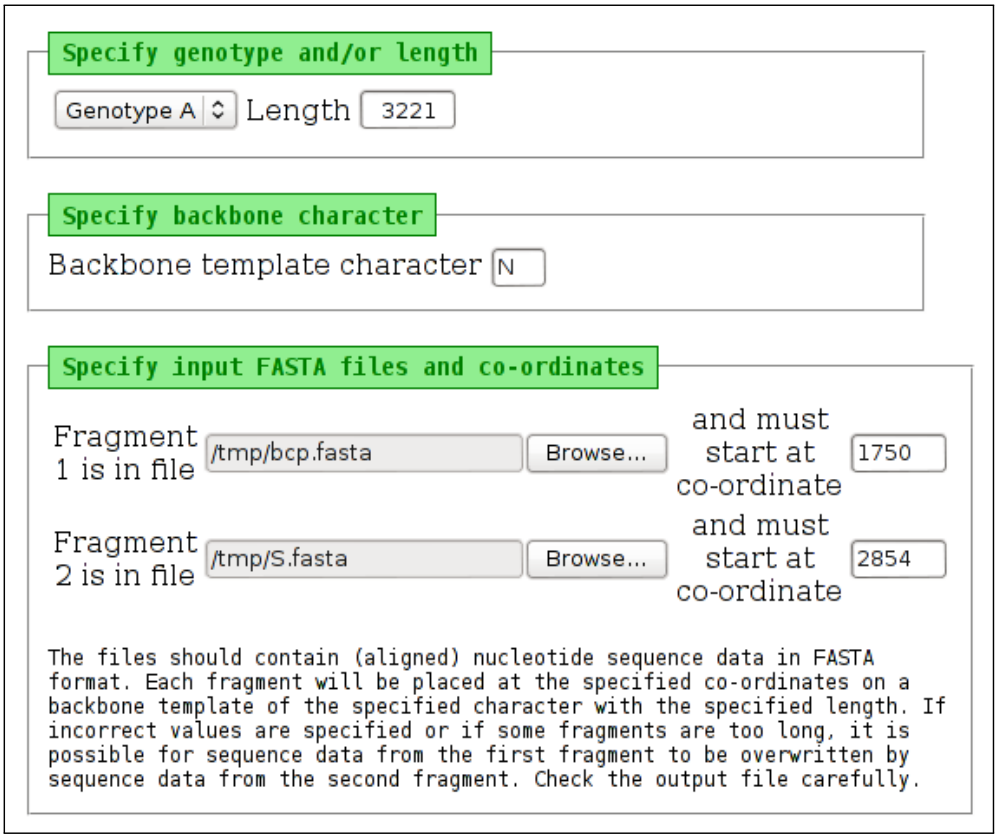


Figure 5.11. *The input page of the "PadSeq Tool". Selecting a genotype from the list will place a default length into the "Length" field. This value can be edited, if necessary. The backbone character can be changed from the default "N". Each of the two input files must be specified, along with the starting position (co-ordinate) at which each fragment should be placed. The tool will "wrap" sequence data, which extends beyond the length specified, as may be required when processing HBV sequence data.*

number of sequences in both files must be the same. Once the tool has completed, an output page is displayed, which contains a link to download one "padded" file of all sequence data in FASTA format. Example output is shown in Figure 5.12.

5.3 Discussion

The bioinformatic infrastructure, which was developed and updated during this study, was used to analyze clinical and molecular sequence data from the Shongwe cohort. This infrastructure includes the PostgreSQL database (Section 3.3), the R script for analysis of clinical data (Section 3.3.3.5), the CLIMB framework (Section 5.1), the online tools described above (Section 5.2) and the tools described in Chapters 6 and 7.

The bioinformatic tools and database developed in this study were focused on addressing the challenges faced by biologists working on HBV in South Africa and Africa. The tools were developed by a researcher with both biological and programming knowledge and experience, but with the aim that the tools be usable by biologists without technological or computer programming expertise. As the tools were used and tested by members of the research group, feedback was received. This was very valuable, as the experiences of researchers running the tools on data can be used to improve the programs.

The suite of tools produced is integrated tightly into the workflow of the wet laboratory. However, each tool is a standalone entity – it is not necessary to use all tools in the suite. As each tool has a distinct purpose and can be used individually, the learning curve for new users is flattened. It is not necessary to learn how to use an entire new software suite – the user can simply learn how to use the specific tool he or she requires.

Molecular characterization of HBV isolates involves analysis of both the BCP/PC region and the S region, as discussed in the two published papers (Chapter 4 and Appendix C). Various tools were used depending on the region being analyzed. Specifically, the "Mutation Reporter Tool" was used extensively on BCP/PC sequence data, and was developed and improved during the course of the study as the need for additional features became apparent. The "Fragment Merger Tool", "Serotyper Tool" and "Pipeline: TreeMail" were used on S region sequence data. Final, curated BCP/PC and S region pairs of sequences were prepared for submission to GenBank via the "PadSeq" tool.

Although these tools were developed while working in an HBV research laboratory, they were

```
>SHH022
TTCCACAGCTTTCCACCAAGCTCTGCAAGATCCCAGAGTCAGGGGCCTGTATTTTCCTGCTGGTGGCTCCAGTTCAGGAACACTCAACCCTGTTCC
AAATACTGCCTCTCACATCTCGTCAATCTCCTCGAGGACTGGGGACCCTGCGTTGAACATGGAGAACATCACATCAGGATTCCTAGGACCCCTGCT
CGTGTTACAGGCGGGGTTTTTCTTGTTGACAAGAATCCTCACAATACCGCAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGGGGGCTC
ACCCGTGTGTCTTGGCCAAAATTCGCAGTCCCCAACCTCCAATCACTCACCAACCTCCTGTCCTCCAATTTGTCCTGGTTATCGCTGGATGTGTCT
GCGGCGTTTTATCATATTCCTCTTCATCCTGCTGCTATGCCTCATCTTCTTGTTGGTTCTTCTGGATTACCAAGGTATGTTGCCCGTTTGTCCTCT
AATTCCAGGATCCACAACAACCAGTACGGGACCCTGCAAAACCTGCACGACTCCTGCTCAAGGCAACTCTATGTTTCCCTCATGTTGCTGTACAAA
ACCTACGGATGGAAATTGCACCTGTATTCCCATCCCATCATCTTGGGCTTTCGCAAAATACCTATGGGAGTGGGCCTCAGTCCGTTTCTCTTGGCT
CAGTTTACTAGTGCCATTTGTTCAGTGGTTCGTAGGGCTTTCCCCCACTGTTTGGCTTTCAGTTATATGGATGATGTGGTATTGGGGGCCAAGTCT
GTACAACATCTTGAGTCCCTTTATACCGCTGTTACCAATTTTCTTTTGTCTTTGGGTATACATTTAANNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNAGATTAGGTTAAAGGTCTTTGTATTAGGAGGCTGTAGGCATAAATTGGTCTGCGCACCATCATCATGCAACTTTT
TCACCTCTGCCTAATCATCTCTTGTACATGTCCCACTGTTCAAGCCTCCAAGCTGTGCCTTGGATGGCTTTGGGGCNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATGGGAGGTTGGTCAGCCAAACCTCGC
AAAGGCATGGGGACGAATCTTTCTGTTCCCAACCCTCTGGGATTCTTTCCCGATCATCAGTTGGACCCTGCATTCGGAGCCAATTCAAACAATCCA
GACTGGGACTTCAACCCCATCAAGGACCACTGGCCACAAGCCAACCAGGTAGGAGTGGGAGCATTCGGCCCAGGGTTCACTCCCCCACACGGAGGT
GTTTTGGGGTGGAGCCCTCAGGCTCAGGGCATATTGGCTACAGTGCCAGCAGTTCCTCCCACTGCCTCCACCAATCGGCAGTCAGGAAGGCAGCCT
ACTCCCATCTCTCCACCTCTAAGAGACACTCATCCTCAGGCCATGCAGTGGAA
```

Figure 5.12. *Example output of the "PadSeq" tool. The two input fragments have been placed at the specified location within a backbone template.*

designed to be as general as possible. This means that data from any organism or source can be submitted to the tools. No limit on the length of data is imposed by any of the tools, with the exception that the final merged sequence, which the "Fragment Merger Tool" produces, should be less than 100,000 bases. The "Serotyper", "Babylon" and "PadSeq" tools are HBV-specific.

The tools developed in the present study are, or will be made, available, online at no cost. Additionally, papers describing two of the tools have been published in open-access journals. Both of these approaches are an advantage in resource-limited settings in Africa. The tools can be accessed from any computer with an Internet connection. License agreements for commercial software programs are typically both expensive and restrictive. Furthermore, the tools can be used by researchers without specific technical expertise. To date, the "Mutation Reporter Tool" has been used and cited in the following papers: Yousif and Kramvis (2012) and Gopalakrishnan *et al.* (2013).

5.4 Scaling and Future Work

Presently, the development versions of the tools, hosted on the server as described above, are used internally by members of the research group. As such, the load on the server has been light. Although the "Mutation Reporter Tool" and "Fragment Merger Tool" have been published (Chapters 6 and 7), it is not anticipated that the load on the server will become unmanageable. As the server is monitored regularly, any prolonged increase in load will be detected. If this occurs, it may be necessary to consider alternate approaches regarding the server infrastructure, such as load-balancing, converting the tools to use WSGI (Web Server Gateway Interface) and/or implementing other frameworks to manage queues and workloads.

Future development work includes completing an implementation of an online HBV genotyper, based on Kramvis and Kew (2005), which is currently in beta, detecting recombinant sequences and building or modifying tools to process amplicon resequencing data from the ultra-deep pyrosequencing platform. Development on the database includes building analyses using both clinical and molecular data.

Availability Note The source code will be available on request, once the tools have been published and the PhD degree has been awarded.

Chapter 6

Paper II

Mutation Reporter Tool: An online tool to interrogate loci of interest, with its utility demonstrated using hepatitis B virus

Bell, T. G. and Kramvis, A. (2013)

Virology Journal **10:** 62 (10.1186/1743-422X-10-62)

References References for literature cited in the following paper appear within the references section of the paper itself, and not necessarily in the references section of this thesis.

Journal *Virology Journal* is a peer-reviewed, ISI-accredited, open-access journal, with an impact factor of 2.34.

Website The "Mutation Reporter Tool" is available online at *http://hvdr.bioinf.wits.ac.za/tools/*

Bell and Kramvis *Virology Journal* 2013, **10**:62
http://www.virologyj.com/content/10/1/62

VIROLOGY JOURNAL

METHODOLOGY Open Access

Mutation Reporter Tool: An online tool to interrogate loci of interest, with its utility demonstrated using hepatitis B virus

Trevor G Bell and Anna Kramvis*

Abstract

Background: An online tool, which extracts and summarises nucleotide or amino acid sequence data at specified loci of interest, was developed and tested using the basic core promoter/precore (BCP/PC) region of the hepatitis B virus (HBV). The tool is aimed at researchers without specialist computer skills.

Methods: The tool consists of a web-based front-end, with a CGI script, which runs Python code to generate an output web-page. The Python code searches the input sequence data for a specified anchor motif, after which it generates summary tables and graphs of residue and motif distributions.

Results: After the user provides an input file in FASTA format containing aligned sequence data (nucleotides or amino acids) and specifies an anchor motif at a known coordinate, the tool summarizes the nucleotides or amino acids at the specified loci, their frequency and analyzes motif patterns of the loci.The tool can output a graph that displays the frequency of mutations relative to a reference sequence. The tool was used to analyze the BCP/PC region of HBV belonging to subgenotypes A1, A2 and subgenotype D and to serotype HBV. The "Discovery Mode" ignores conserved loci and assists in identifying potential loci of interest.

Conclusions: Although HBV was used to demonstrate the utility of the *Mutation Reporter Tool*, the tool has wide application as it is genome-agnostic: nucleotide or amino acid sequence data from any organism can be processed. Rapid characterisation of many sequences can be achieved easily when the loci of interest are known. The tool is available online, without charge, at http://hvdr.bioinf.wits.ac.za/tools.

Keywords: Hepatitis B virus, Mutations, Sequence analysis

Background

Example organism: Hepatitis B virus

Hepatitis B virus (HBV) is one of the most important blood-borne pathogens and is endemic to the sub-Saharan African and southeast Asian regions. Worldwide, around 2 billion people have been exposed to the virus, 240 million are chronically infected, and more than half a million die annually from infection-related liver diseases [1]. At approximately 3,200 nucleotides, the HBV genome is small and has been well-characterized. The genome codes for seven different proteins from four overlapping reading frames (ORFs). To date, nine different genotypes of HBV have been identified: A to D [2,3], E and F [3-5], G [6], H [7], I [8-12] and genotype J has recently been proposed [13]. Subgenotypes have been recogonized in genotypes A to D, F and I, and these are named numerically [14]. Disease progression, clinical manifestation of illness and treatment response differ between these genotypes [15-17].

Mutations (single nucleotide polymorphisms, or "SNPs") in the genetic sequence of HBV are common, as the virus polymerase lacks proof-reading ability [18]. Patterns of mutations at various known loci have been used to characterize the virus [14]. Certain patterns are characterstic of a particular genotype, or subgenotype [14,19], and can therefore be used to identify, or "genotype", a given sample. Patterns at other loci are characteristic of known drug-resistant mutants [20], or indicate other

*Correspondence: Anna.Kramvis@wits.ac.za
Hepatitis B Virus Diversity Research Programme, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, 7 York Road, Parktown, Johannesburg, 2193, South Africa

important characteristics, such as down-regulation of, for example, hepatitis B e antigen (HBeAg) [21]. Therefore, the examination of nucleotides at one or more known loci, either together or individually, is routinely used to characterise HBV sequences. Identification of mutations of interest is not always straighforward, however, for a number of reasons. Firstly, the HBV genome is circular (numbered from position "1" at the Eco*R*1 restriction site), but sequence data is linear, and position "1" lies within a region of interest, which is typically sequenced both downstream and upstream of this position. Secondly, HBV genotypes are not the same length, ranging from 3182, for genotype D, to 3248 nucleotides for genotype G. Thirdly, insertions and/or deletions of varying length may be present in some isolates, or, fourthly, isolates may be recombinants of two or more known genotypes. Thus, automated analysis of the genome is complex and sequence data should be carefully curated.

Basic core promoter/precore (BCP/PC) mutations

HBeAg is a non-particulate secretory protein expressed by HBV. The pre-core/core open reading frame encodes for HBeAg [22]. The basic core promoter (BCP), which covers the distal X region and the proximal pre-core (PC) region, directs transcription of PC mRNA, which is translated into the pre-core/core fusion protein that is the precursor of HBeAg. This protein has a signal peptide at its amino end that targets it to the endoplasmic reticulum, where it is post-translationally modified by truncation at a fixed site on its amino end and at variable sites on its carboxyl end [21]. Various mutations within the BCP and PC regions affect the expression of HBeAg at the transcriptional, translational and post-translational levels [23,24]. The BCP A1762T/G1764A mutations affect transcription of the PC mRNA [25]. Mutations that affect HBeAg expression at the translational level include Kozak sequence (1809-1812) mutations and the G1896A stop codon mutation. Substitutions at 1809-1812 are found mainly in subgenotype A1. HBeAg expression is impaired by Kozak mutations by a leaky scanning mechanism [26]. The classical G1896A transition leads to a tryptophan to stop codon mutation, which results in the truncation of HBeAg precursor and abrogation of HBeAg expression [27]. The emergence of G1896A leads to the stabilization of the encapsidation signal (ε) on the pregenomic RNA in genotypes with 1858T, but is rarely found in strains which have 1858C [28]. At the post-translational level, the G1862T mutation, characteteristic of subgenotype A1, introduces a phenylalanine, which interferes with signal peptide cleavage and maturation of HBeAg [29]. Clinically, HBeAg is used as an index of viral replication, infectivity, severity of disease and response to antiviral treatment. Mutations that affect HBeAg expression are clinically relevant [17] and thus analysis of their distribution is important. We demonstrate the utility of the *Mutation Reporter Tool* using the BCP/PC mutations as an example.

Loci of interest and patterns of residues

Analysis of loci of interest, which may be dispersed across the genome, and the resulting patterns of these loci, has traditionally been a manual, interactive process, which is time-consuming and error-prone. A new online tool, the *Mutation Reporter Tool*, has been developed to rapidly and easily display loci of interest and patterns of residues for any sequence data (nucleotides or amino acids) submitted by the user. Feedback from members of the *Hepatitis Virus Diversity Research Programme*, who used development versions of the tool extensively to analyze HBV sequences, was incorporated into the present version.

Results and Discussion

The *Mutation Reporter Tool* is one component of a larger project currently in progress, and makes use of a common (shared) Python computer language module, consisting of a "Sequence" class, which contains several methods. The tool consists of a web-based front-end, with which the user interacts, and a CGI script, which runs the Python code and generates the output web-page. The tool has been developed to assist scientists with data analysis and does not require any specialist computer skills or installation. A detailed online tutorial is available. HBV sequence data will be used to demonstrate the utility of the tool.

Usage

A section of the input interface of the tool is shown in Figure 1. An input file in FASTA format containing aligned sequence data (nucleotides or amino acids) is specified. The loci of interest are specified relative to a known "anchor motif" at a known genomic co-ordinate. The location of the first occurrence of the whole anchor motif in the first sequence in the file is used as the position from which the *specified* loci are determined. The loci of interest are specified as comma-separated integers without spaces (for example: 1762,1764) and/or dash-separated ranges of integers without spaces (for example: 1809-1812).

For example, the basic core promoter/precore (BCP/PC) region of HBV is routinely sequenced. Within this sequence fragment of approximately 500 nucleotides, the highly-conserved motif "AGATTA" is found at co-ordinate 1750. A file containing aligned BCP/PC sequence data is submitted to the tool with "AGATTA" specified as the anchor motif and "1750" as the anchor position. Loci of interest *downsteam* of 1750 are then specified by their absolute (and known) co-ordinates in the genome. Loci, which are known to affect the expression of

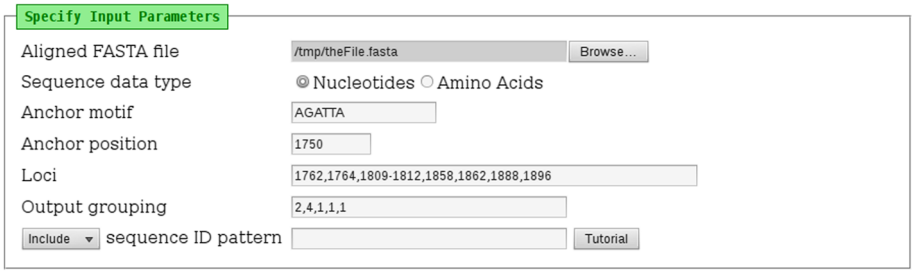


Figure 1 The input interface of the Mutation Reporter Tool. A FASTA input file of sequence data is specified. These sequence data (nucleotides or amino acids) may be genomic or subgenomic fragments. An "anchor motif", which is common to all sequences, is provided. The "anchor position" specifies the genomic co-ordinate of the start of the "anchor motif", such that the downstream loci of interest can be specified accordingly. In this figure, nine loci of interest have been specified. These nine loci will be grouped into columns according to the "output grouping" field: a column containing the first two loci, followed by a column containing the next four loci, followed by a column containing the final locus. Specific sequences in the input file can be included or excluded by entering a regular expression into the appropriate field. This field is blank in the figure, which indicates that all sequences will be included.

HBeAg, are found at 1762, 1764 and 1896. In subgenotype A1, the "Kozak" sequence, which modulates the translation of HBeAg, is located at position "1809-1812". All these loci are therefore entered into the "Loci" field as "1762,1764,1809-1812,1896". Only these loci are extracted from each sequence in the input file and included on the output page.

The loci of interest can optionally be grouped into columns according to the "Output grouping" field. The field accepts a comma-separated list of integers, which indicate the number of loci to group into one output column. If no output grouping is specified, the tool will output all loci into one output column. Using the previous example of loci and an output grouping of "2,4,1", the output would place the nucleotides at 1762 and 1764 together into one column (specified by the output grouping of "2"), the Kozak sequence at 1809-1812 into another column, and nucleotide at 1896 into a third column.

If only some sequences from the input file are to be processed, a "regular expression" can be entered next to the "Include/Exclude" drop-down box. This will then either include (or exclude) sequences for which the FASTA ID matches (or does not match) the regular expression provided. A tutorial describing regular expressions is linked from the input page for reference. Subsets of sequence data stored in one FASTA file can therefore easily be analysed separately, without having to create additional files. FASTA IDs in the output are truncated to the number of characters specified on the input page. If "Output percentages" is not selected, absolute counts are given as output, instead of percentages.

Output

The tool produces several tables of output. The first (Figure 2) shows the residue at each of the specified loci for each sequence in the input file. The loci are grouped into columns as specified by the output grouping. The next output table shows the distribution of each residue at each locus (as a raw count or a percentage, Figure 3). Figure 4 shows part of the next table of output, which reports the number of occurrences (as a percentage, sorted in ascending order) of each unique motif pattern, as created by placing all of the specified loci next to each other in the order specified. This motif pattern can be used to classify sequences into groups and to identify the motif, which occurs most frequently. A graph of the motif distribution is displayed below the table. The raw data used to create this graph can be downloaded as a CSV file. A link below the final output table (Figure 4) opens a new page, which shows the FASTA ID associated with each of the motif patterns. This output is grouped by motif pattern for reference.

Example Usage: HBV serotypes

In addition to genotypic classification, HBV strains can be classified into one of nine serological subtypes (serotypes) [30]. This classification is determined by the amino acids

ID	1762 1764	1809 1810 1811 1812	1858	1862	1888	1896
SHH001A	AG	TCAT	C	G	G	G
SHH009A	TA	TCAT	C	G	A	G
SHH016A	AG	GCAC	T	G	G	A
SHH022A	AG	TCAT	C	G	A	G
SHH027A	AG	TCAC	T	G	G	A
SHH029A	AG	TCAT	C	G	A	G

Figure 2 Loci distribution. The nucleotides at each of the specified loci within the BCP/PC region for each sequence in the input file are shown in this table. The loci are grouped into columns as specified.

Bell and Kramvis *Virology Journal* 2013, **10**:62
http://www.virologyj.com/content/10/1/62

	1762	1764	1809	1810	1811	1812	1858	1862	1888	1896
A	083.67	018.37	004.08	002.04	085.71				059.18	010.20
C				095.92	008.16	022.45	089.80		002.04	
G		081.63	016.33					085.71	032.65	089.80
T	016.33		079.59	002.04	006.12	077.55	010.20	014.29	006.12	

Figure 3 Residue distribution summary. The distribution of residues at each of the loci is shown, either as a percentage or as a raw count, depending on the parameter specified on the input interface (percentages are shown in this figure). Ambiguous (degenerate) bases and gaps, if present in the sequence, would also be included in the table.

present at either three or five known positions within the HBV surface antigen (HBsAg) [31-34]. HBV serotype is loosely correlated with genotype [19]. A published decision tree summarizes the interpretation of the amino acid positions to detemine the HBV serotype [34]. Translated (amino acid) sequence data covering the HBV surface gene can be submitted to the *Mutation Reporter Tool* with the five amino acid positions of interest (122, 160, 127, 159 and 140) specified. An output grouping of "1,1,1,1,1" should be specified to place each amino acid into its own column for easier reading. The amino acids at each position for each sequence can then be examined together with the decision tree to determine the HBV serotype.

Mutation distribution graph

A "Reference motif" can be specified on the input page. This motif should include the reference ("wild-type") residue for each of the specified loci, in order. For example, if loci "1809-1812,1896" are specified and the input file consists of HBV subgenotype A1 sequences, the reference motif would be "TCATG". If a reference motif is specified, the output page will include a graph, which indicates the percentage of non-reference (mutant) residues present at each locus. If the input sequence does not contain the ambiguous base "N", then specifying a reference motif consisting only of "N" characters will result in the tool including all of the residues at each locus, as all residues will not match the reference residue of "N" at each locus. Additional parameters on the input page are used to customize the graph appearance. These include specifying the graph dimensions (in pixels). Loci at which all sequences contain only the reference residue can be suppressed by selecting the appropriate control on the input page. Selecting the "Y-Axis scaled to 100%" control will ensure that the Y-axis of the graph extends from 0% to 100%. This is useful when preparing several graphs which are to be compared with each other. If this control is not selected, the Y-axis will be scaled according to the input data. The raw data used to construct the graph can be downloaded in CSV format from a link on the output page.

Example analysis: Subgenotypes A1 and A2, and genotype D

A comparison of the BCP/PC region of subgenotypes A1, A2 and genotype D is depicted in Figure 5. The nucleotide at position 1858 can differentiate between genotypes A and D. Genotype A has 1858C (Figure 5A and B), whereas genotype D has 1858T (Figure 5C). The presence of 1858C precludes the development of the G1896A mutation because this would destabilize ε and compromise the replication of the virus [28]. On the other hand, in genotype D, with 1858T, the G1896A mutation would stabilize ε because of the formation of a Watson-Crick base pair between 1858T and 1896A. Thus,

Label	Motif	Distribution
L0001	AGTCATCGGG	044.90
L0002	AGTCATCTGG	010.20
L0003	AGGCACCGGG	006.12
L0004	TATCATCGGG	006.12
L0005	AGGCACTGGA	004.08
L0006	TATCTTCGGG	004.08
L0007	AATCCTCGGG	002.04
L0008	AGACACCGGG	002.04
L0009	AGACATCGGG	002.04
L0010	AGGCACCGGA	002.04
L0011	AGGCACTTGG	002.04
L0012	AGTACTTGGA	002.04
L0013	AGTCACTGGA	002.04
L0014	AGTCCTCGGG	002.04
L0015	AGTCTTCGGG	002.04
L0016	TAGCACCTGG	002.04
L0017	TATCACCGGG	002.04
L0018	TATTCTCGGG	002.04

Figure 4 Motif distribution. This table shows the distribution, as a percentage, of each unique motif created by placing all the specified loci next to each other in order.

Bell and Kramvis *Virology Journal* 2013, **10**:62
http://www.virologyj.com/content/10/1/62

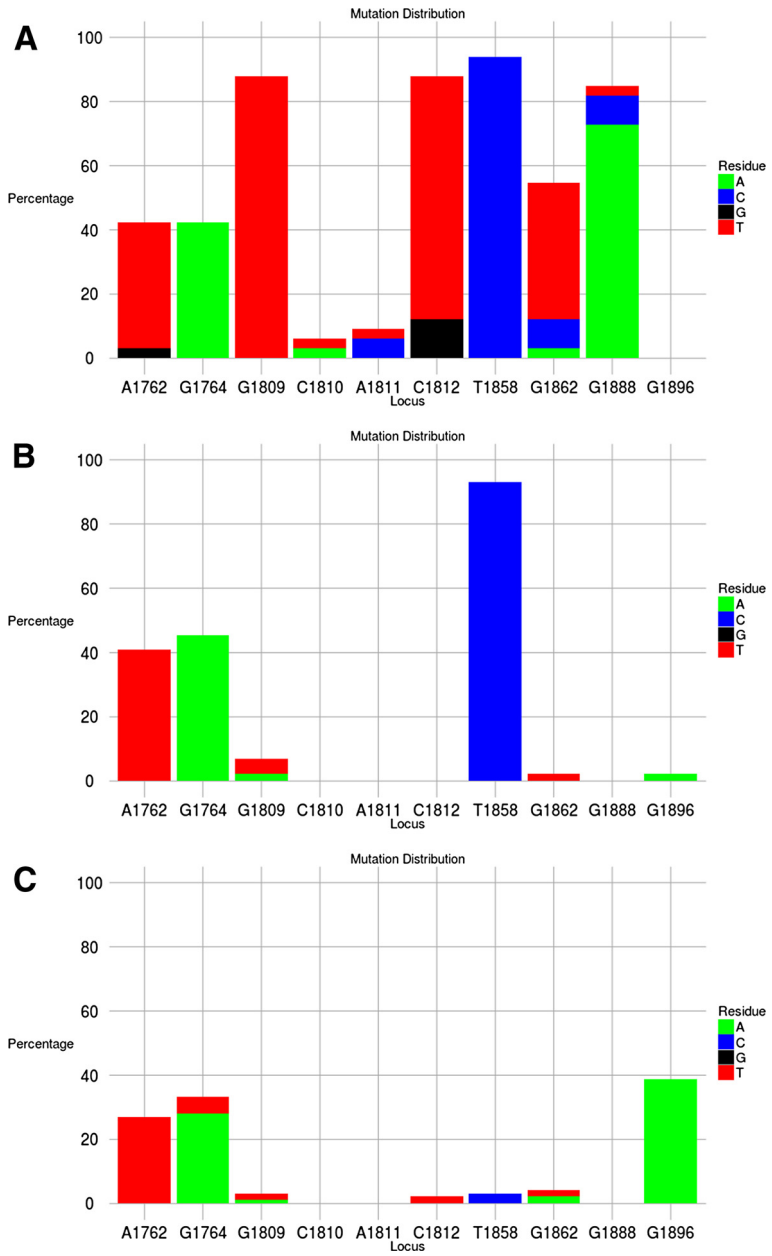


Figure 5 Mutation distribution graph. Mutation distribution graphs showing the percentage of mutant residues relative to the reference motif found at ten loci of interest specified (1762, 1764, 1809-1812, 1858, 1862, 1888, 1896). Three data sets were submitted to the tool to produce the three graphs. Panel **A** shows the mutation distribution for 33 subgenotype A1 samples, panel **B** for 34 subgenotype A2 samples and panel **C** for 93 genotype D samples. The reference motif used was AGGCACTGGG. This is also shown by the letter preceding each locus on the X-axis. To facilitate direct comparisons between the graphs, conserved loci were not suppressd and the Y-axis was scaled to 100% by selecting the appropriate controls on the input page.

as is demonstrated in Figure 5, the G1896A mutation is found in genotype D but not in genotype A. The G1896A leads to a tryptophan to stop codon mutation, which results in the truncation of HBeAg precursor and abrogation of HBeAg expression [27]. G1762T/1764A mutations, which affect the expression of precore mRNA and cause a reduction in HBeAg expression can develop in genotypes A and D. However, because G1896A rarely occurs in genotype A, the frequency of 1762T/1764A is higher in this genotype compared to genotype D (Figure 5). This is the only mutation that can affect HBeAg expression in subgenotype A2 (Figure 5B), whereas in subgenotype A1 there are additional mutations that can modulate HBeAg expression (Figure 5A). In subgenotype A1,

which is also characterized by 1888A, TCAT instead of GCAC occurs at position 1809-1812 (Figure 5A). This change in the Kozak sequence, preceding the precore start codon at 1814, impairs HBeAg expression by a leaky scanning mechanism. The effect of the Kozak mutations on HBeAg expression is comparable with that of the A1762T/G1764A. Co-existence of 1762T1764A and Kozak mutations reduces HBeAg expression in an additive manner [26]. Thus it can be seen that subgenotype A1 has alternative mechanisms for reducing HBeAg expression compared to genotype D, and subgenotype A2 does not develop mutations that can abrogate HBeAg expression. These differences correlate with the findings of an earlier study, which showed that the prevalence of HBeAg in serum was significantly lower in carriers of subgenotype A1 than in carriers of A2 or D [35].

Discovery mode

When the "Discovery Mode" option on the input page is selected, the tool examines the distribution of residues at each of the specified loci and selects for processing only those loci which are *not* conserved across all input sequences. This mode can be used to "discover" loci of interest by specifying a range of loci, such as "1-100" for example, rather than specific, known loci. The tool will then examine the residues at loci 1 to 100, and will include for further processing and output only those loci at which two or more (different) residues are found. Loci at which only one residue is found will be excluded from the analysis entirely.

When "Discovery Mode" is selected, the "Output grouping", "Reference motif" and graphing parameters are disabled, as the number and position of loci, which would be included in the final analysis is only known after the tool has processed the file. Also, as this is a "discovery mode", it will not be known in advance which loci should logically be grouped together as a unit of interest.

Limitations

A limitation, by design, is that the sequence data must be aligned. The position of the anchor motif in the first sequence is taken as the anchor position, and loci in all sequences are referenced according to this position. If the input sequence data are not aligned, or if the anchor motif is incorrect, the tool may return incorrect data. Whilst the number of loci, which can be specified, is not limited, it may not be feasible to enter more than a few dozen loci, as this generates a large amount of output data. All loci values must be greater than the anchor position. Updates to the tool will be made to address limitations as necessary.

Conclusions

As an online tool, available free of charge, no download or installation is required. As demonstrated, this tool can be used for both genotyping and serotyping of HBV without the requirement of computer skills or knowledge of phylogenetics. However, as the tool is genome-agnostic, it has a wide application and nucleotide or amino acid sequence data from any organism can be analysed. Loci of interest, which may be located many hundreds of residues apart, can easily be extracted and their distribution summarised. Rapid characterisation of many sequences, or subsets of sequences, can be achieved easily when the loci of interest are known. Using the "Discovery Mode", conserved and therefore uninformative loci, are automatically ignored, and potential loci of interest can be found and identified.

Methods

The *Mutation Reporter Tool* consists of a web-based front-end ("client" interface) with which the user interacts, and a CGI (common gateway interface) script on a server, which runs Python language [36] code to generate the output web-page. The tool is one component of a larger project currently under development, which makes use of a common, shared Python library. The input FASTA file which the user specifies is saved locally (on the server) by the CGI script and then processed by the Python library. Methods within this library are responsible for loading sequence data from a FASTA file, processing the input parameters, and extracting the requested data from the FASTA file. The output HTML page is written to disk by the Python script. The optional output graphs are generated using the *ggplot2* graphics library [37] in the R statistical programming language [38]. If graphs are requested by the user, the Python script writes the relevant data to disk as a CSV (comma-separated value) file. A short R script, which is customized based on the input parameters specified, is also written to disk. The Python script then calls the R script, which generates the graph and writes it to disk. The images are then linked on the output HTML page. The tool is an online resource, which requires a client browser to connect to the tool's web-server. As such, there is no stand-alone, offline version available for download.

The tool, which assumes that the submitted sequence data is aligned, finds the first occurrence of the anchor motif in the first sequence in the input file. The first character of the anchor motif is then considered to be at the position specified as the anchor position. Sequence data at each of the specified loci for all sequences in the file is then accessed and tabulated. Loci positions are mapped to positions in the sequence data using the anchor motif as an offset value. Data from the loci specified are grouped into columns according to the "output grouping" field. If this field is not specified, all loci are grouped into one output column. If a sequence ID pattern was specified, the tool executes the appropriate regular expression match on

the FASTA IDs in the input file. In "Discovery Mode", loci at which no variation is found are excluded.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
AK is the principle investigator. TB conceived the idea of the tool, wrote the code, established and maintained the server, software and hardware. TB and AK wrote the paper. Both authors read and approved the final manuscript.

Acknowledgements
TB received bursaries from the National Research Foundation (NRF), the Medical Research Foundation (MRC), the Poliomyelitis Research Foundation, National Bioinformatics Network and the University of the Witwatersrand. AK received funding from the NRF (GUN#65530) and the MRC. Mark Keyter and Mukhlid Yousif tested the tool extensively and provided valuable feedback and suggestions.

Received: 16 July 2012 Accepted: 28 January 2013
Published: 23 February 2013

References

1. WHO: **World Health Organization: Hepatitis B fact sheet 204 (July 2012, Revision).** 2012. [Accessed on 04 August 2012]. [http://www.who.int/mediacentre/factsheets/fs204/en/].
2. Okamoto H, Tsuda F, Sakugawa H, Sastrosoewignjo RI, Imai M, Miyakawa Y, Mayumi M: **Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes.** *J Gen Virol* 1988, **69:**2575–2583.
3. Norder H, Couroucé AM, Magnius LO: **Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes.** *Virology* 1994, **198:**489–503.
4. Norder H, Hammas B, Lofdahl S, Courouce AM, Magnius LO: **Comparison of the amino acid sequences of nine different serotypes of hepatitis B surface antigen and genomic classification of the corresponding hepatitis B virus strains.** *J Gen Virol* 1992, **73**(Pt 5):1201–1208.
5. Naumann H, Schaefer S, Yoshida CF, Gaspar AM, Repp R, Gerlich WH: **Identification of a new hepatitis B virus (HBV) genotype from Brazil that expresses HBV surface antigen subtype adw4.** *J Gen Virol* 1993, **74:**1627–1632.
6. Stuyver L, Gendt SD, Geyt CV, Zoulim F, Fried M, Schinazi RF, Rossau R: **A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness.** *J Gen Virol* 2000, **81:**67–74.
7. Arauz-Ruiz P, Norder H, Robertson BH, Magnius LO: **Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America.** *J Gen Virol* 2002, **83:**2059–2073.
8. Olinger CM, Jutavijittum P, Hübschen JM, Yousukh A, Samountry B, Thammavong T, Toriyama K, Muller CP: **Possible new hepatitis B virus genotype, southeast Asia.** *Emerg Infect Dis* 2008, **14:**1777–1780.
9. Tran TTH, Trinh TN, Abe K: **New complex recombinant genotype of hepatitis B virus identified in Vietnam.** *J Virol* 2008, **82:**5657–5663.
10. Arankalle VA, Gandhe SS, Borkakoty BJ, Walimbe AM, Biswas D, Mahanta J: **A novel HBV recombinant (genotype I) similar to Vietnam/Laos in a primitive tribe in eastern India.** *J Viral Hepatitis* 2010, **17:**501–510.
11. Osiowy C, Kaita K, Solar K, Mendoza K: **Molecular characterization of hepatitis B virus and a 9-year clinical profile in a patient infected with genotype I.** *J Med Virol* 2010, **82:**942–948.
12. Yu H, Yuan Q, Ge SX, Wang HY, Zhang YL, Chen QR, Zhang J, Chen PJ, Xia NS: **Molecular and phylogenetic analyses suggest an additional Hepatitis B virus genotype "I".** *PLOS One* 2010, **5:**e9297.
13. Tatematsu K, Tanaka Y, Kurbanov F, Sugauchi F, Mano S, Maeshiro T, Nakayoshi T, Wakuta M, Miyakawa Y, Mizokami M: **A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J.** *J Virol* 2009, **83:**10538–10547.
14. Kramvis A, Arakawa K, Yu MC, Nogueira R, Stram DO, Kew MC: **Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus.** *J Med Virol* 2008, **80:**27–46.
15. Mayerat C, Mantegani A, Frei PC: **Does hepatitis B virus (HBV) genotype influence the clinical outcome of HBV infection?** *J Viral Hepatitis* 1999, **6:**299–304.
16. Sumi H, Yokosuka O, Seki N, Arai M, Imazeki F, Kurihara T, Kanda T, Fukai K, Kato M, Saisho H: **Influence of hepatitis B virus genotypes on the progression of chronic type B liver disease.** *Hepatology* 2003, **37:**19–26.
17. Kramvis A, Kew MC: **Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy.** *J Viral Hepatitis* 2005, **12:**456–464.
18. Steinhauer DA, Holland JJ: **Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA.** *J Virol* 1986, **57:**219–228.
19. Kramvis A, Kew M, François G: **Hepatitis B virus genotypes.** *Vaccine* 2005, **23:**2409–2423.
20. Zoulim F, Locarnini S: **Hepatitis B virus resistance to Nucleos(t)ide analogues.** *Gastroenterology* 2009, **137:**1593–1608.
21. Revill P, Yuen L, Walsh R, Perrault M, Locarnini S, Kramvis A: **Bioinformatic analysis of the hepadnavirus e-antigen and its precursor identifies remarkable sequence conservation in all orthohepadnaviruses.** *J Med Virol* 2010, **82:**104–115.
22. Ou JH, Laub O, Rutter WJ: **Hepatitis B virus gene function: the precore region targets the core antigen to cellular membranes and causes the secretion of the e antigen.** *Proc Nat Acad Sci USA* 1986, **83:**1578–1582.
23. Kramvis A, Kew MC: **Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes.** *Hepatology Res* 2007, **37:**S9–S19.
24. Tong S: **Impact of viral genotypes and naturally occurring mutations on biological properties of hepatitis B virus.** *Hepatology Res* 2007, **37:**S3–S8.
25. Buckwold VE, Xu Z, Chen M, Yen TS, Ou JH: **Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication.** *J Virol* 1996, **70:**5845–5851.
26. Ahn SH, Kramvis A, Kawai S, Spangenberg HC, Li J, Kimbi G, Kew M, Wands J, Tong S: **Sequence variation upstream of precore translation initiation codon reduces hepatitis B virus e antigen production.** *Gastroenterology* 2003, **125:**1370–1378.
27. Carman WF, Hadziyannis S, Mcgarvey MJ, Jacyna MR, Karayiannis P, Makris A, Thomas HC: **Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection.** *Lancet* 1989, **334:**588–591.
28. Lok AS, Akarca U, Greene S: **Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal.** *Proc Nat Acad Sci USA* 1994, **91:**4077–4081.
29. Chen CY, Crowther C, Kew MC, Kramvis A: **A valine to phenylalanine mutation in the precore region of hepatitis B virus causes intracellular retention and impaired secretion of HBe-antigen.** *Hepatology Res* 2008, **38:**580–592.
30. Magnius LO, Norder H: **Genotypes and molecular epidemiology of the Hepatitis B virus as reflected by sequence variability of the S-Gene.** *Intervirology* 1995, **38:**24–34.
31. Wands JR, Wong MA, Shorey J, Brown RD, Marciniak RA, Isselbacher KJ: **Hepatitis B viral antigenic structure: signature analysis by monoclonal radioimmunoassays.** *Proc Nat Acad Sci USA* 1984, **81:**2237–2241.
32. Mimms LT, Floreani M, Tyner J, Whitters E, Rosenlof R, Wray L, Goetze A, Sarin V, Eble K: **Discrimination of hepatitis B virus (HBV) subtypes using monoclonal antibodies to the PreS1 and PreS2 domains of the viral envelope.** *Virology* 1990, **176:**604–619.
33. Swenson PD, Riess JT, Krueger LE: **Determination of HBsAg subtypes in different high risk populations using monoclonal antibodies.** *J Virological Methods* 1991, **33:**27–38.
34. Purdy MA, Talekar G, Swenson P, Araujo A, Fields A: **A new algorithm for deduction of Hepatitis B surface antigen subtype determinants from the amino acid sequence.** *Intervirology* 2007, **50:**45–51.
35. Tanaka Y, Hasegawa I, Kato T, Orito E, Hirashima N, Acharya SK, Gish RG, Kramvis A, Kew MC, Yoshihara N, Shrestha SM, Khan M, Miyakawa Y, Mizokami M: **A case-control study for differences among hepatitis B**

virus infections of genotypes A (subtypes Aa and Ae) and D. *Hepatology* 2004, **40:**747–755.
36. van Rossum G: **Python.** [http://www.python.org].
37. Wickham H: *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer; 2009.
38. R Core Team: *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing; 2012. [http://www.R-project.org/].

doi:10.1186/1743-422X-10-62
Cite this article as: Bell and Kramvis: *Mutation Reporter Tool*: An online tool to interrogate loci of interest, with its utility demonstrated using hepatitis B virus. *Virology Journal* 2013 **10**:62.

Chapter 7

Paper III

Fragment Merger: An online tool to merge overlapping long sequence fragments

Bell, T. G. and Kramvis, A. (2013)

Viruses **5:** 824–833

References References for literature cited in the following paper appear within the references section of the paper itself, and not necessarily in the references section of this thesis.

Journal *Viruses* is a peer-reviewed, ISI-accredited, open-access journal, with an impact factor of 1.50. This journal is part of the MDPI group and is also known as "Viruses Basel".

Website The "Fragment Merger Tool" is available online at *http://hvdr.bioinf.wits.ac.za/tools/*

Viruses **2013**, *5*, 824-833; doi:10.3390/v5030824

OPEN ACCESS

viruses

ISSN 1999-4915

www.mdpi.com/journal/viruses

Article

Fragment Merger: An Online Tool to Merge Overlapping Long Sequence Fragments

Trevor G. Bell and Anna Kramvis *

Hepatitis Virus Diversity Research Programme (HVDRP), Department of Internal Medicine, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, 2050 Johannesburg, South Africa

* Author to whom correspondence should be addressed; Anna.Kramvis@wits.ac.za, Fax: +27-(0)-86-529-6806

Received: 16 January 2013; in revised form: 26 February 2013 / Accepted: 26 February 2013 / Published: 12 March 2013

Abstract: While PCR amplicons extend to a few thousand bases, the length of sequences from direct Sanger sequencing is limited to 500–800 nucleotides. Therefore, several fragments may be required to cover an amplicon, a gene or an entire genome. These fragments are typically sequenced in an overlapping fashion and assembled by manually sliding and aligning the sequences visually. This is time-consuming, repetitive and error-prone, and further complicated by circular genomes. An online tool merging two to twelve long overlapping sequence fragments was developed. Either chromatograms or FASTA files are submitted to the tool, which trims poor quality ends of chromatograms according to user-specified parameters. Fragments are assembled into a single sequence by repeatedly calling the EMBOSS *merger* tool in a consecutive manner. Output includes the number of trimmed nucleotides, details of each merge, and an optional alignment to a reference sequence. The final merge sequence is displayed and can be downloaded in FASTA format. All output files can be downloaded as a ZIP archive. This tool allows for easy and automated assembly of overlapping sequences and is aimed at researchers without specialist computer skills. The tool is genome- and organism-agnostic and has been developed using hepatitis B virus sequence data.

Keywords: sequence data; sequence fragments; chromatograms; DNA assembly; amplicons; hepatitis B virus

1. Introduction

DNA sequencing is a routine procedure in many wet laboratories. This sequencing may include direct, Sanger sequencing, or any of the many "next generation" sequencing technologies. Many sequence assembly software programs are available [1–6]. Most of these programs assemble short reads of next-generation sequencing data, with a small number assembling a query sequence against a reference sequence (mapping assembly). Some of these programs are only available commercially, and some are only available on specific operating system platforms. Comprehensive, integrated bioinformatics software solutions, which may also include such functionality, are typically very expensive and their usage is restricted to licensed workstations only. Such software suites are often complex, requiring training and high levels of computer proficiency to operate effectively. In addition, installing and using many of the available programs can be difficult, may require technical expertise, access to a specific operating system platform or expensive hardware. In resource-limited settings, where capacity and finances are generally limited, the purchase and use of such software is not possible. To address some of these issues, we have developed an online, web-based, sequence assembly tool to easily assemble overlapping long PCR amplicons as sequenced by direct, Sanger sequencing technology. A web-based tool requires no installation and can be used via any web-browser from any operating system platform. No specialist technical skills are required to use the tool, which was developed and tested extensively using hepatitis B virus (HBV) sequence data.

HBV has a partially double-stranded circular DNA genome, which, depending on genotype, ranges in length from 3182 to 3248 nucleotides. By convention, the *EcoR*I restriction enzyme cleavage site (G|AATTC), located within the surface gene, is denoted as nucleotide position 1. The genome contains four partially overlapping reading frames and codes for seven proteins. Nine distinct HBV genotypes are known, with up to 32 subgenotypes described to date [7,8]. Sequence heterogeneity is common in HBV, as the viral polymerase lacks proof-reading ability [9,10]. Furthermore, variant strains, which exhibit insertions, deletions or SNPs, are commonly encountered and reported. In addition, recombination of large or small regions between two or more variants has also been reported [11].

Sequence data from either the surface (S) gene (approximately 1200 nucleotides) or the entire genome are essential to determine the viral genotype or subgenotype, to characterize the virus, and to identify indels and SNPs. Such data are routinely obtained by single or nested PCR amplification of the regions in question [12,13], followed by direct sequencing (typically in the forward direction only) using a number of internal sequencing primers. The resulting amplicons are typically between 500 and 800 nucleotides in length. The HBV genome is circular, but the resulting sequence is a linear fragment. Assembling these fragments into a complete gene or genome has previously been undertaken manually. The chromatogram (trace) file for each fragment is viewed [14,15] and checked. The poor quality ends are trimmed manually. Sequence data for each fragment is imported into an editor, such as GeneDoc [16], and each fragment is slid until it overlaps with another fragment. The complete sequence is then constructed by either entering the sequence of bases or by editing a reference sequence. This process is extremely time-consuming, repetitive and error-prone.

We describe here the implementation of a genome-agnostic, web-based, assembly tool, which has been developed using HBV sequence data.

2. Implementation

2.1. Overview

The online tool we have developed is a Python 2.6.5+ CGI script [17], hosted on an Ubuntu GNU/Linux server [18]. The tool processes between two and twelve overlapping long sequence fragments; that is, a series of fragments in which sequence data between subsequent pairs of fragments overlap. These fragments may be from either chromatograms (trace files in "AB1" format) or FASTA files, or a mixture of both formats. FASTA files (which will not be trimmed by the tool) would typically contain sequence data, which has previously been curated (trimmed and checked). Chromatograms will be trimmed according to the "trim window" and "trim threshold" integer parameters specified for each file. Base calls and quality scores are extracted from each chromatogram using a Python ABIF file reader [19]. A moving window of "trim window" nucleotides is slid over each sequence, from each end, until the quality scores of all nucleotides within this window are greater than or equal to the "trim threshold" value. The default values of 10 for the trim window and 20 for the trim threshold have proven to be suitable for the HBV sequence data analyzed during the development of the tool. After the chromatogram has been trimmed, the base calls in the chromatogram are extracted and used as the sequence data. If a chromatogram is of such poor quality that it is trimmed to a length of zero, the user will be notified and no merge will be performed. As the function of this tool is not to assemble shotgun sequence data, fragment files (chromatograms and/or FASTA files) must be specified in the order in which the fragments should be merged. Sequences, which should be reversed and/or complemented, must be specified before they are merged. This information is readily available from the amplification and sequencing protocols used.

The EMBOSS *merger* program [20], which implements the Needleman–Wunsch global alignment algorithm [21] to align two sequences, is used with default parameters to merge the sequence data. When only two input fragments are specified, the *merger* tool is executed from the Python script with two input files and the output is written directly to an output file. When more than two fragments are specified, a "bash" shell command is built, which calls the *merger* program repeatedly, depending on the number of fragments specified. One of the two input sequences to the *merger* tool can be provided from the UNIX "stdin" pipe and the output can be redirected to the UNIX "stdout" pipe. To avoid creating many temporary files for each of the successive two-fragment merges, the "bash" shell command is constructed such that the output from one merge is piped in as the input to the next merge. The following three commands are used to construct the required command-line:

```
  Command A: merger -asequence SEQ1 -bsequence SEQ2 -auto -outseq
/dev/stdout
Command B: merger -asequence /dev/stdin -bsequence SEQ3 -auto -outseq
/dev/stdout
Command C: merger -asequence /dev/stdin -bsequence SEQ4 -auto -outseq
SEQ99
```

Command A runs the *merger* program to merge sequences *SEQ1* and *SEQ2* and writes the output to "stdout". Command B merges the data from "stdin" with sequence *SEQ3* and writes the output to "stdout". Command C merges the data from "stdin" with *SEQ4* and writes the output to an output file *SEQ99*, which will then contain the complete merged sequence. This process can therefore be generalised into a command-line consisting of "Command A — Command B (repeated $n-3$ times) — Command C", where "n" is the number of fragments and $n \geq 3$. For example, if three fragments are specified, the command-line will consist of "Command A — Command C". Command B will not be included, as it is repeated $(n-3)$ times. If five fragments are specified, the command-line will consist of "Command A — Command B — Command B — Command C". The total number of merges executed is therefore $n-1$.

The *Fragment Merger Tool* consists of a web-based front-end ("client" interface) with which the user interacts, and a CGI (common gateway interface) script on a server, which runs Python language [17] code to generate the output web-page. The tool is one component of a larger project currently under development, which makes use of a common, shared Python library. The input files, which the user specifies, are saved locally (on the server) by the CGI script and then processed by the Python library. Methods within this library are responsible for loading sequence data from the files, processing the input parameters, and executing the merges. The output HTML page is written to disk by the Python script. The tool is an online resource, which requires a client browser to connect to the tool's web-server, and therefore no stand-alone, offline version is available for download.

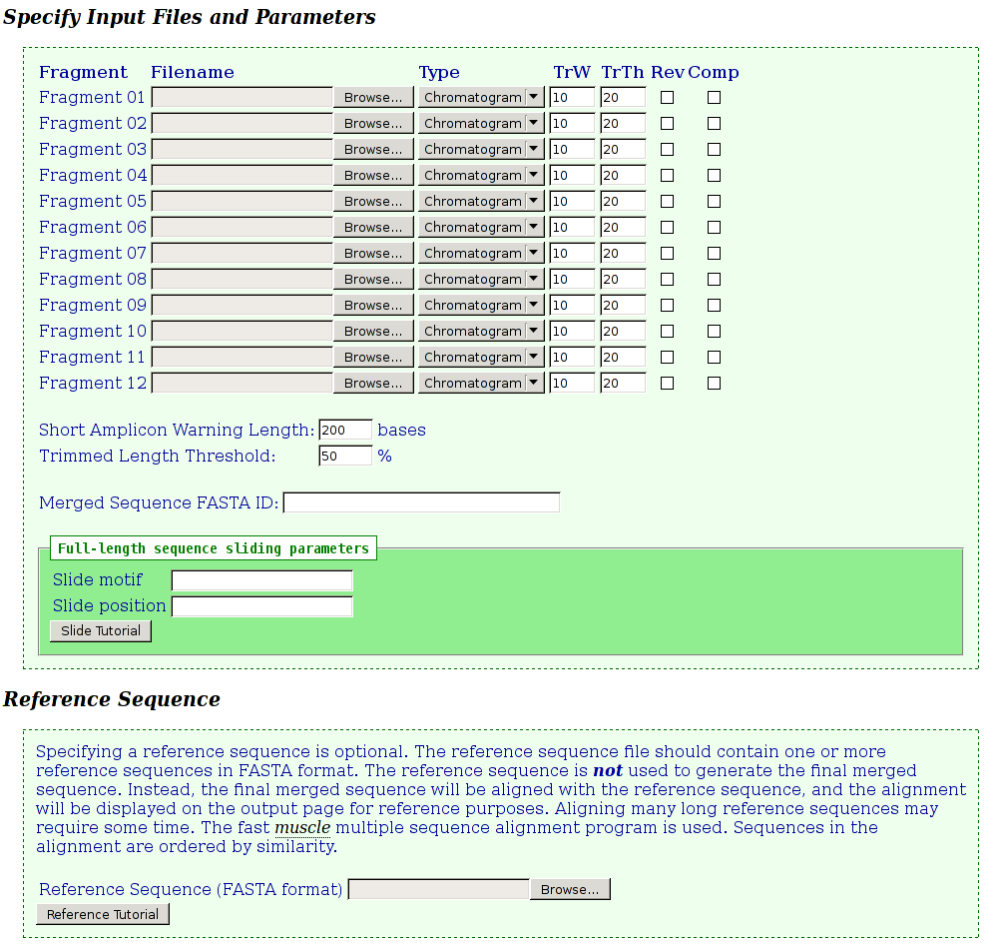
2.2. Input

The input page for the tool is shown in Figure 1. Filenames for two to twelve chromatograms and/or FASTA files are specified. The "type" field will automatically change to indicate the type of the input file based on the filename extension, but can be changed manually, if required. When a chromatogram is specified, the trim window ("TrW") and trim threshold ("TrTh") values can be adjusted, as described previously. Relaxing these parameters may assist in obtaining a merged sequence, by less stringently trimming lower quality ends. Sequence data can be reversed and/or complemented as required, by selecting the appropriate check-boxes for each fragment. The tool will display a yellow "warning" or a red "error" icon on the output page next to fragments, which are below specified lengths, to assist the user in checking the input data. Fragments, which are shorter than the "Short Amplicon Warning Length", will be flagged with the "warning" icon, while those that are trimmed to below the "Trimmed Length Threshold" percentage of their original length will be flagged with the "error" icon. The value entered into the optional "Merged Sequence FASTA ID" field will be used as the FASTA ID of the final merged sequence. If this field is omitted, the FASTA ID of the merged sequence will be that of the first fragment submitted. When sequences are trimmed, the trim parameters will be appended to the existing FASTA ID for reference.

Optional "Slide motif" and "Slide position" parameters may also be specified. These are only applicable when assembling full-length circular genomes. A final merged sequence representing a full genome is unlikely to be in the correct orientation because of the location of the various sequencing primers. However, the nucleotide data of such sequences can be slid into the correct orientation. For

Viruses **2013**, *5* **828**

Figure 1. Fragment Merger Input Page. Each input file (chromatogram or FASTA file) is specified in the order in which the merge is to be performed. Trimming parameters can be specified for each chromatogram. Selecting the "Rev" or "Comp" check-boxes will reverse or complement the sequence data, respectively. Two input boxes specify the thresholds for considering a fragment as "short" and for triggering a warning for potentially poor quality chromatograms. Sliding parameters and a reference sequence file (in FASTA format) can be specified.



HBV, for example, the *EcoR*I restriction enzyme cleavage site (G|AATTC), located within the surface gene, is denoted as nucleotide position 1. When a slide motif and slide position are specified, the tool will slide the final merged sequence so that the specified motif is located at the specified position in the sequence. If the quality or orientation of a merge is not satisfactory, the process can easily be repeated, after adjusting the relevant input parameters and/or specifying sliding parameters.

Optionally, a reference sequence file, containing one or more reference sequences in FASTA format, may be specified. These reference sequences are not used to construct the merged sequence, but to check the merged sequence using multiple sequence alignment as implemented by the fast Muscle program [22]. The alignment will be included on the output page and made available for download.

When fragments derived from a full-length circular genome (such as a viral genome) are specified, it is likely that the final merged sequence will not be in the correct orientation. The same scenario will occur

when sequences straddling the "start" position of a circular genome are specified. In the former case, the merge sequence can be slid into the correct orientation by specifying appropriate "slide" parameters on the input page as described above. In either case, specifying a reference sequence in the standard orientation will result in an alignment with long regions of gaps on one or both ends. This can be avoided by preparing the reference sequence to match the orientation of the merged sequence, or by preparing a double-length reference sequence. Such a sequence would place the end of the first linear sequence adjacent to the start of the second linear sequence, thereby creating an artificial full-length linear sequence.

3. Results and Discussion

Once the merges have run, detailed output is provided to the user as shown in Figures 2–4. The first section of output (Figure 2) displays details for each fragment submitted. When chromatograms are submitted, the number of bases, which were trimmed from each end, is shown, whereas when FASTA files are submitted, no trimming is performed and a value of "100%" is shown. The various possible notification icons are described in Figure 2. Data from fragments flagged with a yellow or red icon are not excluded from the merge, but the user should check the final merged sequence carefully.

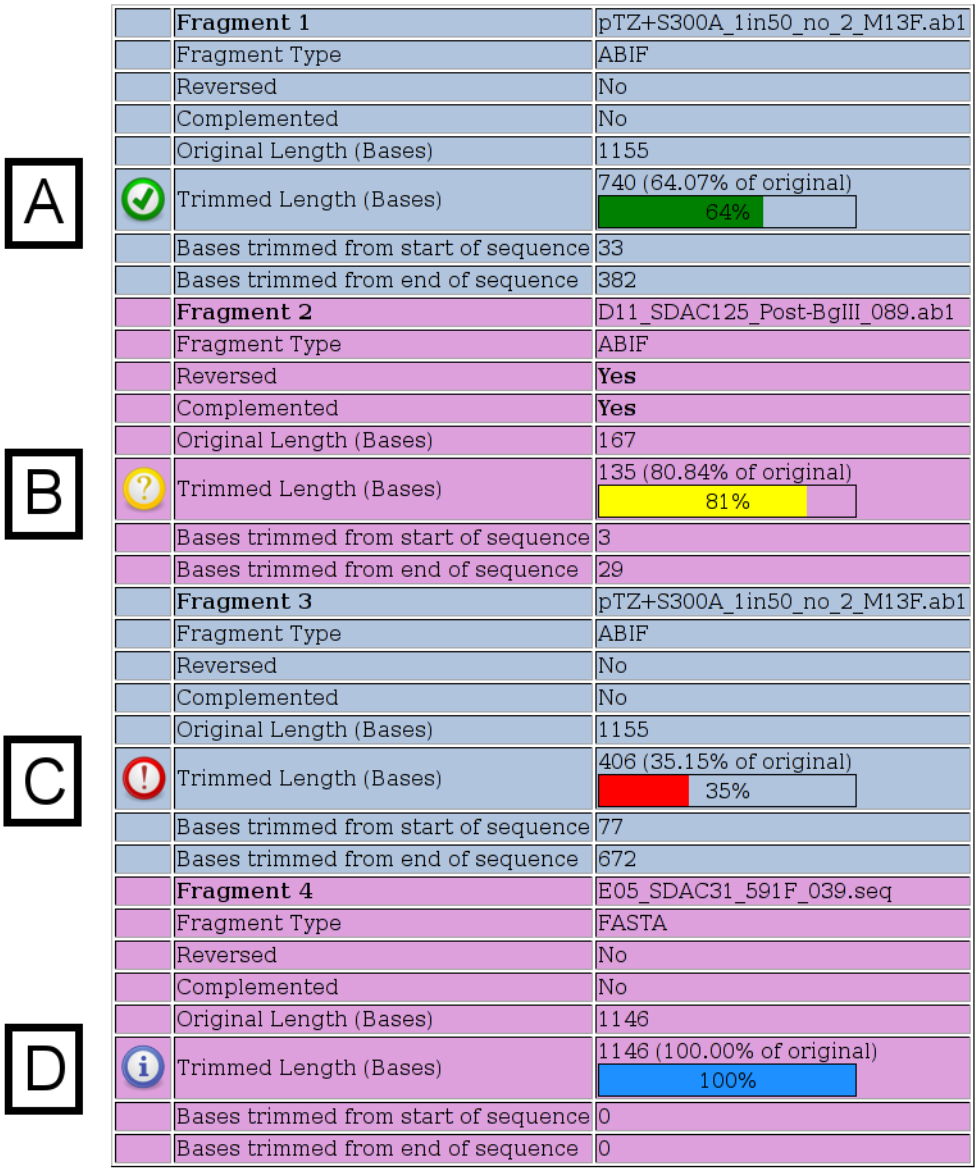
The additional sections of output are shown in Figures 3 and 4. In Figure 3A, the final merged (assembled) sequence is shown and can be downloaded as a file in FASTA format. The length is displayed, as well as any sliding parameters, which were provided. Detailed output of each of the successive merges, as generated by the *merger* program, is provided in a table (Figure 3B). If a reference sequence file was specified, the merged sequence aligned against the reference sequence/s is displayed (Figure 3C). The alignment is intended to be used as a quick check to validate the success and/or accuracy of the merge. If no merge is possible (because of insufficient areas of overlap between two sequences), the *merger* program will simply concatenate the two input sequences, with only one overlapping nucleotide. It is therefore important that the final sequence is checked carefully, preferably against known reference sequences. If the reference sequence file contains several sequences, additional time will be required to generate the alignment. Typically, one or two reference sequences should be sufficient. A download hyperlink to a ZIP archive file containing all input and output files (excluding chromatograms) is provided. The archive (Figure 4A) contains the untrimmed input sequence data, the trimmed input sequence data (files starting with the name "ToMerge"), the final merged sequence (a file starting with the name "Merge0"), the reference sequence file, the unaligned merged and reference sequences, the aligned merged and reference sequences, the output text files from the *merger* program for each merge (files starting with the name "outFile") and a "README" text file describing each of the files, for reference (Figure 4B).

The time taken for the Python CGI script to execute was calculated by subtracting the timestamp when the script completed from the timestamp when the script started. The average execution time, from 195 runs of the tool over several months, merging three overlapping fragments from chromatogram input data, was 0.48 ± 0.12 seconds.

The tool has been used extensively to assemble the complete surface gene of HBV from three overlapping fragments. Although the HBV genome is circular, sequence data (either from direct

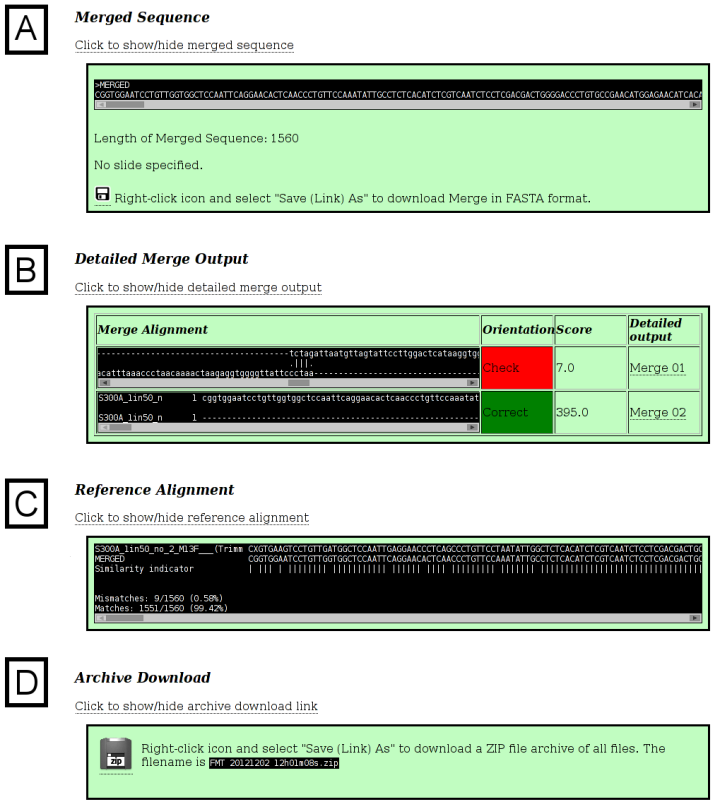
Viruses **2013**, *5* **830**

Figure 2. Fragment Details. Details for each fragment submitted are shown on the output page. All possible notification icons are shown here. (**A**) A green icon indicates that the chromatogram has been trimmed, but is not shorter than any of the specified thresholds. (**B**) A yellow icon indicates that the trimmed chromatogram is shorter than the specified warning length, which has a default value of 200. (**C**) A red icon indicates that the trimmed chromatogram is shorter than the specified percentage of its original length, which has a default value of 50%. (**D**) A blue icon indicates that a FASTA file was specified, in which case, no trimming is performed.



sequencing or from sequencing of clones) is linear. The tool can therefore be used with both circular and linear sequences. The tool has also been used to assemble entire HBV genomes from 6 or 7 overlapping fragments. In this case, fragments in both the forward and reverse direction were used.

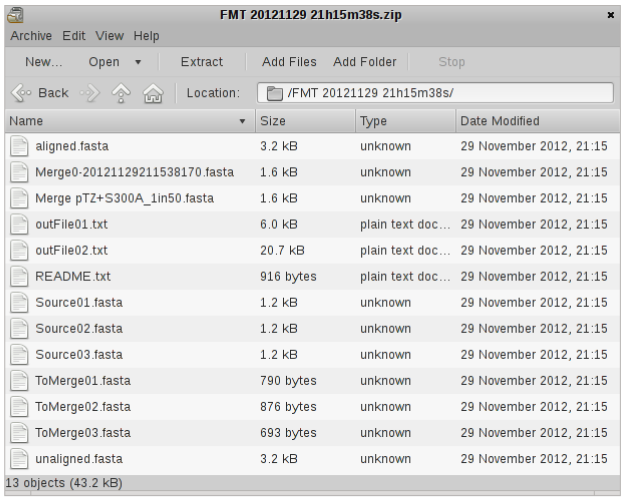
Figure 3. Additional Output. The additional sections of the output page are shown. (**A**) The final merged sequence in FASTA format is displayed and can be downloaded. (**B**) The detailed output of each of the successive merges is provided in a table. The alignment of the two fragments for each merge, and their score (as generated by the *merger* program), are provided. Since the fragments are merged in order sequentially, the expectation is that the end of the first fragment in each merge will overlap with the start of the next fragment. If this is the case, a green table cell with the word "Correct" is shown under the "Orientation" column. If the merge has occurred in the other orientation, the cell is shaded red with the word "Check" displayed. The final column provides a hyperlink to the full, detailed output for each merge, as generated by the *merger* program. This output includes a table detailing any conflicts, which the *merger* program detected, between the two sequences, and the base, which was used in the output sequence. It is advisable to check this detailed output before continuing to use the final merged sequence in any downstream applications or analyses. (**C**) The alignment of the merged sequence against the reference sequence(s) is shown. Conserved loci across all sequences are indicated with a "|" character, mismatches are indicated with a space character, and the total number of matches and mismatches is shown. (**D**) An archive containing all input and output files (excluding chromatograms) can be downloaded. All the files in the archive are in a folder named with the date and time the merge was executed.



The tool has been designed to assist users in processing sequence data and requires that the user exercise discretion when submitting data and interpreting the results. Poor quality data in chromatograms

Figure 4. Archive Contents. (**A**) A list of the files in the archive. (**B**) The "README.txt" file from the downloaded archive. This file provides the filename and a description for all of the files in the archive.

A



B

```
README
------

Fragment Merger Tool (HVDR)
---------------------------

This archive contains all files from the merge

Run completed 2012-11-29 21:15

Source01.fasta  Input sequence in FASTA format
ToMerge01.fasta Sequence used for merge in FASTA format (trimmed if from chromatogram input)
Source02.fasta  Input sequence in FASTA format
ToMerge02.fasta Sequence used for merge in FASTA format (trimmed if from chromatogram input)
Source03.fasta  Input sequence in FASTA format
ToMerge03.fasta Sequence used for merge in FASTA format (trimmed if from chromatogram input)
Merge0-20121129211538170.fasta  Final merged sequence in FASTA format
outFile01.txt   merger output
outFile02.txt   merger output
Merge pTZ+S300A_1in50.fasta     Supplied reference sequence
unaligned.fasta Final merged sequence and reference sequence before alignment
aligned.fasta   Final merged sequence and reference sequence after alignment
```

could result in false indels in the final sequence. The tool does not disambiguate any ambiguous bases in chromatogram data, nor does it search for, or remove, any vector-specific or other primer sequences. If this is required, a FASTA file of the edited data, without the primer sequences, should be submitted to the tool. This is a not a mapping assembler and does not make use of a reference sequence for assembly.

4. Conclusions

This tool allows for easy and automated assembly of long overlapping sequence fragments, which can be input as either chromatograms or FASTA files, and provides detailed visual output. Sequence data from insertion or deletion mutants and recombinants can be assembled, as a reference sequence is not used for assembly. Although it has been developed and tested extensively on HBV sequence data, it is genome- and organism-agnostic. The length of the final, merged sequence should be less than 100,000 nucleotides. The tool does not require technical expertise, or special hardware or software to use, and is aimed at researchers without specialist computer skills.

Acknowledgements

TB received funding from the National Bioinformatics Network (NBN), the National Research Foundation (NRF), the Poliomyelitis Research Foundation, the Medical Research Council (MRC) and the University of the Witwatersrand. AK received funding from the NRF and the MRC. HVDRP team members Mark Keyter, Mukhlid Yousif and Suzanne Nicholson tested the tool extensively and provided valuable feedback and suggestions.

References

1. Software packages for next gen sequence analysis: SEQanswers. Available online: http://seqanswers.com/forums/showthread.php?t=43 (accessed on 14 January 2013).
2. Software/list SEQwiki. Available online: http://seqanswers.com/wiki/Software/list (accessed on 14 January 2013).
3. Sequence Assembly. Available online: http://en.wikipedia.org/wiki/Sequence_assembly#Available_assemblers (accessed on 14 January 2013).
4. Huang, X.; Madan, A. CAP3: A DNA Sequence Assembly Program. *Genome Research* **1999**, *9*, 868–877.
5. Phred, Phrap, and Consed (Green Group). Available online: http://phrap.org/phredphrapconsed.html (accessed on 14 January 2013).
6. Bonfield, J. K.; Smith, K. F.; Staden, R. A new DNA sequence assembly program. *Nucleic Acids Research* **1995**, *23*, 4992–4999.
7. Yu, H.; Yuan, Q.; Ge, S. X.; Wang, H. Y.; Zhang, Y. L.; Chen, Q. R.; Zhang, J.; Chen, P. J.; Xia, N. S. Molecular and phylogenetic analyses suggest an additional hepatitis B virus genotype "I". *PLOS ONE* **2010**, *5(2)*, e9297.
8. Kramvis, A.; Arakawa, K.; Yu, M. C.; Nogueira, R.; Stram, D. O.; Kew, M. C. Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. *Journal of Medical Virology* **2008**, *80*, 27–46.
9. Steinhauer, D.A.; Holland, J.J. Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. *Journal of Virology* **1986**, *57*, 219–228.
10. Yamamoto, K.; Horikita, M.; Tsuda, F.; Itoh, K.; Akahane, Y.; Yotsumoto, S.; Okamoto, H.; Mikakawa, Y.; Mayumi, M. Naturally occurring escape mutants of hepatitis B virus with various mutations in the S gene in carriers seropositive for antibody to hepatitis B surface antigen. *Journal of Virology* **1994**, *68*, 2671–2676.
11. Kramvis, A.; Kew, M.; François, G. Hepatitis B virus genotypes. *Vaccine* **2005**, *23*, 2409–2423.
12. Gunther, S.; Li, B.C.; Miska, S.; Kruger, D.H.; Meisel, H.; Will, H. A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and revelas deletion mutants in immunosupressed patients. *Journal of Virology* **1995**, *69*, 5437–5444.
13. Vermeulen, M.; Dickens, C.; Lelie, N.; Walker, E.; Coleman, C.; Keyter, M.; Reddy, R.; Crookes, R.; Kramvis, A. Hepatitis B virus transmission by blood transfusion during 4 years of individual-donation nucleic acid testing in South Africa: estimated and observed window period risk. *Transfusion* **2012**, *52*, 880–892.

14. Chromas Lite (Technelysium Pty Ltd). Available online: http://www.technelysium.com.au/chromas_lite.html (accessed on 14 January 2013).
15. FinchTV 1.4.0 (Geospiza, Inc.; Seattle, WA, USA).
16. Nicholas, K.B.; B., N.H.; Deerfield, D.W. GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEW.NEWS* **1997**, *4*, 14.
17. van Rossum, G. Python .
18. (anon.). Fragment Merger Tool Available online: http://hvdr.bioinf.wits.ac.za/tools (accessed on 01 March 2013).
19. Interactive Biosoftware. Available online: http://www.interactive-biosoftware.com/software/resources/abif-reader (accessed on 14 January 2013).
20. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite (2000). *Trends in Genetics* **2000**, *16*, 276–277.
21. Needleman, S.B.; Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **1970**, *48*, 443–453.
22. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797.

Chapter 8

Concluding Discussion

During the course of this study, it became evident that several distinctive aspects of HBV/HIV co-infection in southern Africa exist. Both viruses are hyperendemic in the region, and thus the disease burden is high. From published data, the HIV prevalence in Mpumalanga is 15.4% (Shisana *et al.*, 2009), and our study showed a 24% HBV prevalence in the cohort of HIV-infected individuals, almost two-thirds of which were HBsAg-negative. Moreover, the genotypes prevailing for both viruses differ from those found in other regions of the world. The HIV subtype circulating in southern Africa is HIV-1 subtype C (Goudsmit, 1997; Bredell *et al.*, 1998). The HIV-infected individuals in the present cohort were infected with HBV subgenotype A1 (Makondo *et al.*, 2012) (Appendix C), the subgenotype also prevailing in HBV mono-infected individuals in the region (Kimbi *et al.*, 2004). HBV transmission in southern Africa is largely horizontal during early childhood (Barin *et al.*, 1981; Whittle *et al.*, 1983; Kew, 1996; World Health Organization, 2012a), rather than sexually later in life as found in developed countries (World Health Organization, 2012a). However, because of the immunosuppression caused by HIV, HBV infection is now arising later in life, as a result of reactivation, superinfection, or a new infection, typically from sexual contact. Moreover, the socio-economic factors of poverty and migrant labour, as well as a lack of resources at all levels, increases the disease burden. Although the previous paucity of HBV/HIV co-infection studies is starting to be addressed (Burnett *et al.*, 2005; Mphahlele *et al.*, 2006; Firnhaber *et al.*, 2008, 2009; Lukhwareni *et al.*, 2009; Boyles and Cohen, 2011; Makondo *et al.*, 2012; Mayaphi *et al.*, 2012), more studies are required in order to gain a clear and complete understanding of HBV/HIV co-infection in southern Africa, with its unique and challenging environment.

The number of lifetime sexual partners in the Shongwe cohort was the only factor distin-

guishing HBV DNA-positive from HBV DNA-negative individuals, suggesting a sexual transmission of HBV and/or HIV. Similarly, this factor distinguished HIV-positive and HIV-negative groups elsewhere in South Africa (Mayaphi *et al.*, 2012). In a systematic review of 68 epidemiological studies in sub-Saharan Africa, consisting of 17,000 HIV-positive adults and 73,000 controls, HIV-positivity was significantly and linearly associated with a high number of sexual partners (Chen *et al.*, 2007). The high prevalence of concurrent sexual relationships in sub-Saharan Africa compared to other regions, may be partly responsible for the high prevalence of HIV infection in the region (Halperin and Epstein, 2007; Mah and Halperin, 2010). The majority of truck drivers in a recent study in Nigeria, which has an overall HIV prevalence of 4.6%, had more than one sexual partner (Atilola *et al.*, 2010). However, condoms were used in more than 85% of the sexual contact away from home. In contrast, in the present study, 62% of the participants reported never using a condom, with 29% reporting occasional usage.

Sixty-two percent of the Shongwe participants were women, which confirms the finding that in most southern African countries, more women are receiving ART-treatment than men (Muula *et al.*, 2007). Men tend to attend clinics for treatment later than females (Grinsztejn *et al.*, 2011) and therefore potentially transmit HBV and/or HIV for a longer time before seeking treatment. This is supported by the fact that men in the Shongwe cohort were significantly older than women. Many factors drive the HIV epidemic in South Africa, including poor education about HIV risk, the disruption of family stability by migrant workers, polygamous marriages, patriarchal practices, and scenarios where women are unable to refuse intercourse (Couper, 2004). Temporary migration in rural South Africa impacts health and well-being of households (Collinson, 2010). Migrants, who may be exposed to high-risk sex, may infect partners when they return home (Collinson, 2010). Additionally, migrants returning less frequently are more likely to infect partners with HIV than those who return frequently (Collinson, 2010). Evidence suggests that the high rates of HIV prevalence in rural areas may be due to rural-rural migration, rather than urban-rural migration (Coffee *et al.*, 2005).

Participants in the study were impoverished, relatively poorly-educated and lived in a rural area. One third of the cohort had at most a grade 5 education. Almost three-quarters had at most a grade 11 education, with the majority of the remainder having a grade 12 education. Median individual income was R950 (US$100) per month, with a median household income of R1400 (US$150) per month. HIV prevalence is particularly high among poorly-educated women (Hargreaves *et al.*, 2007). Whilst the reduction of poverty is important, HIV incidence may

be lowered by increasing the level of education in the general population (Barnighausen *et al.*, 2007). Changes in behaviour at the population level, as a result of prevention strategies for HIV, have been of limited success (Hargreaves *et al.*, 2007). Although condom usage is increasing, risky behaviour, including early sexual debut and multiple sexual partners, is becoming more common (Hargreaves *et al.*, 2007). In some cases, risky sexual behaviour may be undertaken knowingly for various reasons, including wanting a baby, the promise of money or gifts, or coercion (Moore and Oppong, 2007; Moore *et al.*, 2007a,b).

By applying the *Taormina* criteria of OBI, which include PCR amplification of three genomic regions and determination of HBV viral load, almost two-thirds of the individuals in the Shongwe cohort were HBsAg-negative and HBV DNA-positive. This is the first southern African study to strictly apply the *Taormina* definition of OBI. Only three samples (1% of the cohort and 4% of the HBV-positive group) met the definition of "true occult", with the remaining 43 samples having HBV viral loads >200 IU/ml, and therefore "false occult" (Raimondo *et al.*, 2008). Apart from HBsAg-negativity, these 43 individuals were clinically indistinguishable from HBsAg-positive individuals, and we suggested and coined the term *HBsAg-covert* (Chapter 4) for such samples, because although it is important to distinguish between what was previously called true and false occult, the term "occult" should not be used synonymously for both.

The inability to detect HBsAg in the three true occult samples is a result of the low viral loads and low replication (Raimondo *et al.*, 2013). HBsAg-negativity in the HBsAg-covert group, which had a higher viral load, could, however, be a result of mutations. Firstly, these mutations may alter the S region in such a way that the HBsAg is not detected by commercial assays (Salisse and Sureau, 2009), or impair either virion or S antigen secretion (Huang *et al.*, 2012). A statistically significant increase in mutations in the major hydrophilic region in OBI in genotype B and C has been reported (Huang *et al.*, 2012; Kim *et al.*, 2013). Mutations P120A and Q129R/H, found in Shongwe HBV samples, lie within the HBsAg "a" determinant and may therefore result in non-optimal binding to commercial assays (Makondo *et al.*, 2012). The Y100C mutation, associated with HBsAg-negativity in blood donors (Gutiérrez *et al.*, 2004) and previously reported in OBI in subgenotype A1 (Motta-Castro *et al.*, 2008; Motta *et al.*, 2010), was also found in Shongwe samples (Makondo *et al.*, 2012). The common, classical G145R mutation (Carman *et al.*, 1990) was not found. In the present study, only the S174N mutation occurred more frequently in HBsAg-negative than in overt. Mutations at this position have been reported previously (Weinberger *et al.*, 2000; Song *et al.*, 2005; Ghaziasadi *et al.*, 2012). Mutations

S174N and V168A have been identified as HBV reactivation markers in HBV isolated from a serologically-negative HBV/HIV co-infected patient (Henke-Gendo *et al.*, 2008; Makondo *et al.*, 2012). Shongwe serum samples with higher viral loads were more likely to be infected with HBV with the V168A mutation (Makondo *et al.*, 2012). The profile of mutations differs by HBV genotype and HBsAg status (Allain *et al.*, 2009; Huang *et al.*, 2012; Makondo *et al.*, 2012; Kim *et al.*, 2013). Secondly, the G1862T mutation, found only in HBV from HBsAg-covert infected individuals from Shongwe, effects a valine to phenylalanine missense mutation in the pre-core region, which may interfere with signal peptide cleavage. The mutant HBeAg produced by this variant is not secreted, but rather accumulates in the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) (Chen *et al.*, 2008). This accumulation of HBeAg may interfere with the expression of HBsAg, which is also processed in the ERGIC (Makondo *et al.*, 2012). Thirdly, pre-S deletion mutants, found in four HBsAg-positive and one HBsAg-negative Shongwe sample/s (Makondo *et al.*, 2012), may prevent HBsAg secretion (Melegari *et al.*, 1994). The prevalence of mutations associated with the reduction of antibody binding to HBsAg and/or decreased secretion of HBsAg in HIV-positive patients reportedly ranges between 3% and 34% (Simon *et al.*, 2013; Thibault *et al.*, 2013).

The presence of HBsAg-negative infection and the circulation of mutant strains in the community has important public health implications. As HBsAg-covert infection and OBI are transmissible, either sexually, or via blood or organ donations, or can be reactivated as a result of immunosuppresion, they are public health risk factors (Raimondo *et al.*, 2013). A recent study in Thailand has shown that there is a very low risk of mother to child transmission from HIV-infected women with true OBI (Khamduang *et al.*, 2013). Whilst the clinical implications of HBsAg-negative infection are unclear, they can cause liver disease and HCC in the same way as HBsAg-positive infections (Raimondo *et al.*, 2013). In fact, in the present study, HBsAg-positive and HBsAg-negative groups were clinically indistinguishable (Chapter 4). There was no significant difference in APRI scores between HBsAg-negative and HBsAg-positive groups, with one in ten having elevated APRI scores, indicating advanced fibrosis. Moreover, HBV/HIV co-infected individuals had pre-S deletion mutants, which have also been found in HCC patients infected with HBV subgenotype A1 (Kew *et al.*, 2005; Skelton, 2010). This could be a risk factor for the development of HCC in HIV-positive individuals, who are living longer as a result of the immune reconstitution and the introduction of ART (Makondo *et al.*, 2012). Furthermore, of concern is the fact that drug resistance mutations were found in 10% of isolates (3 of 29) sequenced in the

Shongwe study, even though all patients were treatment naïve (Makondo *et al.*, 2012).

The high prevalence of HBsAg-negative infection in HIV-positive individuals, as shown by this and other studies (Mphahlele *et al.*, 2006; Firnhaber *et al.*, 2008, 2009), makes it imperative that this infection is detected. Prolonging life via ART is resulting in increased incidents of liver-related illness later in life (Thomas, 2006). If HBsAg-negative infection is detected earlier, appropriate treatment targeting HBV can be started sooner, and therapies containing lamivudine can be avoided, as long-term use of lamivudine can induce viral resistance in HBV, followed by treatment failure, in both immunocompetent and immunosuppressed patients (Honkoop *et al.*, 1997; Lindström *et al.*, 2004). The long-term consequences of ignoring HBV infection include the emergence of drug-resistant variants within populations, and severe clinical manifestations of the infection, including liver damage and HCC later in life. It is, therefore, important that NAT be incorporated into testing procedures, as this is the only method of accurately detecting HBsAg-negative infection. Whilst the "gold standard" of three genomic regions is recommended by the *Taormina* panel, a sensitive amplification of a single region of the HBV genome may suffice and this could be tested in a pilot study. The costs of neglecting to test for HBsAg-negative infection will accrue in the future, with the emergence of drug-resistant HBV and an increasing prevalence of HCC later in life.

The mutations found in HBV subgenotype A1 (Makondo *et al.*, 2012) differed from those found in other genotypes. Finding and characterizing these mutations required careful analysis and a thorough knowledge of sequence data. Since processing of sequence data is labour-intensive and time-consuming, tools, which can assist or automate some of the processes, can reduce the time required to analyze data and can streamline the process. The development of new bioinformatic tools facilitated these analyses at all levels, by automating some of the steps involved and avoiding repetition of tasks. Examples of these include the *Fragment Merger Tool*, the *Pipeline: TreeMail* and the *PadSeq Tool*. The *Mutation Reporter Tool* was developed and tested using sequence data from the BCP/PC region of the HBV genome, in this and other studies (Yousif and Kramvis, 2012; Gopalakrishnan *et al.*, 2013). As the tool is a general one, it has recently been used with amino acid data from the S region. The use of these tools means that results are more accurate and comparable, and procedures can easily be repeated to verify results.

The development of a database to store clinical and molecular data means that all members of the research group can easily access accurate, updated data. Storing clinical and molecular data in the same database allows for powerful querying. For example, it was possible to extract

baseline S region sequence data for all the female Shongwe participants who are HBsAg-negative, or for all the Shongwe overt participants with viral loads greater than 10,000 IU/ml. This level of querying means that subsets of the data can be analysed and compared comprehensively and statistically.

At present, database queries are not integrated with the various online tools. Future development could include pipelining of query results into the analysis tools. This would easily enable further powerful analyses to be undertaken. Adding reference sequences and data from additional studies to the database will further enhance the comparisons, which can be undertaken. Follow-up sequence data from the Shongwe cohort, as well as data from other African studies currently in progress, will be added to the database. This will allow both longitudinal analyses and larger, cross-cohort, analyses to be undertaken.

Next-generation sequencing (amplicon resequencing or ultra-deep pyrosequencing, UDPS) has become much more affordable and accessible in recent years. Future studies on the Shongwe cohort include UDPS of regions of interest of HBV. This is currently being piloted on samples from elsewhere in Africa. Simulated UDPS data consisting of 5000 samples of 300 bases each have successfully been processed by the *Mutation Reporter Tool*. Improvements to tools will be made, as required, to process UDPS data, in addition to developing new tools as the need may arise.

The unique combination of viral genotypes, human hosts and socio-economic factors in southern Africa requires further in-depth studies and targeted solutions, which consider the overall well-being of the people involved. The development of bioinformatic tools is a positive step in this direction.

Literature Cited

Adekunle, A. E., Oladimeji, A. A., Temi, A. P., Adeseye, A. I., Akinyeye, O. A. and Taiwo, R. H. (2011). Baseline CD4+ T lymphocyte cell counts, hepatitis B and C viruses seropositivity in adults with Human Immunodeficiency Virus infection at a tertiary hospital in Nigeria. *Pan African Medical Journal* **9**: 6.

Adewole, O. O., Anteyi, E., Ajuwon, Z., Wada, I., Elegba, F., Ahmed, P., Betiku, Y., Okpe, A., Eze, S., Ogbeche, T. and Erhabor, G. E. (2009). Hepatitis B and C virus co-infection in Nigerian patients with HIV infection. *Journal of Infection in Developing Countries* **3**(5): 369–375.

Ahn, S. H., Kramvis, A., Kawai, S., Spangenberg, H. C., Li, J., Kimbi, G., Kew, M., Wands, J. and Tong, S. (2003). Sequence variation upstream of precore translation initiation codon reduces hepatitis B virus e antigen production. *Gastroenterology* **125**: 1370–1378.

Alcantara, L. C., Cassol, S., Libin, P., Deforche, K., Pybus, O. G., Van Ranst, M., Galvao-Castro, B., Vandamme, A. M. and de Oliveira, T. (2009). A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. *Nucleic Acids Research* **37**(Web Server issue): W634–642.

Allain, J.-P., Belkhiri, D., Vermeulen, M., Crookes, R., Cable, R., Amiri, A., Reddy, R., Bird, A. and Candotti, D. (2009). Characterization of occult hepatitis B virus strains in South African blood donors. *Hepatology* **49**: 1868–1876.

Allen, M. I., Deslauriers, M., Andrews, C. W., Tipples, G. A., Walters, K. A., Tyrrell, D. L., Brown, N. and Condreay, L. D. (1998). Identification and characterization of mutations in hepatitis B virus resistant to lamivudine. Lamivudine Clinical Investigation Group. *Hepatology* **27**(6): 1670–1677.

Alter, H. J. and Blumerg, B. S. (1966). Further studies on a "new" human isoprecipitin system (Australia antigen). *Blood* **27**: 297–309.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**(3): 403–410.

Anthony, P. P., Ishak, K. G., Nayak, N. C., Poulsen, H. E., Scheuer, P. J. and Sobin, L. H. (1977). The morphology of cirrhosis: definition, nomenclature, and classification. *Bulletin of the World Health Organization* **55**: 521–540.

Arankalle, V. A., Gandhe, S. S., Borkakoty, B. J., Walimbe, A. M., Biswas, D. and Mahanta, J. (2010). A novel HBV recombinant (genotype I) similar to Vietnam/Laos in a primitive tribe in eastern India. *Journal of Viral Hepatitis* **17**: 501–510.

Arauz-Ruiz, P., Norder, H., Robertson, B. H. and Magnius, L. O. (2002). Genotype H: a new Amerindian genotype

of hepatitis B virus revealed in Central America. *Joural of General Virology* **83**: 2059–2073.

Aspinall, S. (1995). Effectiveness of a low-cost hepatitis B vaccine in the South African EPI programme: field trial results. *Epidemiological Comments* **22**: 8–10.

Atilola, G. O., Akpa, O. M. and Komolafe, I. O. (2010). HIV/AIDS and the long-distance truck drivers in south-west Nigeria: a cross-sectional survey on the knowledge, attitude, risk behaviour and beliefs of truckers. *Journal of Infection and Public Health* **3**(4): 166–178.

Atlamazoglou, V., Thireou, T., Hamodrakas, Y. and Spyrou, G. (2006). MetaBasis: a web-based database containing metadata on software tools and databases in the field of bioinformatics. *Applied Bioinformatics* **5**(3): 187–192.

Balogun, T. M., Emmanuel, S. and Ojerinde, E. F. (2012). HIV, Hepatitis B and C viruses' coinfection among patients in a Nigerian tertiary hospital. *Pan African Medical Journal* **12**: 100.

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological Reviews* **35**: 235–241.

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

Bar-Gal, G. K., Kim, M. J., Klein, A., Shin, D. H., Oh, C. S., Kim, J. W., Kim, T.-H., Kim, S. B., Grant, P. R., Pappo, O., Spigelman, M. and Shouval, D. (2012). Tracing hepatitis B virus to the 16th century in a Korean mummy. *Hepatology* **56**: 1671–1680.

Barin, F., Perrin, J., Chotard, J., Denis, F., N'Doye, R., Diop Mar, I., Chiron, J. P., Coursaget, P., Goudeau, A. and Maupas, P. (1981). Cross-sectional and longitudinal epidemiology of hepatitis B in Senegal. *Prog. Med. Virol.* **27**: 148–162.

Barker, L. F., Shulman, N., Murray, R., Hirschman, R. J., Ratner, F., Diefenbach, W. C. and Geller, H. M. (1970). Transmission of serum hepatitis. *The Journal of the American Medical Association* **276**: 841–844. doi: 10.1001/jama.1996.03540100085042.

Barnighausen, T., Hosegood, V., Timaeus, I. M. and Newell, M. L. (2007). The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. *AIDS* **21 Suppl 7**: 29–38.

Barth, R. E., Huijgen, Q., Tempelman, H. A., Mudrikova, T., Wensing, A. M. and Hoepelman, A. I. (2011). Presence of occult HBV, but near absence of active HBV and HCV infections in people infected with HIV in rural South Africa. *Journal of Medical Virology* **83**: 929–934.

Bell, T. G., Makondo, E., Martinson, N. and Kramvis, A. (2010). Hepatitis B virus infection in HIV-positive adults from rural South Africa: occult or overt - that is the question. *Poster presentation by Bell, T. G. at the 25th Annual Meeting on the Molecular Biology of Hepatitis B Viruses, October 9–13, 2010, Taipei, Taiwan*.

Bell, T. G., Makondo, E., Martinson, N. A. and Kramvis, A. (2012). Hepatitis B virus infection in human immunodeficiency virus infected Southern African adults: occult or overt - that is the question. *PLOS ONE* **7(10)**: e45750.

Benn, J. and Schneider, R. J. (1995). Hepatitis B virus HBx protein deregulates cell cycle checkpoint controls. *Proceedings of the National Academies of Sciences of the United States of America* **92**: 11215–11219.

Benson, D. A., Karsch-Mizrachi, I., Clark, K., Lipman, D. J., Ostell, J. and Sayers, E. W. (2012). GenBank. *Nucleic*

Acids Research **40**(D1): D48–D53.

Bickersteth Medical Society (1968). "Clinical Evening". *South African Medical Journal* **42**: 796.

Blumberg, B. S., Alter, H. J. and Visnich, S. (1965). A "new" antigen in leukemia sera. *Journal of the American Medical Association* **191**: 541–546.

Blumberg, B. S., Gerstley, B. J. S., Hungerford, D. A., Thomas London, W. and Sutnick, A. I. (1967). A serum antigen (Australia antigen) in Down's syndrome, leukemia, and hepatitis. *Annals of Internal Medicine* **66**: 924–931.

Bollyky, P. L., Rambaut, A., Harvey, P. H. and Holmes, E. C. (1996). Recombination between sequences of hepatitis B virus from different genotypes. *Journal of Molecular Evolution* **42**: 97–102.

Bornman, H. and Whittall, J. (1997). *Mpumalanga: the place where the sun rises: a tourist guide*. Rainbird Publishers, Nelspruit.

Botha, J. F., Ritchie, M. J., Dusheiko, G. M., Mouton, H. W. and Kew, M. C. (1984). Hepatitis B virus carrier state in black children in Ovamboland: role of perinatal and horizontal infection. *Lancet* **1**: 1210–1212.

Bowyer, S. M. and Sim, J. G. (2000). Relationships within and between genotypes of hepatitis B virus at points across the genome: footprints of recombination in certain isolates. *Journal of General Virology* **81**: 379–392.

Bowyer, S. M., van Staden, L., Kew, M. C. and Sim, J. G. (1997). A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa. *Journal of General Virology* **78 (Pt 7)**: 1719–1729.

Boyles, T. H. and Cohen, K. (2011). The prevalence of hepatitis B infection in a rural South African HIV clinic. *South African Journal of Medicine* **101**: 470–471.

Bozdayi, A. M., Uzunalimoglu, O., Turkyilmaz, A. R., Aslan, N., Sezgin, O., Sahin, T., Bozdayi, G., Cinar, K., Pai, S. B., Pai, R., Bozkaya, H., Karayalcin, S., Yurdaydin, C. and Schinazi, R. F. (2003). YSDD: a novel mutation in HBV DNA polymerase confers clinical resistance to lamivudine. *Journal of Viral Hepatitis* **10**: 256–265.

Bréchot, C., Hadchouel, M., Scotto, J., Fonck, M., Potet, F., Vyas, G. N. and Tiollais, P. (1981). State of hepatitis B virus DNA in hepatocytes of patients with hepatitis B surface antigen-positive and -negative liver diseases. *Proceedings of the National Academies of Sciences of the United States of America* **78**(6): 3906–3910.

Bréchot, C., Gozuacik, D., Murakami, Y. and Paterlini-Bréchot, P. (2000). Molecular bases for the development of hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC). *Seminars in Cancer Biology* **10**: 211–231.

Bredell, H., Williamson, C., Sonnenberg, P., Martin, D. J. and Morris, L. (1998). Genetic characterization of HIV type 1 from migrant workers in three South African gold mines. *AIDS Research and Human Retroviruses* **14**(8): 677–684.

Buckwold, V. E., Xu, Z., Chen, M., Yen, T. S. and Ou, J. H. (1996). Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication. *Journal of Virology* **70**: 5845–5851.

Buckwold, V. E., Xu, Z., Yen, T. S. and Ou, J. H. (1997). Effects of a frequent double-nucleotide basal core promoter mutation and its putative single-nucleotide precursor mutations on hepatitis B virus gene expression and replication. *Journal of General Virology* **78**: 2055–2065.

Bunz, F. (2008). *Principles of cancer genetics*. Springer. ISBN 978-1-4020-6783-9.

Burnett, R., François, G., Kew, M., Leroux-Roels, G., Meheus, A., Hoose, A. and Mphahlele, M. (2005). Hepatitis B virus and human immunodeficiency virus co-infection in sub-Saharan Africa: a call for further investigation. *Liver International* **25**: 201–213.

Burnett, R. J., Ngobeni, J. M., Francois, G., Hoosen, A. A., Leroux-Roels, G., Meheus, A. and Mphahlele, M. J. (2007). Increased exposure to hepatitis B virus infection in HIV-positive South African antenatal women. *Int J STD AIDS* **18**: 152–156.

Burnett, R. J., Kramvis, A., Dochez, C. and Meheus, A. (2012). An update after 16 years of hepatitis B vaccination in South Africa. *Vaccine* **30 Suppl 3**: 45–51.

Buti, M., Jardi, R., Cotrina, M., Rodriguez-Frias, F., Esteban, R. and Guardia, J. (1998). Transient emergence of hepatitis B variants in a patient with chronic hepatitis B resistant to lamivudine. *Journal of Hepatology* **28**(3): 510 – 513. ISSN 0168-8278. doi: 10.1016/S0168-8278(98)80327-9.

Cabrerizo, M., Bartolome, J., Caramelo, C., Barril, G. and Carreno, V. (2000). Molecular analysis of hepatitis B virus DNA in serum and peripheral blood mononuclear cells from hepatitis B surface antigen-negative cases. *Hepatology* **32**: 116–123.

Carman, W. F., Hadziyannis, S., Mcgarvey, M. J., Jacyna, M. R., Karayiannis, P., Makris, A. and Thomas, H. C. (1989). Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. *The Lancet* **334**: 588–591.

Carman, W. F., Zanetti, A. R., Karayiannis, P., Waters, J., Manzillo, G., Tanzi, E., Zuckerman, A. J. and Thomas, H. C. (1990). Vaccine-induced escape mutant of hepatitis B virus. *Lancet* **336**(8711): 325–329.

Carman, W. F., Fagan, E. A., Hadziyannis, S., Karayianni, P., Tassopoulos, N. C., Williams, R. and Thomas, H. C. (1991). Association of a precore genomic variant of hepatitis B virus with fulminant hepatitis. *Hepatology* **14**: 219–222.

Chadwick, D., Stanley, A., Sarfo, S., Appiah, L., Ankcorn, M., Foster, G., Schwab, U., Phillips, R. and Geretti, A. M. (2013). Response to antiretroviral therapy in occult hepatitis B and HIV co-infection in West Africa. *AIDS* **27**(1): 139–141.

Chang, S. F., Netter, H. J., Bruns, M., Schneider, R., Frölich, K. and Will, H. (1999). A new avian hepadnavirus infecting snow geese (*Anser caerulescens*) produces a significant fraction of virions containing single-stranded DNA. *Virology* **262**: 39–54.

Chauhan, R., Bell, T. G., Gopalakrishnan, D. and Kramvis, A. (2011). Molecular characterization of the surface overlapping polymerase region of hepatitis B virus (HBV) isolated from HIV infected southern Africans experiencing virological breakthrough (VBTH) during lamivudine based antiretroviral treatment. *Poster presentation by Chauhan, R. at the 26th Annual Meeting on the Molecular Biology of Hepatitis B Viruses, October 9–12, 2011, Orlando, Florida, USA*.

Chen, C. Y., Crowther, C., Kew, M. C. and Kramvis, A. (2008). A valine to phenylalanine mutation in the precore region of hepatitis B virus causes intracellular retention and impaired secretion of HBe-antigen. *Hepatology Research* **38**: 580–592.

Chen, L., Jha, P., Stirling, B., Sgaier, S. K., Daid, T., Kaul, R. and Nagelkerke, N. (2007). Sexual risk factors for HIV

infection in early and advanced HIV epidemics in sub-Saharan Africa: systematic overview of 68 epidemiological studies. *PLOS ONE* **2**(10): e1001.

Chen, M. T., Billaud, J. N., Sallberg, M., Guidotti, L. G., Chisari, F. V., Jones, J., Hughes, J. and Milich, D. R. (2004). A function of the hepatitis B virus precore protein is to regulate the immune response to the core antigen. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 14913–14918.

Chun, H. M., Roediger, M. P., Hullsiek, K. H., Thio, C. L., Agan, B. K., Bradley, W. P., Peel, S. A., Jagodzinski, L. L., Weintrob, A. C., Ganesan, A., Wortmann, G., Crum-Cianflone, N. F., Maguire, J. D., Landrum, M. L., Chun, H., Agan, B., Bradley, W., Kortepeter, M., Macalino, G., Decker, C., Ganesan, A., Warkentien, T., Whitman, T., Fraser, S., Hartzell, J., Ressner, R., Waterman, P., Weintrob, A., Wortmann, G., Zapor, M., Hsue, G., Johnson, A., Banks, S., Lalani, T., Maguire, J., Eberly, L., Lifson, A., Hullsiek, K., Roediger, M., Eggleston, C., Jagodzinski, L., O'Connell, R., Peel, S., Landrum, M., Okulicz, J., Merritt, S., Powers, J., Tramont, E., Bavaro, M., Crum-Cianflone, N. and Linfesty, M. (2012). Hepatitis B virus coinfection negatively impacts HIV outcomes in HIV seroconverters. *Journal of Infectious Diseases* **205**(2): 185–193.

Chung, R. T. (2006). Hepatitis C and B viruses: the new opportunists in HIV infection. *Topics in Antiviral Medicine* **14**: 78–âĂŞ83.

Coates, J. A., Cammack, N. and Jenkinson, H. J. (1992). The separated enantiomers of 2'-deoxy-3'-thiacytidine (BCII 189) both inhibit human immunodeficiency virus replication in vitro. *Antimicrobial Agents and Chemotherapy* **36**: 202–205.

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. and de Hoon, M. J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**(11): 1422–1423.

Coffee, M. P., Garnett, G. P., Mlilo, M., Voeten, H. A., Chandiwana, S. and Gregson, S. (2005). Patterns of movement and risk of HIV infection in rural Zimbabwe. *Journal of Infectious Diseases* **191 Suppl 1**: S159–167.

Collinson, M. A. (2010). Striving against adversity: the dynamics of migration, health and poverty in rural South Africa. *Global Health Action* **3**.

Combe, P., La Ruche, G., Bonard, D., Ouassa, T., Faye-Kette, H., Sylla-Koko, F. and Dabis, F. (2001). Hepatitis B and C infections, human immunodeficiency virus and other sexually transmitted infections among women of childbearing age in CÃťte d'Ivoire, West Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **95**(5): 493–496.

Couper, I. D. (2004). Reframing the HIV/AIDS debate in developing countries I: setting the scene. *Rural and Remote Health* **4**(2): 280.

Dane, D. S., Cameron, C. H. and Briggs, M. (1970). Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *The Lancet* **295**: 695–698.

Davis, L. G., Weber, D. J. and Lemon, S. M. (1989). Horizontal transmission of hepatitis B virus. *Lancet* **1**: 889–893.

de Franchis, R., Hadengue, A., Lau, G., Lavanchy, D., Lok, A., McIntyre, N., Mele, A., Paumgartner, G., Pietrangelo, A., Rodes, J., Rosenberg, W. and Valla, D. (2003). EASL International Consensus Conference on Hepatitis B. 13-14 September, 2002 Geneva, Switzerland. Consensus statement (long version). *Journal of Hepatology* **39 Suppl**

1: 3–25.

de Oliveira, T., Deforche, K., Cassol, S., Salminen, M., Paraskevis, D., Seebregts, C., Snoeck, J., van Rensburg, E. J., Wensing, A. M., van de Vijver, D. A., Boucher, C. A., Camacho, R. and Vandamme, A. M. (2005). An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* **21**(19): 3797–3800.

Dejean, A., Bréchot, C., Tiollais, P. and Wain-Hobson, S. (1983). Characterization of integrated hepatitis B viral DNA cloned from a human hepatoma and the hepatoma-derived cell line PLC/PRF/5. *Proceedings of the National Academies of Sciences of the United States of America* **80**: 2505–2509.

Delaney, W. E., Locarnini, S. and Shaw, T. (2001). Resistance of hepatitis B virus to antiviral drugs: current aspects and directions for future investigation. *Antiviral Chemistry and Chemotherapy* **107**: 449–445.

Dickens, C., Kew, M. C., Purcell, B. H. and Kramvis, A. (2013). Occult Hepatitis B Virus infection the Chacma Baboon (*Papio ursinus orientalis*). *Emerging Infectious Diseases* doi: dx.doi.org/10.3201/eid1904.121107.

DiMattia, M. A., Watts, N. R., Stahl, S. J., Grimes, J. M., Steven, A. C., Stuart, D. I. and Wingfield, P. T. (2013). Antigenic switching of hepatitis B virus by alternative dimerization of the capsid protein. *Structure* **21**(1): 133–142.

Diop-Ndiaye, H., Toure-Kane, C., Etard, J. F., Lo, G., Diaw, P., Ngom-Gueye, N. F., Gueye, P. M., Ba-Fall, K., Ndiaye, I., Sow, P. S., Delaporte, E. and Mboup, S. (2008). Hepatitis B, C seroprevalence and delta viruses in HIV-1 Senegalese patients at HAART initiation (retrospective study). *Journal of Medical Virology* **80**(8): 1332–1336.

Doong, S. L., Tsai, C. H., Schinazi, R. F., Liotta, D. C. and Cheng, Y. C. (1991). Inhibition of the replication of hepatitis B virus in vitro by 2',3'-dideoxy-3'-thiacytidine and related analogues. *Proceedings of the National Academies of Sciences of the United States of America* **88**: 8495–8499.

Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* **8**(3): 186–194.

Ewing, B., Hillier, L., Wendl, M. C. and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* **8**(3): 175–185.

Fares, M. A. and Holmes, E. C. (2002). A revised evolutionary history of hepatitis B virus (HBV). *Journal of Molecular Evolution* **54**: 807–814.

Farrell, J. and Denstag, J. (2002). Epidemiology and prevention of hepatitis B virus infection. In: **Lai, C. and Locarnini, S.** (Eds.), *Hepatitis B virus*, pp. 115–130. International Medical Press, London, UK, first edition.

Felsenstein, J. (1989). PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics* **5**: 164–166.

Findlay, G. M. and MacCallum, F. O. (1937). Note on acute hepatitis and and yellow fever immunization. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **31**: 297–308.

Firnhaber, C., Reyneke, A., Schulze, D., Malope, B., Maskew, M., MacPhail, P., Sanne, I. and Di Bisceglie, A. (2008). The prevalence of hepatitis B co-infection in a South African urban government HIV clinic. *South African Medical Journal* **98**: 541–544.

Firnhaber, C., Viana, R., Reyneke, A., Schultze, D., Malope, B., Maskew, M., Di Bisceglie, A., MacPhail, P., Sanne, I. and Kew, M. (2009). Occult hepatitis B virus infection in patients with isolated core antibody and HIV co-infection

in an urban clinic in Johannesburg, South Africa. *Int. J. Infect. Dis.* **13**: 488–492.

Firnhaber, C., Chen, C. Y., Evans, D., Maskew, M., Schulz, D., Reyneke, A. and Kramvis, A. (2012). Prevalence of hepatitis B virus (HBV) co-infection in HBV serologically-negative South African HIV patients and retrospective evaluation of the clinical course of mono- and co-infection. *Int. J. Infect. Dis.* **16**: e268–272.

François, G., Dochez, C., Mphahlele, J., Burnett, R., Van Hal, G. and Meheus, A. (2008). Hepatitis B vaccination in Africa: mission accomplished? *The Southern African Journal of Epidemiology and Infection* **23**: 24–28.

Fu, L. and Cheng, Y. C. (1998). Role of additional mutations outside the YMDD motif of hepatitis B virus polymerase in L(-)SddC (3TC) resistance. *Biochemical Pharmacology* **55**(10): 1567–1572.

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

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

Ganem, D. and Prince, A. M. (2004). Hepatitis B Virus Infection - Natural History and Clinical Consequences. *New England Journal of Medicine* **350**(11): 1118–1129.

Ganem, D. and Varmus, H. E. (1987). The molecular biology of the hepatitis B viruses. *Annu. Rev. Biochem.* **56**: 651–693.

GenBank (2013). GenBank Release Notes. *ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt*. Accessed 20 February 2013.

Georgi-Geisberger, P., Berns, H., Loncarevic, I. F., Yu, Z. Y., Tang, Z. Y., Zentgraf, H. and Schröder, C. (1992). Mutations on free and integrated hepatitis B virus DNA in a hepatocellular carcinoma: footprints of homologous recombination. *Oncology* **49**: 386–395.

Ghaziasadi, A., Ziaee, M., Norouzi, M., Malekzadeh, R., Alavian, S. M., Saberfar, E., Judaki, M. A., Ghamari, S., Khedive, A., Namazi, A., Rahimnia, R. and Jazayeri, S. M. (2012). The prevalence of hepatitis B virus surface antigen (HBsAg) variations and correlation with the clinical and serologic pictures in chronic carriers from Khorasan Province, North-East of Iran. *Acta Medica Iranica* **50**(4): 265–272.

Gibas, C. and Jambeck, P. (2001). *Developing Bioinformatics Computer Skills*. O'Reilly & Associates, Sebastopol, first edition.

Gilbert, C. and Feschotte, C. (2010). Genomic Fossils Calibrate the Long-Term Evolution of Hepadnaviruses. *PLOS Biology* **8**: e1000495.

Gilbert, D. (2004). Bioinformatics software resources. *Briefings in Bioinformatics* **5**(3): 300–304.

Goodman, Z. (2002). Histopathology of hepatitis B virus infection. In: **Lai, C. and Locarnini, S.** (Eds.), *Hepatitis B virus*, pp. 131–143. International Medical Press, London, UK, first edition.

Gopalakrishnan, D., Bell, T. G. and Kramvis, A. (2011). Quasispecies evolution in the basic core promoter/precore region of hepatitis B virus (HBV) isolated from HBV/HIV co-infected South Africans following initiation of anti-retroviral treatment. *Oral presentation by Gopalakrishnan, D. at the 26th Annual Meeting on the Molecular Biology of Hepatitis B Viruses, October 9–12, 2011, Orlando, Florida, USA*.

Gopalakrishnan, D., Keyter, M., Shenoy, K. T., Balakumaran, L. K., Barathi, T., Thomas, V., Vinayakumar, K. R.,

Panackel, C., Korah, A. T., Nair, R. and Kramvis, A. (2013). Predominance of Subgenotype A1 of Hepatitis B Virus in Liver Disease Patients from the State of Kerala, Southern India. *World Journal of Gastroenterology* **(Submitted)**.

Goudsmit, J. (1997). *Viral Sex: The Nature of AIDS*. Oxford University Press, New York, New York.

Grethe, S., Heckel, J. O., Rietschel, W. and Hufert, F. T. (2000). Molecular epidemiology of hepatitis B virus variants in nonhuman primates. *Journal of Virology* **74**: 5377–5381.

Grinsztejn, B., Smeaton, L., Barnett, R., Klingman, K., Hakim, J., Flanigan, T., Kumarasamy, N., Campbell, T. and Currier, J. (2011). Sex-associated differences in pre-antiretroviral therapy plasma HIV-1 RNA in diverse areas of the world vary by CD4(+) T-cell count. *Antiviral Therapy (London)* **16**(7): 1057–1062.

Guidotti, L. G., Matzke, B., Pasquinelli, C., Shoenberger, J. M., Rogler, C. E. and Chisari, F. V. (1996). The hepatitis B virus (HBV) precore protein inhibits HBV replication in transgenic mice. *Journal of Virology* **70**: 7056–7061.

Guo, H., Mason, W. S., Aldrich, C. E., Saputelli, J. R., Miller, D. S., Jilbert, A. R. and Newbold, J. E. (2005). Identification and characterization of Avihepadnaviruses isolated from exotic anseriformes maintained in captivity. *Journal of Virology* **79**: 2729–2742.

Gust, I. D., Burrell, C. J., Coulepis, A. G., Robinson, W. S. and Zuckerman, A. J. (1986). Taxonomic classification of human hepatitis B virus. *Intervirology* **25**: 14–29.

Gutiérrez, C., Devesa, M., Loureiro, C. L., Leon, G., Liprandi, F. and Pujol, F. H. (2004). Molecular and serological evaluation of surface antigen negative hepatitis B virus infection in blood donors from Venezuela. *Journal of Medical Virology* **73**(2): 200–207.

H. P. W. (1970). Book Review. *South African Medical Journal* **44**: 164.

Halperin, D. T. and Epstein, H. (2007). Why is HIV prevalence so severe in southern Africa? *South African Journal of HIV Medicine* **8**(1): 19–25.

Harania, R. S., Karuru, J., Nelson, M. and Stebbing, J. (2008). HIV, hepatitis B and hepatitis C coinfection in Kenya. *AIDS* **22**(10): 1221–1222.

Hargreaves, J. R., Bonell, C. P., Morison, L. A., Kim, J. C., Phetla, G., Porter, J. D., Watts, C. and Pronyk, P. M. (2007). Explaining continued high HIV prevalence in South Africa: socioeconomic factors, HIV incidence and sexual behaviour change among a rural cohort, 2001-2004. *AIDS* **21 Suppl 7**: 39–48.

Harries, E. H. R., Mitman, M. and Taylor, I. (1940). *Clinical practice in infectious diseases*. E. & S. Livingstone, Ltd., Edinburgh, first edition.

Harries, E. H. R., Mitman, M. and Taylor, I. (1951). *Clinical practice in infectious diseases*. E. & S. Livingstone, Ltd., Edinburgh, fourth edition.

Hawkins, C., Christian, B., Ye, J., Nagu, T., Aris, E., Chalamilla, G., Spiegelman, D., Mugusi, F., Mehta, S. and Fawzi, W. (2012). Prevalence of Hepatitis B co-infection and response to antiretroviral therapy among HIV-infected patients in urban Tanzania. *AIDS* .

Hayer, J., Jadeau, F., Deleage, G., Kay, A., Zoulim, F. and Combet, C. (2013). HBVdb: a knowledge database for Hepatitis B Virus. *Nucleic Acids Research* **41**(Database issue): D566–570.

Heiberg, I. L., Hoegh, M., Ladelund, S., Niesters, H. G. and Hogh, B. (2010). Hepatitis B virus DNA in saliva from

children with chronic hepatitis B infection: implications for saliva as a potential mode of horizontal transmission. *Pediatr. Infect. Dis. J.* **29**: 465–467.

Henke-Gendo, C., Amini-Bavil-Olyaee, S., Challapalli, D., Trautwein, C., Deppe, H., Schulz, T. F., Heim, A. and Tacke, F. (2008). Symptomatic hepatitis B virus (HBV) reactivation despite reduced viral fitness is associated with HBV test and immune escape mutations in an HIV-coinfected patient. *Journal of Infectious Diseases* **198**(11): 1620–1624.

Hino, O., Tabata, S. and Hotta, Y. (1991). Evidence for increased in vitro recombination with insertion of human hepatitis B virus DNA. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 9248–9252.

Hoffmann, C. J., Minkah, N., Leipzig, J., Wang, G. P., Arens, M. Q., Tebas, P. and Bushman, F. D. (2007). DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic Acids Research* **35**: e91.

Hoffmann, C. J., Charalambous, S., Martin, D. J., Innes, C., Churchyard, G. J., Chaisson, R. E., Grant, A. D., Fielding, K. L. and Thio, C. L. (2008). Hepatitis B Virus Infection and Response to Antiretroviral Therapy (ART) in a South African ART Program. *Clinical Infectious Diseases* **47**: 1479–1485. doi: 10.1086/593104.

Hogg, R. V. and Tanis, E. A. (2005). *Probability and Statistical Inference*. Printice Hall, seventh edition.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. and van de Pol, S. (1982). Rapid evolution of RNA genomes. *Science* **215**: 1577–1585.

Hollinger, F. B. and Sood, G. (2010). Occult hepatitis B virus infection: a covert operation. *Journal of Viral Hepatitis* **17**: 1–15.

Honkoop, P., Niesters, H. G., de Man, R. A., Osterhaus, A. D. and Schalm, S. W. (1997). Lamivudine resistance in immunocompetent chronic hepatitis B: indicence and patterns. *Journal of Hepatology* **26**: 1393–1395.

Hoofnagle, J. H., Seeff, L. B., Bales, Z. B. and Zimmerman, H. J. (1978). Type B hepatitis after transfusion with blood containing antibody to hepatitis B core antigen. *New England Journal of Medicine* **298**: 1379–1383.

Hu, K. Q. (2002). Occult hepatitis B virus infection and its clinical implications. *Journa of Viral Hepatitis* **9**: 243–257.

Huang, C. H., Yuan, Q., Chen, P. J., Zhang, Y. L., Chen, C. R., Zheng, Q. B., Yeh, S. H., Yu, H., Xue, Y., Chen, Y. X., Liu, P. G., Ge, S. X., Zhang, J. and Xia, N. S. (2012). Influence of mutations in hepatitis B virus surface protein on viral antigenicity and phenotype in occult HBV strains from blood donors. *Journal of Hepatology* **57**(4): 720–729.

Hunt, C. M., McGill, J. M., Allen, M. I. and Condreay, L. D. (2000). Clinical relevance of hepatitis B viral mutations. *Hepatology* **31**: 1037–1044. doi: 10.1053/he.2000.6709.

Hunt, L. T. and Dayhoff, M. O. (1983). Margaret O. Dayhoff 1925-1983. *DNA* **2**: 97–98.

Idoko, J., Meloni, S., Muazu, M., Nimzing, L., Badung, B., Hawkins, C., Sankale, J. L., Ekong, E., Murphy, R., Kanki, P. and Thio, C. L. (2009). Impact of hepatitis B virus infection on human immunodeficiency virus response to antiretroviral therapy in Nigeria. *Clinical Infectious Diseases* **49**(8): 1268–1273.

Ijaz, S., Arnold, C., Dervisevic, S., Mechurova, J., Tatman, N., Tedder, R. S. and Naoumov, N. V. (2008). Dynamics of Lamivudine-resistant hepatitis B virus during Adefovir monotherapy versus Lamivudine plus Adefovir combination therapy. *Journal of Medical Virology* **80**: 1160–1170.

Ilboudo, D., Karou, D., Nadembega, W. M., Savadogo, A., Djeneba, O., Pignatelli, S., Pietra, V., Bere, A., Simpore,

J. and Traore, A. S. (2007). Prevalence of human herpes virus-8 and hepatitis B virus among HIV seropositive pregnant women enrolled in the Mother-to-Child HIV Transmission Prevention Program at Saint Camille Medical Centre in Burkina Faso. *Pakistan Journal of Biological Sciences* **10**(17): 2831–2837.

Ilboudo, D., Simpore, J., Ouermi, D., Bisseye, C., Sagna, T., Odolini, S., Buelli, F., Pietra, V., Pignatelli, S., Gnoula, C., Nikiema, J. B. and Musumeci, S. (2010). Towards the complete eradication of mother-to-child HIV/HBV coinfection at Saint Camille Medical Centre in Burkina Faso, Africa. *Brazilian Journal of Infectious Diseases* **14**(3): 219–224.

Jilbert, A., Burrell, C., Triatni, M. and Kann, M. (2002). Hepatitis B virus replication. In: **Lai, C. and Locarnini, S.** (Eds.), *Hepatitis B virus*, pp. 43–53. International Medical Press, London, UK, first edition.

Junker-Niepmann, M., Bartenschlager, R. and Schaller, H. (1990). A short cis-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. *The EMBO Journal* **9**: 3389–3396.

Kania, D., Sangare, L., Sakande, J., Koanda, A., Nebie, Y. K., Zerbo, O., Combassere, A. W., Guissou, I. P. and Rouet, F. (2009). A new strategy to improve the cost-effectiveness of human immunodeficiency virus, hepatitis B virus, hepatitis C virus, and syphilis testing of blood donations in sub-Saharan Africa: a pilot study in Burkina Faso. *Transfusion* **49**(10): 2237–2240.

Kann, M. (2002). Structural and molecular virology. In: **Lai, C. and Locarnini, S.** (Eds.), *Hepatitis B virus*, pp. 9–22. International Medical Press, London, UK, first edition.

Kao, J.-H. and Chen, D.-S. (2002). The natural history of hepatitis B virus infection. In: **Lai, C. and Locarnini, S.** (Eds.), *Hepatitis B virus*, pp. 161–172. International Medical Press, London, UK, first edition.

Kao, J. H., Chen, P. J., Lai, M. Y. and Chen, D. S. (2003). Basal core promoter mutations of hepatitis B virus increase the risk of hepatocellular carcinoma in hepatitis B carriers. *Gastroenterology* **124**: 327–334.

Kapembwa, K. C., Goldman, J. D., Lakhi, S., Banda, Y., Bowa, K., Vermund, S. H., Mulenga, J., Chama, D. and Chi, B. H. (2011). HIV, Hepatitis B, and Hepatitis C in Zambia. *Journal of Global Infectious Diseases* **3**(3): 269–274.

Kew, M. C. (1996). Progress towards the comprehensive control of hepatitis B in Africa: a view from South Africa. *Gut* **38 Suppl 2**: S31–S36.

Kew, M. C., Kramvis, A., Yu, M. C., Arakawa, K. and Hodkinson, J. (2005). Increased hepatocarcinogenic potential of hepatitis B virus genotype A in Bantu-speaking sub-saharan Africans. *Journal of Medical Virology* **75**(4): 513–521.

Kfutwah, A. K., Tejiokem, M. C. and Njouom, R. (2012). A low proportion of HBeAg among HBsAg-positive pregnant women with known HIV status could suggest low perinatal transmission of HBV in Cameroon. *Virolology Journal* **9**: 62.

Khamduang, W., Ngo-Giang-Huong, N., Gaudy-Graffin, C., Jourdain, G., Suwankornsakul, W., Jarupanich, T., Chalermpolprapa, V., Nanta, S., Puarattana-Aroonkorn, N., Tonmat, S., Lallemant, M., Goudeau, A. and Sirirungsi, W. (2013). Prevalence, risk factors and impact of isolated antibody to hepatitis B core antigen and occult HBV infection in HIV-1 infected pregnant women. *Clinical Infectious Diseases* doi: 10.1093/cid/cit166.

Kiire, C. F. (1996). The epidemiology and prophylaxis of hepatitis B in sub-Saharan Africa: a view from tropical and

subtropical Africa. *Gut* **38 Suppl 2**: 5–12.

Kim, H., Lee, S. A., Kim, D. W., Lee, S. H. and Kim, B. J. (2013). Naturally occurring mutations in large surface genes related to occult infection of hepatitis B virus genotype C. *PLOS ONE* **8**(1): e54486.

Kim, H. N., Scott, J., Cent, A., Cook, L., Morrow, R. A., Richardson, B., Tapia, K., Jerome, K. R., Lule, G., John-Stewart, G. and Chung, M. H. (2011). HBV lamivudine resistance among hepatitis B and HIV coinfected patients starting lamivudine, stavudine and nevirapine in Kenya. *Journal of Viral Hepatitis* **18**(10): e447–452.

Kimbi, G. C., Kramvis, A. and Kew, M. C. (2004). Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa. *Journal of General Virology* **85**(Pt 5): 1211–1220.

Kottilil, S., Jackson, J. O. and A., P. M. (2005). Hepatitis B and hepatitis C in HIV-infection. *Indian Journal of Medical Research* **121**: 424–450.

Kramvis, A. and Kew, M. C. (1998). Structure and function of the encapsidation signal of hepadnaviridae. *Journal of Viral Hepatitis* **5**: 357–367.

Kramvis, A. and Kew, M. C. (1999). The core promoter of hepatitis B virus. *Journal of Viral Hepatitis* **6**: 415–427.

Kramvis, A. and Kew, M. C. (2005). Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy. *Journal of Viral Hepatitis* **12**: 456–464.

Kramvis, A. and Kew, M. C. (2007). Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes. *Hepatology Research* **37**: S9–S19.

Kramvis, A., Bukofzer, S., Kew, M. C. and Song, E. (1997). Nucleic acid sequence analysis of the precore region of hepatitis B virus from sera of southern African black adult carriers of the virus. *Hepatology* **25**: 235–240.

Kramvis, A., Kew, M. and François, G. (2005). Hepatitis B virus genotypes. *Vaccine* **23**: 2409–2423.

Kramvis, A., Arakawa, K., Yu, M. C., Nogueira, R., Stram, D. O. and Kew, M. C. (2008). Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. *Journal of Medical Virology* **80**: 27–46.

Kumar, S. and Dudley, J. (2007). Bioinformatics software for biologists in the genomics era. *Bioinformatics* **23**(14): 1713–1717.

Lai, C. (2002). Discovery and classification. In: **Lai, C. and Locarnini, S.** (Eds.), *Hepatitis B virus*, pp. 43–53. International Medical Press, London, UK, first edition.

Lai, C. and Locarnini, S. (Eds.) (2002). *Hepatitis B virus*. International Medical Press, London, UK, first edition.

Lamberts, C., Nassal, M., Velhagen, I., Zentgraf, H. and Schroder, C. H. (1993). Precore-mediated inhibition of hepatitis B virus progeny DNA synthesis. *Journal of Virology* **67**: 3756–3762.

Lanford, R. E., Chavez, D., Brasky, K. M., Burns, R. B. and Rico-Hesse, R. (1998). Isolation of a hepadnavirus from the woolly monkey, a New World primate. *Proceedings of the National Academies of Sciences of the United States of America* **95**: 5757–5761.

Laurent, C., Bourgeois, A., Mpoudi-Ngole, E., Kouanfack, C., Ciaffi, L., Nkoue, N., Mougnutou, R., Calmy, A., Koulla-Shiro, S., Ducos, J. and Delaporte, E. (2010). High rates of active hepatitis B and C co-infections in HIV-1

infected Cameroonian adults initiating antiretroviral therapy. *HIV Med.* **11**(1): 85–89.

Leung, N. W. Y., Lai, C. L., Chang, T. T., Guan, R., Lee, C. M., Ng, K. Y., Lim, S. G., Wu, P. C., Dent, J. C., Edmundson, S., Condreay, L. D. and Chien, R. N. (2001). Extended lamivudine treatment in patients with chronic hepatitis B enhances hepatitis B e antigen serconversion rates: results after 3 years of therapy. *Hepatology* **33**: 1527–1532.

Levy, V. and Grant, M. (2006). Antiretroviral therapy for hepatitis B virus-HIV-coinfected patients: Promises and Pitfalls. *Clinical Infectious Diseases* **43**: 904–910.

Li, J. S., Tong, S. P., Wen, Y. M., Vitvitski, L., Zhang, Q. and Trepo, C. (1993). Hepatitis B virus genotype A rarely circulates as an HBe-minus mutant: possible contribution of a single nucleotide in the precore region. *Journal of Virology* **67**(9): 5402–5410.

Libin, P., Deforche, K., Laethem, K., Camacho, R. and Vandamme, A. (2007). RegaDB: An Open Source, Community-Driven HIV Data and Analysis Management Environment. In: *Fifth European HIV Drug Resistance Workshop, Cascais, Portugal*.

Lin, C.-L. and Kao, J.-H. (2010). Clinical Implications of Hepatitis B Virus Variants. *Journal of the Formosan Medical Association* **109**: 321–325.

Lindström, A., Odeberg, J. and Albert, J. (2004). Pyrosequencing for detection of Lamivudine-resistant hepatitis B Virus. *Journal of Clinical Microbiology* **42**: 4788–4795.

Locarnini, S., McMillan, J. and Bartholomeusz, A. (2003). The hepatitis B virus and common mutants. *Seminars in Liver Disease* **23**: 5–20.

Lok, A. S., Akarca, U. and Greene, S. (1994). Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal. *Proceedings of the National Academies of Sciences of the United States of America* **91**: 4077–4081.

Lok, A. S. F., Conjeevaram, H. S. and Negro, F. (2007). Hepatitis B and D. In: **Schiff, E. R., Sorrell, M. F. and Maddrey, W. C.** (Eds.), *Schiff's Diseases of the Liver*, pp. 746–806. Lippincott Williams and Wilkins, Philadelphia, tenth edition.

Lukhwareni, A., Burnett, R. J., Selabe, S. G., Mzileni, M. O. and Mphahlele, M. J. (2009). Increased detection of HBV DNA in HBsAg-positive and HBsAg-negative South African HIV/AIDS patients enrolling for highly active antiretroviral therapy at a Tertiary Hospital. *Journal of Medical Virology* **81**: 406–412.

Lürman, A. (1885). Eine Icterusepidemie. *Berliner Klinische Wochenschrift* **22**: 20–23.

MacCallum, F. O. (1947). Editorial: Homologous serum hepatitis. *The Lancet* **2**: 691–692.

MacCallum, F. O. (1953). Hepatitis. *British Medical Bulletin* **9**: 221–226.

MacCallum, F. O. (1971). A review of the present position of Australia antigen. *Postgraduate Medical Journal* **47**: 481–483.

MacCallum, F. O. (1972a). 1971 International Symposium on Viral Hepatitis: Historial Perspectives. *Canadian Medical Association Journal, Special Issue* **106**: 423–426.

MacCallum, F. O. (1972b). Early studies of viral hepatitis. *British Medical Bulletin* **28**: 105–108.

MacDonald, D. M., Holmes, E. C., Lewis, J. C. M. and Simmonds, P. (2000). Detection of Hepatitis B Virus Infection in Wild-Born Chimpanzees (*Pan troglodytes verus*): Phylogenetic Relationships with Human and Other Primate Genotypes. *Journal of Virology* **74**: 4253–4257. doi: 10.1128/JVI.74.9.4253-4257.2000.

MacNalty, A. S. (1938). *Annual report for the Chief Medical Officer of the Ministry of Health for the Year 1937*. HMSO, London.

MacNalty, A. S. (1939). *Annual report for the Chief Medical Officer of the Ministry of Health for the Year 1938*. HMSO, London.

Magnius, L. O. and Norder, H. (1995). Genotypes and Molecular Epidemiology of the Hepatitis B Virus as Reflected by Sequence Variability of the S-Gene. *Intervirology* **38**: 24–34.

Mah, T. L. and Halperin, D. T. (2010). Concurrent sexual partnerships and the HIV epidemics in Africa: evidence to move forward. *AIDS and Behaviour* **14**(1): 11–16.

Makondo, E., Bell, T. G. and Kramvis, A. (2010). Genotyping and molecular characterization of hepatitis B virus (HBV) from human immunodeficiency virus infected South African patients. *Poster presentation by Makondo, E. at the 25th Annual Meeting on the Molecular Biology of Hepatitis B Viruses, October 9–13, 2010, Taipei, Taiwan*.

Makondo, E., Bell, T. G. and Kramvis, A. (2012). Genotyping and molecular characterization of hepatitis B virus from human immunodeficiency virus-infected individuals in Southern Africa. *PLOS ONE* **7(9)**: e46345.

Makuwa, M., Souquière, S., Telfer, P., Bourry, O., Rouquet, P., Kazanji, M., Roques, P. and Simon, F. (2006). Hepatitis viruses in non-human primates. *Journal of Medical Primatology* **35**: 384–387. doi: 10.1111/j.1600-0684.2006.00163.x.

Mamadou, S., Ide, M., Maazou, A. R., Aoula, B., Labo, S. and Bozari, M. (2012). HIV infection and hepatitis B seroprevalence among antenatal clinic attendees in Niger, West Africa. *HIV AIDS (Auckland, N.Z.)* **4**: 1–4.

Marion, P. L., Oshiro, L. S., Regnery, D. C., Scullard, G. H. and Robinson, W. S. (1980). A virus in Beechey ground squirrels that is related to hepatitis B virus of humans. *Proceedings of the National Academies of Sciences of the United States of America* **77**: 2941–2945.

Marshall, J. (1949). Correspondence. *South African Medical Journal* **23**: 665.

Mason, W. S., Seal, G. and Summers, J. (1980). Virus of Pekin ducks with structural and biological relatedness to human hepatitis B virus. *Journal of Virology* **36**: 829–836.

Matee, M. I., Magesa, P. M. and Lyamuya, E. F. (2006). Seroprevalence of human immunodeficiency virus, hepatitis B and C viruses and syphilis infections among blood donors at the Muhimbili National Hospital in Dar es Salaam, Tanzania. *BMC Public Health* **6**: 21.

Matthews, G., Manzini, P., Hu, Z., Khabo, P., Maja, P., Matchaba, G., Sangweni, P., Metcalf, J., Pool, N., Orsega, S. and Emery, S. (2011). Impact of lamivudine on HIV and hepatitis B virus-related outcomes in HIV/hepatitis B virus individuals in a randomized clinical trial of antiretroviral therapy in southern Africa. *AIDS* **25**: 1727–1735.

Mavenyengwa, R. T., Moyo, S. R. and Nordbo, S. A. (2010). Streptococcus agalactiae colonization and correlation with HIV-1 and HBV seroprevalence in pregnant women from Zimbabwe. *European Journal of Obstetrics & Gynecology and Reproductive Biology* **150**(1): 34–38.

Mayans, M. V., Hall, A., Inskip, H., Chotard, J., Lindsay, S., Alonso, P., Coromina, E., Mendy, M. and Whittle, H. (1990). Risk factors for transmission of hepatitis B virus to Gambian children. *The Lancet* **336**: 1107–1109. doi: 10.1016/0140-6736(90)92580-B.

Mayaphi, S. H., Roussow, T. M., Masemola, D. P., Olorunju, S. A., Mphahlele, M. J. and Martin, D. J. (2012). HBV/HIV co-infection: the dynamics of HBV in South African patients with AIDS. *South African Medical Journal* **102**(3 Pt 1): 157–162.

Mayerat, C., Mantegani, A. and Frei, P. C. (1999). Does hepatitis B virus (HBV) genotype influence the clinical outcome of HBV infection? *Journal of Viral Hepatitis* **6**: 299–304.

McMahon, B. J. (2009). The natural history of chronic hepatitis B virus infection. *Hepatology* **49**: 45–55.

McMahon, M. A., Jilek, B. L., Brennan, T. P., Shen, L., Zhou, Y., Wind-Rotolo, M., Xing, S., Bhat, S., Hale, B., Hegarty, R., Chong, C. R., Liu, J. O., Siliciano, R. F. and Thio, C. L. (2007). The HBV drug Entecavir – effects on HIV-1 replication and resistance. *The New England Journal of Medicine* **356**: 2614–2621. PDF version not available. Sunday 26 October 2008.

Melegari, M., Bruno, S. and Wands, J. R. (1994). Properties of hepatitis B virus pre-S1 deletion mutants. *Virology* **199**(2): 292–300.

Mimms, L. (1995). Hepatitis B virus escape mutants: pushing the envelope of chronic hepatitis B virus infection. *Hepatology* **21**: 884–887.

Mimms, L. T., Floreani, M., Tyner, J., Whitters, E., Rosenlof, R., Wray, L., Goetze, A., Sarin, V. and Eble, K. (1990). Discrimination of hepatitis B virus (HBV) subtypes using monoclonal antibodies to the PreS1 and PreS2 domains of the viral envelope. *Virology* **176**: 604–619.

Minuk, G. Y., Shaffer, E. A., Hoar, D. I. and Kelly, J. (1986). Ground squirrel hepatitis virus (GSHV) infection and hepatocellular carcinoma in the Canadian Richardson ground squirrel (Spermophilus richardsonii). *Liver* **6**: 350–356.

Moges, F., Kebede, Y., Kassu, A., Mulu, A. and Tiruneh, M. (2006). Seroprevalence of HIV, hepatitis B infections and syphilis among street swellers in Gondar city, northwest Ethiopia. *Ethiopian Journal of Health Development* **20**: 160–165.

Moore, A. M., Awusabo-Asare, K., Madise, N., John-Langba, J. and Kumi-Kyereme, A. (2007a). Coerced first sex among adolescent girls in sub-Saharan Africa: prevalence and context. *African Journal of Reproductive Health* **11**(3): 62–82.

Moore, A. M., Biddlecom, A. E. and Zulu, E. M. (2007b). Prevalence and meanings of exchange of money or gifts for sex in unmarried adolescent sexual relationships in sub-Saharan Africa. *African Journal of Reproductve Health* **11**(3): 44–61.

Moore, A. R. and Oppong, J. (2007). Sexual risk behavior among people living with HIV/AIDS in Togo. *Social Science and Medicine* **64**(5): 1057–1066.

Motta, J. S., Mello, F. C., Lago, B. V., Perez, R. M., Gomes, S. A. and Figueiredo, F. F. (2010). Occult hepatitis B virus infection and lamivudine-resistant mutations in isolates from renal patients undergoing hemodialysis. *Journal*

of Gastroenterology and Hepatology **25**: 101–106.

Motta-Castro, A., Martins, R., Araujo, N., Niel, C., Facholi, G., Lago, B., Mello, F. and Gomes, S. (2008). Molecular epidemiology of hepatitis B virus in an isolated Afro-Brazilian community. *Archives of Virology* **153**: 2197–2205. ISSN 0304-8608. doi: 10.1007/s00705-008-0237-0.

Mphahlele, J., Lukhwareni, A., Burnett, R., Moropeng, L. and Ngobeni, J. (2006). High risk of occult hepatitis B virus infection in HIV-positive patients from South Africa. *Journal of clinical virology* **35**: 14–20.

Mutwa, P. R., Boer, K. R., Rusine, J. B., Muganga, N., Tuyishimire, D., Reiss, P., Lange, J. M. and Geelen, S. P. (2012). Hepatitis B Virus Prevalence and Vaccine Response in HIV-Infected Children and Adolescents on Combination Antiretroviral Therapy in Kigali, Rwanda. *Pediatric Infectious Dissease Journal* .

Muula, A. S., Ngulube, T. J., Siziya, S., Makupe, C. M., Umar, E., Prozesky, H. W., Wiysonge, C. S. and Mataya, R. H. (2007). Gender distribution of adult patients on highly active antiretroviral therapy (HAART) in Southern Africa: a systematic review. *BMC Public Health* **7**: 63.

Myers, R., Clark, C., Khan, A., Kellam, P. and Tedder, R. (2006). Genotyping Hepatitis B virus from whole- and sub-genomic fragments using position-specific scoring matrices in HBV STAR. *Journal of General Virology* **87**(Pt 6): 1459–1464.

Myers, R. E., Gale, C. V., Harrison, A., Takeuchi, Y. and Kellam, P. (2005). A statistical model for HIV-1 sequence classification using the subtype analyser (STAR). *Bioinformatics* **21**(17): 3535–3540.

Nagalo, M. B., Sanou, M., Bisseye, C., Kabore, M. I., Nebie, Y. K., Kienou, K., Kiba, A., Dahourou, H., Ouattara, S., Zongo, J. D. and Simpore, J. (2011). Seroprevalence of human immunodeficiency virus, hepatitis B and C viruses and syphilis among blood donors in Koudougou (Burkina Faso) in 2009. *Blood Transfusion* **9**(4): 419–424.

Nagu, T. J., Bakari, M. and Matee, M. (2008). Hepatitis A, B and C viral co-infections among HIV-infected adults presenting for care and treatment at Muhimbili National Hospital in Dar es Salaam, Tanzania. *BMC Public Health* **8**: 416.

Nakamura, T., Tokino, T., Nagaya, T. and Matsubara, K. (1988). Microdeletion associated with the integration process of hepatitis B virus DNA. *Nucleic Acids Research* **16**: 4865–4873.

National Department of Health (2011). The National Antenatal Sentinel HIV and Syphilis Prevalence Survey, South Africa, 2010. *http://www.doh.org.za*. Accessed on 29 September 2012.

Naumann, H., Schaefer, S., Yoshida, C. F., Gaspar, A. M., Repp, R. and Gerlich, W. H. (1993). Identification of a new hepatitis B virus (HBV) genotype from Brazil that expresses HBV surface antigen subtype adw4. *Joural of General Virology* **74**: 1627–1632.

N'Dri-Yoman, T., Anglaret, X., Messou, E., Attia, A., Polneau, S., Toni, T., Chenal, H., Seyler, C., Gabillard, D., Wakasugi, N., Eholie, S. and Danel, C. (2010). Occult HBV infection in untreated HIV-infected adults in CÃt'te d'Ivoire. *Antiviral Therapy (London)* **15**(7): 1029–1034.

Needleman, S. B. and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48**: 443–453.

Niesters, H. G., Honkoop, P., Haagsma, E. B., de Man, R. A., Schalm, S. W. and Osterhaus, A. D. (1998). Identi-

fication of more than one mutation in the hepatitis B virus polymerase gene arising during prolonged lamivudine treatment. *Juornal of Infectious Diseases* **177**(5): 1382–1385.

Noppornpanth, S., Haagmans, B. L., Bhattarakosol, P., Ratanakorn, P., Niesters, H. G., Osterhaus, A. D. and Poovorawan, Y. (2003). Molecular epidemiology of gibbon hepatitis B virus transmission. *Journal of General Virology* **84**: 147–155.

Norder, H., Hammas, B., Lofdahl, S., Couroucé, A. M. and Magnius, L. O. (1992). Comparison of the amino acid sequences of nine different serotypes of hepatitis B surface antigen and genomic classification of the corresponding hepatitis B virus strains. *Journal of General Virology* **73 (Pt 5)**: 1201–1208.

Norder, H., Couroucé, A. M. and Magnius, L. O. (1994). Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. *Virology* **198**: 489–503.

Norder, H., Ebert, J. W., Fields, H. A., Mushahwar, I. K. and Magnius, L. O. (1996). Complete Sequencing of a Gibbon Hepatitis B Virus Genome Reveals a Unique Genotype Distantly Related to the Chimpanzee Hepatitis B Virus. *Virology* **218**: 214–223. doi: 10.1006/viro.1996.0181.

Noubiap, J. J., Joko, W. Y., Nansseu, J. R., Tene, U. G. and Siaka, C. (2013). Sero-epidemiology of human immunodeficiency virus, hepatitis B and C viruses, and syphilis infections among first-time blood donors in Edéa, Cameroon. *International Journal of Infectious Diseases* .

Nyirenda, M., Beadsworth, M. B., Stephany, P., Hart, C. A., Hart, I. J., Munthali, C., Beeching, N. J. and Zijlstra, E. E. (2008). Prevalence of infection with hepatitis B and C virus and coinfection with HIV in medical inpatients in Malawi. *Jouranl of Infection* **57**(1): 72–77.

Ocama, P., Opio, K. C., Kagimu, M., Seremba, E., Wabinga, H. and Colebunders, R. (2011). Hepatitis B virus and HIV infection among patients with primary hepatocellular carcinoma in Kampala, Uganda. *African Health Sciences* **11 Suppl 1**: S20–23.

Okamoto, H., Imai, M., Kametani, M., Nakamura, T. and Mayumi, M. (1987). Genomic heterogeneity of hepatitis B virus in a 54 year old woman who contracted the infection through materno-fetal transmission. *Japanese Journal of Experimental Medicine* **57**: 231–236.

Okamoto, H., Tsuda, F., Sakugawa, H., Sastrosoewignjo, R. I., Imai, M., Miyakawa, Y. and Mayumi, M. (1988). Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *Journal of General Virology* **69**: 2575–2583.

Olinger, C. M., Jutavijittum, P., Hübschen, J. M., Yousukh, A., Samountry, B., Thammavong, T., Toriyama, K. and Muller, C. P. (2008). Possible new hepatitis B virus genotype, southeast Asia. *Emerging Infectious Diseases* **14**: 1777–1780.

O'Meara, D., Wilbe, K., Leitner, T., Hejdeman, B., Albert, J. and Lundberg, J. (2001). Monitoring resistance to human immunodeficiency virus type 1 protease inhibitors by pyrosequencing. *Journal of Clinical Microbiology* **39**: 464–473.

Open Bioinformatics Foundation (2012). Open Bioinformatics Foundation. *http://www.open-bio.org/wiki/Main_Page*. Accessed 13 December 2012.

Orito, E., Mizokami, M., Ina, Y., Moriyama, E. N., Kameshima, N., Yamamoto, M. and Gojobori, T. (1989). Host-independent evolution and a genetic classification of the hepadnavirus family based on nucleotide sequences. *Proceedings of the National Academies of Sciences of the United States of America* **86**: 7059–7062.

Osiowy, C., Giles, E., Tanaka, Y., Mizokami, M. and Minuk, G. Y. (2006). Molecular evolution of hepatitis B virus over 25 years. *Journal of Virology* **80**: 10307–10314.

Osiowy, C., Kaita, K., Solar, K. and Mendoza, K. (2010). Molecular characterization of hepatitis B virus and a 9-year clinical profile in a patient infected with genotype I. *Journal of Medical Virology* **82**: 942–948.

Ou, J. H. (1997). Molecular biology of hepatitis B virus e antigen. *Journal of Gastroenterology and Hepatology* **12**(9-10): S178–187.

Ouermi, D., Simpore, J., Belem, A. M., Sanou, D. S., Karou, D. S., Ilboudo, D., Bisseye, C., Onadja, S. M., Pietra, V., Pignatelli, S., Gnoula, C., Nikiema, J. B. and Kabre, G. B. (2009). Co-infection of Toxoplasma gondii with HBV in HIV-infected and uninfected pregnant women in Burkina Faso. *Pakistan Journal of Biological Sciences* **12**(17): 1188–1193.

Owiredu, W. K., Kramvis, A. and Kew, M. C. (2001). Hepatitis B virus DNA in serum of healthy black African adults positive for hepatitis B surface antibody alone: possible association with recombination between genotypes A and D. *Journal of Medical Virology* **64**: 441–454.

Panjaworayan, N., Roessner, S. K., Firth, A. E. and Brown, C. M. (2007). HBVRegDB: annotation, comparison, detection and visualization of regulatory elements in hepatitis B virus sequences. *Virology Journal* **4**: 136.

Paraskevis, D., Magiorkinis, G., Magiorkinis, E., Ho, S. Y., Belshaw, R., Allain, J. P. and Hatzakis, A. (2012). Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology* .

Patel, P., Davis, S., Tolle, M., Mabikwa, V. and Anabwani, G. (2011). Prevalence of hepatitis B and hepatitis C coinfections in an adult HIV centre population in Gaborone, Botswana. *Americal Journal of Tropical Medicine and Hygiene* **85**(2): 390–394.

Paulij, W. P., de Wit, P. L., Sunnen, C. M., van Roosmalen, M. H., Petersen-van Ettekoven, A., Cooreman, M. P. and Heijtink, R. A. (1999). Localization of a unique hepatitis B virus epitope sheds new light on the structure of hepatitis B virus surface antigen. *Journal of General Virology* **80 (Pt 8)**: 2121–2126.

Peters, M. G. (2007). *Topics in HIV Medicine*. International AIDS Society, San Francisco.

Pillay, D., Cane, P. A., Ratcliffe, D., Atkins, M. and Cooper, D. (2000). Evolution of Lamivudine-resistant hepatitis B virus and HIV-1 in co-infected individuals: an analysis of the CAESAR study. *AIDS* **14**: 1111–1116.

Pippi, F., Bracciale, L., Stolzuoli, L., Giaccherini, R., Montomoli, E., Gentile, C., Filetti, S., De Luca, A. and Cellesi, C. (2008). Serological response to hepatitis B virus vaccine in HIV-infected children in Tanzania. *HIV Med.* **9**(7): 519–525.

Pirillo, M. F., Bassani, L., Germinario, E. A., Mancini, M. G., Vyankandondera, J., Okong, P., Vella, S. and Giuliano, M. (2007). Seroprevalence of hepatitis B and C viruses among HIV-infected pregnant women in Uganda and Rwanda. *Journal of Medical Virology* **79**(12): 1797–1801.

Pollack, J. R. and Ganem, D. (1994). Site-specific RNA binding by a hepatitis B virus reverse transcriptase initiates

two distinct reactions: RNA packaging and DNA synthesis. *Journal of Virology* **68**: 5579–5587.

Prassolov, A., Hohenberg, H., Kalinina, T., Schneider, C., Cova, L., Krone, O., Frölich, K., Will, H. and Sirma, H. (2003). New Hepatitis B Virus of Cranes That Has an Unexpected Broad Host Range. *Journal of Virology* **77**: 1964–1976. doi: 10.1128/JVI.77.3.1964-1976.2003.

Prince, A. M. (1968). An antigen detected in the blood during the incubation period of serum hepatitis. *Proceedings of the National Academies of Sciences of the United States of America* **60**: 814–821.

Propert, S. A. (1938). Hepatitis after prophylactic serum. *The British Medical Journal* **2**: 677–678.

Pult, I., Netter, H. J., Bruns, M., Prassolov, A., Sirma, H., Hohenberg, H., Chang, S. F., Frölich, K., Krone, O., Kaleta, E. F. and Will, H. (2001). Identification and analysis of a new hepadnavirus in white storks. *Virology* **289**: 114–128.

Purdy, M. A., Talekar, G., Swenson, P., Araujo, A. and Fields, A. (2007). A New Algorithm for Deduction of Hepatitis B Surface Antigen Subtype Determinants from the Amino Acid Sequence. *Intervirology* **50**: 45–51.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raimondo, G., Pollicino, T., Cacciola, I. and Squadrito, G. (2007). Occult hepatitis B virus infection. *Journal of Hepatology* **46**: 160–170.

Raimondo, G., Allain, J. P., Brunetto, M. R., Buendia, M. A., Chen, D. S., Colombo, M., Craxi, A., Donato, F., Ferrari, C., Gaeta, G. B., Gerlich, W. H., Levrero, M., Locarnini, S., Michalak, T., Mondelli, M. U., Pawlotsky, J. M., Pollicino, T., Prati, D., Puoti, M., Samuel, D., Shouval, D., Smedile, A., Squadrito, G., Trepo, C., Villa, E., Will, H., Zanetti, A. R. and Zoulim, F. (2008). Statements from the Taormina expert meeting on occult hepatitis B virus infection. *Journal of Hepatology* **49**: 652–657.

Raimondo, G., Caccamo, G., Filomia, R. and Pollicino, T. (2013). Occult HBV infection. *Seminars in Immunopathology* **35**(1): 39–52.

Ramos, J. M., Toro, C., Reyes, F., Amor, A. and Gutiérrez, F. (2011). Seroprevalence of HIV-1, HBV, HTLV-1 and Treponema pallidum among pregnant women in a rural hospital in Southern Ethiopia. *Journal of Clinical Virology* **51**(1): 83–85.

Revill, P., Yuen, L., Walsh, R., Perrault, M., Locarnini, S. and Kramvis, A. (2010). Bioinformatic analysis of the hepadnavirus e-antigen and its precursor identifies remarkable sequence conservation in all orthohepadnaviruses. *Journal of Medical Virology* **82**: 104–115.

Rice, P., Longden, I. and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite (2000). *Trends in Genetics* **16**: 276–277.

Robson, S. C., Schoub, B. and Abdool Karim, S. S. (1994). Viral hepatitis B – an overview. *South African Medical Journal* **84**: 530–535.

Rodriguez-Frias, F., Buti, M., Jardi, R., Cotrina, M., Viladomiu, L., Esteban, R. and Guardia, J. (1995). Hepatitis B virus infection: precore mutants and its relation to viral genotypes and core mutations. *Hepatology* **22**: 1641–1647.

Rossner, M. T. (1992). Review: hepatitis B virus X-gene product: a promiscuous transcriptional activator. *Journal of*

Medical Virology **36**: 101–117.

Rouet, F., Chaix, M. L., Inwoley, A., Msellati, P., Viho, I., Combe, P., Leroy, V., Dabis, F. and Rouzioux, C. (2004). HBV and HCV prevalence and viraemia in HIV-positive and HIV-negative pregnant women in Abidjan, Côte d'Ivoire: the ANRS 1236 study. *Journal of Medical Virology* **74**(1): 34–40.

Rouet, F., Chaix, M. L., Kpozehouen, A., Inwoley, A., Anaky, M. F., Fassinou, P., Rouzioux, C., Blanche, S. and Msellati, P. (2009). Relationship of CD4+ T-cell counts and plasma HIV-1 RNA levels with serological HBeAg/anti-HBe patterns obtained in West-African HBV-HIV-1-co-infected children. *Journal of Tropical Pediatrics* **55**(6): 409–412.

Sa-nguanmoo, P., Thongmee, C., Ratanakorn, P., Pattanarangsan, R., Boonyarittichaikij, R., Chodapisitkul, S., Theamboonlers, A., Tangkijvanich, P. and Poovorawan, Y. (2008). Prevalence, whole genome characterization and phylogenetic analysis of hepatitis B virus in captive orangutan and gibbon. *Journal of Medical Primatology* **37**: 277–289. doi: 10.1111/j.1600-0684.2008.00290.x.

Sagoe, K. W., Agyei, A. A., Ziga, F., Lartey, M., Adiku, T. K., Seshi, M., Arens, M. Q. and Mingle, J. A. (2012). Prevalence and impact of hepatitis B and C virus co-infections in antiretroviral treatment naÃŕve patients with HIV infection at a major treatment center in Ghana. *Journal of Medical Virology* **84**(1): 6–10.

Salisse, J. and Sureau, C. (2009). A function essential to viral entry underlies the hepatitis B virus "a" determinant. *Journal of Virology* **83**: 9321–9328.

Sanchez-Tapias, J. M., Costa, J., Mas, A., Bruguera, M. and Rodes, J. (2002). Influence of hepatitis B virus genotype on the long-term outcome of chronic hepatitis B in western patients. *Gastroenterology* **123**: 1848–1856.

Santos, E. A., Sucupira, M. V. F., Arabe, J. and Gomes, S. A. (2004). Hepatitis B virus variants in an HIV-HBV co-infected patient at different periods of antiretroviral treatment with and without lamivudine. *BMC Infectious Diseases* **4**: 29.

Sawyer, W. A., Meyer, K. F. and Eaton, M. D. (1944). Jaundice in army personnel in the western region of the United States and its relation to vaccination against yellow fever. *American Journal of Hygiene* **39**: 337–436.

Scaglioni, P. P., Melegari, M. and Wands, J. R. (1997). Posttranscriptional regulation of hepatitis B virus replication by the precore protein. *Journal of Virology* **71**: 345–353.

Schettler, C. H. (1971). Goose virus hepatitis in the Canada Goose and Snow Goose. *Journal of Wildlife Diseases* **7**: 147–148.

Schneider, M. V., Watson, J., Attwood, T., Rother, K., Budd, A., McDowall, J., Via, A., Fernandes, P., Nyronen, T., Blicher, T., Jones, P., Blatter, M. C., De Las Rivas, J., Judge, D. P., van der Gool, W. and Brooksbank, C. (2010). Bioinformatics training: a review of challenges, actions and support requirements. *Briefings in Bioinformatics* **11**(6): 544–551.

Schuppan, D. and Afdhal, N. H. (2008). Liver cirrhosis. *Lancet* **371**: 838–851.

Severini, A., Liu, X. Y., Wilson, J. S. and Tyrrell, D. L. (1995). Mechanism of inhibition of duck hepatitis B virus polymerase by (-)-$beta$-L-2'-3'-dideoxy-3'-thiacytidine. *Antimicrobial Agents and Chemotherapy* **39**: 1430–1435.

Sharp, P. M. and Hahn, B. H. (2010). The evolution of HIV-1 and the origin of AIDS. *Philosophical Transactions of*

the Royal Society B: Biological Sciences **365**: 2487–2494.

Shisana, O., Rehle, T., Simbayi, L., Zuma, K., Jooste, S., Pillay-van Wyk, V., Mbelle, N., van Zyl, J., Parker, W., Zungu, N. P., Pezi, S. and the SABSSM III Implementation Team (2009). *South African national HIV prevalence, incidence, behaviour and communication survey 2008: A turning tide among teenagers?* HSRC Press, Cape Town.

Simon, B., Kundi, M. and Puchhammer, E. (2013). Analysis of mutations in the S gene of hepatitis B virus strains in patients with chronic infection by online bioinformatics tools. *Journal of Clinical Microbiology* **51**(1): 163–168. doi: 10.1128/JCM.01630-12.

Simpore, J., Granato, M., Santarelli, R., Nsme, R. A., Coluzzi, M., Pietra, V., Pignatelli, S., Bere, A., Faggioni, A. and Angeloni, A. (2004). Prevalence of infection by HHV-8, HIV, HCV and HBV among pregnant women in Burkina Faso. *Journal of Clinical Virology* **31**(1): 78–80.

Simpore, J., Savadogo, A., Ilboudo, D., Nadambega, M. C., Esposito, M., Yara, J., Pignatelli, S., Pietra, V. and Musumeci, S. (2006). Toxoplasma gondii, HCV, and HBV seroprevalence and co-infection among HIV-positive and -negative pregnant women in Burkina Faso. *Journal of Medical Virology* **78**(6): 730–733.

Skelton, M. (2010). *Characterization of mutants and splice variants of hepatitis B virus isolated from South African black hepatocellular carcinoma patients*. Ph.D. thesis, University of the Witwatersrand.

Song, B. C., Kim, H., Kim, S. H., Cha, C. Y., Kook, Y. H. and Kim, B. J. (2005). Comparison of full length sequences of hepatitis B virus isolates in hepatocellular carcinoma patients and asymptomatic carriers of Korea. *Journal of Medical Virology* **75**(1): 13–19.

Song, E., Dusheiko, G. M., Bowyer, S. and Kew, M. C. (1984). Hepatitis B virus replication in southern Africa blacks with HBsAg-positive hepatocellular carcinoma. *Hepatology* **4**: 608–610.

Sprengel, R., Varmus, H. E. and Ganem, D. (1987). Homologous recombination between hepadnaviral genomes following in vivo DNA transfection: implications for studies of viral infectivity. *Virology* **159**: 454–456.

Sprengel, R., Kaleta, E. F. and Will, H. (1988). Isolation and characterization of a hepatitis B virus endemic in herons. *Journal of Virology* **62**: 3832–3839.

Starkman, S., MacDonald, D., Lewis, J., Holmes, E. and Simmonds, P. (2003). Geographic and species association of hepatitis B virus genotypes in non-human primates. *Virology* **314**: 381–393. doi: 10.1016/S0042-6822(03)00430-6.

Steinhauer, D. A. and Holland, J. J. (1986). Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. *Journal of Virology* **57**: 219–228.

Stewart, B., Jobarteh, M. L., Sarge-Njie, R., Alabi, A., de Silva, T., Peterson, K., Peterson, I., Whittle, H., Rowland-Jones, S., Jaye, A., Cotten, M. and Mendy, M. (2011). Emergence of HBV resistance to lamivudine (3TC) in HIV/HBV co-infected patients in The Gambia, West Africa. *BMC Research Notes* **4**: 561.

Stuyver, L., Gendt, S. D., Geyt, C. V., Zoulim, F., Fried, M., Schinazi, R. F. and Rossau, R. (2000). A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *Joural of General Virology* **81**: 67–74.

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

19–26.

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

Summers, J., Smolec, J. M. and Snyder, R. (1978). A virus similar to human hepatitis B virus associated with hepatitis and hepatoma in woodchucks. *Proceedings of the National Academies of Sciences of the United States of America* **75**: 4533–4537.

Sutcliffe, S., Taha, T. E., Kumwenda, N. I., Taylor, E. and Liomba, G. N. (2002). HIV-1 prevalence and herpes simplex virus 2, hepatitis C virus, and hepatitis B virus infections among male workers at a sugar estate in Malawi. *Journal of Acquired Immune Deficiency Syndrome* **31**(1): 90–97.

Suzuki, F., Suzuki, Y., Tsubota, A., Akuta, N., Someya, Y., Kobayashi, M., Saitoh, S., Arase, Y., Ikeda, K. and Kumanda, H. (2002). Mutations of polymerase, precore and core promotor gene in hepatitis B virus during 5-year lamivudine therapy. *Journal of Hepatology* **37**: 824–830.

Suzuki, F., Kumanda, H. and Nakamura, H. (2006). Changes in viral loads of Lamivudine-resistant mutants and evolution of HBV sequences during Adefovir Dipivoxil therapy. *Journal of Medical Virology* **78**: 1025–1034.

Suzuki, F., Suzuki, Y., Akuta, N., Yatsuji, H., Sezaki, H., Arase, Y., Kawamura, Y., Hosaka, T., Kobayashi, M., Ikeda, K., Watahiki, S. and Kumada, H. (2008). Changes in viral loads of Lamivudine-resistant mutants during Entecavir therapy. *Hepatology Research* **38**: 132–140.

Swenson, P. D., Riess, J. T. and Krueger, L. E. (1991). Determination of HBsAg subtypes in different high risk populations using monoclonal antibodies. *Journal of Virological Methods* **33**: 27–38.

Takahashi, K., Aoyama, K., Ohno, N., Iwata, K., Akahane, Y., Baba, K., Yoshizawa, H. and Mishiro, S. (1995). The precore/core promoter mutant (T1762A1764) of hepatitis B virus: clinical significance and an easy method for detection. *Journal of General Virology* **76 (Pt 12)**: 3159–3164.

Takahashi, K., Brotman, B., Usuda, S., Mishiro, S. and Prince, A. M. (2000). Full-Genome Sequence Analyses of Hepatitis B Virus (HBV) Strains Recovered from Chimpanzees Infected in the Wild: Implications for an Origin of HBV. *Virology* **267**: 58–64. doi: 10.1006/viro.1999.0102.

Tan, T. W., Tong, J. C., Khan, A. M., de Silva, M., Lim, K. S. and Ranganathan, S. (2010). Advancing standards for bioinformatics activities: persistence, reproducibility, disambiguation and Minimum Information About a Bioinformatics investigation (MIABi). *BMC Genomics* **11 Suppl 4**: S27.

Tanaka, Y., Hasegawa, I., Kato, T., Orito, E., Hirashima, N., Acharya, S. K., Gish, R. G., Kramvis, A., Kew, M. C., Yoshihara, N., Shrestha, S. M., Khan, M., Miyakawa, Y. and Mizokami, M. (2004). A case-control study for differences among hepatitis B virus infections of genotypes A (subtypes Aa and Ae) and D. *Hepatology* **40**: 747–755.

Tang, H., Delgermaa, L., Huang, F., Oishi, N., Liu, L., He, F., Zhao, L. and Murakami, S. (2005). The transcriptional transactivation function of HBx protein is important for its augmentation role in hepatitis B virus replication. *Journal of Virology* **79**: 5548–5556.

Tassopoulos, N. C., Volpes, R., Pastore, G., Heathcote, J., Buti, M., Goldin, R. D., Hawley, S., Barber, J., Condreay,

L., Fraser Gray, D. and Group, T. L. P. M. S. (1999). Efficacy of lamivudine in patients with hepatitis B e antigen-negative/hepatitis B virus DNA-positive (precore mutant) chronic hepatitis B. *Hepatology* **29**: 889–896.

Tatematsu, K., Tanaka, Y., Kurbanov, F., Sugauchi, F., Mano, S., Maeshiro, T., Nakayoshi, T., Wakuta, M., Miyakawa, Y. and Mizokami, M. (2009). A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. *Joural of Virology* **83**: 10538–10547.

Telatela, S. P., Matee, M. I. and Munubhi, E. K. (2007). Seroprevalence of hepatitis B and C viral co-infections among children infected with human immunodeficiency virus attending the paediatric HIV care and treatment center at Muhimbili National Hospital in Dar-es-Salaam, Tanzania. *BMC Public Health* **7**: 338.

Tessema, B., Yismaw, G., Kassu, A., Amsalu, A., Mulu, A., Emmrich, F. and Sack, U. (2010). Seroprevalence of HIV, HBV, HCV and syphilis infections among blood donors at Gondar University Teaching Hospital, Northwest Ethiopia: declining trends over a period of five years. *BMC Infectious Diseases* **10**: 111.

Testut, P., Renard, C. A., Terradillos, O., Vitvitski-Trepo, L., Tekaia, F., Degott, C., Blake, J., Boyer, B. and Buendia, M. A. (1996). A new hepadnavirus endemic in arctic ground squirrels in Alaska. *Journal of Virology* **70**: 4210–4219.

Thibault, V., Benhamou, Y., Seguret, C., Bochet, M., Katlama, C., Bricaire, F., Opolon, P., Poynard, T. and Agut, H. (1999). Hepatitis B virus (HBV) mutations associated with resistance to lamivudine in patients co-infected with HBV and human immunodeficiency virus. *Journal of Clinical Microbiology* **37**: 3013–3016.

Thibault, V., Gaudy-Graffin, C., Colson, P., Gozlan, J., Schnepf, N., Trimoulet, N., Pallier, C., Saune, K., Branger, M., Coste, M. and Thoraval, F. R. (2013). Epidemiological, virological and clinical characteristics of HBV infection in 223 HIV co-infected patients: a French multi-centre collaborative study. *Virology Journal* **10**(87). doi: 10.1186/1743-422X-10-87.

Thio, C. (2009). Hepatitis B and human immunodeficiency virus coinfection. *Hepatology* **49**: S138–S145.

Thio, C. L., Sulkowski, S. and Thomas, D. L. (2005). Treatment of chronic hepatitis B in HIV-infected persons: Thinking outside the black box. *Clinical Infectious Diseases* **41**: 1035–1040.

Thomas, D. L. (2006). Growing importance of liver disease in HIV-infected persons. *Hepatology* **43**(2 Suppl 1): S221–229.

Thomas London, W., Sutnick, A. I. and Blumberg, B. S. (1969). Australia antigen and acute viral hepatitis. *Annals of Internal Medicine* **70**: 55–59.

Tisdale, M., Kemp, S. D., Parry, N. R. and Larder, B. A. (1993). Rapid in vitro selection of human immunodeficiency virus type 1 resistant to 3'-thiacytidine inhibitors due to a mutation in the YMDD region of reverse transcriptase. *Proceedings of the National Academies of Sciences of the United States of America* **90**: 5653–5656.

Torbenson, M. and Thomas, D. L. (2002). Occult hepatitis B. *The Lancet Infectious Diseases* **2**(8): 479–486.

Tran, A., Kremsdorf, D., Capel, F., Housset, C., Dauguet, C., Petit, M. A. and Bréchot, C. (1991). Emergence of and takeover by hepatitis B virus (HBV) with rearrangements in the pre-S/S and pre-C/C genes during chronic HBV infection. *Journal of Virology* **65**: 3566–3574.

Tran, T. T. H., Trinh, T. N. and Abe, K. (2008). New complex recombinant genotype of hepatitis B virus identified in

Vietnam. *Joural of Virology* **82**: 5657–5663.

Truant, R., Antunovic, J., Greenblatt, J., Prives, C. and Cromlish, J. A. (1995). Direct interaction of the hepatitis B virus HBx protein with p53 leads to inhibition by HBx of p53 response element-directed transactivation. *Journal of Virology* **69**: 1851–1859.

Tsebe, K. V., Burnett, R. J., Hlungwani, N. P., Sibara, M. M., Venter, P. A. and Mphahlele, M. J. (2001). The first five years of universal hepatitis B vaccination in South Africa: evidence for elimination of HBsAg carriage in under 5-year-olds. *Vaccine* **19**: 3919–3926.

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

UNAIDS (2010a). UNAIDS Global Report Fact Sheet Sub-Saharan Africa 2010. *http://www.unaids.org*. Accessed on 29 September 2012.

UNAIDS (2010b). UNAIDS Report on the Global AIDS Epidemic 2010. *http://www.unaids.org*. Accessed on 29 September 2012.

UNAIDS (2011). UNAIDS Data Tables 2011. *http://www.unaids.org*. Accessed on 29 September 2012.

Uneke, C. J., Ogbu, O., Inyama, P. U., Anyanwu, G. I., Njoku, M. O. and Idoko, J. H. (2005). Prevalence of hepatitis-B surface antigen among blood donors and human immunodeficiency virus-infected patients in Jos, Nigeria. *Memórias de Instituto Oswaldo Cruz* **100**(1): 13–16.

Urban, S., Schulze, A., Dandri, M. and Petersen, J. (2010). The replication cycle of hepatitis B virus. *Journal of Hepatology* **52**: 282–284.

Visvanathan, K., Skinner, N. A., Thompson, A. J., Riordan, S. M., Sozzi, V., Edwards, R., Rodgers, S., Kurtovic, J., Chang, J., Lewin, S., Desmond, P. and Locarnini, S. (2007). Regulation of Toll-like receptor-2 expression in chronic hepatitis B by the precore protein. *Hepatology* **45**: 102–110.

Wainberg, M. A., Salmon, H., Gu, Z., Montaner, J. S., Cooley, T. P., McCaffrey, R., Ruedy, J., Hirst, H. M., Cammack, N. and Cameron, J. (1995). Development of HIV-1 resistance to (-)2'-deoxy-3'-thiacytidine in patients with AIDS or advanced AIDS-related complex. *AIDS* **9**: 351–357.

Wands, J. R., Wong, M. A., Shorey, J., Brown, R. D., Marciniak, R. A. and Isselbacher, K. J. (1984). Hepatitis B viral antigenic structure: signature analysis by monoclonal radioimmunoassays. *Proceedings of the National Academies of Sciences of the United States of America* **81**: 2237–2241.

Wang, C., Mitsuya, Y., Gharizadeh, B., Ronaghi, M. and Shafer, R. W. (2007). Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome Research* **17**: 1195–1201.

Wang, G. and Seeger, C. (1993). Novel mechanism for reverse transcription in hepatitis B viruses. *J of Virology* **67**: 6507–6512.

Wang, X. W., Forrester, K., Yeh, H., Feitelson, M. A., Gu, J. R. and Harris, C. C. (1994). Hepatitis B virus X protein inhibits p53 sequence-specific DNA binding, transcriptional activity, and association with transcription factor ERCC3. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 2230–2234.

Wang, X. W., Gibson, M. K., Vermeulen, W., Yeh, H., Forrester, K., Sturzbecher, H. W., Hoeijmakers, J. H. and

Harris, C. C. (1995). Abrogation of p53-induced apoptosis by the hepatitis B virus X gene. *Cancer Res.* **55**: 6012–6016.

Warren, K. S., Heeney, J. L., Swan, R. A., Heriyanto and Verschoor, E. J. (1999). A new group of hepadnaviruses naturally infecting orangutans (*Pongo pygmaeus*). *Journal of Virology* **73**: 7860–7865.

Weinberger, K. M., Bauer, T., Bohm, S. and Jilg, W. (2000). High genetic variability of the group-specific a-determinant of hepatitis B virus surface antigen (HBsAg) and the corresponding fragment of the viral polymerase in chronic virus carriers lacking detectable HBsAg in serum. *Journal of General Virology* **81**(Pt 5): 1165–1174.

Westland, C., Delaney, W., Yang, H., Chen, S. S., Marcellin, P., Hadziyannis, S., Gish, R., Fry, J., Brosgart, C., Gibbs, C., Miller, M. and Xiong, S. (2003). Hepatitis B virus genotypes and virologic response in 694 patients in phase III studies of adefovir dipivoxil1. *Gastroenterology* **125**: 107–116.

Whittle, H. C., Bradley, A. K., McLauchlan, K., Ajdukiewicz, A. B., Howard, C. R., Zuckerman, A. J. and McGregor, I. A. (1983). Hepatitis B virus infection in two Gambian villages. *Lancet* **1**: 1203–1206.

Will, H., Reiser, W., Weimer, T., Pfaff, E., Buscher, M., Sprengel, R., Cattaneo, R. and Schaller, H. (1987). Replication strategy of human hepatitis B virus. *Journal of Virology* **61**: 904–911.

Wolters, L. M. M., Niesters, H. G. M., Hansen, B. E., van der Ende, M. E., Kroon, F. P., Richter, C., Brinkman, K., Meenhorst, P. L. and de Man, R. A. (2002). Development of hepatitis B virus resistance for lamivudine in chronic hepatitis B patients co-infected with the human immunodeficiency virus in a Dutch cohort. *Journal of Clinical Virology* **24**: 173–181.

World Health Organization (2009). World Health Organization: Hepatitis B Fact Sheet 204 (August 2008 Revision). *http://www.who.int/mediacentre/factsheets/fs204/en/*. Accessed on 06 January 2011.

World Health Organization (2012a). World Health Organization: Hepatitis B Fact Sheet 204 (July 2012 Revision). *http://www.who.int/mediacentre/factsheets/fs204/en/*. Accessed on 04 August 2012.

World Health Organization (2012b). World Health Organization: HIV/AIDS Fact Sheet 360 (July 2012 Revision). Accessed on 27 September 2012.

World Health Organization (2012c). World Health Organization: Immunization, Vaccines and Biologicals. *http://www.who.int/immunization/topics/hepatitis_b/en/*. Accessed on 22 October 2012.

Wren, J. D. (2004). 404 not found: the stability and persistence of URLs published in MEDLINE. *Bioinformatics* **20**(5): 668–672.

Yamamoto, K., Horikita, M., Tsuda, F., Itoh, K., Akahane, Y., Yotsumoto, S., Okamoto, H., Mikakawa, Y. and Mayumi, M. (1994). Naturally occurring escape mutants of hepatitis B virus with various mutations in the S gene in carriers seropositive for antibody to hepatitis B surface antigen. *Journal of Virology* **68**: 2671–2676.

Yami, A., Alemseged, F. and Hassen, A. (2011). Hepatitis B and C Viruses Infections and Their Association with Human Immunodeficiency Virus: A Cross-Sectional Study among Blood Donors in Ethiopia. *Ethiopian Journal of Health Sciences* **21**(1): 67–75.

Yan, H., Zhong, G., Xu, G., He, W., Jing, Z., Gao, Z., Huang, Y., Qi, Y., Peng, B., Wang, H., Fu, L., Song, M., Chen, P., Gao, W., Ren, B., Sun, Y., Cai, T., Feng, X., Sui, J. and Li, W. (2012). Sodium taurocholate cotransporting

polypeptide is a functional receptor for human hepatitis B and D virus. *eLife* **1**: e00049.

Yang, C. H. and Cho, M. (2012). Hepatitis B virus X gene differentially modulates cell cycle progression and apoptotic protein expression in hepatocyte versus hepatoma cell lines. *Journal of Viral Hepatitis* **20**: 50–58. doi: 10.1111/j.1365-2893.2012.01625.x.

Yim, H. J. and Lok, A. S. (2006). Natural history of chronic hepatitis B virus infection: what we knew in 1981 and what we know in 2005. *Hepatology* **43**: S173–181.

Ymele, F. F., Keugoung, B., Fouedjio, J. H., Kouam, N., Mendibi, S. and Mabou, J. D. (2012). High Rates of Hepatitis B and C and HIV Infections among Blood Donors in Cameroon: A Proposed Blood Screening Algorithm for Blood Donors in Resource-Limited Settings. *Journal of Blood Transfusion* doi: 10.1155/2012/458372.

Yousif, M. and Kramvis, A. (2012). Genotype D of hepatitis B virus and its subgenotypes: An update. *Hepatology Research* doi: doi:10.1111/j.1872-034X.2012.01090.x.

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

Yuen, L. K., Ayres, A., Littlejohn, M., Colledge, D., Edgely, A., Maskill, W. J., Locarnini, S. A. and Bartholomeusz, A. (2007). SeqHepB: a sequence analysis program and relational database system for chronic hepatitis B. *Antiviral Research* **75**(1): 64–74.

Zhu, T., Korber, B. T., Nahmias, A. J., Hooper, E., Sharp, P. M. and Ho, D. D. (1998). An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**: 594–597.

Zlotnick, A., Tan, Z. and Selzer, L. (2013). One protein, at least three structures, and many functions. *Structure* **21**(1): 6–8.

Zoufaly, A., Onyoh, E. F., Tih, P. M., Awasom, C. N. and Feldt, T. (2012). High prevalence of hepatitis B and syphilis co-infections among HIV patients initiating antiretroviral therapy in the north-west region of Cameroon. *International Journal of STD & AIDS* **23**(6): 435–438.

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

Zuckerman, A. J., Thornton, A., Howard, C. R., Tsiquaye, K. N., Jones, D. M. and Brambell, M. R. (1978). Hepatitis B outbreak among chimpanzees at the London Zoo. *Lancet* **2**: 652–654.

Appendix A

Lurman (1885) Translation

An article by Lürman (1885), published in German, is considered the first report of hepatitis B virus, although the cause of the epidemic was not known at the time. The article was kindly translated into English in 2012 by family friends, Lothar and Marion Hüfner, who are not medical professionals. Translating technical, medical German, written in the 19^{th} century, proved to be challenging. Editing was undertaken by T. G. Bell. The short list of references in the article is not included. It is interesting to see how scientific research was undertaken and reported some 130 years ago. Lürman is thorough in his investigation and in the application of scientific methods, and is cautious in inferring too much from his findings.

II. A jaundice epidemic.

Reported by

Dr. ***Lürman*** *in Bremen.*

Jaundice epidemics have been reported at various times from various places, and are not rare. However, I make the following observations cautiously. Fröhlich observed more than 30 yellow jaundice epidemics and has published papers describing them. The majority took place in army barracks and prisons – in places where many people were forced to live in confined spaces for a long time. These jaundice epidemics were caused by the surroundings and all those infected experienced diarrhea. Only in one case does Fröhlich speak about a jaundice epidemic being transmitted via an infection or disease. It can be inferred that the following epidemic is of the same type.

The jaundice epidemic observed from October 1883 to April 1884 in Bremen affected the staff of a company called Actien-Gesellschaft "Weser". This company, which is involved in the production of ships, machinery and

foundry work, lies on the western end of the city. The ground on which the factory is situated is sand and is next to the River Weser. Even when water levels are high, the factory is not normally flooded. No large alterations to the factory buildings have been undertaken recently. No phosphor is used in the factory.

Within the factory area, there is a well which has been used since 1845. The water has been chemically tested many times and has been found to be fine. Only during the summer months is water distributed to the different parts of the factory via five wooden water tanks, which are, according to requirements, refilled with fresh water a few times during the day. During winter, water is taken directly from the well. The bucket system is used for toilets, and these are located in six different places in the factory. For disinfection, carbolic chalk is used, and the waste is taken away efficiently. During the winters of 1883 and 1884, the company employed between 1200 and 1500 people. Some staff obtained their food from nearby canteens, whilst the rest would eat at home or bring their food to work. In previous years, no jaundice epidemics had been noticed in the company.

At the end of October 1883, the first few cases of jaundice symptoms appeared, and the number of cases rose to 33 by the end of November 1883. During December, another 137 cases were noticed, and by the end of January 1884, a further 14 cases appeared. By February and March, a further 5 cases appeared, and in April 1884, another two cases came to light. The total reached 191, but it can be taken that the number of sick people was, in reality, much higher, because a number of sick people did not go to the doctor. In the whole city, during that epidemic at the factory, only a few cases of jaundice symptoms came to light. This is not unusual for this time of the year though.

Illness was noticed amongst all levels of workers, from the ordinary labourer to masters, and the illness was also noticed in staff who lived in quite different circumstances outside the factory. No trend was noticed with respect to the age of those infected. People who came from the country were infected, as were those who lived in the city. Those who worked in different departments, and even those who worked in the open air, were also infected. There was no trend in infections related to location.

Symptoms started with the stomach and bowels. This lasted for at least 8 days, but sometimes for weeks, until the patients turned yellow. During this time, the patients complained of pressure and fullness in the stomach area. They had no appetite, felt sick, experienced vomiting and dizziness, and lacked energy for work. Most of the time, they were constipated, and only seldom had diarrhea. After turning yellow, a process that took approximately one day, some of the symptoms described above became less violent. But soon afterwards, their discomfort worsened. Additionally, patients had itchy skin and, in rare cases, exhibited yellow eyes, and the constipation was then replaced with very heavy diarrhea. In most cases, this lead to an enlargement of the epigastrium, which was very sensitive to pressure. An enlargement of the liver was never noticed. A clear enlargement of the gallbladder could not be proven, and even less did I manage to diagnose the same as a growth. The whole process ran without fever. In most instances, a distinct slowing of the pulse was noticed. The excretion at the beginning, when patients turned yellow, was always discoloured and yellowish, but often returned to normal colouration within a few days, even though the patients' skin and eyes remained

yellow. Urine showed its known yellowish colour. We could always establish the "Gmelinsche" reaction on gall colouring, while the proof of bile, which was sometimes tested by the local health authority, could not be established. Subjective and objective symptoms were very changeable and depended on the intensity and duration of the illness. Only in one instance was cholemia noticed, which was accompanied by heavy brain symptoms, with reduced diuresis, and with after-effects of hydrops ascites and anasarca. Even this very sick person recovered after a 5 month illness. None of the sick people died.

Different shades of skin discolouring were noticed, from very light, to lemon-yellow, to a darker yellow-brown to olive. In a few cases, the illness lasted only between 8 and 14 days. In most cases, it lasted from 4 to 6 weeks. Several patients were sick for longer than 6 weeks. The ones who were ill for a long time had yellow-green skin and eyes, together with extreme loss of weight and strength. During the illness, almost everyone lost weight. Some lost between 12 and 15 pounds. Most of the sick people were able to return to work after a short break. Only in extreme cases, were some people required to rest for several weeks.

As far as the history of the this epidemic is concerned, we can conclude that all known literature and cases of jaundice epidemics are not relevant. Atmospheric influences can be disregarded, whilst the epidemic was definitely localized to the premises of the Weser company. Two other factories, which employed 600 people, did not report any illness during the epidemic. There is very little reason to suspect a miasmatic origin of the epidemic, because the factory is relatively high off the ground, with natural ventilation, and no changes of the terrain took place during 1883.

One of the more important points to mention, is that all the symptoms which normally produce diarrhea, such as poor or spoiled food, were absent. The staff at the factory live under very different conditions, and with a few exceptions, in different parts of the city and the countryside. The food consumed by the various people who fell ill was quite different from one case to the next. Also, family members of married people reported no cases of jaundice. The brandy which was consumed by most of the workers comes from different sources. Also, several people did not drink alcohol. In reference to the drinking water, which could be assumed as the most likely cause, it must be noticed that in the months of August and December it was tested and found to be fine. A few people who become ill did not drink any water from the well, but brought their own supplies from home.

We can determine from the above, that none of the people who become ill had common food habits, but we still have to mention one important aspect. On 13 August 1883, all workers and officials were re-vaccinated against smallpox, because of some cases which appeared in the workforce. The inoculation took place in three different locations on the factory premises, and was performed by six doctors. The serum was supplied from a local chemist, who obtained it third-hand. The names of all those inoculated were recorded. The doctors inoculated people with a scalpel, which was disinfected with a 1% carbolic acid mixture.

In Hall A, 540 people were inoculated; in Hall B, 466 people were inoculated, and in Hall C, 283 people were inoculated. Four iron containers were used, each one containing 100 lymph tubes. Apart from the people which were inoculated at the factory, a further 87 were inoculated by different doctors at different places outside the factory, and different lymph tubes were used. Another 50 of the people who were not present on the

day of the inoculation were inoculated at a later time at the factory. Of the 540 people inoculated in Hall A, 141 became ill. Of the 466 people inoculated in Hall B, 35 became ill. Of the 283 people who were inoculated in Hall C, only 14 became ill. Of the other 50 people who were inoculated at a later date, only 1 became ill. In total, 191 people became ill. None of the other 87 people, who were inoculated by different doctors at difference places, from different batches, became ill. None of the 500 people employed by the factory in April 1884, after the inoculations, became ill.

Since the smallpox inoculation was not successful in most of the workers, it was noticed that more people in that category fell sick, compared to the category where the smallpox inoculation was successful. It is still my opinion that a successful or unsuccessful smallpox inoculation had no great influence on the yellow jaundice epidemic. Until now, there were no reported cases of ill workers, who had not received the inoculation, but had left the service of the company. One case was observed concerning a worker who, 14 days before inoculation, started work and who took part in the inoculation. Otherwise, 9 cases came to light of workers who left the company after they received inoculation. One worker, for example, who was injured in the middle of October and was discharged, became ill with jaundice. Two former workers who joined the army in October, but who were employed by the company on 13 August and received the inoculation, became ill on the 10 and 11 December. Another ex-worker of the company, who was also inoculated on 13 August, also became ill. Even so, nowhere else in the army, or in other places in and around Bremen, were any cases of yellow jaundice reported.

It is also worth mentioning that, amongst the senior staff, six families were living on the premises and were inoculated. The wife and son of one of these, a porter, became ill, even though they had no physical contact with the rest of the workers. From the protocol of re-vaccination, it can be established that these sick people were not inoculated with the same serum that came in one of the iron containers used in Halls A, B and C. Between the time of the smallpox re-vaccination and the onset of yellow jaundice symptoms, sick people mentioned they were quite well up to that point. The smallpox scars did not look unusual.

In light of the above, is can be concluded that this epidemic is the result of an infectious disease in a specific place at a specific time, and has an incubation period between 2 and 8 months. The localization of infections at the company has been clearly stated. The timing also clearly shows that, of the people who left the company before 13 August, as well as those who were employed after 1 September, no case of jaundice came about. Also, no cases of jaundice were reported in people who were employed after 13 August, or made redundant shortly after 13 August.

The yellow jaundice epidemic started at the beginning of October, reached its peak in December, and then slowly decreased until April. The question of etiology cannot be answered with the above facts. None of the agents known to be able to produce a yellow jaundice epidemic were present. Even the drinking water, which could have been the cause, was ruled out because of the reasons stated above and because it was tested. It must also be mentioned that there is always a container of drinking water in the offices and technical department. Of 12 people there, 2 developed jaundice, but they were also part of the mass inoculation, while the others

who remained healthy had been inoculated privately.

It can be suggested that the moment of infection was 13 August, when the re-vaccination took place. To recap quickly: none of the 87 people inoculated outside the factory became ill, and none of the 500 people employed by the company after the inoculation became ill. Also, 9 people who left the factory shortly after the re-vaccination became ill, as did workers, who were employed by the factory shortly before the inoculation.

I am unable to give a clear explanation of this strange causal connection.

Appendix B

Shongwe Documentation

Ethical Clearance

Ethical approval documentation from the Mpumalanga Department of Health and from the Human Ethics Research Committee of the Faculty of Health Sciences of the University of the Witwatersrand are provided on the following pages.

health
Department:
Health
MPUMALANGA PROVINCE

No. 7 Government Boulevard
Riverside Park
Extention 2
NELSPRUIT
1200

Private Bag X 11285
NELSPRUIT
1200
Tel.: +27 13 766 3429
Fax: +27 13 766 3458

Litiko LeteMphilo UmNyango WezaMaphilo Departement van Gesondheid

Enquiries: Molefe Machaba (013) 766 3009/3235

02 February 2009

Prof Anna Kramvis
Wits
No 7, York Road
Parktown
JHB
2193

Dear Prof Anna Kramvis

APPLICATION FOR RESEARCH & ETHICS APPROVAL: MOLECULAR AND FUNCTIONAL CHARACTERISATION OF HEPATITIS B VIRUS (HBV) GENOTYPES ISOLATED FROM HUMAN IMMUNODEFICIENCY VIRUS (HIV) INFECTED SOUTH AFRICANS.

The Provincial Research and Ethics Committee has approved your research proposal in the latest format that you sent. No Issues of ethical consideration were identified.

Kindly ensure that you provide us with the report once your research has been completed.

Kind regards,

Molefe Machaba
Research and Epidemiology

17-02-2009
Date

Mpumalanga PHREC
Chairperson: Mosa Moshabela

17-02-2009
Date

MPUMALANGA PROVINCE
DEPARTMENT OF HEALTH
PLANNING & INFORMATION
2009 -02- 17
PRIVATE BAG X 11285
NELSPRUIT 1200

We care Do you?

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Mr Trevor Bell

CLEARANCE CERTIFICATE	**Protocol M090107**
PROJECT	Internal medicine Molecular Evolution of Hepatitis B Virus in Antiretrovial-Treated Human Immunodeficiency Virus Infected Southern Africans
INVESTIGATORS	Mr Trevor Bell.
DEPARTMENT	Internal Medicine
DATE CONSIDERED	09.01.30
DECISION OF THE COMMITTEE*	Approved unconditionally

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

DATE 09-02-16 **CHAIRPERSON** .....
(Professor P E Cleaton Jones)

*Guidelines for written 'informed consent' attached where applicable

cc: Supervisor : Dr A kramvis

DECLARATION OF INVESTIGATOR(S)

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

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES...

.....

Hepatitis Virus Diversity Research Program
Department of Internal Medicine
Area 551
Charlotte Maxeke Johannesburg Academic Hospital
7 York Road, Parktown

07 May 2009

AMENDMENT TO INFORMED CONSENT DOCUMENT FOR PROTOCOL M090107

I have previously been granted ethics approval for my PhD study (protocol M090107, issued on 16 February 2009). Since the submission of the documentation for that application, we have recruited a research nurse and expect to start the study in June 2009. I have amended the Informed Consent document (from Version 1.0 to Version 1.5) and have attached it to this covering letter for your reference. Apart from minor formatting, spelling and typographical errors which I have been corrected, the following changes have been effected:

1. "Introduction": The name of the nurse has been inserted and the affiliation of the nurse has been changed.
2. The fourth point under "Study Procedures" has been reworded to indicate that studies relating to liver cancer *may* be carried out in the future, rather than the previous implication that such studies *would* be carried out.
3. "Persons to Contact for Problems or Questions": The previous medical officer has been replaced with Dr Precious Gabashane at Shongwe Hospital.

TG Bell
Student Number 9300466/E
Trevor.Bell@students.wits.ac.za

University
of the Witwatersrand,
Johannesburg

Human Research Ethics Committee (Medical)
(formerly Committee for Research on Human Subjects (Medical))
Secretariat: Research Office, Room SH10005, 10th floor, Senate House • Telephone: +27 11 717-1234 • Fax: +27 11 339-5708
Private Bag 3, Wits 2050, South Africa

13 May 2009

Dr Trevor Bell
Hepatitis Virus Diversity Research Program
Department of Internal Medicine
Area 551
Charlotte Maxeke Johannesburg Academic Hospital
University

Dear Dr Bell

RE: Protocol M090107-Amendments to Informed Consent Document

This letter serves to confirm that the Chairman of the Human Research Ethics Committee (Medical) has reviewed and approved the revised Informed Consent Form Version 1.5 on the abovementioned protocol.

Thank you for keeping us informed and updated.

Yours sincerely,

Anisa Keshav
Secretary
Human Research Ethics Committee (Medical)

Informed Consent

The Informed Consent document, which participants signed, is provided on the following pages. The document was made available in English, isiZulu, Sesotho, Siswati and Xitsonga.

PARTICIPANT INFORMATION LEAFLET AND ENROLLMENT INFORMED CONSENT

The molecular and functional characterization of hepatitis B virus (HBV) genotypes isolated from human immunodeficiency virus (HIV) infected Southern Africans
(Shongwe Hospital HIV / HBV Study)

Version 1.5a
May 2009

INVESTIGATORS
Prof. Anna Kramvis
Hepatitis Virus Diversity Research Programme
Department of Internal Medicine
Faculty of Health Sciences
University of the Witwatersand
Tel: (011) 488-3100
Email: **Anna.Kramvis@wits.ac.za**

Dr Neil A Martinson
Perinatal HIV Research Unit
Wits Health Consortium
University of the Witwatersrand
and
Chris Hani Baragwanath Hospital
Old Potch Road, Soweto
Tel: (011) 989 9703
E-mail: martinson@hivsa.com

INTRODUCTION

Good day, my name is **Agatha Nkosi**. I am a research nurse assisting in a study being carried out by the University of the Witwatersrand in Johannesburg. I would like to invite you be part of a research study that will assess whether or not you have been exposed to hepatitis B virus (HBV).

Before you decide whether to take part in the study, we would like to explain why we are doing the study, the risks and benefits, and what would be expected of you if you agree to be in the study. This study is sponsored by the Medical Research Council (MRC) of South Africa.

PURPOSE OF THE STUDY

This study is researching hepatitis B virus (HBV). A virus is a very small germ that gets caught up in the cells of the person who has it and can cause severe illness or maybe no illness at all. HBV is a very common infection in some parts of Africa and HBV causes liver damage in some people who are infected by it. People who have been exposed to HBV can either cure themselves of the infection over time, or be unable to cure

themselves and will continue to be infected with HBV. Ongoing HBV infection may cause serious liver problems later, including liver cancer. Because HIV-infection is also very common, we want to look at people who are infected with both HIV and HBV.

At the first visit we want to:

- Count how many HIV-infected people (people who have the virus that causes AIDS) also have been exposed to HBV
- Count how many of the ones who are HBV infected also have HBV that could be transmitted to other people
- How many people have cleared the infection themselves
- Check how many people have the genetic makeup that allows liver cancer

If you have HBV, we also would like to see you another 4 times: 3, 6, 12 and 18 months after you started antiretroviral (ARV) treatment, to see what happens to HBV when HIV is treated with ARVs. If we can find hepatitis B virus (HBV) in your blood, we will see what family or type of HBV it is. We will ask about 250 people to give blood at the first visit and we estimate that about 30 of you will have hepatitis B infection and will be followed up while taking ARVs.

VOLUNTARY PARTICIPATION

It is important that you know the following:

- You don't have to be in this study if you do not want to. You will still receive medical care at the HIV clinic here.
- You may decide to withdraw from the study at any time, without affecting your normal medical care.

STUDY PROCEDURES

This study will not give you better care than if you were not part of it. We are only studying HBV. Your doctor or nurse here at the clinic will still decide what is the best treatment for you. They may use some of the results we find to decide.

At the first visit and and at each scheduled visit you will:

- Have an interview with the research nurse or doctor. You will be asked to answer questions about yourself and your health.
- You will have blood taken for HIV viral load and CD4 count at enrollment and at the last study visit 6 months later, either as part of routine care or research participation.
- We will take blood for studies of HBV and to see how your liver is working at the first visit. If your blood shows that you have active HBV, we will let the doctor in charge know and we will ask you if we can see you again after 3, 6, 12 and 18 months after you started ARVs. At each of these visits we will take blood for HBV studies and to check how your liver is working. We will store your samples at the Medical School in Johannesburg, but will only do tests on them related to HBV and HIV.
- Sometimes HBV causes liver cancer. One of the tests we may do in the future will be a test of the chemicals that determine how your body works (the chromosomes). This test checks if you may be more likely to get liver cancer related to hepatitis B infection (in medical terms, we will be looking to see if you have the p53 gene on chromosome 17).

- You will be asked for updated information on where you live and how to keep in contact with you. If you miss a visit, the study staff will try to contact you by telephone. They also may visit your home to find you. They will try to reach you through the contact people that you provide. If they talk to these people that know you, they will not tell them why they are trying to reach you.

RISKS AND/OR DISCOMFORTS
We will take about 5 teaspoons (25ml) of blood from you. This may make you feel discomfort or pain when your blood is drawn. Occasionally people may feel dizzy or faint. You may get a bruise where the needle goes in.

BENEFITS
You will get no direct health benefit from being in this study, but we will be informing your doctor or nurse if you have active hepatitis B infection and that may be of indirect benefit to you. The doctor may change the HIV antiretrovirals you receive if he/she thinks that this is required.

LEAVE THE STUDY
You can leave this study at any time without causing a problem. If you decide to leave the study, please tell the research nurse or clinic staff why you wish to leave.

COSTS TO YOU
There is no cost to you for being part of this study. Study investigations are free of charge. You will be reimbursed R25 for your travel costs if you agree to be part of the study and you will receive R50 each time when you come to your follow up visits.

CONFIDENTIALITY
We will keep your information private and confidential. Your information may only be disclosed if required by law. With your permission, study staff may visit you at home to check on your health or to contact you if you are late for a study visit. Any publication of this study will not use your name or identify you personally. Your name will not be stored in any electronic form in a database or computer. Your study records may be reviewed by study staff and people who check that we are actually doing this study the way we said we would do it.

RESEARCH-RELATED INJURY
If you are injured as a result of participation in this study, the study clinic will give you immediate necessary treatment for your injuries. The cost of this treatment will not be charged to you.

PERSONS TO CONTACT FOR PROBLEMS OR QUESTIONS
If you ever have questions about this study or in case you are injured as a result of participation in this research study, you should contact Dr Precious Gabashane at Shongwe Hospital (013 781 0219) or Dr Neil Martinson at the PHRU (011 989 9703 during office hours).

If you have questions about your rights as a research subject you can contact Professor Cleaton Jones on 011 717 2301, Wits Research Ethics Committee, 10th Floor, Senate House, University of the Witwatersrand.

STATEMENT OF CONSENT AND SIGNATURES

I have read this form in its entirety, or I have had it read to me in the event that I am not able to read. I have discussed the information with study staff. My questions have been answered. I understand that my decision whether or not to take part in the study is my own. I understand that I may withdraw at any time. A witness has been present for the entire consent process.

Participant Name	Participant Signature or Thumb-print	Date

Interviewer Name	Interviewer Signature	Date

Consent Discussion Witness	Witness Signature or Thumb-print	Date

CRF

The Clinical Report Form, which the research nurse completed for each participant, is provided. This is followed by the follow-up and termination forms used in the study.

University of the Witwatersrand
Medical School
Shongwe Hospital HIV / HBV Study

BARCODE STICKER

ENROLLMENT
Yellow

PLEASE ANSWER ALL QUESTIONS

VARIABLE	QUESTION	RESPONSE CODES	RESPONSE
	Informed Consent completed and signed? If "No", STOP	[N]=No (**STOP**) [Y]=Yes	____
	HIV-infected? If "No", STOP	[N]=No (**STOP**) [Y]=Yes	____
	Assessed as about to initiate ARVs at this or the next visit? If "No", STOP	[N]=No (**STOP**) [Y]=Yes	____
	Vaccinated against HBV? If "Yes", STOP	[N]=No [Y]=Yes (**STOP**)	____
	CD4 Count (If greater than 350, STOP) *(from Shongwe Hospital Records)*	___ ___ ___	
	CD4 Count Date *(Most Recent Test)*	[d][d][m][m][y][y]	__ __ __ __ __ __
	HIV Viral Load *(from Shongwe Hospital Records)*	___ ___ ___ ___ ___	
	Language of Informed Consent	[E]=English [S]=Sesotho [X]=Xitsonga [Z]=Zulu	____
	Participant Initials *(for filing purposes only)*	First Name: __	Surname: __
	Date of Enrollment	[d][d][m][m][y][y]	__ __ __ __ __ __
	Date of Birth	[d][d][m][m][y][y]	__ __ __ __ __ __
	Age in years *(only if Date of Birth is unknown)*	____ ____	
	Gender	[F]=Female [M]=Male	
	Ethnicity *(for research purposes only)*	[B]=African/Black [A]=Asian [C]=Coloured [I]=Indian [W]=White [?]=Other/Unknown	

University of the Witwatersrand
Medical School
Shongwe Hospital HIV / HBV Study

BARCODE STICKER

ENROLLMENT
Yellow

VARIABLE	QUESTION	RESPONSE CODES	RESPONSE
	Ethnicity of Mother *(for research purposes only)*	(Save as above)	
	Ethnicity of Father *(for research purposes only)*	(Save as above)	
	Height in centimetres		___ ___ ___ cm
	Weight in kilograms (today)		___ ___ ___ kg
	Ever had yellow eyes or jaundice?	[N]=No [Y]=Yes	____
	If "Yes", when was the last time?	[m][m] [y][y]	___ ___ ___ ___
	Ever had pale stools and/or dark urine?	[N]=No [Y]=Yes	____
	If "Yes", when was the last time?	[m][m] [y][y]	___ ___ ___ ___
	Previously diagnosed with HBV?	[N]=No [Y]=Yes	____
	If "Yes", when?	[m][m] [y][y]	___ ___ ___ ___
	If "Yes", what treatment was prescribed?		
	Any traditional scarification marks (Xaba)?	[N]=No [Y]=Yes	____
	If "Yes", specify number of episodes		_____
	Mother/Father or Sister/Brother (from same mother) ever have liver disease? *(More than one can be selected)*	[M]=Mother [F]=Father [S]=Sister [B]=Brother [N]=None	__ __ __ __
	If "Yes", specify type of disease	[H]=Hepatitis [C]=Liver Cancer [A]=Alcohol-rel. [D]=Drug/Toxin *[Specify other]*	___

University of the Witwatersrand
Medical School
Shongwe Hospital HIV / HBV Study

BARCODE STICKER

ENROLLMENT
Yellow

VARIABLE	QUESTION	RESPONSE CODES	RESPONSE
	Smoking Habits	[N]=No [Y]=Yes [Q]=Quit	____
	Alcohol Consumption Frequency	[0]=Never [1]=Weekends Only [2]=Daily	____
	List any other previous illnesses of a serious nature. For example: Cancer, TB, Syphilis, Herpes, Hepatitis A/C/D		
	Total number of sexual partners		___ ___ ___
	Age when first had sex		___ ___
	When you have sex, do you use a condom	[0]=Always [1]=Sometimes [2]=Never [-]=Not Applicable	____
	Have you ever in your life received a blood transfusion (blood or blood products)?	[N]=No [Y]=Yes	____
	Highest level of school completed	[0]=None [1]=Grade 0 to 5 [2]=Grade 6 to 11 [3]=Grade 12 [4]=Diploma/Degree [?]=Other/Unknown	____
	Employment status	[0]=Employed [1]=Self-employed [2]=Unemployed [?]=Other/Unknown	____
	Personal income per month	[-][-][-][-][-] =Unknown	R __ __ __ __ __
	Total family income per month	[-][-][-][-][-] =Unknown	R __ __ __ __ __
	Current living situation	[0]=House [1]=Flat [2]=Shack [3]=Shelter [4]=Homeless [?]=Other/Unknown	____

University of the Witwatersrand
Medical School
Shongwe Hospital HIV / HBV Study

BARCODE STICKER

ENROLLMENT
Yellow

VARIABLE	QUESTION	RESPONSE CODES	RESPONSE
	Number of people in living situation		___ ___
	Flush-toilet in living situation?	[0]=No [1]=Yes [?]=Other/Unknown	____
	Total number of children under five who lived in the same house as you when you were under five? *(Count those who died, those who moved away and those who stayed)*		_____
	Your Country of Birth		
	Your Town of Birth		
	Country of Birth of your Mother		
	Town of Birth of your Mother		
	Your Current Country of Residence		
	ARV Treatment Regime	[1]=d4T/3TC/NVP [2]=d4T/3TC/ Efavirenz [3]=AZT/DDI/ Lopoinavir/ Ritonavir [4]=Truvada [?]=Unknown/Other	____
	ARV Start Date	[d][d][m][m][y][y]	__ __ __ __ __ __
	ALT (Alanine transaminase)		
	AST (Aspartate transaminase)		
	Hepatitis B Serology Result		
	Hepatitis B DNA Result		
	Date Blood Drawn	[d][d][m][m][y][y]	__ __ __ __ __ __
	Date of this Interview	[d][d][m][m][y][y]	__ __ __ __ __ __
	Signature of Interviewer		

University of the Witwatersrand
Medical School
Shongwe Hospital HIV / HBV Study

BARCODE STICKER

FOLLOW-UP
Pink

PLEASE ANSWER ALL QUESTIONS

VARIABLE	QUESTION	RESPONSE CODES	RESPONSE
	Follow-Up Visit Number	[1]=3 Month Visit [2]=6 Month Visit [3]=12 Month Visit [4]=18 Month Visit	____
	Date of Follow-Up	[d][d][m][m][y][y]	__ __ __ __ __ __
	ARV Treatment Regime	[1]=d4T/3TC/NVP [2]=d4T/3TC/ Efavirenz [3]=AZT/DDI/ Lopoinavir/ Ritonavir [4]=Truvada [?]=Unknown/Other	____
	If ARV Regimen has changed, indicate new regimen	[Same as above]	____
	Treatment Adherence Estimation	[0]=Taken >90% [1]=Taken 80-90% [2]=Taken <80%	____
	Number of doses missed in the preceding 3 days	____	
	Indicate any side-effects of treatment, if applicable		
	Indicate any liver-related problems the attending doctor may have identified at this visit		
	Weight in kilograms (today)	___ ___ ___ kg	
	CD4 Count	___ ___ ___	
	HIV Viral Load	___ ___ ___ ___ ___	
	Date Blood Drawn	[d][d][m][m][y][y]	__ __ __ __ __ __
	Signature of Interviewer		

University of the Witwatersrand
Medical School
Shongwe Hospital HIV / HBV Study

BARCODE STICKER

TERMINATION
Blue

PLEASE ANSWER ALL QUESTIONS

VARIABLE	QUESTION	RESPONSE CODES	RESPONSE
	Date of Termination	[d][d][m][m][y][y]	__ __ __ __ __ __
	Reason for Termination	[0]=Death [1]=Presumed dead [2]=Withdrew [3]=Not returned [4]=Untraceable [5]=HBV Negative [6]=Study Finished	_____
	If "0", date of death (if known)	[d][d][m][m][y][y]	__ __ __ __ __ __
	If "0", cause of death (if known)		
	If "0", place of death (if known)	[0]=Home [1]=Hospital [2]=Clinic [3]=Healer's Home [4]=Hospice [5]=Other/Unknown	____
	All documentation completed correctly and filed	[Y]=Yes [N]=No *[Comments Below]*	____
	Signature of Interviewer		

Comments

Supporting Documents

Examples of additional documentation, which the research nurse completed during the study.

University of the Witwatersrand Medical School
Shongwe Hospital HIV / HBV Study
BLOOD TUBE TRACKING – BLOOD DRAW TO EXTRACTION

THIS FORM MUST BE PACKAGED TOGETHER WITH THE BLOOD TUBES FOR DELIVERY

Participant Number	A=Enrollment B=Follow-up 1 C=Follow-up 2 etc.	Number of Blood Tubes Dispatched for Participant	Date Blood Drawn (DD/MM/YY)	Date Blood Dispatched for Delivery (DD/MM/YY)	Date Blood Received by Lab (DD/MM/YY)	Number of Blood Tubes Received for Participant

University of the Witwatersrand Medical School
Shongwe Hospital HIV / HBV Study
PARTICIPANT COMPENSATION RECORD

THIS FORM MUST BE PACKAGED TOGETHER WITH THE BLOOD TUBES FOR DELIVERY

Date (DD/MM/YYYY)	Participant Number	A=Enrollment B=Follow-up 1 C=Follow-up 2, etc.	Amount (R25 for Enrollment; R50 for Follow-ups)	Research Nurse Signature	Participant Signature / Initials / Thumbprint

University of the Witwatersrand Medical School
Shongwe Hospital HIV / HBV Study
RESEARCH NURSE TIME SHEET

Date (DD/MM/YYYY)	Start Time (HH:MM)	End Time (HH:MM)	Number of Hours	Number of New Enrollments	Number of Follow-Up Interviews	Signature

Appendix C

Paper IV

Genotyping and molecular characterization of hepatitis B virus from human immunodeficiency virus-infected individuals in Southern Africa

Makondo, E., **Bell, T. G.** and Kramvis, A. (2012)

PLOS One **7**(9): e46345.

References References for literature cited in the following paper appear within the references section of the paper itself, and not necessarily in the references section of this thesis.

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

OPEN ACCESS Freely available online

PLOS ONE

Genotyping and Molecular Characterization of Hepatitis B Virus from Human Immunodeficiency Virus-Infected Individuals in Southern Africa

Euphodia Makondo, Trevor G. Bell, Anna Kramvis*

Hepatitis Virus Diversity Research Programme, Department of Internal Medicine, University of the Witwatersrand, Johannesburg, South Africa

Abstract

Hepatitis B virus (HBV) and human immunodeficiency virus (HIV) are hyperendemic in sub-Saharan Africa. The HBV genotypes prevailing in HIV-infected Africans are unknown. Our aim was to determine the HBV genotypes in HIV-infected participants and to identify clinically significant HBV mutations. From 71 HBV DNA^{+ve} HIV-infected participants, 49 basic core promoter/precore (BCP/PC) and 29 complete S regions were successfully sequenced. Following phylogenetic analysis of 29 specimens in the complete S region, 28 belonged to subgenotype A1 and one to D3. Mutations affecting HBeAg expression at the transcriptional (1762T1764A), translational (Kozak 1809–1812, initiation 1814–1816, G1896A with C1858T), or post translational levels (G1862T), were responsible for the high HBeAg-negativity observed. The G1862T mutation occurred only in subgenotype A1 isolates, which were found in one third (7/21) of HBsAg^{-ve} participants, but in none of the 18 HBsAg^{+ve} participants ($p<0.05$). Pre-S deletion mutants were detected in four HBsAg^{+ve} and one HBsAg^{-ve} participant/s. The following mutations occurred significantly more frequently in HBV isolated in this study than in strains of the same cluster of the phylogenetic tree: ps1F25L, ps1V88L/A; ps2Q10R, ps2 R48K/T, ps2A53V and sQ129R/H, sQ164A/V/G/D, sV168A and sS174N ($p<0.05$). ps1I48V/T occurred more frequently in females than males ($p<0.05$). Isolates with sV168A occurred more frequently in participants with viral loads >200 IU per ml ($p<0.05$) and only sS174N occurred more frequently in HBsAg^{-ve} than in HBsAg^{+ve} individuals ($p<0.05$). Prior to initiation of ART, ten percent, 3 of 29 isolates sequenced, had drug resistance mutations rtV173L, rtL180M+rtM204V and rtV214A, respectively. This study has provided important information on the molecular characteristics of HBV in HIV-infected southern Africans prior to ART initiation, which has important clinical relevance in the management of HBV/HIV co-infection in our unique setting.

Citation: Makondo E, Bell TG, Kramvis A (2012) Genotyping and Molecular Characterization of Hepatitis B Virus from Human Immunodeficiency Virus-Infected Individuals in Southern Africa. PLoS ONE 7(9): e46345. doi:10.1371/journal.pone.0046345

Editor: Sheila Mary Bowyer, University of Pretoria/NHLS TAD, South Africa

Received May 29, 2012; **Accepted** August 30, 2012; **Published** September 28, 2012

Funding: This study was supported by grants received from the South African Medical Research Council and the National Research Foundation (NRF; GUN#65530) awarded to AK. TGB received bursaries from the National Bioinformatics Network, Poliomyelitis Research Foundation (PRF) and NRF. EM received a Belgium Technical Cooperation Fellowship and bursaries from the NRF and PRF. Research was supported by funding from Deutsch Forschungsgemeinschaft (German Research Foundation) and the Cancer Association of South Africa. The funders had no role in study design, data collection and analysis, decision to publish.

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

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

Introduction

Hepatitis B virus (HBV), with a genome of ~3,200 base pairs, is the smallest DNA virus infecting humans, yet it is one of the most important human pathogens, causing major health problems globally. HBV, the prototype member of the family *Hepadnaviridae*, is endemic in several parts of the world, including sub-Saharan Africa, which accounts for at least 65 of the 360 million people in the world chronically infected with the virus [1]. HBV causes chronic and acute infections, associated with severe liver diseases, including hepatitis, hepatic fibrosis, cirrhosis, and hepatocellular carcinoma (HCC). Moreover, of the 33.3 million adults and children living with HIV globally, 22.5 million reside in sub-Saharan Africa [2]. HIV infection leads to the acquired immunodeficiency syndrome [3], opportunistic infections, and premature death.

The two viruses share a common mode of transmission and can co-exist in the same host [4], and thus HBV and HIV co-infections are frequent in sub-Saharan Africa [5]. Because HBV infection precedes HIV infection in sub-Saharan Africa [5], the HBV exposure rate does not differ from that found in HBV mono-infected [6–9]. Even though direct comparison between studies is difficult because of differences in study design and geographical regions, a range of 28% to 99.8% exposure to HBV and 0.4% to 23% HBsAg prevalence have been found in HIV-infected South African cohorts [10–20]. Moreover, comparisons between HBV mono-infected and HBV/HIV co-infected individuals are further confounded by the fact that since the introduction of universal HBV vaccination in April 1995, no comprehensive studies have been carried out in South Africa to determine either the exposure or prevalence rates of HBV infection.

We have recently shown that of approximately 300 HIV-infected individuals from a rural cohort in Mpumalanga Province, 77.5% had at least one HBV marker, with 53.7% being HBVDNA^{-ve} (having resolved the infection) and 23.8% being HBVDNA^{+ve} [20]. HBV DNA without HBsAg, was detected in 15.1% of the participants [20], which is within the 8% to 18% range for other South African HIV^{+ve} cohorts [11–14,16].

However, only three of these $HBsAg^{-ve}$ participants met the "Taormina" definition of true occult HBV infection (HBV viral load <200 IU/ml) [21], whereas the remaining were *HBsAg-covert* (HBsAg-***c***ryptic ***o***vert) infections, having higher viral loads (>200 IU/ml) [20].

HBV replicates by reverse transcription of the pregenomic RNA using a viral-encoded polymerase that lacks proof-reading activity. The genome may evolve at an estimated error rate of $1.4–5\times10^{-5}$ nucleotide substitutions/site/year, which results in genetic heterogeneity [22–24]. As a result of this genetic variability, genotypes of HBV have been identified defined by inter-genotypic differences of more than 7.5% in the complete nucleotide sequence [25–27]. To date, phylogenetic analysis of the HBV genome has lead to recognition of nine genotypes of HBV: A to D [26,28], genotype E to F [25,28,29], genotype G [30], genotype H [31] genotype I [32–36]; and a tenth genotype J has been proposed [37]. The genotypes have distinct geographical distributions [38]. In Africa, genotype A is found predominantly in southern, eastern and central Africa, genotype D in northern Africa, whereas the majority of isolates from western Africa belong to genotype E [1]. Subgenotypes have been identified within genotypes A and D [26]. Genotypes A and D coexist in southern Africa, with genotype A predominating, with the dominant subgenotypes being A1 and D3 [39].

Very few studies have been conducted on the genotypes and molecular characterization of HBV in HIV-infected individuals in the Africa [39–42]. Such studies are important because the natural history of infection and response to antiviral therapy are influenced by the HBV genotype [40]. Thus the aim of this study was to molecularly characterize HBV isolated from HIV-infected southern Africans from the Mpumalanga Province cohort prior to the initiation of antiretroviral therapy (ART) [20], in order to determine the genotypes, possibly explain the high level of HBsAg- and HBeAg-negativity and to identify clinically relevant mutations.

Methods

Serum samples

Of the 298 samples obtained from HIV^{+ve} individuals prior to the initiation of antiretroviral therapy (ART), 71 plasma samples were shown to be positive for HBV DNA [20] and were used in this study. All participants from which the plasma samples were obtained signed informed consent. The study was approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand and Mpumalanga Department of Health Research Ethics Committee. Although an attempt was made to sequence all 71 samples, the serological profile of those that were successfully sequenced is shown in Table 1.

Polymerase chain reaction (PCR)

DNA was extracted from 200 µl blood plasma with the QIAamp DNA Blood Mini Kit (QIAGEN Gmbh, Hilden, Germany) and eluted into 75 µl of best-quality water (BQW). Known positive and negative sera and BQW were used as controls for the extraction. The basic core promoter/precore (BCP/PC) region and complete S open reading frame (ORF) were amplified in a MyCycler^{TM} thermocycler (Bio-Rad, Hercules, Ca, USA) using Promega Taq DNA polymerase (Promega, Madison, WI) as described previously in detail [20]. The BCP/PC region of HBV isolates was amplified using a nested PCR: primers 1606 (+) and 1974 (−) were used for the first round and 1653(+) and 1959(−) for the second round to yield an amplicon 1606–1974 from *EcoR*I site [41,42]. A nested PCR reaction was carried out to amplify the complete S open reading frame: primers 2410(+)/1314(−)were used for the first round and 2451 (+) and 1280 (−)for the second round (2451–1260 from *EcoR*I site) [42].

Sequencing

The amplicons were prepared for direct sequencing using the BigDye Terminator v3.0 Cycle Sequencing Ready Reaction Kit (Applied Biosystems., Foster City, USA) and sequencing was performed by the Central Analytical Facility, Stellenbosch University, South Africa, using the ABI 3130XL Genetic analyser (Applied Biosystems, Foster City, CA). BCP/PC sequences were analysed in both the forward and reverse directions of a single fragment, whilst the complete S sequences were analysed in 3 overlapping fragments. In addition to the primers used for amplification, HBV-specific primers [42] were used for sequencing.

Phylogenetic analysis

Both BCP/PC (160 nt, 1742–1901 from *EcoR*I site) and complete surface DNA sequences (1203 nt, 2854–835from *EcoR*I site) were assembled and aligned manually using GeneDoc [43] and fed into MEGA5 [44]. The sequences were compared with corresponding subgenotype A1 sequences of HBV from GenBank. Nucleotide divergence calculations were carried out using Dambe [45]. The evolutionary history was inferred using the Neighbor-Joining method [46] and the evolutionary distances computed using the Kimura 2-parameter method [47]. Bootstrapping was performed using 1 000 replicates in order to determine the support for the specific nodes. The accession numbers of HBV isolates sequenced in this study have been deposited in GenBank/EMBL/DDBJ as JX144270–JX144323.

Results

Using detection of HBV DNA by amplification of at least two of three regions of the HBVgenome, 71 of 298 HIV infected individuals were found to be co-infected with HBV [20]. The basal core promoter/precore (BCP/PC) region of 49 HBV isolates and the complete S region of 29 isolates were successfully sequenced. The relatively longer amplicon of the S region compared to the BCP/PC region, meant that fewer samples could be sequenced in that region successfully. The results are summarized in Table 1.

Analysis of the basal core promoter/precore (BCP/PC) region

Using the criteria of Kramvis *et al* [26], the genotypes/subgenotypes were deduced from the BCP/PC region sequence of 48 isolates (Table 1). The genotype for SHH027 could not be deduced. Four isolates (SHH016, SHH042, SHH053, SHH167), with 1858T, did not belong to genotype A. Four isolates (SHH032, SHH060, SHH217, and SHH249) had Kozak sequence GCAC at 1809–1812, 1858C and 1888G, generally found in subgenotype A2. Thirty nine sequences belonged to subgenotype A1 with Kozak sequence TCAT (or mutant) at nucleotides1809–1812, 1888A and 1858C/T. Although SHH221, with 1858C belonged to genotype A, its subgenotype could not be deduced from the BCP/PC region because it had GCAC at 1809–1812 and 1888A.

Of the 49 cases, 5 were $HBeAg^{+ve}$ and the remaining 44 $HBeAg^{-ve}$. Three of the five isolates from $HBeAg^{+ve}$ cases (SHH121, SHH159, SHH255) did not show any BCP/PC mutations, which can down-regulate or abolish the expression of the HBeAg. These had wild type sequences relative to the consensus for subgenotype A1 of HBV, that is, 1762A/1764G, 1809–1812 (TCAT Kozak), 1862G and 1888A. One isolate from

Table 1. Characteristics of HBV isolated from HIV^{+ve} ART-naïve southern Africans.

Characteristics of Participant							Molecular characteristics of HBV isolates									Mutations[a]
							Basic core promoter/precore region								S region	
Serology	Sample	Sex	Age	ALT (IU/L)	CD4 (cells/ml)	HBVVL (IU/ml)	1762, 1764	1809–1812	1814–1816	1858	1862	1888	1896	Genotype	Subgenotype	
HBsAg^{+ve} HBeAg^{+ve} anti-HBc^{+ve}	SHH121	M	33	22	125	1.02e+07	AG	TCAT	ATG	C	G	A	G	A	A1	-
	SHH159	F	23	38	105	1.33e+03	AG	TCAT	ATG	C	G	A	G	A	A1	ps1F25L, ps1V88L, ps2R48T
	SHH253	M	39	20	223	5.94e+02	AG	**TCCT**	ATG	C	G	A	G	A	A1	ps1V88A, ps2R48T, sE164V
	SHH255	F	36	17	34	1.82e+02	AG	TCAT	ATG	C	G	A	G	A	A1	ps1V88A, ps2R48T, sP120A, sE164V
	SHH274	F	52	39	182	1.10e+04	**TA**	TCAT	**ACG**	C	G	G	G	A	A1	***psDel***, ps1F25L, ps2Q10del, ps2R48T, ps2A53V, sP120T, sV168A
HBsAg^{+ve} HBeAg^{-ve} anti-HBc^{+ve}	SHH001	M	23	38	172	3.59e+08	AG	TCAT	**CTG**	C	G	G	G	A	NS	NS
	SHH009	F	28	138	179	2.67e+05	**TA**	TCAT	ATG	C	G	A	G	A	NS	NS
	SHH011	F	25	21	156	1.39e+03	NS	NS	NS	NS	NS	NS	NS	NS	A1	***psDel***, ps1I48V
	SHH014	M	50	5	184	5.28e+03	NS	NS	NS	NS	NS	NS	NS	NS	A1	ps1F25L, ps1V88L, ps2Q10R, ps2R48T, ps2A53V, sQ129R, sV168A
	SHH016	M	34	55	9	6.25e+03	AG	GCAC	ATG	T	G	G	**A**	Non-A	NS	NS
	SHH022	F	19	28	132	3.35e+03	AG	TCAT	ATG	C	G	A	G	A	A1	-
	SHH048	F	29	7	255	1.58e+02	AG	TCAT	**TTG**	C	G	A	G	A	A1	-
	SHH070	F	41	290	246	5.16e+03	AG	TCAT	**CTG**	C	G	T	G	A	A1	-
	SHH100	M	27	86	148	6.31e+02	AG	TCAT	ATG	C	G	A	G	A	A1	**ps2M1L**
	SHH109	F	25	18	144	1.25e+07	AG	TCAT	ATG	C	G	A	G	A	A1	ps1V88A, ps2R48K
	SHH126	M	35	22	202	1.07e+07	AG	TCAT	ATG	C	G	A	G	A	NS	NS
	SHH148	M	47	32	99	1.49e+04	**TA**	**TCAC**	**AAG**	C	G	G	G	A	A1	-
	SHH167	M	38	129	49	2.86e+04	AG	TACT	ATG	T	G	A	**A**	Non-A	A1	***psDel***, ps1F25L, ps1V88L, ps2Q10del, ps2R48T, ps2A53V, sY100C

Table 1. Cont.

Characteristics of Participant							Molecular characteristics of HBV isolates									Mutations[a]
							Basic core promoter/precore region								S region	
Serology	Sample	Sex	Age	ALT (IU/L)	CD4 (cells/ml)	HBVVL (IU/ml)	1762, 1764	1809–1812	1814–1816	1858	1862	1888	1896	Genotype	Subgenotype	
	SHH180	M	53	60	106	9.09e+04	**TA**	**TTCT**	ATG	C	G	G	G	A	NS	NS
	SHH240	M	35	35	27	2.93e+05	**AA**	**TCCT**	ATG	C	G	A	G	A	NS	NS
	SHH256	M	34	153	201	4.61e+02	AG	TCAT	ATG	C	G	A	G	A	NS	NS
	SHH300	M	33	24	98	3.30e+02	AG	**ACAT**	**ACG**	C	G	G	G	A	A1	***psDel***, ps1F25L, ps1V88L, ps2M1T, ps2Q10del, ps2R48T, ps2A53V
HBsAg^{+ve} HBeAg^{-ve} anti-HBc^{-ve}	SHH042	F	30	11	289	1.98e+03	AG	GCAC	ATG	T	G	G	**A**	Non-A	NS	NS
	SHH055	F	24	8	261	4.00e+01	NS	NS	NS	NS	NS	NS	NS	NS	D3	-
HBsAg^{-ve} HBeAg^{-ve} anti-HBc^{+ve}	SHH027	F	28	13	9	7.39e+02	AG	**TCAC**	ATG	T	G	G	**A**	?	NS	NS
	SHH029	F	27	28	245	1.14e+09	AG	TCAT	ATG	C	G	A	G	A	NS	NS
	SHH032	F	30	23	265	1.26e+03	AG	GCAC	ATG	C	G	G	**A**	A	A1	ps1I48T
	SHH039	M	38	30	46	2.79e+04	AG	TCAT	ATG	C	G	A	G	A	A1	ps1V88A, ps2R48T, sE164A, **sS174N**[¶]
	SHH043	F	41	115	58	1.92e+03	NS	NS	NS	NS	NS	NS	NS	NS	A1	ps1I48V
	SHH044	F	64	35	103	1.43e+05	AG	TCAT	ATG	C	**T**	A	G	A	NS	**pcG1862T**[¶]
	SHH045	F	18	8	151	6.33e+03	NS	NS	NS	NS	NS	NS	NS	NS	A1	***psDel***, ps1I48V, ps2M1I, ps2R48K
	SHH061	F	24	17	227	1.44e+03	**TA**	**TCTT**	ATG	C	G	T	G	A	NS	NS
	SHH074	M	47	4	68	9.47e+03	NS	NS	NS	NS	NS	NS	NS	NS	A1	ps1F25L, ps1V88L, ps2Q10R, ps2A53V
	SHH083	F	33	24	279	6.15e+05	AG	TCAT	ATG	C	G	A	G	A	NS	NS
	SHH094	F	34	24	196	7.17e+05	AG	**ACAC**	ATG	C	G	G	G	A	NS	NS
	SHH107	M	36	122	185	5.00e+01[b]	AG	TCAT	ATG	C	G	A	G	A	A1	sQ129R
	SHH110	F	25	20	148	1.01e+04	AG	TCAT	**CTG**	C	G	A	G	A	A1	sV168A, **sS174N**[¶]
	SHH128	F	38	36	97	2.08e+03	AG	TCAT	ATG	C	G	A	G	A	NS	NS
	SHH130	M	61	36	47	6.57e+03	AG	TCAT	**CTG**	C	G	C	G	A	A1	ps1V88A, ps2M1I, sQ129H, sE164D, sV168A, **sS174N**[¶]
	SHH131	F	28	7	255	1.12e+03	AG	TCAT	ATG	C	G	A	G	A	A1	-

Table 1. Cont.

Characteristics of Participant							Molecular characteristics of HBV isolates									Mutations[a]
							Basic core promoter/precore region								S region	
Serology	Sample	Sex	Age	ALT (IU/L)	CD4 (cells/ml)	HBVVL (IU/ml)	1762, 1764	1809–1812	1814–1816	1858	1862	1888	1896	Genotype	Subgenotype	
	SHH132	M	45	17	44	4.98e+03	AG	TCAT	ATG	C	G	A	G	A	NS	NS
	SHH162	F	43	17	182	1.69e+03	AG	TCAT	ATG	C	G	G	G	A	NS	NS
	SHH193	F	23	25	12	4.39e+03	AG	**TCTT**	ATG	C	G	A	G	A	A1	ps1F25L, ps1V88L, ps2Q10R, ps2R48T, ps2A53V, sY100C, sQ129R, sV168A, **sS174N**[¶]
	SHH217	F	23	16	173	1.01e+04	AG	GCAC	ATG	C	G	G	G	A	NS	NS
	SHH221	F	41	18	74	7.57e+03	**TA**	GCAC	**CTG**	C	**T**	A	G	A	A1	**pcG1862T**[¶], ps1I48V, sV168A
	SHH246	M	37	111	87	3.26e+04	AG	TCAT	ATG	C	**T**	A	G	A	NS	**pcG1862T**[¶]
	SHH249	F	65	25	143	1.30e+04	AG	GCAC	ATG	C	G	G	G	A	NS	NS
	SHH264	M	28	24	101	9.02e+06	**TA**	TCAT	ATG	C	G	G	G	A	NS	NS
	SHH270	F	54	12	234	4.14e+03	NS	NS	NS	NS	NS	NS	NS	NS	A1	ps1F25L, ps1V88L, ps2Q10R, ps2R48T, ps2A53V
HBsAg^{-ve} HBeAg^{+ve} anti-HBc^{-ve}	SHH187	M	63	14	87	9.38e+02	AG	TCAT	ATG	C	**T**	A	G	A	NS	**pcG1862T**[¶]
HBsAg^{-ve} HBeAg^{-ve} anti-HBc^{-ve}	SHH037	F	35	35	277	2.97e+06	NS	NS	NS	NS	NS	NS	NS	NS	A1	ps1I48V
	SHH053	F	37	45	196	8.60e+03	AG	GCAC	ATG	T	**T**	G	G	Non-A	NS	**pcG1862T**[¶]
	SHH060	F	31	15	180	1.96e+02[b]	AG	GCAC	ATG	C	G	G	G	A	NS	NS
	SHH123	M	35	32	168	1.02e+05	AG	TCAT	ATG	C	G	A	G	A	NS	NS
	SHH173	M	34	8	110	5.86e+03	AG	TCAT	ATG	C	**T**	A	G	A	NS	**pcG1862T**[¶]
	SHH184	F	51	5	225	6.45e+03	**TA**	**TCTT**	ATG	C	G	T	G	A	NS	-
	SHH219	F	39	4	214	1.98e+03	AG	TCAT	ATG	C	**T**	A	G	A	NS	**pcG1862T**[¶]

[a]: -: no mutations, NS: no sequence, pc: precore, ps1:pre-S1, ps2:pre-S2, s:HBsAg,*psDel*: preS deletion,
[¶]: occurred at significantly higher frequency in isolates from HBsAg^{-ve} individuals compared to HBsAg^{+ve} ones.
[b]: HBsAg^{-ve} individuals with HBV viral loads (HBVVL) <200 IU/ml representing true occult infection [21].
doi:10.1371/journal.pone.0046345.t001

a $HBeAg^{+ve}$ participant, SHH253, had Kozak sequence TCCT, which would down- regulate but not abolish, HBeAg expression. Another isolate, SHH274, from a $HBeAg^{+ve}$ participant, had a start codon mutation, together with 1762T/1764A. It is possible that HBeAg was expressed from a minor population.

The 44 BCP/PC sequences derived from $HBeAg^{-ve}$ individuals were analysed for mutations (Table 1), which are known to down-regulate the synthesis of HBeAg or abolish its expression. In subgenotype A1, these mutations may be at the transcriptional, translational, or at post translational levels [48]. Mutations 1762T/1764A known to down-regulate HBeAg expression at the transcriptional level occurred in eight isolates (SHH009, SHH061, SHH148, SHH180, SHH184, SHH221 and SHH264, SHH274), in five cases occurring together with the T1753C. The Kozak mutations (1809–1812), affecting the expression of HBeAg at translational level were found in ten isolates, in two occurring together with the precore initiation codon mutations (1814–1816). Four isolates (SHH027, SHH094, SHH148 and SHH300) had either a single or double mutation or six isolates (SHH061, SHH180, SHH184, SHH193, SHH240 and SHH253) had a triple mutation at the Kozak sequence. Eight isolates had precore initiation codon mutations (table 1), which completely abolish HBeAg expression at translational level. The G1862T, which interferes with post-translational modification of the HBeAg-precursor and affects HBeAg expression [49] occurred in isolates from 7 $HBeAg^{-ve}$ sera. The classical G1896A mutation occurred in five isolates and in four cases it occurred together with C1858T. An unusual mutation, T1884C, was detected in three HBV isolates (SHH0187, SHH219 and SHH246). Two of three isolates obtained from true occult infections were sequenced in this region: SHH060 and SHH107 and were wild-type for subgenotype A2 and A1, respectively (table 1).

Of the 39 samples belonging to subgenotype A1, for which the BCP/precore region was sequenced successfully, 18 were from $HBsAg^{+ve}$ and 21from $HBsAg^{-ve}$ sera. There was no difference in the frequency of the various BCP/PC mutations between the $HBsAg^{+ve}$ and $HBsAg^{-ve}$ groups, except for G1862T, which occurred in HBV from one third of $HBsAg^{-ve}$ participants (7/21), but in none of the isolates from 18 $HBsAg^{+ve}$ participants ($p<0.05$).

Phylogenetic and molecular analysis of the complete S region

The complete S region (position 2854–835 from the *EcoR*I site) was sequenced successfully for HBV isolates from 29 participants. Following phylogenetic analysis, 28 of 29 isolates clustered with subgenotype A1, whereas one isolate, SHH055, clustered with genotype D (Figure 1) and belonged to subgenotype D3 (*data not shown*). For the isolate from SHH167, a discordant result was obtained between the HBV genotyping deduced using BCP/PC sequences and the S region phylogenetic analysis of HBV i.e. "not genotype A"and "subgenotype A1", respectively.

Subgenotype A1 isolates from HBV/HIV co-infected individuals were compared to subgenotype A1 isolates from Asian and African countries using a circular unrooted phylogenetic tree (Figure 1). Seventeen isolates (SHH011, SHH014, SHH037, SHH039, SHH043, SHH045, SHH074, SHH109, SHH159, SHH167, SHH193, SHH221, SHH253, SHH255, SHH270, SHH274 and SHH300) clustered with the "Asian" cluster (*red*) and the remaining 11 isolates clustered with the African cluster (*green*). There was no clustering with respect to whether the isolates were derived from $HBsAg^{+ve}$ or $HBsAg^{-ve}$ samples.

Upon translation of the pre-S1/pre-S2, the majority of the subgenotype A1 isolates showed distinct subgenotype A1 amino acids Q54, V74, A86 and V91 in the pre-S1 region and L32 in the preS2 region [39,50]. Twenty isolates belonged to serological subtype *adw*2, eight, including the genotype D isolate, to *ayw*2 and one to *adr* (Figure 1). The majority of the isolates in the "Asian" cluster (*red*) had S5, S6, F25 in the pre-S1. The isolates in the African cluster (*green*) displayed greater variation with three geographically distinct clades: the largest consisting of southern African strains with S5, A6, F25 in the pre-S1, a second consisting of eastern African strains with L5, P6, F25 and a third one consisting of central African strains, which like the "Asian" strains, had S5, S6, F25 in the pre-S1 (Figure 1). The majority of the isolates from the HIV-infected individuals sequenced in the present study displayed great variation at these signature positions (figure 1, table 1). The mean intragroup divergence ± standard deviation (%) of the complete S sequences of the newly sequenced strains from HIV-infected individuals was 2.43+0.12, whereas for the previously sequenced South African HBV subgenotype A1 isolates from HBV mono-infected individuals it was 1.92±0.75 [39].

Five newly sequenced HBV isolates had deletions in the pre-S1/pre-S2 region (figure 2):

1. *SHH011 pre-S1 and pre-S2 deletion mutant:* This mutant strain had a double deletion. The first, a 30 nucleotide deletion found in the preS1 region at position 2900 to 2929 from the *EcoR*I site, leading to a 10 amino acid deletion of the pre-S1 amino acids 16–26. The second, a 66 nucleotide deletion in the pre-S2 region at nucleotide position 3211 to 55 from the *EcoR*I site, leading to a 22 amino acid deletion of the pre-S2 amino acids 1–22.
2. *SHH045 pre-S2 deletion mutant*: This mutant had a 33 nucleotide deletion at position 23 to 55 from the *EcoR*I site, leading to an 11 amino acid deletion of the pre-S2 amino acids 11–22.
3. *SHH167 pre-S2 deletion mutant*: This mutant had a 45 nucleotide deletion at position 9 to 54 from the *EcoR*I site, leading to a 15 amino acid deletion in the preS2 region.
4. *SHH274 and SHH300 pre-S2 deletion mutants:* These mutant strains had a 54 nucleotide deletion at position 2 to 55 from the *EcoR*I site, leading to an 18 amino acid deletion in the pre-S2 region.

All deletion mutants, except for SHH045, were from $HBsAg^{+ve}$ participants.

The following mutations occurred significantly more frequently in HBV isolated from HIV-co-infected individuals in this study than in strains of the same cluster of the phylogenetic tree: in the pre-S1, ps1F25L, ps1V88L/A; in the pre-S2, ps2Q10R, ps2 R48K/T, ps2A53V and in the S region sQ129R/H, sQ164A/V/G/D, sV168A and sS174N ($p<0.05$). In the pre-S1, ps1I48V/T occurred more frequently in females than males ($p<0.05$). Isolates with sV168A occurred more frequently in participants with viral loads greater than 200 IU per ml ($p<0.05$). A mutation in ps2M1 of the preS2 abolished the start codon in four isolates. When comparing the frequency of mutations in $HBsAg^{+ve}$ and $HBsAg^{-ve}$ individuals, only sS174N occurred more frequently in $HBsAg^{-ve}$ individuals ($p<0.05$). sQ129R was the only mutation detected in the only isolate sequenced from a true occult HBV infection (SHH107, viral load <200 IU/ml) [21]. The relevant S region mutations are shown relative to the 'a' determinant (Figure 3).

Molecular analysis of the polymerase region

In the polymerase region, the following mutations occurred significantly more frequently in HBV isolated from HIV-co-infected individuals than in isolates of the same cluster on the

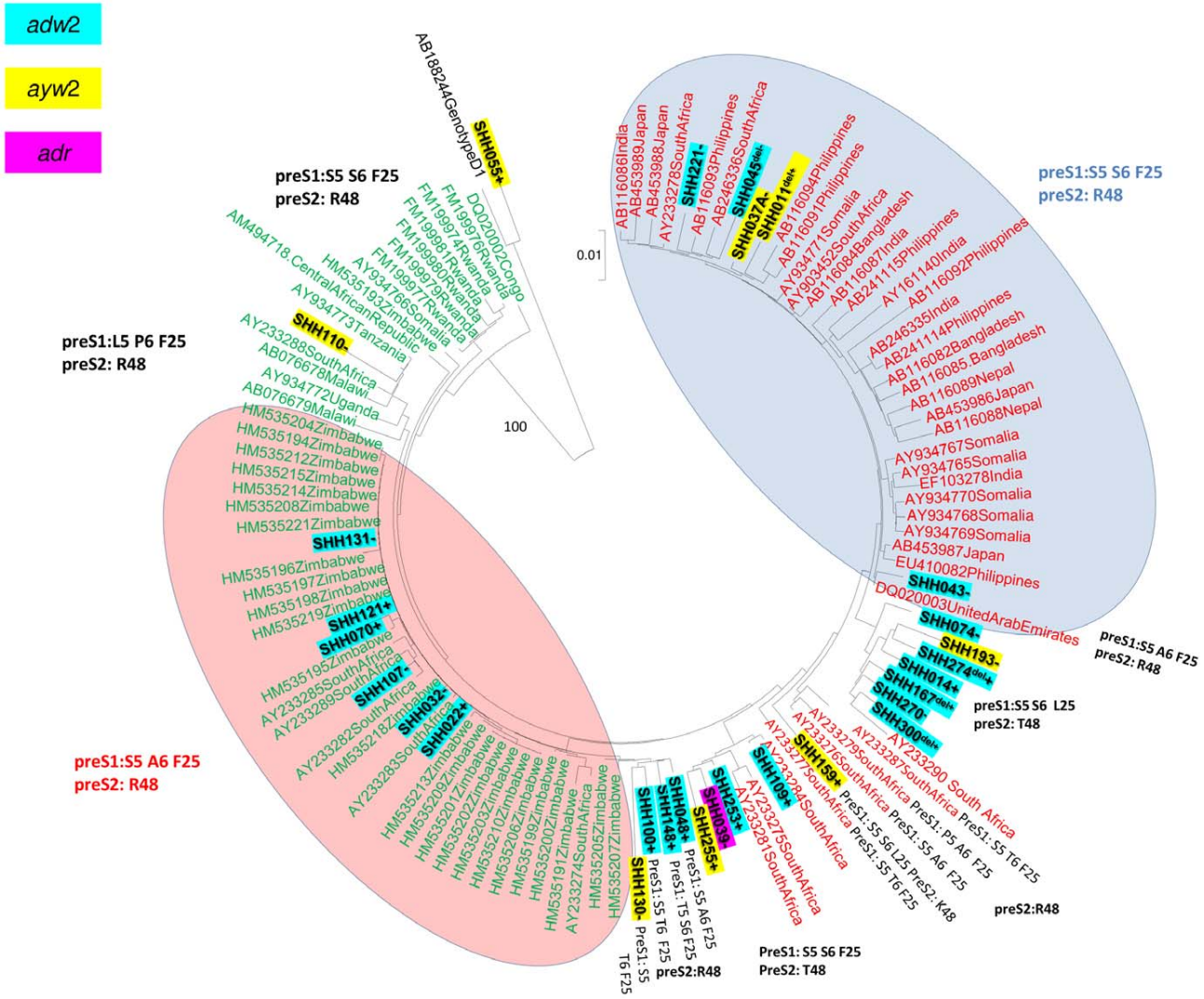


Figure 1. Phylogenetic relationship of complete pre-S1/pre-S2/S sequences (nt 2854–835 from the *EcoR*I site, numbering according to GenBank accession #AY233274) of 29 HBV isolates from HIV infected particpants [isolate number in bold, +: HBsAg^{+ve}, −:HBsAg^{-ve}, del: deletion mutant] to sequences of other African (*green*) and ''Asian'' (*red*) subgenotype A1 HBV isolates obtained from GenBank established using neighbour-joining. Bootstrap statistical analysis was performed using 1000 data sets and the numbers on the nodes indicate the percentage of occurrences. Each sequence obtained from GenBank is designated by its accession number and its country of origin. The characteristic amino acids in S open reading frame are indicated next to the sequences or relevant clades.
doi:10.1371/journal.pone.0046345.g001

phylogenetic tree: spQ23K, spL28P, spS91I, spP132Q, spQ125E and rtE1D ($p<0.05$). The unusual start codon mutation, rtE1D, was seen in eight isolates, which grouped in the ''Asian'' cluster following phylogenetic analysis. This mutation occurred together with rtS105T+rtH122N in 7 isolates and with rtQ125E in three isolates. Analysis of 457 sequences from GenBank revealed that the rtE1D mutation was found in genotype B and G isolates from Asian countries. Glutamic acid (E) and aspartic acid (D) have similar chemical structures and properties; therefore this mutation is not expected to introduce a significant functional change to the reverse transcriptase polymerase and the isolates are probably replicative. Three isolates had drug resistant mutations: SHH011 had rtV214A, SHH074 had rtL180M+rtM204V and SHH130 had rtV173L. There was no significant difference in the frequency of polymerase mutations in HBV from HBsAg^{+ve} and HBsAg^{-ve} individuals. Only one of three isolates (SHH107) obtained from true occult infections was sequenced in this region and found to contain mutations only in the spacer region of the polymerase.

Discussion

Compared to areas of low endemicity, where HIV and HBV are most likely transmitted at the same time during sexual maturity, in southern Africa, where both viruses are endemic, HBV infection occurs before the age of 5 years and these children become chronic carriers of HBV in adulthood [8]. Therefore, the majority of South Africans are naturally protected by antibodies to HBV by the time they acquire HIV at the age of sexual maturity [5]. In a recently completed study of 298 participants, 231 (77.5%) showed at least one HBV marker, 134 (45%) were anti-HBs^{+ve}, either alone (11; 3.7%) or together with anti-HBc (123; 41.3%) [20]. However, immunosuppression, as a result of HIV infection, can lead to HBV

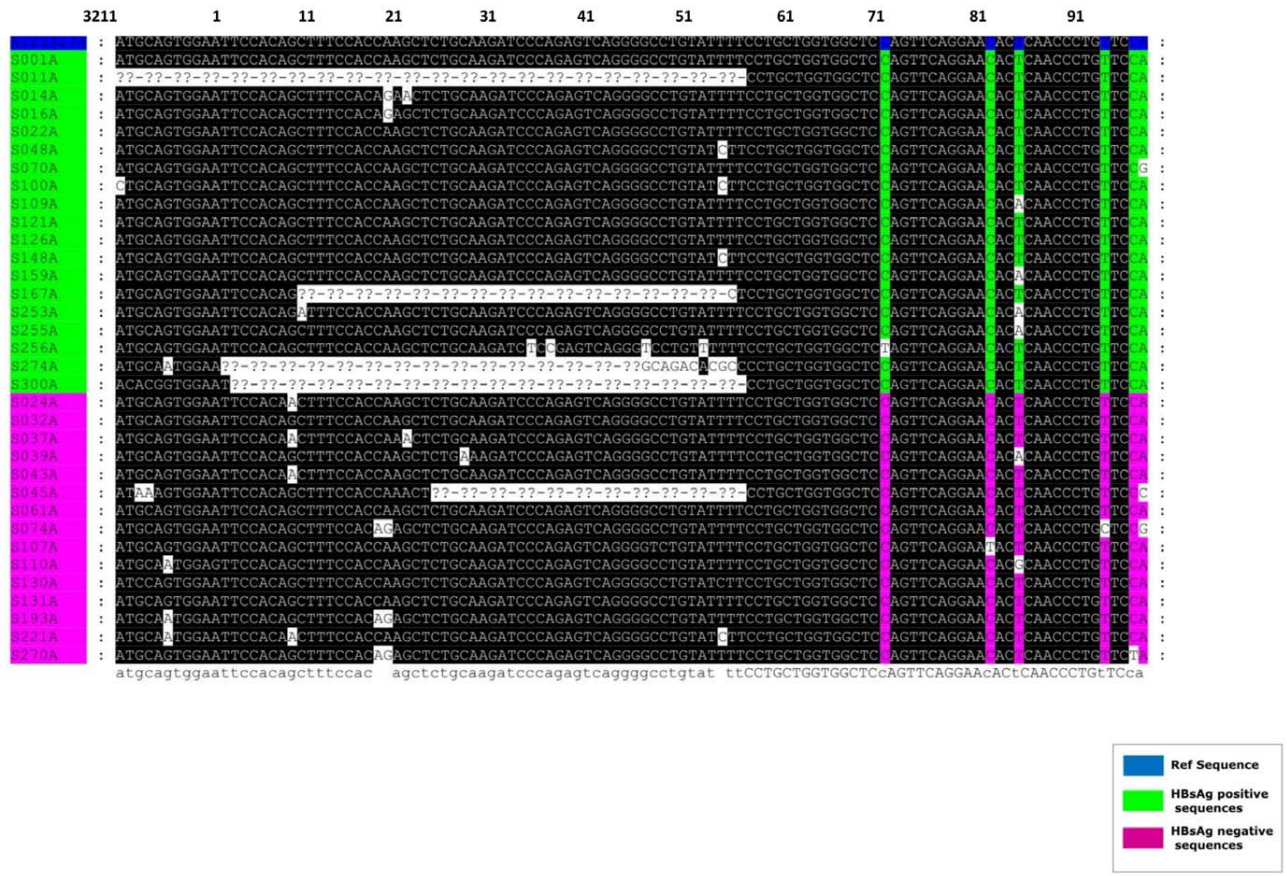


Figure 2. Deletions detected in the pre-S1/pre-S2 region of the HBV belonging to subgenotype A1 and isolated from HIV co-infected southern Africans. Position 1 corresponds to the *EcoR*I cleavage site of the HBV genome.
doi:10.1371/journal.pone.0046345.g002

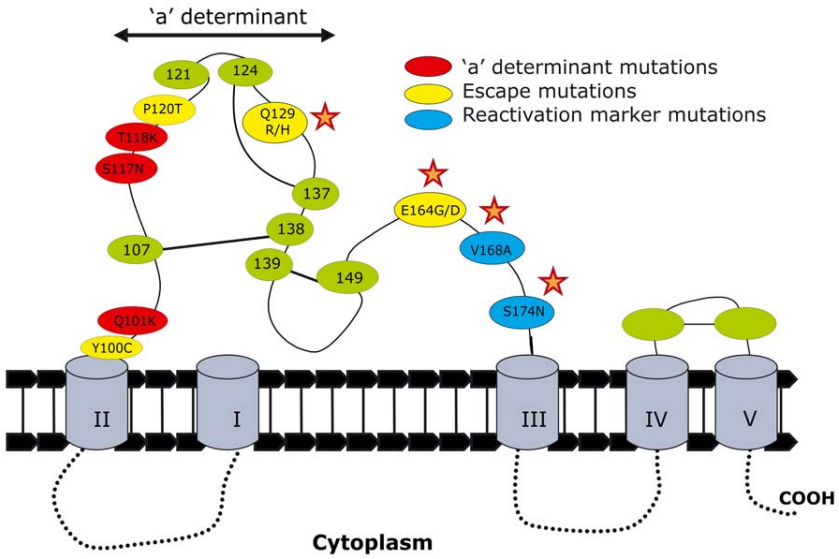


Figure 3. Graphic representation of mutations found within the small envelope protein of the HBV isolated from HIV infected participants. This is a hypothetical representation [91]. The mutations marked with a star occurred significantly more frequently in HBV isolated from HIV-co-infected individuals in this study than in strains of the same cluster of the phylogenetic tree (figure 1).
doi:10.1371/journal.pone.0046345.g003

infection and/or reactivation [51–53], increasing the frequency of HBV infection in HIV^{+ve} participants with previously resolved HBV infection. Of the 231 participants exposed to HBV, 53.7% were HBV DNA^{-ve} (resolved) and 23.8% HBV DNA^{+ve} (current) [8.7% $HBsAg^{+ve}$: 15.1% $HBsAg^{-ve}$] [20]. In the present study we determined the HBVgenotypes and molecularly characterized the HBV isolated from these HBV and HIV- co-infected southern Africans.

Subgenotype A1 was found to be the most prevalent subgenotype in rural South African HIV-infected individuals. In agreement with others, who found a predominance of genotype A of HBV in HIV infected individuals [54,55], we found that the ratio of genotype A to non-A (97% to 3%) was higher in the HBV/HIV co-infected individuals compared to mono-infected individuals. Previous genotyping showed a 75%:25% ratio of genotype A to D in South African mono-infected asymptomatic carriers and liver disease participants [39] and $HBsAg^{-ve}$ blood donors [56], whereas this was higher in $HBsAg^{+ve}$ blood donors (90%:10%). Discordant results between the genotypes deduced when the BCP/PC region and when the S region was sequenced were obtained for isolate SHH0167 (table 1). It is possible that this participant was infected with both genotypes A and D or with a genotype A/D recombinant, which has been shown to occur in South Africa [40]. The only way that this could be differentiated would be by carrying out full genome amplification to identify the recombinant and/or cloning to identify the mixed population.

Upon phylogenetic analysis, the HBV isolates were found in both African and "Asian" clusters (Figure 1), intimating a high diversity in the strains circulating in the rural cohort residing in a relatively small geographical region. This high diversity may be indicative of the high mobility of populations to and from this region. The cohort site, in the Shongwe region, is close to the borders with Swaziland and Mozambique. High levels of migration from these surrounding countries [57], may lead to higher risk of sexually transmitted infections including HBV and further increased genetic variability of HBV. A number of mutations occurred significantly more frequently in HBV isolated from HIV-co-infected individuals in this study than in strains of the same cluster of the phylogenetic tree (figure 1). This was probably as a result of the immunosuppression, which can alter the evolutionary rate of the virus [58]. The isolates from the African cluster showed greater variation than the "Asian" cluster, which comprised mostly Asian subgenotype A1 isolates, with some South African and Somalian isolates. This concurs with the hypothesis that subgenotype A1 has been endemic in the African population for a long period of time [1].

The HBeAg negativity found in 44/49 Shongwe participants (89,7%) could be accounted for by the following HBV mutations: the basic core promoter mutations A1762T/G1764A, which can down-regulate transcription of precore mRNA [59]; the Kozak sequence mutants that affect HBeAg translation [60]; precore start codon mutations that abolish HBeAg expression [61], the G1862T mutation,which interferes with post-translational modification of the HBeAg-precursor [49,62], and the classical G1896A stop codon mutation with C1858T [63]. The 1762T/1764A mutations have been closely related to progression of chronic liver disease [41,64,65] and together with T1753C, found in five Shongwe HBV isolates, have been described as markers for HCC [66–68]. The G1862T mutation occurred in HBV from $HBsAg^{-ve}$ participants but not in HBV from $HBsAg^{+ve}$ participants ($p<0.05$). This mutation causes intracellular retention of the HBeAg precursor, aggresome formation and impaired secretion of HBeAg [49]. It is possible that the intracellular retention of the HBeAg interferes with the expression of HBsAg, leading to HBsAg-negativity.

Pre-S/S sequence data were analyzed in an attempt to explain the high HBsAg-negativity, which was a feature of this rural cohort of HBV/HIVco-infected individuals [20]. Five HBV strains isolated from HBV/HIV co-infected participants had pre-S deletions. Only one of the five participants from which these strains were isolated was $HBsAg^{-ve}$. All five had pre-S2 deletions, ranging in size from 11 to 22 amino acids, with one isolate having, in addition, a 10 amino acid pre-S1 deletion. Deletion mutants in the pre-S region have been previously found to occur more frequently in HIV-coinfected individuals [69] and in HCC patients [70]. Deletion mutants in the pre-S region have been reported to cause overproduction and accumulation of LHBs protein in the endoplasmic reticulum (ER), which causes significant ER stress that may induce DNA damage and genomic instability and hence play a possible role in hepatocarcinogenesis [71,72]. The pre-S mutants may contribute to viral oncogenesis by transcriptional activation of the viral promoter elements [73]. A higher oncogenic potential of pre-S2 deletions has been found compared with that of pre-S1 deletion mutants [74], with pre-S1 deletion mutants displaying different phenotypes to pre-S2 deletion mutants, when transfected in Huh-7 cells [75].

Although, with the exception of sS174N, there was no significant difference between the frequency of the mutations in the S region from $HBsAg^{+ve}$ and $HBsAg^{-ve}$ participants, there were a number of mutations that could account for the inability to detect HBsAg. These included ps1F25L, ps1I48V, ps1V88L/A in the pre-S1 region, ps2M1I, ps2Q10R, ps2R48K/T, ps2A53V in the pre-S2 and sY100C, sP120T/A, sQ129R/H, sE164D and sS174N in the S region (figure 3). Pre-S1 residues 21–96 contain the virus neutralising epitope and a hepatocellular binding site, and the preS2 residues 1–11 and 21–47 have been shown to mediate HBV attachment to hepatocytes [76]. Therefore mutations in these regions may lead to conformational changes in the LHBs and MHBs, which may result in HBsAg negativity. ps2Q10R is in the major pre-S2 antigenic region known to carry numerous B cell, T-helper cell and cytotoxic T-lymphocyte epitopes [77] and may therefore reduce the binding ability of the antibodies to the epitope and interfere with their neutralising effect. sY100C has previously been detected in subgenotype A1 isolated from occult hepatitis participants [70,71] and has been associated with HBsAg-negativity in blood donors [78]. sP120T, which also leads to an rt128N mutation in the polymerase, has been detected in participants with severe hepatitis following lamivudine (LAM) and HBIg treatment [79] and can partially restore the replicative capacity of LAM-resistant HBV *in vitro* [80]. sP120T and sQ129R/H fall within the 'a' determinant and would therefore affect its antigenic ability and infectivity [81]. sQ164A, has been shown to reduce the antigenicity of HDV particles [81] and sE164D, with the concomitant rtV173L, a lamivudine escape mutant [82], has reduced affinity for anti-HBs antibodies *in vitro*, similar to that of the classical G145R [83].

Of interest was the presence of previously identified reactivation markers [84]. The reactivation markers V168A occurred together with S174N, which was only found in isolates from $HBsAg^{-ve}$ but not $HBsAg^{+ve}$ participants. HBV with V168A occurred more frequently in participants with higher viral loads. These have previously been detected in a serologically-negative HBV/HIV co-infected patient following a symptomatic HBV reactivation [84].

The S region mutations detected in the present study differed from those detected previously in $HBsAg^{-ve}$ blood donors, with true occult HBV infection and low viral loads of subgenotype A1 [56]. The HBV mutants in the present study had viral loads

>200 IU/ml, escaped detection by HBsAg assays and probably arose during reactivation or were transmitted to these unvaccinated individuals, in the absence of an immune response or exogenous selective pressure such as vaccination or ART. The presence of minority populations in the quasispecies expressing wild-type HBsAg may account for the fact that these mutations were also isolated from $HBsAg^{+ve}$ individuals. This possibility is being investigated using ultra-deep pyrosequencing.

Even though the participants in the present study had not initiated ART, ten percent, 3 of 29 isolates sequenced, had drug resistance mutations rtV173L, rtL180M+rtM204V and rtV214A, respectively. Mutants rt173L and rt180M have been shown to restore viral replication in the presence of LAM, whereas rt204V results in reduced replication in *in vitro* transfection studies [85,86]. In South Africa, rtM204I has been detected in therapy-naïve HBV/HIV co-infected individuals [87] and rtM204V in treated HBV mono-infected participants [88]. All the mutations described occurred in genotype A. Compared to other genotypes, genotype A in HBV-HIV co-infected participants has been shown to be more prone to immune/vaccine escape mutants, pre-S mutants associated with immune suppression, drug associated mutations and HCC [69,89,90].

In conclusion, the study showed that subgenotype A1 predominates in HBV/HIV co-infected individuals from rural South Africa. Subgenotype A1 HBV isolates had mutations that can affect HBeAg-expression at the transcriptional, translational and posttranslational levels and these mutations can account for the HBeAg negativity seen in the majority of HBV/HIV infected individuals in this cohort. Although there were no significant differences between all S region mutations occurring in HBV from $HBsAg^{+ve}$ and $HBsAg^{-ve}$ individuals, pre-S region mutations and deletions, and 'a'determinant or immune/vaccine escape mutants may account for HBsAg negativity seen in some participants. Moreover, these mutants may escape detection by standard commercial serological tests currently used in South African health services. HBV infection will remain undetected unless nucleic acid testing, which can detect HBV DNA in the presence or absence of HBsAg in the serum, is implemented. Deletion mutants, previously shown to occur in HBV from HCC patients, were detected in the present study. These could be a risk factor for the development of HCC in HIV patients, whose lifespan is being increased and immune system reconstituted following the introduction of ART. Thus more studies are necessary to functionally characterize these deletion mutants. Furthermore, the presence of mutants resistant to LAM, prior to the initiation of LAM-containing ART, has important implications and repercussions, and highlights the need for the inclusion in the treatment regimen of tenofovir (TDF), to which these mutants are sensitive. This study has provided important information on the molecular characteristics of HBV in HIV-infected South Africans prior to the initiation of ART. Our findings have important clinical relevance in the management of HBV and HIV co-infection in our unique setting, where subgenotype A1 HBV is hyperendemic and usually transmitted horizontally in childhood, before HIV infection occurs in adulthood.

Acknowledgments

The authors thank Dr Neil Martinson of the Perinatal HIV Unit, University of the Witwatersrand for help in establishing the cohort; Sisters Ntombikayise Elizabeth Agatha Nkosi and Rosalina Candlovu for enrolling participants and drawing blood samples; Dr Precious Gabashane, chief medical officer, Shongwe Hospital and all the study participants.

Author Contributions

Conceived and designed the experiments: AK EM. Performed the experiments: EM TGB. Analyzed the data: EM TGB AK. Contributed reagents/materials/analysis tools: AK TGB. Wrote the paper: EM TGB AK.

References

1. Kramvis A, Kew MC (2007) Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes. Hepatol Res 37: S9–S19.
2. UNAIDS (2010) UNAIDS report on the global AIDS epidemic 2010. Available: http://www.unaids.org/en/dataanalysis/. Accessed 2011 November 16.
3. Acar A, Kemahli S, Altunay H, Kosan E, Oncul O, et al. (2010) HBV, HCV and HIV seroprevalence among blood donors in Istanbul, Turkey: how effective are the changes in the national blood transfusion policies? Braz J Infect Dis 14: 41–46.
4. Kottilil S, Jackson JO, Polis MA (2005) Hepatitis B & hepatitis C in HIV-infection. Indian J Med Res 121: 424–450.
5. Burnett RJ, Francois G, Kew MC, Leroux-Roels G, Meheus A, et al. (2005) Hepatitis B virus and human immunodeficiency virus co-infection in sub-Saharan Africa: a call for further investigation. Liver Int 25: 201–213.
6. Botha JF, Ritchie MJ, Dusheiko GM, Mouton HW, Kew MC (1984) Hepatitis B virus carrier state in black children in Ovamboland: role of perinatal and horizontal infection. Lancet 1: 1210–1212.
7. Di Bisceglie AM, Dusheiko GM, Kew MC (1985) Detection of markers of hepatitis B virus infection in urine of chronic carriers. J Med Virol 16: 337–341.
8. Kew MC (1996) Progress towards the comprehensive control of hepatitis B in Africa: a view from South Africa. Gut 38 Suppl 2: S31–36.
9. Vardas E, Mathai M, Blaauw D, McAnerney J, Coppin A, et al. (1999) Preimmunization epidemiology of hepatitis B virus infection in South African children. J Med Virol 58: 111–115.
10. Hoffmann CJ, Charalambous S, Martin DJ, Innes C, Churchyard GJ, et al. (2008) Hepatitis B virus infection and response to antiretroviral therapy (ART) in a South African ART program. Clin Infect Dis 47: 1479–1485.
11. Barth RE, Huijgen Q, Tempelman HA, Mudrikova T, Wensing AM, et al. (2011) Presence of occult HBV, but near absence of active HBV and HCV infections in people infected with HIV in rural South Africa. J Med Virol 83: 929–934.
12. Burnett RJ, Ngobeni JM, Francois G, Hoosen AA, Leroux-Roels G, et al. (2007) Increased exposure to hepatitis B virus infection in HIV-positive South African antenatal women. Int J STD AIDS 18: 152–156.
13. Lukhwareni A, Burnett RJ, Selabe SG, Mzileni MO, Mphahlele MJ (2009) Increased detection of HBV DNA in HBsAg-positive and HBsAg-negative South African HIV/AIDS patients enrolling for highly active antiretroviral therapy at a Tertiary Hospital. J Med Virol 81: 406–412.
14. Mphahlele MJ, Lukhwareni A, Burnett RJ, Moropeng LM, Ngobeni JM (2006) High risk of occult hepatitis B virus infection in HIV-positive patients from South Africa. J Clin Virol 35: 14–20.
15. Firnhaber C, Reyneke A, Schulze D, Malope B, Maskew M, et al. (2008) The prevalence of hepatitis B co-infection in a South African urban government HIV clinic. S Afr Med J 98: 541–544.
16. Firnhaber C, Viana R, Reyneke A, Schultze D, Malope B, et al. (2009) Occult hepatitis B virus infection in patients with isolated core antibody and HIV co-infection in an urban clinic in Johannesburg, South Africa. Int J Infect Dis 13: 488–492.
17. Firnhaber C, Chen CY, Evans D, Maskew M, Schulz D, et al. (2012) Prevalence of hepatitis B virus (HBV) co-infection in HBV serologically-negative South African HIV patients and retrospective evaluation of the clinical course of mono- and co-infection. Int J Infect Dis 16: e268–272.
18. Boyles TH, Cohen K (2011) The prevalence of hepatitis B infection in a rural South African HIV clinic. S Afr Med J 101: 470–471.
19. Matthews GV, Manzini P, Hu Z, Khabo P, Maja P, et al. (2011) Impact of lamivudine on HIV and hepatitis B virus-related outcomes in HIV/hepatitis B virus individuals in a randomized clinical trial of antiretroviral therapy in southern Africa. AIDS 25: 1727–1735.
20. Bell TG, Makondo E, Martinson MA, Kramvis A (2012) Hepatitis B Virus Infection in Human Immunodeficiency Virus Infected Southern African Adults: Occult or Overt - That is the Question. PlosOne (in press): e45750. doi:10.1371/journal.pone.0045750.
21. Raimondo G, Allain JP, Brunetto MR, Buendia MA, Chen DS, et al. (2008) Statements from the Taormina expert meeting on occult hepatitis B virus infection. J Hepatol 49: 652–657.
22. Fares MA, Holmes EC (2002) A revised evolutionary history of hepatitis B virus (HBV). J Mol Evol 54: 807–814.
23. Orito E, Mizokami M, Ina Y, Moriyama EN, Kameshima N, et al. (1989) Host-independent evolution and a genetic classification of the hepadnavirus family based on nucleotide sequences. Proc Natl Acad Sci U S A 86: 7059–7062.

24. Gunther S, Sommer G, Von Breunig F, Iwanska A, Kalinina T, et al. (1998) Amplification of full-length hepatitis B virus genomes from samples from patients with low levels of viremia: frequency and functional consequences of PCR-introduced mutations. J Clin Microbiol 36: 531–538.
25. Norder H, Courouce AM, Magnius LO (1992) Molecular basis of hepatitis B virus serotype variations within the four major subtypes. J Gen Virol 73 (Pt 12): 3141–3145.
26. Okamoto H, Tsuda F, Sakugawa H, Sastrosoewignjo RI, Imai M, et al. (1988) Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. J Gen Virol 69 (Pt 10): 2575–2583.
27. Kramvis A, Arakawa K, Yu MC, Nogueira R, Stram DO, et al. (2008) Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. J Med Virol 80: 27–46.
28. Norder H, Courouce AM, Magnius LO (1994) Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. Virology 198: 489–503.
29. Naumann H, Schaefer S, Yoshida CF, Gaspar AM, Repp R, et al. (1993) Identification of a new hepatitis B virus (HBV) genotype from Brazil that expresses HBV surface antigen subtype adw4. J Gen Virol 74 (Pt 8): 1627–1632.
30. Stuyver L, De Gendt S, Van Geyt C, Zoulim F, Fried M, et al. (2000) A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. J Gen Virol 81: 67–74.
31. Arauz-Ruiz P, Norder H, Robertson BH, Magnius LO (2002) Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. J Gen Virol 83: 2059–2073.
32. Yu H, Yuan Q, Ge SX, Wang HY, Zhang YL, et al. (2010) Molecular and phylogenetic analyses suggest an additional hepatitis B virus genotype "I". PLoS One 5: e9297.
33. Arankalle VA, Gandhe SS, Borkakoty BJ, Walimbe AM, Biswas D, et al. (2010) A novel HBV recombinant (genotype I) similar to Vietnam/Laos in a primitive tribe in eastern India. J Viral Hepat: 501–510.
34. Tran TT, Trinh TN, Abe K (2008) New complex recombinant genotype of hepatitis B virus identified in Vietnam. J Virol 82: 5657–5663.
35. Olinger CM, Jutavijittum P, Hubschen JM, Yousukh A, Samountry B, et al. (2008) Possible new hepatitis B virus genotype, southeast Asia. Emerg Infect Dis 14: 1777–1780.
36. Osiowy C, Kaita K, Solar K, Mendoza K (2010) Molecular characterization of hepatitis B virus and a 9-year clinical profile in a patient infected with genotype I. J Med Virol 82: 942–948.
37. Tatematsu K, Tanaka Y, Kurbanov F, Sugauchi F, Mano S, et al. (2009) A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. J Virol 83: 10538–10547.
38. Kramvis A, Kew M, Francois G (2005) Hepatitis B virus genotypes. Vaccine 23: 2409–2423.
39. Kimbi GC, Kramvis A, Kew MC (2004) Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa. J Gen Virol 85: 1211–1220.
40. Kramvis A, Kew MC (2005) Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy. J Viral Hepat 12: 456–464.
41. Takahashi K, Aoyama K, Ohno N, Iwata K, Akahane Y, et al. (1995) The precore/core promoter mutant (T1762A1764) of hepatitis B virus: clinical significance and an easy method for detection. J Gen Virol 76 (Pt 12): 3159–3164.
42. Vermeulen M, Dickens C, Lelie N, Walker E, Coleman C, et al. (2012) Hepatitis B virus transmission by blood transfusion during 4 years of individual-donation nucleic acid testing in South Africa: estimated and observed window period risk. Transfusion 52: 880–892.
43. Nicholas K, Nicholas HBJ (1997) GeneDoc: Analysis and Visualization of Genetic Variation. Available: http://www.psc.edu/biomed/genedoc. Accessed 2010 November.
44. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.
45. Xia X (2000) Data analysis in molecular biology and evolution Boston, Dordrecht, London: Kluwer Academic Publishers.
46. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425.
47. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111–120.
48. Kramvis A, Kew MC (2007) Molecular characterization of subgenotype A1 (subgroup Aa) of hepatitis B virus. Hepatol Res 37: S27–32.
49. Chen CY, Crowther C, Kew MC, Kramvis A (2008) A valine to phenylalanine mutation in the precore region of hepatitis B virus causes intracellular retention and impaired secretion of HBe-antigen. Hepatol Res 38: 580–592.
50. Bowyer SM, van Staden L, Kew MC, Sim JG (1997) A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa. J Gen Virol 78 (Pt 7): 1719–1729.
51. Pastore G, Santantonio T, Monno L, Milella M, Luchena N, et al. (1988) Effects of HIV superinfection on HBV replication in a chronic HBsAg carrier with liver disease. J Hepatol 7: 164–168.
52. Thio CL (2009) Hepatitis B and human immunodeficiency virus coinfection. Hepatology 49: S138–145.
53. Vento S, Di Perri G, Garofano T, Concia E, Bassetti D (1989) Reactivation of hepatitis B in AIDS. Lancet 2: 108–109.
54. Audsley J, Arrifin N, Yuen LK, Ayres A, Crowe SM, et al. (2009) Prolonged use of tenofovir in HIV/hepatitis B virus (HBV)-coinfected individuals does not lead to HBV polymerase mutations and is associated with persistence of lamivudine HBV polymerase mutations. HIV Med 10: 229–235.
55. Quarleri J, Moretti F, Bouzas MB, Laufer N, Carrillo MG, et al. (2007) Hepatitis B virus genotype distribution and its lamivudine-resistant mutants in HIV-coinfected patients with chronic and occult hepatitis B. AIDS Res Hum Retroviruses 23: 525–531.
56. Allain JP, Belkhiri D, Vermeulen M, Crookes R, Cable R, et al. (2009) Characterization of occult hepatitis B virus strains in South African blood donors. Hepatology 49: 1868–1876.
57. Collinson MA, Tollman SM, Kahn K (2007) Migration, settlement change and health in post-apartheid South Africa: triangulating health and demographic surveillance with national census data. Scand J Public Health Suppl 69: 77–84.
58. Zhou Y, Holmes EC (2007) Bayesian estimates of the evolutionary rate and age of hepatitis B virus. J Mol Evol 65: 197–205.
59. Buckwold VE, Xu Z, Yen TS, Ou JH (1997) Effects of a frequent double-nucleotide basal core promoter mutation and its putative single-nucleotide precursor mutations on hepatitis B virus gene expression and replication. J Gen Virol 78 (Pt 8): 2055–2065.
60. Ahn SH, Kramvis A, Kawai S, Spangenberg HC, Li J, et al. (2003) Sequence variation upstream of precore translation initiation codon reduces hepatitis B virus e antigen production. Gastroenterology 125: 1370–1378.
61. Laras A, Koskinas J, Avgidis K, Hadziyannis SJ (1998) Incidence and clinical significance of hepatitis B virus precore gene translation initiation mutations in e antigen-negative patients. J Viral Hepat 5: 241–248.
62. Hou J, Lin Y, Waters J, Wang Z, Min J, et al. (2002) Detection and significance of a G1862T variant of hepatitis B virus in Chinese patients with fulminant hepatitis. J Gen Virol 83: 2291–2298.
63. Carman WF, Jacyna MR, Hadziyannis S, Karayiannis P, McGarvey MJ, et al. (1989) Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. Lancet 2: 588–591.
64. Baptista M, Kramvis A, Kew MC (1999) High prevalence of 1762(T) 1764(A) mutations in the basic core promoter of hepatitis B virus isolated from black Africans with hepatocellular carcinoma compared with asymptomatic carriers. Hepatology 29: 946–953.
65. Orito E, Mizokami M, Sakugawa H, Michitaka K, Ishikawa K, et al. (2001) A case-control study for clinical and molecular biological differences between hepatitis B viruses of genotypes B and C. Japan HBV Genotype Research Group. Hepatology 33: 218–223.
66. Yeh CT, So M, Ng J, Yang HW, Chang ML, et al. (2010) Hepatitis B virus-DNA level and basal core promoter A1762T/G1764A mutation in liver tissue independently predict postoperative survival in hepatocellular carcinoma. Hepatology 52: 1922–1933.
67. Chen WN, Oon CJ (2000) Mutations and deletions in core promoter and precore stop codon in relation to viral replication and liver damage in Singaporean hepatitis B virus carriers. Eur J Clin Invest 30: 787–792.
68. Zhang D, Ma S, Zhang X, Zhao H, Ding H, et al. (2012) Prevalent HBV point mutations and mutation combinations at BCP/preC region and their association with liver disease progression. BMC Infect Dis 10: 271.
69. Audsley J, Littlejohn M, Yuen L, Sasadeusz J, Ayres A, et al. (2010) HBV mutations in untreated HIV-HBV co-infection using genomic length sequencing. Virology 405: 539–547.
70. Kew MC, Kramvis A, Yu MC, Arakawa K, Hodkinson J (2005) Increased hepatocarcinogenic potential of hepatitis B virus genotype A in Bantu-speaking sub-saharan Africans. J Med Virol 75: 513–521.
71. Hsieh YH, Su IJ, Wang HC, Chang WW, Lei HY, et al. (2004) Pre-S mutant surface antigens in chronic hepatitis B virus infection induce oxidative stress and DNA damage. Carcinogenesis 25: 2023–2032.
72. Chen BF, Liu CJ, Jow GM, Chen PJ, Kao JH, et al. (2006) High prevalence and mapping of pre-S deletion in hepatitis B virus carriers with progressive liver diseases. Gastroenterology 130: 1153–1168.
73. Hildt E, Saher G, Bruss V, Hofschneider PH (1996) The hepatitis B virus large surface protein (LHBs) is a transcriptional activator. Virology 225: 235–239.
74. Yeung P, Wong DK, Lai CL, Fung J, Seto WK, et al. (2011) Association of hepatitis B virus pre-S deletions with the development of hepatocellular carcinoma in chronic hepatitis B. J Infect Dis 203: 646–654.
75. Lin CM, Wang GM, Jow GM, Chen BF (2012) Functional analysis of hepatitis B virus pre-s deletion variants associated with hepatocellular carcinoma. J Biomed Sci 19: 17.
76. Kim HS, Ryu CJ, Hong HJ (1997) Hepatitis B virus preS1 functions as a transcriptional activation domain. J Gen Virol 78 (Pt 5): 1083–1086.
77. Paulij WP, de Wit PL, Sunnen CM, van Roosmalen MH, Petersen-van Ettekoven A, et al. (1999) Localization of a unique hepatitis B virus epitope sheds new light on the structure of hepatitis B virus surface antigen. J Gen Virol 80 (Pt 8): 2121–2126.
78. Gutierrez C, Devesa M, Loureiro CL, Leon G, Liprandi F, et al. (2004) Molecular and serological evaluation of surface antigen negative hepatitis B virus infection in blood donors from Venezuela. J Med Virol 73: 200–207.

79. Bock CT, Tillmann HL, Torresi J, Klempnauer J, Locarnini S, et al. (2002) Selection of hepatitis B virus polymerase mutants with enhanced replication by lamivudine treatment after liver transplantation. Gastroenterology 122: 264–273.
80. Torresi J, Earnest-Silveira L, Civitico G, Walters TE, Lewin SR, et al. (2002) Restoration of replication phenotype of lamivudine-resistant hepatitis B virus mutants by compensatory changes in the "fingers" subdomain of the viral polymerase selected as a consequence of mutations in the overlapping S gene. Virology 299: 88–99.
81. Salisse J, Sureau C (2009) A function essential to viral entry underlies the hepatitis B virus "a" determinant. J Virol 83: 9321–9328.
82. Torresi J (2002) The virological and clinical significance of mutations in the overlapping envelope and polymerase genes of hepatitis B virus. J Clin Virol 25: 97–106.
83. Torresi J, Earnest-Silveira L, Deliyannis G, Edgtton K, Zhuang H, et al. (2002) Reduced antigenicity of the hepatitis B virus HBsAg protein arising as a consequence of sequence changes in the overlapping polymerase gene that are selected by lamivudine therapy. Virology 293: 305–313.
84. Henke-Gendo C, Amini-Bavil-Olyaee S, Challapalli D, Trautwein C, Deppe H, et al. (2008) Symptomatic hepatitis B virus (HBV) reactivation despite reduced viral fitness is associated with HBV test and immune escape mutations in an HIV-coinfected patient. J Infect Dis 198: 1620–1624.
85. Allen MI, Deslauriers M, Andrews CW, Tipples GA, Walters KA, et al. (1998) Identification and characterization of mutations in hepatitis B virus resistant to lamivudine. Lamivudine Clinical Investigation Group. Hepatology 27: 1670–1677.
86. Delaney WEt, Yang H, Westland CE, Das K, Arnold E, et al. (2003) The hepatitis B virus polymerase mutation rtV173L is selected during lamivudine therapy and enhances viral replication in vitro. J Virol 77: 11833–11841.
87. Selabe SG, Lukhwareni A, Song E, Leeuw YG, Burnett RJ, et al. (2007) Mutations associated with lamivudine-resistance in therapy-naive hepatitis B virus (HBV) infected patients with and without HIV co-infection: implications for antiretroviral therapy in HBV and HIV co-infected South African patients. J Med Virol 79: 1650–1654.
88. Selabe SG, Song E, Burnett RJ, Mphahlele MJ (2009) Frequent detection of hepatitis B virus variants associated with lamivudine resistance in treated South African patients infected chronically with different HBV genotypes. J Med Virol 81: 996–1001.
89. Thio CL, Seaberg EC, Skolasky R, Phair J, Visscher B, et al. (2002) HIV-1, hepatitis B virus, and risk of liver-related mortality in the Multicenter Cohort Study (MACS). Lancet 360: 1921–1926.
90. Iacomi F, Vincenti D, Vairo F, Solmone M, Mariano A, et al. (2009) Effect of HIV co-infection on mutation patterns of HBV in patients with lamivudine-resistant chronic hepatitis B. J Med Virol 81: 1151–1156.
91. Gous N, Bhimma R, Kew M, Kramvis A (2010) Retrospective characterization of the S open reading frame of HBV isolated from children with membranous nephropathy treated with interferon-alpha2b. Antivir Ther 15: 61–69.

Appendix D

Source Code for "Consolidated Data Table" Method

The source code, which generates the "Consolidated Data Table", is shown below. This method is called by the Python CGI script, when the user selects to view the consolidated data table.

Listing D.1: Source code for "Consolidated Data Table" method

```
1
2  def viewDataTable():
3          import copy
4
5          databaseconnection = psycopg2.connect(user='scryer', password='snow', database='hbvdb', port='5432', host='localhost')
6          minion = databaseconnection.cursor()
7
8          query = "select␣samplingevent.samplingeventcode,␣round(samplingevent.deltadays/30.4,␣1),␣participant.sex,␣(samplingevent.visitdate␣-␣participant.datebirth)/365␣as
               ␣age,␣bmi,␣weight,␣arvregimen␣as␣arv,␣serumvolume,␣plasmavolume,␣aliquotbox,␣extractedbox␣from␣samplingevent,␣participant,␣sample␣where␣samplingevent.
               participant␣=␣participant.participantcode␣and␣participant.followup='t'␣and␣samplingevent.samplingeventcode␣=␣sample.samplingevent␣order␣by␣samplingevent.
               samplingeventcode␣;"
9
10         minion.execute(query)
11         s = []
12         for l in minion:
13                 s.append([l[0], l[1], l[2], l[3], l[4], l[5], l[6], l[7], l[8], l[9], l[10]])
14
15         query = "select␣samplingevent.samplingeventcode,␣wetlab.test,␣wetlab.result,␣wetlab.operator␣from␣samplingevent␣left␣outer␣join␣wetlab␣on␣samplingevent.
               samplingeventcode␣=␣wetlab.participant␣||␣chr(wetlab.timepoint+65)␣and␣include␣=␣'t'␣where␣samplingevent.samplingeventcode␣in␣(select␣samplingevent.
               samplingeventcode␣from␣samplingevent,␣participant␣where␣samplingevent.participant␣=␣participant.participantcode␣and␣participant.followup='t'␣order␣by␣
               samplingevent.samplingeventcode)␣order␣by␣samplingevent.samplingeventcode,␣test,␣result␣;"
16         minion.execute(query);
17
18         t = []
19         for l in minion:
20                 t.append(l)
21
22         # ---------- Process IU/ml ----------
23         iu = []          # list for IU/ml
24         for i in t:
25                 if i[1] == 'QVL':
26                         iu.append((i[0], 'QVI', str(int(round(int(i[2])/4.7))), i[3]))  # added as a tuple; convert to interger, divide by 4.7, round, convert to integer
                               (again) and convert to string
27
28         for i in iu:
29                 t.append(i)      # add IU/ml to 'master' list
30         # ---------------------------------
```

```
31
32          # I could not include the person lookup in the left outer join query for some reason
33          # Therefore, I will query the codes and names from the person table and use them to lookup the name from a dictionary
34          query = "select␣personcode,␣name␣from␣person;"
35          minion.execute(query)
36          Person = {}
37          Person[None] = 'Unknown'
38          for l in minion:
39                  Person[l[0]] = l[1]
40
41          r = ['SAG', 'SAB', 'CAB', 'EAG', 'EAB', 'N01', 'N02', 'N05', 'N03', 'APO', 'ALT', 'CD4', 'QVL', 'QVI', 'CO2', 'APRI']
42
43          op = copy.deepcopy(s)           # x = y[:] is not even a deep copy; changing elements in x changes them in y
44
45          for i in s:
46                  for j in r:
47                          for k in t:
48                                  found = False
49                                  if k[0] == i[0]:        # for speed; if SEs don't match, iterate
50                                          if k[1] == j:  # tests match
51                                                  op[s.index(i)].append(k[3])     # append to the right place in 'op'
52                                                  if j == 'QVL' or j == 'QVI':
53                                                          i.append('%03.2e' % int(k[2]))  # PostgreSQL does not support scientific notation
54                                                  else:
55                                                          i.append(k[2])
56                                                  found = True
57                                                  # print op[s.index(i)], '<br>'
58                                                  break
59                          if not found:
60                                  i.append(None)
61                                  op[s.index(i)].append(None)
62
63          BGCOLOR1 = "one"
64          BGCOLOR2 = "two"
65          BGCOLOR = BGCOLOR1
66
67          old = s[0][0][0:6]      # ignore final character (timepoint) of SE
68          # f = open('/tmp/table.html', 'w')
```

```
69          print 'Content-Type: text/html'
70          print
71          print('<head><title>Shongwe Consolidated Data Table</title><style type="text/css">\n')
72          print('p{font-family:monospace}\ntd{color:black; text-align:center; font-family:monospace; font-size:125%}\ntd.one{background:lightblue}\ntd.two{background:
                lightgreen}\ntd.header{font-weight:bold}\ntd.missing1{background:cornflowerblue}\ntd.missing2{background:forestgreen}\n</style>')
73          print ('<script type="text/javascript">function show_detail(detail) { alert(detail); } </script>')
74          print('</head>')
75
76          print('<html>')
77          print('<p>Data snapshot as at %s.</p>' % (datetime.datetime.now().strftime("%H:%M on %A %d %B %Y")))
78          print('<table border="1">\n')
79
80          HEADERMAPPING = {'N01':'P7P8', 'N02':'BCP', 'N03':' S ', 'N05':' C ', 'QVL':'VL Cp', 'QVI':'VL IU', 'C02':'B Cl'}
81
82          def showHeader():
83                  print('<tr><td class="header">SE</td><td class="header">&#x394;M</td><td class="header">Sex</td><td class="header" >Age</td><td class="header">BMI</td><td
                        class="header">Wt</td><td class="header">ARV</td><td class="header">SV</td><td class="header">PV</td><td class="header">AB</td><td class="header">EB
                        </td>')
84                  for i in r:
85                          if i in HEADERMAPPING:
86                                  out = HEADERMAPPING[i]
87                          else:
88                                  out = i
89                          print('<td class="header">%s</td>' % out)
90                  print('</tr>\n')
91          # --- showHeader ---
92
93          showHeader()
94          c = 0
95          cellName = ['Sampling Event', 'Delta Months', 'Sex', 'Age', 'BMI', 'Weight (kg)', 'ARV (D001: d4T 3TC NVP; D002: d4T 3TC EFV)', 'Serum Volume', 'Plasma Volume', '
                Aliquot Box', 'Extracted Box', 'HBsAg', 'Anti-HBs', 'Anti-HBc', 'HBeAg', 'Anti-HBe', 'P7/P8', 'BCP Region', 'Core Region', 'Surface Region', 'Apoptosense', '
                ALT', 'CD4', 'Viral Load (copies/ml)', 'Viral Load (IU/ml)', 'BCP Clone', 'APRI']
96          for i in s:
97                  print('<tr>')
98                  if i[0][0:6] != old:
99                          c += 1
100                         old = i[0][0:6]
101                         if BGCOLOR == BGCOLOR1:
```

```
102                                 BGCOLOR = BGCOLOR2
103                         else:
104                                 BGCOLOR = BGCOLOR1
105                         if c % 10 == 0:         # repeat header every 10 lines around participant change
106                                 showHeader()
107                 cell = 0
108                 for j in i:
109                         if j == None:
110                                 if BGCOLOR == BGCOLOR1:
111                                         missing="missing1"
112                                 else:
113                                         missing="missing2"
114                                 print('<td class="%s" title="%s [no data]">%s</td>' % (missing, cellName[cell], ' '))
115                         else:
116                                 tt = op[s.index(i)][cell]
117                                 if Person.has_key(tt):
118                                         personName = Person[tt].replace(',','').split(' ')[0]
119                                 else:
120                                         personName = Person[None]
121
122                                 print('<td class="%s" title="%s [%s]">%s</td>' % (BGCOLOR, cellName[cell], personName, j))
123                         cell += 1
124                 print('</td>\n')
125         print('</table>')
126         print('<p>Ends.</p></body></html')
127
128         databaseconnection.close()
129
130         import csv
131         f = open('/tmp/CDT.csv', 'w')
132         w = csv.writer(f)
133         header = ['SE', 'Months', 'Sex', 'Age', 'BMI', 'Wt', 'ARV', 'Serum Volume', 'Plasma Volume', 'Aliquot Box', 'Extracted Box'] + r
134         w.writerow(header)
135         w.writerows(s)
136         f.close()
137
138 # viewDataTable
```

Colophon

This document was prepared using the TEX Live distribution of the LATEX 2$_\varepsilon$ version of the LATEX document preparation system on Debian and Debian-based GNU/Linux distributions. Body text is "Bitstream Charter"[1] and sectional headings are sans serif. Output was compiled directly to PDF via *pdflatex*, using the *latexmk* script. Donald Knuth and Leslie Lamport are acknowledged for producing TEX and LATEX, respectively. Andy Buckley is acknowledged for the hepthesis document class, parts of which were modified and incorporated into this document.

Text was edited in AUCTeX mode in Emacs. The natbib package was used to manage citations. The reference list was maintained by hand, with many BibTeX-formatted citations obtained from TeXMed (*http://www.bioinformatics.org/texmed/*). Some graphics were prepared with Ti*k*Z. Assistance was obtained from *http://tex.stackexchange.com*.

The following packages, listed alphabetically, were used: *babel*, *bibentry*, *calc*, *caption*, *colortbl*, *comment*, *datetime*, *enumitem*, *fancyhdr*, *fixfoot*, *fontenc*, *footmisc*, *geometry*, *graphicx*, *listings*, *longtable*, *lscape*, *mathdesign*, *mdframed*, *multicol*, *multirow*, *natbib*, *pdflscape*, *pdfpages*, *rotating*, *sectsty*, *setspace*, *texshade*, *textcomp*, *tikz*, *url*, *verbatim* and *xcolor*.

[1] (c) Copyright 1989-1992, Bitstream Inc., Cambridge, MA. You are hereby granted permission under all Bitstream propriety rights to use, copy, modify, sublicense, sell, and redistribute the 4 Bitstream Charter (r) Type 1 outline fonts for any purpose and without restriction; provided, that this notice is left intact on all copies of such fonts and that Bitstream's trademark is acknowledged as shown below on all unmodified copies of the 4 Charter Type 1 fonts. BITSTREAM CHARTER is a registered trademark of Bitstream Inc.