

# **The Molecular Evolution and Origins of Hepatitis B Virus in Humans and Non-human Primates**

**Sofie Elisabeth Starkman**

**Thesis submitted for the degree of PhD**

**The University of Edinburgh, 2005**



## **Declaration**

I, Sofie Starkman, hereby declare that this thesis submitted for examination for the degree of PhD has been composed by me, that the work described within it was carried out by me, and that the work described within it has not been submitted for any other degree than that specified above.

4<sup>th</sup> of July 2005, Malmö, Sweden

## **Acknowledgements**

First I would like to thank Peter Simmonds for giving me the opportunity to carry out this research. My only regret is that I didn't get to go to Africa and Asia to collect the samples for the study myself, but I still had fun reading about all the different species of monkeys and apes.

Thanks to John Lewis and Andrew Greenwood for supplying the samples for the primate work.

I would also like to thank Ben Searle and Fiona Davidson for teaching me all the skills I needed to carry out the work.

Thanks also to Deborah MacDonald, Ting-Huei Wang, Matthew Marsden, and Lisa Jarvis and the other BTS staff, for further advice and guidance in the lab.

I also have to thank Gary, for being there with me throughout, especially for putting up with me as the stress of writing up drove me mad.

## Abstract

Infection with hepatitis B virus (HBV) has been detected in most populations throughout the world, as well as in a number of non-human primate species. In humans HBV infection represents a major global health problem, with an estimated 1 million deaths per year due to hepatocellular carcinoma and chronic hepatitis. HBV variants infecting humans can be classified into at least 7 different genotypes differing from each other by 11-13% in nucleotide sequences. A range of distinct genotypes also infect African apes, Asian apes and possibly New World monkeys. Studies of HBV epidemiology and the geographical species associations of different HBV genotypes have led to a number of hypotheses for the origin of HBV in humans and primates. These are the "Out of Africa" hypothesis, the hypothesis that HBV originated in South America and spread in Africa and Western countries in the last 200-300 years, and more recently, proposed origins from cross-species transmission and/or co-evolution of HBV in their current host. The main aim of this thesis is to investigate the molecular evolution of human and non-human primate HBV to gain further insights into the origin of HBV in these species. This investigation was carried out in three main sections.

The first comprised an extensive and detailed genetic analysis of the distribution of human HBV genotypes in HBV endemic areas in sub-Saharan Africa and South East Asia. In the second section complete genome sequences of HBV variants were analysed for recombination between different HBV genotypes. This analysis included the use of a novel method based on the calculation of association scores for phylogenetic groups, an approach that helps resolve many of the uncertainties and

difficulties of interpretation of results arising from conventional methods, such as SimPlot.

The third section investigated the frequencies of HBV infection in non-human primates, and the relationship between HBV genotype, primate species and geographical range. In my survey, HBV infection was confined to African and Asian apes, and uniformly absent from a wide range of African monkey species. Phylogenetic analysis of chimpanzee-, gibbon- and orangutan-derived HBV variants indicated that a geographical rather than a species correlation with genotypes, implying the co-circulation and cross-species transmission of HBV between species of overlapping habitats. However, in no cases were primate-associated HBV variants found in humans, nor human genotypes in non-human primates. These findings and the interspersed nature of human and non-human primate HBV genotypes deepens the mystery of HBV origins and evolution in humans. The findings, however, provide a context for ongoing studies of HBV biological variability and genotype-associated differences in pathogenicity and outcomes of infection.

# The Molecular Evolution and Origins of Hepatitis B Virus in Humans and Nonhuman Primates

## Table of contents

<b><u>1. Introduction</u></b>	<b>10</b>
<u>1.1 History and discovery of hepatitis B virus</u>	11
<u>1.2 The HBV genome</u>	13
<u>1.2.1 Genetic structure and organisation</u>	
<u>1.2.2 Replication of the viral genome</u>	
<u>1.2.3 Classification of HBV</u>	
<u>1.3 Disease association and epidemiology</u>	19
<u>1.4 HBV in other animals- primates, rodents and birds</u>	26
<u>1.4.1 HBV in rodents</u>	
<u>1.4.2 HBV in birds</u>	
<u>1.5 Evolution of viruses and theories of HBV origins</u>	37
<u>1.5.1 Out of Africa hypothesis</u>	
<u>1.5.2 South American origin of HBV</u>	
<u>1.5.3 Cross-species transmission and co-evolution</u>	
<u>1.5.4 Alternative hypotheses</u>	
<u>1.6 Aims and implications for the future</u>	48
<b><u>2. Materials and Methods</u></b>	<b>51</b>
<u>2.1 Extraction of DNA from serum and plasma samples</u>	52
<u>2.2 Amplification of HBV DNA sequences by polymerase chain reaction</u>	53
<u>2.3 PCR product analysis</u>	55
<u>2.4 Sequencing</u>	56
<u>2.4.1 Primers used for HBV amplification and sequencing</u>	
<u>2.4.2 Sequencing of HBV using radiolabelled dNTP's</u>	
<u>2.4.3 Sequencing of HBV using an automated protocol</u>	
<u>2.5 Alignment and analysis of sequence data</u>	62

<u>2.6 Phylogenetic analysis</u>	64
<u>2.7 Assessment of HBV grouping using association indices</u>	66
<u>2.8 Detection of HBV recombination</u>	68
<b><u>3. Genotypic Variation Within the Human Hepatitis B Virus Genotypes</u></b>	<b>70</b>
<u>3.1 Introduction</u>	71
<u>3.2 Materials and Methods</u>	77
<u>3.2.1 Human samples for HBV study</u>	
<u>3.2.2 PreS PCR to investigate primate-like sequences present in humans</u>	
<u>3.2.3 S gene PCR for genotyping</u>	
<u>3.2.4 Genotype-specific phylogenetic analysis of S gene sequences</u>	
<u>3.2.5 Association Index analysis</u>	
<u>3.3 Results</u>	80
<u>3.3.1 Frequency of HBV viraemia</u>	
<u>3.3.2 PreS PCR to detect possible primate-derived virus variants</u>	
<u>3.3.3 Genotype distribution</u>	
<u>3.3.4 Genotype-specific S gene phylogenetic analysis and Association Index analysis</u>	
<u>3.4 Discussion</u>	106
<b><u>4. Recombination between HBV variants of different genotypes</u></b>	<b>115</b>
<u>4.1 Introduction</u>	116
<u>4.2 Materials and Methods</u>	120
<u>4.2.1 Sequences</u>	
<u>4.2.2 Association Index bootscanning</u>	
<u>4.2.3 Bootscanning using SimPlot</u>	
<u>4.2.4 Phylogenetic analysis of possible recombinants</u>	
<u>4.3 Results</u>	125
<u>4.3.1 Phylogenetic segregation of genotypes</u>	
<u>4.3.2 Bootscanning and phylogenetics results</u>	

<u>4.4 Discussion</u>	149
<b><u>5. Survey of HBV prevalence in non-human primates</u></b>	<b>153</b>
<u>5.1 Introduction</u>	154
<u>5.1.1 History of HBV infection in primates</u>	
<u>5.1.2 History of HBV infection in chimpanzees</u>	
<u>5.1.3 History of HBV infection in gibbons</u>	
<u>5.1.4 History of HBV infection in orang-utans</u>	
<u>5.1.5 HBV in other primates</u>	
<u>5.1.6 Cross-species transmission of HBV</u>	
<u>5.2 Materials and Methods</u>	161
<u>5.2.1 Primate serum and plasma samples</u>	
<u>5.2.2 Screening, sequencing and phylogenetics</u>	
<u>5.2.3 Highly conserved primers for screening Old World monkeys</u>	
<u>5.2.4 Mitochondrial 12S region PCR and sequencing of primate samples</u>	
<u>5.2.5 Association Index analysis</u>	
<u>5.2.6 Recombination analysis</u>	
<u>5.3 Results</u>	168
<u>5.3.1 Frequency of viraemia in primates</u>	
<u>5.3.1.1 Frequency of viraemia seen in study samples</u>	
<u>5.3.1.2 Overall frequency of viraemia seen in primates</u>	
<u>5.3.2 Phylogenetic analysis of primate HBV isolates</u>	
<u>5.3.3 Mitochondrial PCR and sequencing of primate samples</u>	
<u>5.3.4 Geographical distribution of primates in Africa and Asia</u>	
<u>5.3.5 Association Index analysis of primate HBV genomes</u>	
<u>5.3.6 Recombination analysis</u>	
<u>5.4 Discussion</u>	199
<b><u>6. Discussion and Conclusions</u></b>	<b>207</b>
<u>6.1 Frequency of viraemia and epidemiology</u>	208
<u>6.2 Geographical Distribution of Genotypes</u>	210
<u>6.3 Recombination within HBV genomes</u>	212
<u>6.4 Origins and Evolution of HBV within Primates, Human and Non-human</u>	214

<u>6.5 Conclusions</u>	217
<b><u>7. Bibliography</u></b>	<b>218</b>
<b><u>Appendix A</u></b>	<b>236</b>
<b><u>Appendix B</u></b>	<b>239</b>
<b><u>Appendix C</u></b>	<b>247</b>
<b><u>Appendix D</u></b>	<b>255</b>
<b><u>Appendix E</u></b>	<b>285</b>
<b><u>Appendix F</u></b>	<b>298</b>

# **Chapter 1**

## **Introduction**

## 1.1 History and discovery of hepatitis B virus

Hepatitis as a disease has been known for a very long time, possibly even since the time of the Babylonians (Hollinger, 1996). Hepatitis transmitted by blood, however, was not described or established until the 1930's. This thesis will investigate one of the viruses that have since been identified as a causative agent of hepatitis in humans, hepatitis B virus (HBV).

HBV represents a major global health problem, with an estimated 1 million deaths from hepatocellular carcinoma and chronic liver disease each year (Thomas and Jacyna, 1993). HBV was the first virus from a family of viruses to be discovered. These are the hepadnaviruses, and the other members of this family will be discussed in more details in later sections. In the 1960's B.S. Blumberg was involved in some research into the antibody responses mounted by people who had received blood transfusions, and through this he stumbled upon an antigen which was given the name Australia antigen, as it had been isolated from sera from Aborigines (Blumberg, 2003). It was not known at the time, but what had in fact been found was the surface antigen of HBV. Studies into the distribution of this antigen identified patients suffering from hepatitis as the major carrier group. This in turn led to the idea that part of a hepatitis virus had been discovered, and the term hepatitis B virus was coined, before the existence of such a virus was known. The name assignment was alphabetical; hepatitis A was used to describe the infectious form of hepatitis and hepatitis B described the transfusion-related form of hepatitis. Any virus associated with this type of hepatitis would therefore be the hepatitis B virus. We

now know that transfusion-related hepatitis, or type B hepatitis, is in fact caused by one of two viruses, HBV or hepatitis C virus (HCV).

A major breakthrough came with the discovery of the Dane particle in 1970 (Dane et al., 1970), as the agent responsible for causing type B hepatitis revealed itself. The Dane particle turned out to be the virus particle that they had been looking for, and the Australia antigen was renamed hepatitis B surface antigen (HBsAg). Once the virus responsible for causing hepatitis had been identified it became possible to try to prevent the spread of hepatitis through transfusions. The introduction of screening tests for HBsAg in the late 1960's led to a dramatic drop in the incidence of transfusion-associated hepatitis, but it was not until the 1980's and the discovery of hepatitis C virus (HCV) that this type of hepatitis virtually disappeared (Blumberg, 2003). The rapid development of molecular biology techniques in the last 10-20 years has led to a huge increase in what we now know about HBV, the infection it causes, and its epidemiology throughout the world. HBV-like viruses have been identified in species other than humans, and multiple genotypes exist within the human HBV group. These aspects of HBV infection will be covered in following sections and chapters.

## 1.2 The HBV genome

In this section the genetic structure, organisation, and replication of the HBV genome is going to be discussed.

### 1.2.1 Genetic structure and organisation

The genome of HBV is around 3200 nucleotides in size, consists of circular partially double-stranded DNA (Tillolais et al., 1985), and is found inside a virion which is 42 nm in diameter and has an envelope. The inner core contains the core antigen and the envelope contains the HBsAg. The HBV genome consists of four open reading frames (ORFs) which overlap (Fig. 1-1.). These have been given the names S (surface), C (core), P (polymerase) and X (X protein) (Tillolais et al., 1985).

The S region is where the viral envelope is encoded, and this is in turn divided into three regions. These are the S gene, PreS1 region and PreS2 region and are initiated by three separate start codons and share a common stop codon. The PreS1 region may vary in length between genotypes, human genotype D and the non-human primate genotypes have a deletion in this region that distinguishes them from the other genotypes. The PreS1 and PreS2 regions are also more variable in sequence compared to the S gene, implying a more important role for this region that requires sequence conservation. This may also be due to less immune pressure being placed on the S region than the rest of the genome. Three envelope proteins are produced by this open reading frame, S (S gene), M (PreS2), and L (PreS1). Antibodies against S are able to prevent HBV infection (Seeger and Mason, 2000), suggesting an important role for this protein in the attachment and entry of HBV into cells. The S

gene has also been found to be useful for the genotyping of HBV (Norder et al., 1993).

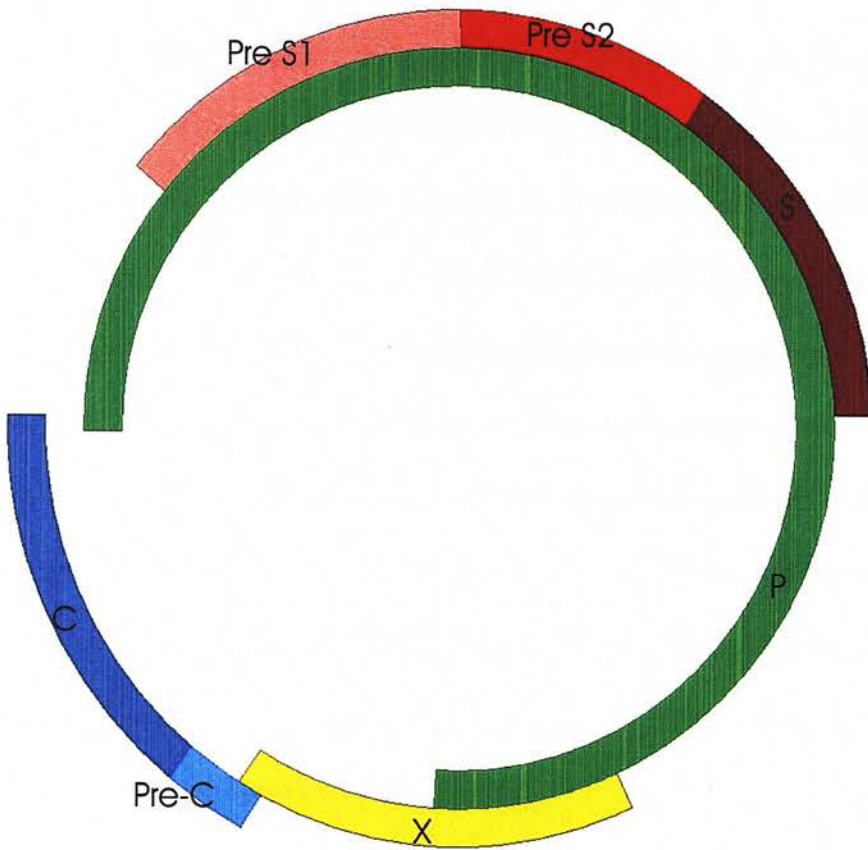


Fig. 1-1. The circular genome of hepatitis B virus consists of four open reading frames (ORFs) which overlap, meaning most of the genome codes for more than ORF. The S ORF (pre S1, pre S2, and S) is completely contained within the P ORF, which also overlaps with ORF X and ORF C (pre-C and C).

The C region encodes the viral core protein which is the major component of the nucleocapsid (Tillolais et al., 1985). The C region contains two start codons. Translation starting at the first of these leads to the production of e antigen and translation starting at the second results in core antigen production. The e antigen is thought to play a role in the attachment of the core to the envelope. Hepatitis e antigen is a secretory core gene product. E antigen attached to cells may serve as a potential target for antibodies involved in eliminating HBV (Schlicht and Schaller, 1989). Mutations in the PreC promoter appear to have a major impact on viral DNA replication and may result in a lower level or stop of e antigen production (Parekh et al., 2003), the significance of which will be discussed in section 1.3. Mutagenesis studies of the core gene have shown that deletions or insertions in the central fold of the protein destabilizes the protein (Koschel et al., 1999), and this indicates that this region of the genome has constraints when it comes to the amount of variation that is possible without impairing replication of the virus. This in turn will place evolutionary constraints on HBV.

The P region encodes the viral polymerase protein, which also has a reverse transcriptase function. It has no proof-reading activity and errors may therefore be introduced when the genome is replicated (Seeger and Mason, 2000). The mutation rate of this gene has been found to be higher than that of the rest of the genome, possibly due to the large extent of overlap with other genes (see Fig.1-1). This may be important in the development of drug-resistant HBV variants (Park et al., 2003).

The X region is the least characterised of the HBV coding regions. The X gene overlaps both the pol gene and the precore region. It appears to be well conserved between genotypes and species (Kidd-Ljunggren et al., 1995). The X protein may be involved in the attachment of HBV to cells (Paran et al., 2003), as well as having transactivating activities. It appears to be essential for viral replication, but the mechanisms involved are unknown. It is not clear what role it plays during HBV infection (Seeger and Mason, 2000). The X protein has been implicated in the development of hepatocellular carcinoma. It has been proposed that X mutants may be selectively replicated in tumour tissues and thereby modifying X protein activity and aiding carcinogenesis (Iavarone et al., 2003; Tu et al., 2001). Other studies have shown that X protein induces the expression of interleukin-18, a pro-inflammatory cytokine, in the liver and this may be associated with liver injury during the course of HBV infection (Lee et al., 2002).

### 1.2.2 Replication of the viral genome

When HBV has entered a new host it needs to replicate to establish an infection and produce progeny viruses to enable its eventual spread to new susceptible hosts. Attachment to the cell surface of hepatocytes is followed by entry into the cell and uncoating of the viral genome prior to replication (Cooper et al., 2003).

The first step in HBV replication is the conversion of the relaxed circular viral genome into a covalently closed circular genome (cccDNA). The mechanisms involved in this are unknown. cccDNA then accumulates in the in the nucleus and is used as the transcription template for the viral messenger RNAs (mRNAs). Four viral

mRNAs are produced. The largest of these is the pre-genomic RNA which will be reverse transcribed to produce new viral genomes. This is also the C mRNA and P mRNA. Separate mRNAs are then produced for translation into the viral envelope proteins and the X protein. HBV transcription is thought to be dependent on transcription factors found in the hepatocytes. Transcription is followed by assembly. Following translation of the viral polymerase the pre-genomic RNA is reverse transcribed to form a new viral genome which is then packaged into the virion. The reverse transcriptase is involved in the packaging together with host factors (Cooper et al., 2003; Seeger and Mason, 2000). New virus particles can now be transported to the cell surface and released, and may then go on to infect new cells, and possibly new hosts.

### 1.2.3 Classification of HBV

HBV genomes have been classified into 9 serotypes and 8 genotypes. The serotypes are based on the amino acid sequences of the surface antigen. The genotype distribution follows a geographical pattern and will be described in more detail in chapter 3. The serotypes also have geographical distributions, and show different associations with particular genotypes. For example, serotypes ayw1 and adw2 are associated with genotype A, ayw1 with genotype B, ayr, adr and adrq- with genotype C, ayw2 and ayw3 with genotype D, ayw4 with genotype E and adw4p- with genotype F (Norder et al., 1992b). The genotypes were identified using phylogenetic analysis of large numbers of HBV isolates (Kidd-Ljunggren et al., 1994; Norder et al., 1992b; Norder et al., 1992a; Norder et al., 1993; Norder et al., 1994; Okamoto et al., 1988). The genotyping of HBV may be important when diagnosing patients, and

deciding which course of treatment to choose, as genotypes are associated with different prognoses (see section 1.3). HBV can be genotyped either by polymerase chain reaction (PCR) followed by sequencing and phylogenetic analysis, but a number of alternative strategies have been developed for use in diagnostic laboratories around the world. Restriction fragment length polymorphism (RFLP) analysis following PCR amplification of viral DNA has proven a useful tool in the relatively rapid determination of genotypes (Lindh et al., 1997; Lindh et al., 1998; Mizokami et al., 1999). This type of method is very useful when it is necessary to determine which genotype an individual is infected with so that the most effective treatment can be chosen. Type-specific PCR amplification is another rapid genotyping method (Kato et al., 2001; Naito et al., 2001).

### 1.3 Disease association and epidemiology

As mentioned earlier HBV is a major public health problem, causing high levels of morbidity and mortality throughout the world. Certain areas of the world are endemic for HBV infection, such as Southeast Asia, sub-Saharan Africa, and South and Central America, and here prevalence rates of active infection of 8-15% can be seen. Even larger proportions of the populations of these countries show evidence of past infection with HBV (Andre, 2000; Ding et al., 2001; Ding et al., 2003a; Liu et al., 2002; Murhekar et al., 2002; Nakai et al., 2001; Quintero et al., 2002; Sirisena et al., 2002; Sugauchi et al., 2002a; Tanaka, 2000). High rates of infection have also been identified in Russia (Flodgren et al., 2000) and among the Inuits in Greenland (Langer et al., 1997). In traditionally low prevalence countries like England and Italy high infection rates can be detected in immigrant populations from high endemicity areas (Chironna et al., 2000; Kawsar and Goh, 2002). This is leading to a change in the HBV epidemiology of the UK, with high prevalence rates seen in certain population groups. This is also leading to the introduction of HBV genotypes not traditionally found in Europe.

The modes of transmission of HBV include parenteral, sexual, and perinatal routes (Hollinger, 1996), but mosquitoes have also been suggested as a possible vector for transmission (Blow et al., 2002). Unsafe injections have been identified as a major factor for new HBV infections in the developing world, and it has been estimated that 8-16 million new infections could occur each year by this transmission route (Kane et al., 1999). Hepatitis B can cause acute infection which is then cleared (transient) or chronic infections (Blumberg et al., 1975a).

The majority of cases are transient when infection occurs in adults, whereas infection in children tends to lead to chronic infection (Hollinger, 1996; Seeger and Mason, 2000). It is this that makes it a difficult infection to eradicate, as there is a huge reservoir of potential infection spreaders. One possible explanation for the difference seen in the type of infection that results depending on the age of the individual is that the immune system needs to be fully developed in order to mount a successful response to HBV and eliminate it. In Asia perinatal infection is predominant, whereas in Africa horizontal spread between young children is the main mechanism for maintaining HBV infection within the population.

Even though most infections are transient they still have a major effect, as these infections usually last for 1-6 months. The first few weeks after infection are asymptomatic, which again has an impact when it comes to spreading infection (Seeger and Mason, 2000). After this initial asymptomatic period a highly viraemic stage follows, and virus is then gradually eliminated until it becomes undetectable after 4 to 6 months. HBsAg can be detected in the blood from a month post infection until 5 months post infection (see Fig.1-2 below). Once acute disease is seen antibodies against the core antigen start to appear, and these remain detectable for years after infection has disappeared. Antibodies against the surface antigen become detectable once in the convalescent phase (Hollinger, 1996).

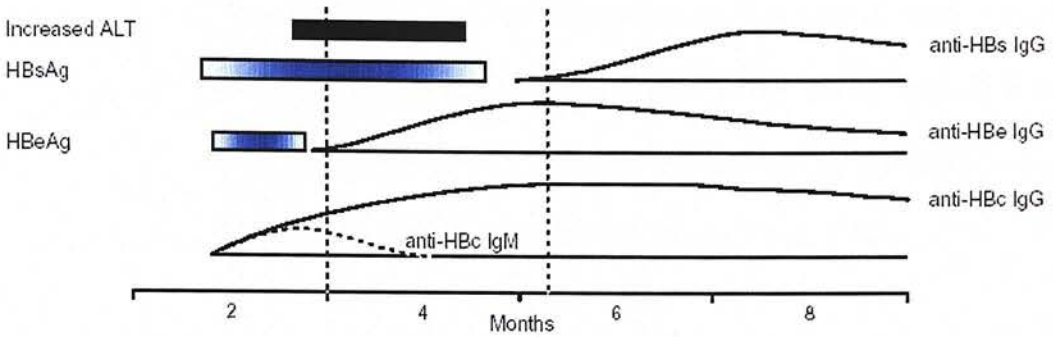


Figure 1-2. Schematic diagram showing the serological marker profile of HBV infection . The diagram was taken from a PHLS communicable diseases report (Teo, 1992).

These serological markers are important in the diagnosis and prognosis of HBV, but they leave a window of around one month after infection where there are no detectable signs of infection. This is where the development of molecular detection techniques has made a dramatic impact. The routine use of PCR screening identifies individuals who are in the very early stages of HBV infection. Hepatitis e antigen positivity can be used as an indicator of how infectious a carrier is. Patients positive for HBeAg have been shown to have high levels of infectious virions and HBV DNA (Hollinger, 1996). As described earlier core promoter mutations may lead to reduced e antigen production, and this would influence this type of test. A patient may appear to be less infectious than he/she really is.

Chronic infection with hepatitis B virus follows the same initial disease course as that described for acute and transient infection above (Hollinger, 1996). Patients who develop chronic infection in adulthood tend to have a poorer prognosis than those who get infected as children, as there may be outbreaks of acute liver disease,

cirrhosis of the liver may occur and hepatocellular carcinoma may develop (Blumberg et al., 1975b; Seeger and Mason, 2000). HBV DNA can be found integrated in the cancerous cells in the liver of an infected patient although this is a rare event and the role in carcinogenesis is unknown (Hino et al., 1984).

The risk of developing liver cancer is 10-15% for chronically infected patients. Certain mutations in the core promoter have been identified that increase this risk (Kao et al., 2003). Mutations within the PreS2 region are also found at a high frequency (50%) in cancer patients (Bläckberg and Kidd-Ljunggren, 2003). Mutations associated with carcinoma have also been identified as coinciding with a T cell epitope of the surface antigen (Tai et al., 1997). Certain genotypes seem to be more closely associated with the development of hepatocellular carcinoma than others, for example genotype C (Ding et al., 2001; Ding et al., 2003b; Nakayoshi et al., 2003; Sugauchi et al., 2002a). This association is made more complex due to the fact that genotype C causes more severe disease in Japan than it does in China (Sugauchi et al., 2002b). It has also been shown that immune-mediated cell injury can trigger tumour development even in the absence of viral factors, and this suggests that a combination of host and viral factors are involved in the carcinogenesis due to HBV infection (Koike et al., 2002; Nakamoto et al., 1998).

There are no treatments that are able to clear HBV infection once the chronic state has been established. The aim for treatment is therefore damage limitation. Alpha interferon and lamivudine are the two most common treatments that are used with HBV today, although tenofovir is also being used (Seeger and Mason, 2000). These

treatments suppress viral replication and reduce the symptoms of liver disease. The alternative to these treatments in severe disease is transplantation of the liver. Recurring infection after transplantation is a problem (Roche et al., 2003) and mutants able to evade the drugs used to suppress symptoms may be selected for in these situations (Brind et al., 1997). Treatment of HBV with different drugs, those mentioned above and new ones under development, can also give rise to the emergence of resistant mutant viruses. For example, a mutation in the pol gene has been detected that is associated with the use of adefovir dipivoxil (Angus et al., 2003), a treatment that has been shown to give good reductions in serum levels of HBV DNA (Westland et al., 2003). Mutations have also been identified in B and T cell epitopes of the X protein, and this may also be involved in evasion of the immune responses (Hwang et al., 2003).

Even though HBV infection cannot be cured, it can be prevented. The development of HBV vaccines has made liver cancer the first cancer that can be prevented by vaccination. Vaccination is the process where a vaccine is injected into the body to induce immunity against the agent in question. This is a form of immunisation. Immunisation is the exposure of an individual to an antigen in order to raise antibodies against a particular agent to improve the ability to resist infection. Plasma-derived vaccines were developed in the 1970's , and were then replaced by recombinant yeast-derived vaccines in the 1980's (Hollinger, 1996). Vaccination has been very successful in reducing the numbers of new infections and the development of chronic infections in countries like The Gambia (Fortuin et al., 1993; Whittle et al., 1991) where HBV infection is endemic. It has also been successful in Italy

(Stroffolini et al., 2000) and in the Pacific Islands (Vryheid et al., 2001). As has been seen before for other infections, the introduction of vaccination has brought with it the emergence of vaccine-escape mutants. These mainly have mutations within the S gene that prevent neutralization by antibodies by altering the a determinant of the surface antigen (Hollinger, 1996). Research into the development of new alternative vaccines that will prevent the emergence of such mutants is ongoing.

As mentioned above, different genotypes of HBV are associated with different disease outcomes and serological profiles. A recent study from China found an association between more severe disease in genotype C than genotype B infected patients. They also found that genotype B patients sero-converted to HBeAg earlier than those infected with C, and suggested this might explain the less active infection seen for this genotype (Chu et al., 2002). These results were seen also in Shanghai and Harbin provinces, China (Ding et al., 2001; Ding et al., 2003a). Genotype C is also associated with increased mortality in Japan, where genotypes B and C are the most commonly isolated genotypes (Orito et al., 2001). In Thailand there also seems to be a bias for severe disease and development of liver cancer in genotype C patients when compared to genotype B infected patients (Sugauchi et al., 2002a). Similar results were reported from Spain, where genotype F was associated with a more frequently fatal outcome compared to genotypes A and D, and suppression of symptoms was more successful in genotype A (Sánchez-Tapias et al., 2002).

In contrast to this, researchers found that there was no difference in the severity of disease caused by genotypes A and D in Uzbekistan (Kato et al., 2002c). In contrast

to the association between genotype C and severe disease described above for China, another study also from China found the opposite, genotype B was associated with higher mortality (Yuen et al., 2003). This clearly shows that host factors as well as viral factors play a major role in determining the outcome of HBV disease. For example, mutations in the p53 gene have been suggested as playing a role in the development of hepatocellular carcinoma (Kirk et al., 2000). Viral factors may be influencing changes in the host, as well as the host immune response driving changes in the viral genome.

#### 1.4 HBV in other animals- primates, rodents, and birds

Apart from the 8 human HBV genotypes, HBV-like viruses have been isolated from a number of other species of primate, rodent and bird. Primates were early on found to be positive for HBV during post-capture screening programs, although these infections were thought to be of human origin, either through vaccinations or other contacts (Deinhardt, 1976). Primates captured in the wild were routinely vaccinated with pooled human sera to protect them from human infections, and as this occurred prior to the identification of the hepatitis B virus this is a possible route of introduction of HBV into primate populations. Primates may also catch human infections through biting and scratching accidents.

Species-specific HBV variants were then reported for chimpanzees (Vaudin et al., 1988; Zuckerman et al., 1978), gibbons (Mimms et al., 1993; Norder et al., 1996), orang-utans (Warren et al., 1999), and the New World primate woolly monkey (Lanford et al., 1998). Since these first reports several more isolates from these species and also a gorilla (Grethe et al., 2000) have been reported (Aiba et al., 2003; Hu et al., 2000; Hu et al., 2001; Lanford et al., 2000; MacDonald et al., 2000; Noppornpanth et al., 2003; Starkman et al., 2003; Takahashi et al., 2000; Takahashi et al., 2001; Vartanian et al., 2002; Verschoor et al., 2001). Primate HBV variants and their history are described in more detail in Chapter 5.

##### 1.4.1 HBV in rodents

Rodent HBV-like viruses have been isolated from woodchucks (*Marmota monax*) (Summer et al., 1978), Beechey ground squirrels (*Spermophilus beecheyi*) (Marion et

al., 1980), tree squirrels (*Sciurus carolinensis pennsylvanicus*) (Feitelson et al., 1986), and Arctic ground squirrels (*Spermophilus parryi kennicotti*) (Testut et al., 1996). Woodchuck hepatitis virus (WHV) may be a cause of hepatic disease in the host species, as it was identified in animals that had died of hepatocellular carcinoma or acute hepatitis. Two animals with WHV surface antigen positive sera that died were found to have malignant liver tumours. The level of inflammation seen in the livers of these animals was high, whereas a surface antigen positive animal that died of heart failure was shown to have minimal inflammation. This difference in inflammation is seen also in humans with HBV infection (Summers et al., 1978). These woodchucks were all part of a research colony, but wild-caught woodchucks have also been found to have a high rate of carriage of WHV (Tyler et al., 1981). Due to the similarities between human HBV and WHV, the woodchuck is a useful model system for research into viral hepatitis and hepatocellular carcinoma (HCC) (Summers et al., 1978).

The woodchuck has been extensively used as a research model for HBV infection. One example of the use of the WHV model is in antiviral therapy development. From this type of research it has emerged that responses to toxicity of compounds in woodchucks is similar to human responses, and the woodchuck may therefore be a useful model to use more widely in drug development (Tennant and Gerin, 2001). The WHV model has also been applied to DNA vaccine research. DNA vaccines based on WHV core antigen and surface antigen have been shown to induce an immune response that inhibited infection with WHV in vaccinated animals (Lu et al., 1999). Mutagenesis studies into the biological function of the WHV X protein have

shown that HBx functional domains are biologically important. Mutated viruses are able to replicate as well as wild-type viruses and appear to induce a strong immune response, leading to immunity to subsequent challenge with infectious wild-type WHV. These studies identify the X protein as a potential target for future research into novel treatments and vaccines (Zhang et al., 2001).

In a recent study it was shown in the WHV model that an antibody to the hepatic asialoglycoprotein receptor, frequently found in patients with hepatic damage, could affect the severity of WHV infection (Diao et al., 2003). Antibody present prior to infection appeared to be indicative of a chronic infection outcome, and induction of antibody production during the course of infection increased the severity of the existing disease. Work on the WHV model has revealed the presence of genomes in serum with deletions and rearrangements like those found in integrated genomes in HCC. This suggests these mutation events may take place prior to the integration and these genomes maybe involved in HCC pathology once integrated. The study implicated topoisomerase I in the production of the mutants (Kew et al., 1993). Virus escape mutants normally have alterations in MHC Class I restricted T cell epitopes of the S gene (Tai et al., 1997). This has been shown to occur also in WHV infection (Botta et al., 2000).

Although WHV infection is similar to human HBV infection in certain aspects, there are also differences between the two. Human core deletion mutants have been shown to act like defective interfering particles, meaning they replicate preferentially to wild-type virus although they require the presence of the wild-type virus to replicate

(Sahu et al., 2002). Defective mutants exist also in WHV infection, but here there appears to be no defective interference activity of these mutant viruses (Sahu et al., 2002). Mutagenesis was used to create in WHV the most common precore mutation of HBV to produce a stop codon. In contrast to HBV the precore gene was found not to be essential for viral replication, though it may be important in chronic infection (Chen et al., 1992), as precore mutations are frequently found in chronically infected woodchucks (Botta et al., 2000), a further parallel between WHV and HBV. There may be other such differences between other similar mutant strains, and further research will be needed to evaluate the use of woodchucks as a model for human disease. The genomes of WHV and HBV are related, but aligning their sequences for phylogenetic analysis is difficult. The most conserved area of the genome that allows alignment is found in the S gene region, where it is possible to design primers for PCR conserved enough to allow detection and amplification of all known primate and rodent hepadnavirus variants (Starkman et al., 2003).

Beechey ground squirrel hepatitis virus (GSHV) was the second HBV-like virus to be identified in rodents. The virion is similar to HBV. The animals it was isolated from showed no signs of disease, although it was impossible to follow up whether disease developed at a later stage in infection, as the animals were released after sampling. GSHV appears to be enzootic, with 44-55% of animals in different locations being infected with the virus. The serology results also suggest a persistent infection in these animals, so it is possible that there may be an association with liver pathology similar to WHV even though this was not observed in the study (Marion et al., 1980).

Tree squirrel hepatitis virus (THBV) also appears to be associated with hepatitis in its host through the establishment of a chronic infection. A carrier rate of around 50% was reported, again suggesting it is enzootic within the host species (Feitelson et al., 1986).

Arctic squirrel hepatitis virus (ASHV) is also enzootic in wild populations (carriage rate of 14%), and again appears to be associated with hepatic disease in the host (Testut et al., 1996). There is around 16% nucleotide difference between the rodent hepadnaviruses, and ASHV is most closely related to GSHV. The rodent viruses have been shown to be around 37% different from the human hepatitis B viruses at the nucleotide level. Phylogenetic analysis has also revealed the presence of geographically defined clusters of WHV which had not been reported before (Testut et al., 1996).

#### 1.4.2 HBV in birds

Hepadnaviruses have also been isolated from large number of species of bird, for example ducks (*Anas domesticus*), herons (*Ardea cinerea*), snow geese (*Anser caerulescens*), storks (*Ciconia ciconia*), and cranes (*Anthropoides virgo*, *Balearica regulorum*) (Chang et al., 1999b; Mason et al., 1980; Prassolov et al., 2003; Pult et al., 2001b; Sprengel et al., 1988). Duck hepatitis B virus (DHBV) was the first avian hepadnavirus to be discovered and is also the most widely studied, partly due to availability of animals for study. As for WHV, DHBV was first isolated from captive populations, in this case commercially bred ducks in the USA (Mason et al., 1980).

Subsequent studies have also identified DHBV in wild bird populations in France with a relatively high rate of carriage (3-12%) (Cova et al., 1986). It is not known whether liver disease is associated with DHBV infection, as commercial flocks are usually killed before they reach the age of 2 years (Mason et al., 1980), and more recent studies have failed to find any evidence of significant liver damage in naturally infected ducks (Jilbert and Kotlarski, 2000).

Again like the WHV, DHBV has been used as a model for HBV infection, including investigations into the immune responses to hepadnaviral infection. Inoculation experiments suggest that both viral factors and host factors play a role in determining disease outcome. Viral factors may be the infective dose, the viral replication kinetics and tropism, and host factors being of the immune response. Low virus dose appeared to induce a protective immune response, whereas a high dose induced a non-protective response (Jilbert and Kotlarski, 2000). The duck humoral immune response to DHBV infection is similar to that in humans with anti-HBc being detected in chronic carriers, but dissimilar in the fact that anti-HBs does not appear to be formed (Jilbert and Kotlarski, 2000).

Mutagenesis studies of DHBV envelope protein have identified a structural determinant that plays an important role in the translocation of virus particles during assembly, and that may be common to all hepadnaviruses (Grgacic, 2002). The DHBV genome has mostly been reported as lacking the X open reading frame found in all mammalian hepadnaviruses, but traces of an X-like protein coding sequence have now been reported that corresponds in position to the mammalian X gene (Lin

and Anderson, 2000). An X-like protein has now also been found to be expressed *in vivo* in DHBV infection, encoded by a hidden open reading frame. There are functional similarities between this protein and the mammalian X proteins, suggesting a common ancestry for these genes with the divergence in sequence being mediated by host adaptation (Chang et al., 2001).

The availability of ducks makes them an attractive model for research. DHBV has also been used to estimate the natural mutation rate of hepadnaviruses, something there is little information about. Mutagenesis/reversion studies found the mutation rate of DHBV to be  $1 \times 10^{-5}$  to  $1 \times 10^{-6}$  (Pult et al., 2001a), which is lower than the previous estimate of  $2.1 \times 10^{-5}$ , which was based on the HBV evolution within a chronically infected carrier (Hannoun et al., 2000a).

The DHBV model has proved a useful one for investigating the mechanisms of human HBV entry. The narrow host ranges and also the tissue specificity have been major obstacles. DHBV work has shown that uptake into cells is a process with at least two steps. First- binding to a primary receptor that appears to be common to all avian hepadnaviruses, and then as a receptor-virus complex binding to a secondary species-specific receptor. A PreS domain was identified as the binding site for the primary receptor. Mutational studies revealed that it was not the sequence that determined the receptor interaction directly, but the tertiary structure of the domain and hence a common cell interaction with related receptors for avian viruses. As the authors of this study suggested the conservation of a structural element within the surface antigen, rather than a specific sequence brings the advantage for

hepadnaviruses of allowing a high level of diversity to escape host immune responses (Urban et al., 1998).

The primary receptor has now been identified as gp180, or carboxypeptidase D. The proposed mechanism of cell entry is that gp180 binds the virus at the cell surface and following binding of a secondary or co-receptor internalizes the bound particles (Breiner et al., 1998). An analogous protein to gp180 has now been identified in humans- p80. The PreS 1 region contains the binding site for p80, indicating again that the S gene may be involved in the species specificity of the hepadnaviruses (Ryu et al., 2000). Phylogenetic analysis has now revealed the presence of DHBV subgroups defined by the geographical origin of the species of duck from which the virus was isolated (Pult et al., 2001b).

Heron HBV (HHBV) appears to be enzootic in grey herons, and also very host-specific, as it is not possible to infect ducks with this virus. As for DHBV infection it does not appear that HHBV is associated with hepatocellular carcinoma in its host. It is 21.6% different from DHBV at the nucleotide level over the entire genome, and again it appears to lack the X ORF found in the mammal hepadnaviruses (or orthohepadnaviruses). The highest level of divergence between DHBV and HHBV is found in the PreS region of the viral genome (50.3%) and this is suggested as the likely reason HHBV will not infect ducks and may be an important host species/range determinant (Sprengel et al., 1988).

Snow goose HBV (SGHBV) infection also appears to have a high prevalence, although it is not known whether this occurs in nature, as the geese from which it was isolated were captive. It has a genome identical in length to DHBV, and it does appear to have an X gene. The putative protein encoded by this X gene does not show any significant similarity to those produced by the orthohepadnaviral X genes. Phylogenetic analysis of the avian hepadnaviruses has revealed that they stem from a common ancestor. SGHBV differs from DHBV by 11-13% and HHBV by 27% respectively at the nucleotide level (Chang et al., 1999b).

Storks have a close evolutionary relationship with herons, and mirroring this relationship the white stork HBV (STHBV) is more closely related to HHBV than any of the other avian hepadnaviruses (14.2-14.5% difference in nucleotide sequence compared to 22.2-23.6% difference between STHBV and DHBV). It also appears to be enzootic, at least in captive stork populations. The most variable region of the genome was again found to be the PreS region, and there again appears to be an X gene present encoding a short X protein (Pult et al., 2001b).

Crane HBV (CHBV) is most closely related to a hepadnavirus isolated from a Ross's goose (RGHBV), even though evolutionarily cranes are closely related to herons and storks. CHBV was able to infect duck hepatocytes, which STHBV and HHBV are not able to do. The PreS region, thought to be involved in host species determination, of CHBV differed from those of all other avian hepadnaviruses with the highest level of divergence seen throughout the genome. An X gene which is also divergent from

all other avian hepadnavirus X genes is present (Prassolov et al., 2003). This suggests that the PreS and X regions of hepadnaviruses are species specific.

Due to the severity of HBV disease in humans it is highly desirable to carry out research into disease course and mechanisms, treatments, and prevention using animal models. The presence of HBV-like viruses in several other species of animal makes this possible. Some of these are more attractive as potential models than others. Ducks and squirrels are attractive due to their availability, but there are also limits to their usefulness. For example, there is still a lot that is unknown about the duck immune system (Jilbert and Kotlarski, 2000), and it may be that these differences make the model unsuitable for human research. Primates are attractive due to their close relationship to humans in evolutionary terms, and although some vaccine studies have been carried out using chimpanzees (Prince et al., 1997) and orang-utans (Davis et al., 2000) it has to be remembered that these species are on the endangered list, and using them in primary research is not feasible. The morality of using our closest relatives for experimentation also has to be considered. The WHV model may be more useful for this type of research (Lu et al., 1999).

The hepadnaviruses discovered to date all show species-specificity when it comes to which host they are able to infect. Infectivity studies using HHBV and DHBV have shown that the host range is determined at the virus entry level, either at the binding step or the internalization step, as determined by the pre-S region variability of the large surface protein (Ishikawa and Ganem, 1995). Both similarities and differences between the different hepadnaviruses give us important insights into the evolution of

these viruses. Studies on the X protein of human, rodent, and avian hepatitis B viruses have shown that their transactivating and pro-apoptotic functions are conserved (Schuster et al., 2002). Experimental infection of non-hepatoma cell lines with HBV, DHBV, and WHV indicates that there are differences in transcription factor usage between these mammalian and avian hepadnaviruses during viral replication. This may reflect the adaptation of the viruses to their host organisms (Tang and McLachlan, 2002).

### 1.5 Evolution of viruses and theories of HBV origins

The viral genome evolves just like the genomes of bacteria and eukaryotes through the fixation of naturally occurring mutations, or nucleotide changes. As is the case for HBV, fixation of these mutations may be subject to constraints imposed by fitness selection. In the case of HBV, the existence of four overlapping open reading frames, and 67% of the Pol gene also coding for other proteins severely limits the number of neutral sites, and therefore limits the amount of variation possible (Mizokami et al., 1997). Mechanisms involved in the evolutionary genetic change are numerous and varied. These mechanisms are discussed in more detail in Simmonds 2001(a and b). Pressure from the immune system of the host may be responsible for driving genetic change in B or T cell epitopes, in order to continue infecting the host and be spread to other new hosts, rather than being neutralised and eliminated. Sequence change may also be driven by the widespread use of HBV vaccines that may lead to changes in the a determinant in HBsAg. Antiviral treatment frequently leads to changes in the Pol gene that confer antiviral resistance (Angus et al., 2003). Rapid sequence change under these selection pressures is favoured by large population sizes in HBV-infected individuals, and by the nature of the replication mechanisms of HBV, in which transcription of the genome is carried out by enzymes without proof-reading activity, and consequently high nucleotide misincorporation rates (see section 1.2).

For viruses like HBV where multiple genotypes or subtypes exist, the possibility of one host being infected with more than one of these types at one time introduces another mechanism to drive the evolution of the viral genome- recombination.

Recombination can also occur with parts of the host genome in viruses such as HBV that can integrate into the host genome (see section 1.2). Recombination with genomes from other related or unrelated viruses infecting the same host is also a possibility. The crossing of the species barrier by a virus can drive its evolution by all the above mentioned mechanisms, but especially immune pressure and recombination. The use of molecular biology to investigate the genomic nucleic acid composition of viruses is an important tool in virology today, especially when coupled with statistical analysis and phylogenetics. By obtaining the genetic sequences of viruses one can attempt to discover their origin and evolutionary history.

Three main hypotheses regarding the evolutionary origin of HBV have been proposed. These will be discussed separately below. Alternative hypotheses will also be discussed.

#### 1.5.1 Out of Africa

This hypothesis proposes that HBV originated in Africa and was then spread to the rest of the World as man moved out of Africa 100 000 to 150 000 years ago (Magnius and Norder, 1995; Norder et al., 1994). This hypothesis was mainly based on the observed distribution of serotypes of HBV, which appeared to be determined by geographical origin of the host population. The distribution of the genotypes of HBV, however, do not show the same distribution pattern, and each of the serotypes may be found in more than one genotype, including the primate genotypes. The phylogenetic outlier position of human genotype F and H can also not be explained.

South America was the last main landmass to be populated, yet the viruses here are the most divergent. Apart from these two genotypes there is no association between the HBV genotypes and population groups, even though the genotypes do exhibit differences in geographical distributions (see chapter 3 for more details about genotypes and their distributions).

The fact that HBV from other primates exists and are phylogenetically interspersed with the human genotypes (see chapter 5) also does not agree with this hypothesis, as they would be expected to be much more distantly related to human viruses and to each other, indeed, to include the different primate genotypes in this hypothesis one needs to propose a process of co-evolution element in which the origin of would date back to the time of ape speciation around 18 million years ago (Stewart and Disotell, 1998). In summary, the poor match between human HBV genotypes and the population groups they infect, and the “wrong” position for the non-human primate genotypes, indicates major flaws with this as an explanation for HBV diversity.

Even though to date this hypothesis does not seem to fit the available data on HBV molecular evolution this type of scenario has been proposed for other viruses such as hepatitis G virus/GBV-C (HGV/GBV-C) (Charrel et al., 1999). The distribution of HGV/GBV-C genotypes matches the population groups that emerged following the spread out of Africa. Furthermore, there are also genotypes found in chimpanzees and New World monkeys, and these do group separately from the human genotypes as expected (Simmonds, 2001b; Simmonds, 2001a). The polyoma virus JC virus (JVC) is a virus that follows the out of Africa model also very well.

This virus can be found as several genotypes with distinct geographical distributions as well as distinct ethnic host populations (Chang et al., 1999a; Guo et al., 1996; Hattwell and Sharp, 2000; Sugimoto et al., 1997). JCV has in fact become so adapted to its host populations that there does not appear to be any transmission across ethnic groups (Kato et al., 1997). This was demonstrated by the absence of Japanese derived isolates in American soldiers based in Japan, and vice versa, though it might have been expected to be some transmission between the two ethnic groups when in such close contact. This has enabled researchers to determine that the distribution of genotypes does mirror the spread of modern humans. Another polyoma virus, BK virus (BKV) may also have evolved following this model (Laura Jackson, unpublished data). The fact that this hypothesis can be applied to other virus infections means that it will have to be reconsidered and re-evaluated as more HBV sequence data becomes available.

#### 1.5.2 New World Origin

This was the second hypothesis to be proposed. It stipulates that the hepatitis B virus originated in South America, and subsequently spread to Europe and the rest of the World through contact with returning Spanish colonisers around 400 years ago (Bollyky et al., 1997). The outlying position of genotypes F and H, which are found mainly in aboriginal populations of South America, and the high prevalence of HBV genotype F within native populations in South America has also been put forward in support of this hypothesis (Bollyky and Holmes, 1999). Other evidence proposed to support this hypothesis includes the existence of woolly monkey HBV, and rodent HBV. All the rodent species from which HBV variants have been isolated are

indigenous to the Americas, indicating that cross-species transmission from these to other primates may represent the origin of HBV in the New World (Bollyky and Holmes, 1999).

However, this theory was proposed at a time when most researchers still assumed that any HBV infections seen in primates were of human origin (Bollyky and Holmes, 1999; Lanford et al., 1998; Norder et al., 1996; Zuckerman et al., 1978). In this thesis, and in other published studies, the overwhelming evidence for natural infections with species-specific primate HBV genotypes in the Old World makes this type of scenario unlikely. To include the primate HBV genotypes in this hypothesis one would have to consider multiple cross-species transmission from humans to primate populations in the wild in Africa and Asia over the last 400 years. This is the length of time that populations from Europe and South America have been in contact.

The amount of variation seen within each of the genotypes is also inconsistent with such a recent origin. The rate of mutation (evolution) of HBV has been estimated as  $2.1 \times 10^{-5}$  substitutions/site/year (Hannoun et al., 2000a). This rate was based on the rate of change seen for a HBeAg carrier. The rate of change in such carriers is slower than that seen in individuals who are able to mount an effective immune response to the infection. This places the origin of human HBV at 2300-3100 years ago. This rate of mutation is slow to generate the levels of diversity seen in only 400 years as suggested by this hypothesis. To generate such a large amount of diversity the rate of change would have to be much faster. Therefore, the geographical distribution of the genotypes is poorly explained by this hypothesis. If the genotypes existed before HBV spread to the rest of the World, then why are genotypes F and H the only ones

represented in the New World, and consequently, why are they not found elsewhere(Simmonds, 2001b)?

### 1.5.3 Cross-species transmission and co-evolution

This was the last of the three hypotheses to be proposed (MacDonald et al., 2000). This hypothesises that the human genotypes seen today originate from multiple cross-species transmission events from primate carriers following co-evolution of HBV and these primate hosts over 10-35 million years. The existence of more divergent HBV variants in mammals and birds opens up the possibility that the co-evolution may stretch even further back, as primates and rodents diverged around 110 million years ago (Simmonds, 2001b). Human endemic areas show specific genotypes, for example F and H in South America, E is sub-Saharan western Africa, and B and C in Southeast Asia. These areas are also where the primates species that to date have been found to carry HBV are found, i.e. woolly monkeys in South America, chimpanzees and gorillas in Africa and gibbons and orang-utans in Southeastern Asia. One possibility is that the genotypes in these areas arose through cross-species transmission, and the human genotypes are the result of recent endemic spread (Simmonds, 2001a) as is suggested for human immunodeficiency virus (HIV) (see below).

Zoonosis as a factor in the evolution of HBV has been suggested by other groups (Hu et al., 2000). Hu (2000) argued that the relationship between the phylogenetic and geographical separation of human genotype F and the woolly monkey isolate from all other primate HBV variants indicated a possible zoonosis event. The discovery of

orang-utan HBV and subsequently genomic variants in geographically separated populations (Verschoor et al., 2001; Warren et al., 1999) also suggest co-evolution from a common ancestor for ape and human genotypes A-E with possible multiple cross-species transmissions responsible for creating the different genotypes. An ancient origin for HBV is implied by the existence of interspecies variants of HBV determined by species/geography. Genotype F and the woolly monkey virus may represent even older viruses (Verschoor et al., 2001). This is also suggested by gibbon studies (Grethe et al., 2000) where five genomic groups were described based on the geographical origin of the animal from which virus was isolated.

Arguments against this scenario for HBV evolution have also been proposed (Hu et al., 2000). Gibbon HBV and chimpanzee HBV are in fact more closely related to each other than to the human genotypes in Africa and Asia respectively, and this is considered indicative of an absence of co-evolution and cross-species transmission.

This means that the HBV genotype phylogeny does not reflect the host phylogeny, as would be expected if co-evolution had taken place. Also, any species involved in these events remain unidentified. And finally; to date no genotypes have been found to be shared between humans and primates. One chimpanzee was detected that carried a human genotype E isolate, but this may be due to transmission in captivity (Takahashi et al., 2000). For more detail about this isolate see section 1.5.4.

Despite the currently conflicting data surrounding the cross-species transmission hypothesis for the origin of HBV infection in humans, such a process is now believed to explain the presence of human immunodeficiency virus type 1 (HIV-1) and type 2

(HIV-2) in human populations. HIV-2 origins were reported first, and were described as a minimum of 6 separate transmission events from sooty mangabeys (*Cercocebus atys*) to humans in West Africa (Feng et al., 1992). HIV-1 is believed to have crossed from troglodytes chimpanzees (see chapter 5 for further information about chimpanzee subspecies and geographical distribution) of Central African origin to humans at least 3 times (Gao et al., 1999), and the HIV-1 genotypes seen distributed widely across the World are due to the recent epidemic spread of the infection within its new host populations. The common ancestor for HIV-1 type M and SIVcpz has been proposed to date back to 1675, with zoonotic transmission of SIVcpz introducing HIV-1 to human populations occurring sometime before the 1920's, as the subtypes of HIV-1 had already begun to diverge by then. The major social changes in Africa in the last century then contributed to the epidemic spread of infection (Salemi et al., 2001).

Co-evolution of SIV and its various host primates over thousands of years, with recombination and cross-species transmission events have also been described by researchers (Georges-Courbot et al., 1998; Souquiere et al., 2001). The finding of sub-species specific HBV variants within chimpanzee HBV, as can also be seen in SIV, taken together with the fact that the modes of transmission of HBV and HIV are similar suggest that cross-species transmission of HBV is a distinct possibility (Hu et al., 2001).

Co-evolution between an infectious agent such as a virus and its host is not unidirectional. Changes will occur in the host to counteract changes in the virus.

Consequently, further changes will occur in the virus to counteract these new host changes. The host and virus drive the evolution of the other forwards.

#### 1.5.4 Alternative hypotheses for HBV origins

In stark contrast to the “Out of Africa” hypothesis and cross-species transmission/co-evolution hypothesis discussed above, it has also been proposed that the evolution of HBV is independent of the evolution of the host species (Orito et al., 1989). Calculated divergence times for HBV are much more recent than the host divergence times. For example, DHBV was calculated as diverging 30 000 years ago, whereas birds diverged 300 million years ago. With the more recent dates of divergence this hypothesis does, however, not rule out the possibility of cross-species transmission as described above. Most alternative hypotheses that have been proposed do suggest cross-species transmission events have occurred in one direction or the other.

The discovery of genotype E in a chimpanzee (Takahashi et al., 2000) gave rise to a number of speculations as to the origin of this virus isolate. The first was that it was human in origin, and was transmitted to the animal either in the wild or in captivity. The second was it came from another infected chimpanzee in the wild, assuming that genotype E is shared between humans and chimpanzees. This last suggestion led to the hypothesis that genotype E was once a chimpanzee strain that was transmitted to humans and subsequently evolved into the genotype E seen today. The lack of genotype E sequences available on GenBank makes this hypothesis difficult to investigate, but as to date there is only one confirmed report of a chimpanzee

genotype E carrier, and as there have been no reports of human carrying chimpanzee-like viruses, there is currently no further evidence to support this hypothesis.

Lanford (2000) suggested three alternative explanations for the close relationship between gibbon HBV, chimpanzee HBV and human HBV that became apparent from his phylogenetic studies of several new isolates from gibbons in captivity. The first of these is that there was a common ancestor for these viruses, but they diverged too recently for there to be much distinction between them to show this ancestry. As Lanford himself points out, the outlying position of genotype F within phylogenetic trees makes this unlikely. His second explanation suggests a human origin for the other primate HBV infections. Humans infected gibbons and this new infection was then established in the wild, and started to diverge into a new genotype. The third explanation is that there are no primate genotypes, but that these infections represent current cross-species transmissions from humans to gibbons in endemic areas. If this was the case, where is the evidence for these genotypes in human populations?

A more recent origin for HBV, with frequent cross-species transmission, has also been proposed (Fares and Holmes, 2002). This places the origin for hominid HBV no earlier than 6000 years ago and the origin for woolly monkey HBV at 9000 years ago. Cross-species transmission is supported by the close relationship between gibbon and orang-utan HBV isolates and the fact that their habitats overlap (see chapter 5 for more information). A similar relationship is seen between chimpanzee and gorilla HBV. The interspersed positions of human and primate genotypes in phylogenetic trees also suggests this type of transmission. Humans are put forward as

the vector most important in spreading HBV across the vast geographical distances necessary to seed HBV infection in human populations throughout the world, and to other primate species in these same areas.

Fares (Fares and Holmes, 2002) suggest that Old World and New World viruses must have diverged either more than 15 000 years ago or less than 500 years ago for humans to have been able to spread the viruses. Two alternative explanations are subsequently offered as to the origin of human genotype F and woolly monkey HBV. His first suggestion gives genotype F an Old World origin with entry into the New World recently, and the woolly monkey isolate dates back to when modern humans first migrated to the New World. The second supports the New World origin hypothesis described earlier. This suggests that genotype F has a New World origin and in fact represents the first virus to diverge from a New World monkey ancestral virus. HBV then spread out of South America post-Columbus, and seeded all subsequent outbreaks of HBV, both human and primate. As for the New World origin hypothesis described earlier, the level of diversity within the genotypes, and the absence of genotypes other than F and H in South America, and conversely F and H in the rest of the world, makes this unlikely. There is now a large amount of HBV genotype data available on GenBank, and to date there have been very few reports of these two genotypes outside of South America.

## 1.6 Aims and implications for the future

The main aim of this thesis is to investigate the molecular evolution of human and non-human primate HBV to gain further insights into the origin of HBV in these species. This type of extensive study has not been carried out before, as most studies concentrate on either human or non-human HBV. The working hypothesis that the research design is based on is the cross-species transmission and co-evolution hypothesis described earlier. Each component of this project is centred around its own set of aims.

The survey of HBV genotype distribution and variation in humans is designed around three aims;

- 1- to investigate further the genotype distributions in endemic areas,
- 2- to investigate further the diversity within the genotypes, including investigating the existence of geographically defined clusters within the genotypes,
- 3- to search for evidence of primate-like sequences in humans that would imply cross-species transmission or a common origin.

The analysis of recombination between HBV genotypes is carried out to assess the extent of recombination within HBV, and the likely impact of this on the evolution of HBV.

The survey of HBV genotype prevalence in non-human primates is also based on three aims.

- 1- to assess the overall frequency of infection in different species of primate, to allow for closer comparison with human HBV infection,
- 2- to assess the existence of subtypes within each of the primate genotypes,
- 3- to look for evidence of recombination within the primate HBV genotypes in the search for evidence of cross-species transmission.

It is important to understand the origin and evolutionary mechanisms employed by a virus such as HBV. The data collected from this project will greatly increase the amount of sequence data available to researchers in the field of HBV biology. This has implications for both epidemiology and other research into for example novel treatment strategies and prevention. Knowledge of which regions of the genome that are most likely to change can target investigations towards more stable regions that may be more suitable and susceptible to intervention. To epidemiologists it may provide clues as to why the huge differences in incidence and severity of disease seen in different populations and parts of the world exist. This may in turn be useful when trying to devise new prevention strategies for countries where HBV is still of high endemicity and a major public health problem, such as sub-Saharan Africa and Southeastern Asia.

By understanding the mechanisms involved in the evolution of HBV more effective treatments and vaccines may be designed to target potentially critical structures within the viral genome. Through molecular studies regions of the viral genome can be identified that undergo little or much change, and these can then be investigated as potential new targets for antiviral drugs, or for use in vaccine development where a

strong immune response to the vaccine is critical for its success. This assumes that any such regions will continue to change at the rate currently seen, so mutation rates will need to be monitored over time to ensure any new strategies devised do not drive mutation within the viral genome.

# Chapter 2

## Materials and Methods



## 2.1 Extraction of DNA from serum and plasma samples

Nucleic acid was extracted from serum or plasma samples using a standard phenol/chloroform extraction protocol. 100µl volumes of serum were incubated at 37°C for 2 hours in 400µl lysis buffer (made with 1M Tris-HCl pH8.4, 0.5M EDTA, 5M NaCl, 10% SDS and protein kinase) in 1.5 ml Eppendorf tubes. Following incubation tubes were centrifuged briefly to remove condensation from the lids. 45µl phenol (water saturated, Rathburn) were added to the samples and tubes were placed in a shaker for 20 minutes. Tubes were then centrifuged for 5 minutes at 15000 rpm at room temperature in a refrigerated Hereaus bench top centrifuge to separate the layers. The aqueous layer was transferred to a new tube containing 450µl chloroform/isoamyl alcohol (50:1 mix, BDH) and placed on the shaker for 15 minutes. Tubes were centrifuged for another 5 minutes to separate the layers. Nucleic acid was then precipitated by adding the aqueous layer to 800µl 100% ethanol containing 40µl 3mM sodium acetate and 1µl glycogen and stored at -20°C overnight. To collect the nucleic acid samples were centrifuged for 20 minutes at 15000 rpm at 0°C. Pellets were then washed by adding 600µl 80% ethanol and centrifuging for another 15 minutes at 0°C. All traces of ethanol were then removed by careful pipetting and the pellets dried for 10 minutes at 42°C in a hotblock. Pellets were then resuspended in 30µl nuclease free water (Promega) and heated to 65°C for 10 minutes to ensure that nucleic acid was completely dissolved prior to polymerase chain reaction (PCR) set up. Tubes were centrifuged briefly to remove condensation from the lids. One positive serum control and one negative serum control were extracted alongside the samples to check for success of extraction and possible contamination during the process. The sample used as a positive control was also sequenced to ensure there was no contamination.

## 2.2 Amplification of HBV DNA sequences by polymerase chain reaction

Samples were subjected to two separate nested (two rounds of amplification) polymerase chain reactions (PCR reactions) in the screening process. Nested PCR increases the specificity and sensitivity of PCR in comparison with single round amplification reactions. PCR 1 amplifies a segment of the PreS2/S region, and PCR 2 amplifies a segment covering the overlap region between ORF P and the PreS1 region (see Fig. 2-1 below).

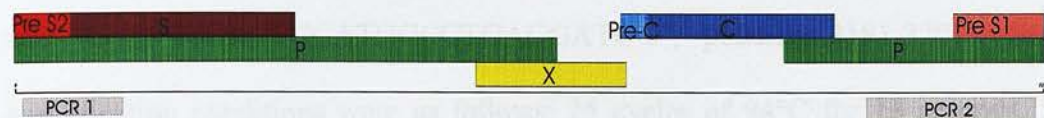


Fig 2-1. Simplified genome map of the hepatitis B virus genome. The 4 ORF's and their relative positions along the 3200 bp long genome are shown. The segments amplified by PCR 1 and PCR 2 are shown in the grey boxes.

First round PCR reactions were set up using 2 $\mu$ l of extracted DNA in a total reaction volume of 50 $\mu$ l. The PCR buffer was made up containing 10 $\mu$ l 10x PCR buffer without magnesium, 10 $\mu$ l MgCl<sub>2</sub>, 10 $\mu$ l 3mM dNTP's (Promega), 1.5 $\mu$ l of each primer (Oswel and Life Technologies) at 100  $\mu$ M concentration and 0.5 U of *Taq* polymerase (Promega) per sample. Prior to cycling the reactions were overlaid with liquid paraffin. PCR 1 primary reaction was carried out using primers S1 (5'-CATCAGGAYTCCTAGGACCCCT-3', position 171-192) and S5 (5'-GAGGCATAGCAGCAGGATGMAGAGG-3', position 406-430). PCR 2 primary reaction was carried out using primers 41 (5'-CGTCGCMGAAGATCCAATCT-3', position 2423-2443) and 19 (5'-CKGAACTGGAGCCACCARCAG-3', position 69-89). First round amplification was carried out under the following conditions; 30 cycles of 94°C for 18 seconds, 55°C for 21 seconds and 68°C for 90 seconds, followed by a final extension at 68°C

for 7 minutes and a holding step at 20°C for 10 minutes. Second round PCR reactions were set up using 1.5µl of first round product in a total reaction volume of 26.5µl. PCR buffer was made up as described above and reactions were overlaid with liquid paraffin before cycling. PCR 1 second round was carried out using primers S3 (5'-CGTGTTACAGGCGGKGTCTTTCTTGT-3', position 196-221) and S6 (5'-ATGATAAAACGCCGCAGACACATC-3', position 379-402). PCR 2 second round was carried out using primers 42 (5'-GTATYCCTTGGACTCATAAGG-3', position 2461-2481) and 43 (5'-CCACTGCATGGCCTGAGGATG-3', position 3181-3201). Second round amplification conditions were as follows; 25 cycles of 94°C for 18 seconds, 50°C for 21 seconds and 72°C for 90 seconds, followed by a final extension at 72°C for 6 minutes and a holding step at 20°C for 10 minutes. PCR amplifications were carried out in a Techne Genius thermal cycler (Techne, Cambridge UK).

### 2.3 PCR product analysis

A 2% agarose gel was made by dissolving 6g agarose (SeaKemLE agarose, FMC Bioproducts) in 300ml 1xTBE (diluted from 10x TBE made with 108g Tris base (BDH), 55g Boric acid (BDH), 40ml 0.5M EDTA pH8 in 1l distilled water) by heating in a microwave at high level for 5 minutes. The agarose solution was then cooled by stirring on a magnetic stirrer for approximately 15 minutes. The gellified top layer containing any bubbles was then removed and Ethidium bromide (final concentration  $6.67 \times 10^4$  mg/ml) was added. The agarose solution was then poured into a gel tray and combs were inserted to create wells in the gel. The gel was left to set for a minimum of 30 minutes at room temperature. 15 $\mu$ l of second round PCR product were then loaded into the wells and run for 20-40 minutes at 150V in a gel tank containing 1xTBE. Size markers of 1kb or 100 bp (Promega) were run alongside the PCR products to check the amplified fragment sizes to verify that PCR reactions had worked. PCR products amplified for the purpose of sequencing were always electrophoresed for 40 minutes to ensure accurate sizing. The gel was then illuminated under UV light and a photograph was taken to record the results.

## 2.4 Sequencing

Sequencing of samples was done using radiolabelled dNTP's in the first instance, and when facilities became available, using an automated fluorescent label protocol.

### 2.4.1 Primers used for HBV amplification and sequencing

For sequencing of primate samples overlapping primers were designed to cover the entire genome of HBV (MacDonald et al., 2000). Primers 44, 45, 46, and 47 were designed specifically for this study (for all primer positions and sequences see Fig. 2-3 and Table 2-1 below). Primers 44 and 45 were used to investigate size differences of PreS fragments (see Fig.3-x), and primers 46 and 47 were used to amplify a fragment of the HBV genome isolated from chimpanzee Osang (see Fig.2-2 below). Primers were synthesized by Oswel and Life Technologies. All fragments for sequencing were amplified using a nested PCR protocol with buffers and PCR conditions as described above for the screening of samples.

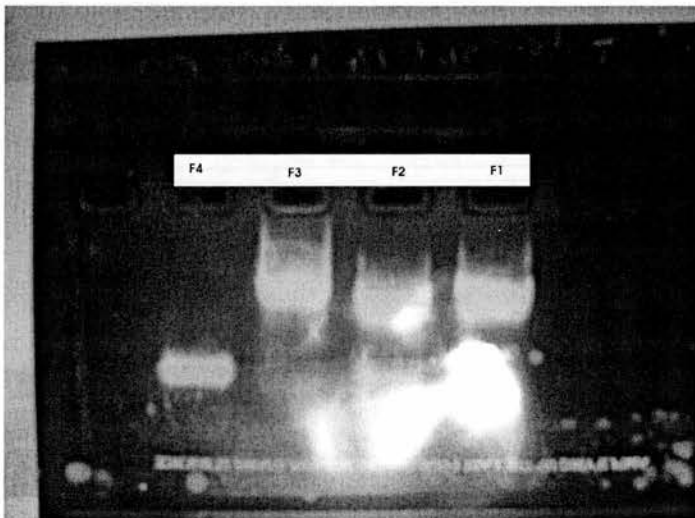


Figure 2-2. Gel image showing the fragment amplified by primers 46 & 47, 400bp in size (F2). Fragment 1 is 450 bp long, fragment 3 is 520 bp long, and fragment 4 is 300 bp long.

PRIMER NO.	POSITION	SEQUENCE	DIRECTION
1	2301-2320	GATAGGGGCATTTGGTGGTC	R
2	1862-1881	ACTGTTCAAGCCTCCAAGCT	F
3	2271-2287	AGTGCGAATCCACACTC	R
*4	1932-1953	GAGCTWCTGTGGAGTTACTCTC	F
*11	592-615	AAWTGCACWTGTATTCCCATCCC	F
*12	800-822	AGACAMAAGAAAATTGGTAAYAG	R
13	2892-2913	CAAYCCNYTGGGATTCTTYCCC	F
14	2953-2975	AATCCMGATTGGGACTTCAAYMC	F
15	245-267	GTCCACCACGAGTCTAGAYTCTK	R
16	265-587	CCCCTRGAAAAYTGAGAGAAGTC	R
17	300-322	GTCYTGCCAAAATTTCGAGTCC	F
18	3181-3201	CATCCTCAGGCCATGMAGTGG	F
19	69-89	CKGAACTGGAGCCACCARCAG	R
21	265-287	GACTTCTCTCARTTTTCYAGGGG	F
*22	379-402	GATGTRTCTGCGGCGTTTTATCAT	F
23	641-664	ACGGRCTRAGGCCCACTCCCATAG	R
24	661-684	AGTAAAYTGAGCCARGAGAAACGG	R
25	752-772	TGGTATTGGGGGCCAARTCTG	F
*26	1171-1191	AGCAAACACTTGGCACAGVCC	R
27	1192-1212	CCARCCAGTGGGDGTTGCRTC	R
28	1102-1122	TCGCCAACTTAYAAGGCCTTT	F
29	1153-1173	CCTTTACCCCGTTGCYCGGCA	F
*30	1583-1603	CACGGWCCGGCAGATGAGAAG	R
31	1583-1603	GCAGAGGTGAAGCGAAGTGCA	R
32	1522-1541	ACGGGGCGCACCTCTCTTTA	F
*33	1542-1561	CGCGGWCTCCCCGTCTGTGC	F
34	2027-2046	CAATGYTCNGGAGACTCTAA	R
35	2055-2075	ACCATACNGCACTCAGGCAAG	F
39	2700-2718	CTGGATAATAAGGTTTAAT	R
40	2726-2746	GTTTGGAARTAATGATTAAC	R
41	2423-2443	CGTCGCMGAAGATCCAATCT	F
*42	2461-2481	GTATYCCTTGGACTCATAAGG	F
*43	3181-3201	CCACTGCATGGCCTGAGGATG	R
*44	2959-2985	GTCCTTGWTGGGRRTGAAGTCCCAATC	R
*45	2824-2848	GGTCACMWTAYTCYTGGGAACAAGA	F
*46	1778-1801	ACCAWTTTATGCYRCAGCCTCCTA	R
47	1415-1442	GGACGTCCTTTGKTACGTCCC	F
S1	171-192	CATCAGGAYTCCTAGGACCCCT	F
*S3	196-221	CGTGTTACAGGCGGKGTKTTTCTTGT	F
*S5	406-430	GAGGCATAGCAGCAGGATGMAGAGG	R
S6	379-402	ATGATAAAACGCCGCAGACACATC	R

Table 2-1. Primers used for PCR amplification and sequencing of hepatitis B virus DNA positive samples. Primers marked with a \* were used in the sequencing reactions. The directions of the primers have been abbreviated to F for forward and R for reverse. The positions of the primers are also shown in Fig. 2-1. Positions are described with reference to HBV isolate accession number V00866 (HBVADW).

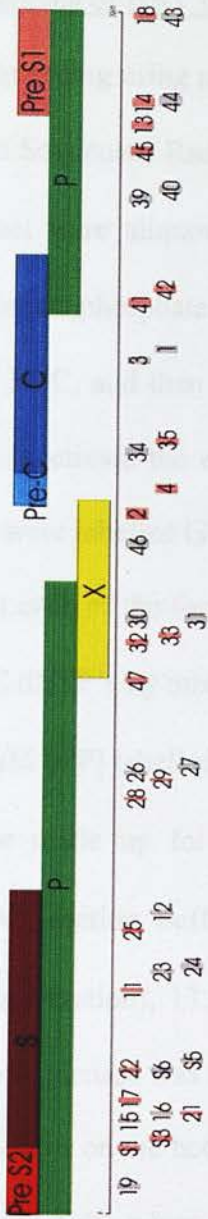


Fig. 2-3. Simplified genome map of the hepatitis B virus genome. The 4 ORF's and their relative positions along the 3200 bp long genome are shown. The positions and direction of the primers used are shown. Forward primers are shown in pink boxes and reverse primers in grey boxes.

#### 2.4.2 Sequencing of HBV using radiolabelled dNTP's

Manual sequencing using radiolabelled dNTP's was carried out using the Sanger method with the Thermo Sequenase Radiolabeled Terminator Sequencing Kit (USB). 5µl of second round PCR product were aliquoted into 0.5ml Eppendorf tubes. 1µl each of exonuclease 1 and shrimp alkaline phosphatase was added and mixed well. This was then incubated for 15 minutes at 37°C, and then quickly transferred to a hotblock at 80°C and left for another 15 minutes to inactivate the enzymes. Four 0.5ml tubes for each sample and primer that was being used were labelled G (guanine), A (adenine), T (thymine), and C (cytosine) respectively to represent each of the four radiolabelled bases. Termination mixes were made up for the G, A, T, and C dNTP's by mixing 2µl T mastermix (7.5µM dGTP, dATP, dTTP, and dCTP) and 0.25µl 0.3µM [<sup>33</sup>P] labelled dNTP (Amersham Pharmacia Biotech) per sample. Then reaction mixes were made up for each of the primers to be used, containing per sample 2µl concentrated reaction buffer (260mM Tris-HCl pH9.5, 65mM MgCl<sub>2</sub>), 0.5µl primer (at 100µM concentration), 13.5µl distilled water and 2µl thermo sequenase DNA polymerase. The thermo sequenase was kept in the freezer and added just before the buffer was to be used. After 15 minutes on the hotblock the tubes were removed and centrifuged to collect the DNA at the bottom of the tubes. 3.5µl of the treated DNA were then added to the G tube of each PCR strip. 18µl of reaction mix were then added to this tube and mixed by careful pipetting. 5µl of this were then aliquoted into the G, A, T, and C tubes of the PCR strip. 2.5µl of the appropriate dNTP termination mix were then added to each tube in the PCR strips and the reactions were then overlaid with liquid paraffin. The reactions were then thermocycled on a Techne Genius for 30 cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 60 seconds. To stop the reactions after cycling has finished 8µl of the sequencing reaction were added to a well in a 96 well microtitre plate containing 5µl stop solution (95% formamide,

20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF). This was either directly loaded onto a polyacrylamide gel or stored at -20°C until needed. If stored at -20°C, reactions were defrosted on a hotblock at 100°C for 1-3 minutes before being loaded onto the gel.

A 6% polyacrylamide gel was made to visualise sequencing products. 21g Urea (BDH) and 0.05g APS were mixed with 6ml Sequagel concentrate and 5ml 10xSanger TBE and made up to a volume of 50ml with distilled water and dissolved under gentle heating in a conical flask. The solution was then left to cool to room temperature and 21µl of TEMED were added. The gel was then poured between clean glass plates, a comb to make the main well inserted, and left to set for approximately 2 hours. The gel was fastened into the sequencing rig and 1xSanger TBE (diluted from 10x TBE with distilled water) added to the top to cover the edge of the inner plate and the bottom to cover the foot of the gel. A comb was inserted to make the individual wells, and then 6µl of sequencing reaction was added to the wells. The gel was then run for approximately 1.5 hours at 75W. The sequencing rig was then drained of buffer and the gel removed and left to cool until hand temperature. The plates were then separated and the gel transferred to Whatman paper. The gel was then covered with Saran wrap and dried at 80°C under a vacuum. When dry the gel was placed in a cassette with X-ray film (Biomax MR, Kodak) and left for a minimum of 24 hours. The film was then developed in an automated developer. Sequence data was manually entered into the SIMMONIC sequence analysis program for editing and alignment.

#### 2.4.3 Sequencing of HBV using an automated protocol

Automated sequencing was carried out using either the dRhodamine or BigDye sequencing kits (Applied Biosystems). For the dRhodamine protocol PCR product was used to give a concentration of between 30-90ng of DNA in each reaction, whereas for the BigDye protocol the size of fragment to be sequenced determined the amount of PCR product to add to each reaction (as described in manufacturer's manual). The required amount of PCR product was aliquoted into PCR strips into which 2µl of the appropriate primer (3.2mM) had already been aliquoted. 4µl dRhodamine/BigDye ready reaction mix was then added and the volume made up to 20µl with milliQ water (prepared on site at the Royal (Dick) Veterinary School, Edinburgh). The reactions were then overlaid with a drop of liquid paraffin and cycled in a Techne Genius thermocycler for 25 cycles; rapid thermal ramp to 96°C, 96°C for 30 seconds, rapid thermal ramp to 50°C, 50°C for 15 seconds, rapid thermal ramp to 60°C, 60°C for 4 minutes, followed by a rapid thermal ramp to 4°C and held there until ready to perform purification reaction. DNA was purified by removing all the reaction with a pipette and adding to 0.5ml Eppendorf tubes containing 1µl 3M NaAc (pH5.2) and 25µl 100% ethanol. Tubes were then left at room temperature for 15 minutes to precipitate the DNA, followed by centrifugation at 13000 rpm at room temperature in a Hereaus bench centrifuge for 30 minutes. The ethanol was then removed using a pipette and the pellet was washed by centrifugation for 5 minutes with 125µl 70% ethanol. The ethanol was then removed by pipetting and any residual ethanol removed by pulse-spinning and a final pipetting. The pellet was dried for 2 minutes at 90°C. (In September 2002 a robotic system for the clean up of sequencing reactions was introduced at King's Buildings, and manual precipitation and cleaning was no longer necessary.) Samples were then sent for analysis on an ABI 3100 genetic analyser (Applied Biosystems) at the automated sequencing service at King's buildings, The University of Edinburgh. Results were returned via email and then viewed

using Chromas. Sequences were directly imported from Chromas into SIMMONIC for editing and alignment.

### 2.5 Alignment and analysis of data

The SIMMONIC software was originally developed to investigate the structural constraints on the evolution of hepatitis G virus/GBV-C (Simmonds & Smith, 1999). This software allows the alignment and analysis of multiple sequences. Data generated using SIMMONIC can be exported in several different formats, such as PHYLIP, MEGA and FASTA, for use in further downstream applications.

Once imported into the SIMMONIC sequence analysis program sequences were aligned to 61 sequences from GenBank (see Table 2-2) representing the human genotypes, and 26 previously published primate isolates representing chimpanzee HBV, gorilla HBV, gibbon HBV, orangutan HBV, and woolly monkey HBV. For primate samples, aligned fragments representing the entire HBV genome were merged using a function of the SIMMONIC program to create complete genomes. Manually obtained sequences were entered into the SIMMONIC program from two separate readings and then compared and ambiguities checked. All Chromas sequence traces were viewed during alignment of sequences and ambiguities were checked. Following alignment sequences were analysed phylogenetically.

ACCESSION NUMBER	GENOTYPE	ACCESSION NUMBER	GENOTYPE
AF297625	A	X75658	F
U87742	A	X75663	F
X51970	A	AB036910	F
AF090638	A	AB036916	F
L13994	A	AB036908	F
AJ344115	A	AF223962	F
AF297624	A	AB036913	F
Z35717	A	AB036905	F
E00010	A	AB056516	G
AF090639	A	AF405706	G
D00331	B	AB056515	G
AY033073	B	AB056514	G
D23679	B	AF160501	G
D50522	B	AB056513	G
D00329	B	AY090454	H
AJ131133	B	AY090457	H
AF282918	B	AY090460	H
AF121249	B	AF305327	CHIMPANZEE
D00330	B	AJ131575	CHIMPANZEE
AF121243	B	AF242585	CHIMPANZEE
AF411508	C	AF242586	CHIMPANZEE
AF241359	C	AB032433	CHIMPANZEE
AF411412	C	AB032432	CHIMPANZEE
AF363961	C	AB032431	CHIMPANZEE
D00630	C	AF222323	CHIMPANZEE
AF330110	C	AF222322	CHIMPANZEE
M12906	C	AB046525	CHIMPANZEE
AB049610	C	D00220	CHIMPANZEE
X75665	C	AY077735	GIBBON
AY057948	C	AY077736	GIBBON
AF297619	D	AB037928	GIBBON
X68292	D	AB037927	GIBBON
U95551	D	U46935	GIBBON
X65257	D	AJ131568	GIBBON
X80925	D	AJ131569	GIBBON
X72702	D	AJ131571	GIBBON
Z35716	D	AJ131572	GIBBON
Y07587	D	AJ131573	GIBBON
AJ344116	D	AJ131574	GIBBON
AF121240	D	AJ131267	GORILLA
X75664	E	NC_002168	ORANGUTAN
X75667	E	AF193864	ORANGUTAN
X69798	F	AF046996	WOOLLY MONKEY
AF223965	F		

Table 2-2. Accession numbers of sequences obtained from GenBank used to align sequence data.

## 2.6 Phylogenetic analysis

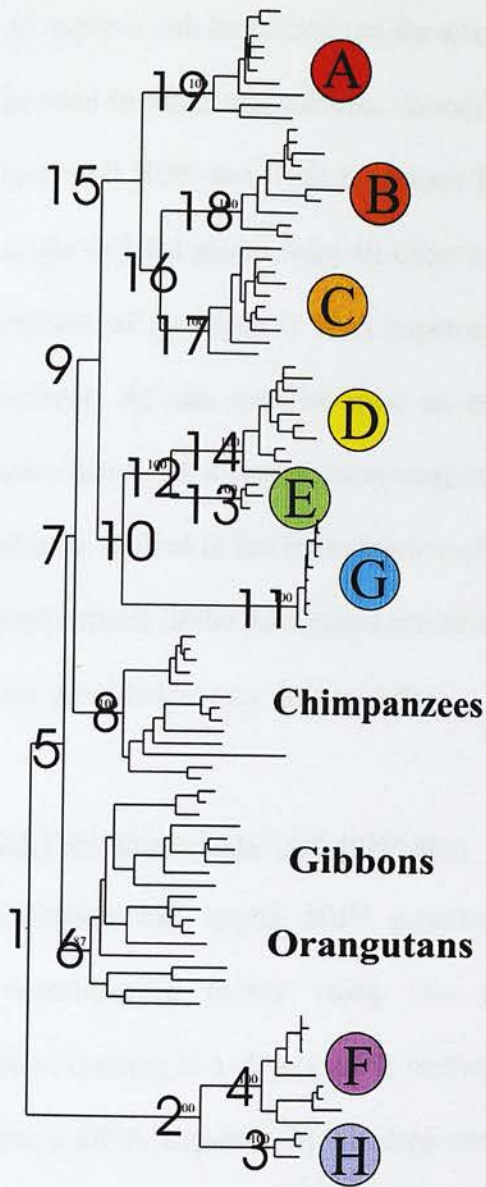
Phylogenetic analysis of aligned sequences was carried out using the MEGA2 program (Kumar et al., 2001). Phylogenetic trees were constructed using the Jukes-Cantor model and neighbour-joining analysis with 100 bootstrap replications. For analysis of primate samples the complete genomes were compared to the previously published sequences described for the amplification and sequencing of the samples. A larger data set of S gene sequences was created by searching GenBank and downloading further sequence data (see Appendix A). This data set was then split into genotype-specific sequence alignments and sample sequence data was re-analysed. Primate samples were also analysed in the S gene region of the HBV genome, as a number of further isolates from chimpanzees, orangutans and gibbons are available with sequences from this region. S gene sequences used in this analysis include those used for aligning and genotyping the data and these further partial sequences (see Table 2-3).

ACCESSION NUMBER	GENOTYPE	ACCESSION NUMBER	GENOTYPE
AF222312	CHIMPANZEE	AF477482	GIBBON
AF222313	CHIMPANZEE	AF477483	GIBBON
AF222314	CHIMPANZEE	AF477484	GIBBON
AF222316	CHIMPANZEE	AF477485	GIBBON
AF222317	CHIMPANZEE	AF477486	GIBBON
AF222318	CHIMPANZEE	AF477487	GIBBON
AF222319	CHIMPANZEE	AF477488	GIBBON
AF222320	CHIMPANZEE	AF477489	GIBBON
AF222321	CHIMPANZEE	AF477490	GIBBON
AF305326	CHIMPANZEE	AF477491	GIBBON
AF305328	CHIMPANZEE	AF477492	GIBBON
AF305329	CHIMPANZEE	AF477493	GIBBON
AF305330	CHIMPANZEE	AF477494	GIBBON
AF213010	GIBBON	Y17559	ORANGUTAN
AF213009	GIBBON	Y17561	ORANGUTAN
AF213008	GIBBON	Y17562	ORANGUTAN
AF274495	GIBBON	Y17564	ORANGUTAN
AF274496	GIBBON	Y17565	ORANGUTAN
AF274499	GIBBON		

Table 2-3. Additional accession numbers of further partial primate sequences obtained from GenBank. These sequences were used in the phylogenetic analysis of the S gene region of primate HBV isolates.

## 2.7 Assessment of HBV groupings using association indices

Intragenotype groupings visible after phylogenetic analysis (see chapter 3) were investigated using Association Index (AI) analysis. This is a novel phylogenetic analytical method, which was initially developed to investigate the segregation of populations of HIV-1 isolates from different body tissues from the same patient (Wang et al., 2001), and it couples the SIMMONIC package to the PHYLIP program package (Felsenstein, 1993). The degree of genetic segregation between variants from different locations is scored based on phylogenetic analysis. An association value ( $d$ ) is calculated using the formula  $d = (n - f)/2^{n-1}$ , where  $f$  is the frequency of the most common sample type within each cluster within the tree, and  $n$  is the total number of sequences within each cluster. Values for  $d$  are calculated for each node within the tree and added together to give a total  $d$  for the tree. Random re-sampling of sequences into groups is carried out 10 times to provide a control value based on the null hypothesis of no segregation. 100 bootstrap replications are carried out as part of the control analysis, thus yielding 1000 control values. The control  $d$  of the tree is the mean of these 1000 values. The Association Index value is obtained by the ratio of  $d$  total for the tree divided by the mean  $d$  for the control analysis. The maximum score possible is 1, and this would indicate no segregation of variants, whereas a score of 0 would indicate complete segregation. A score of 0-0.5 indicates a significant degree of segregation, 0.5-0.7 indicates a lower degree of segregation, and 0.7-1 indicates no significant segregation of groups. The segregation of HBV into genotypes will be used as an example to demonstrate this method (see Fig. 2-4).



Node <i>d</i>	Value
1	$98-19/2^{97} = 4.99 \times 10^{-28}$
2	$7.32 \times 10^{-4}$
3	0
4	0
5	$3.41 \times 10^{-24}$
6	0
7	$1.52 \times 10^{-18}$
8	0
9	$1.28 \times 10^{-15}$
10	$3.09 \times 10^{-6}$
11	0
12	$7.32 \times 10^{-4}$
13	0
14	0
15	$3.73 \times 10^{-8}$
16	$1.91 \times 10^{-5}$
17	0
18	0
19	0

Total *d* for tree =  $1.49 \times 10^{-3}$

Mean control *d* = 22.24

Association Index for tree =  $6.68 \times 10^{-5}$

Fig. 2-4. Example AI calculation for the complete genome phylogenetic tree of HBV genotypes. The human genotypes and the primate genotypes are marked. There are 19 main nodes within this tree, and these are marked with the numbers 1-19. The calculation to obtain the *d* value is shown for the root node (node 1). For all remaining nodes only the *d* values are shown. The control value was calculated using the SIMMONIC program.

AI analysis can be carried out for whole tree analysis as shown above in Fig. 2-4, but can also be used in other applications. Associations between groups can be investigated, for example how well HBV genotype D isolates from Europe and Africa segregate. The segregation of a single defined group from all other sequences within a tree can be investigated, for example isolates of genotype C from hepatocellular carcinoma patients versus all other genotype C isolates. AI can also be used as an alternative genotyping method by investigating the associations of a single query sequence to the different genotypes. This last application can also be applied in the investigation of recombination (see section 2.8.1 below). All these four applications of the AI method can be carried out on the complete sequences or as bootscans as described below (see section 2.8).

### 2.8 Detection of HBV recombination

Complete full length HBV genomes were investigated for evidence of inter-genotype recombination events using two different bootscanning methods, SimPlot and AI. Bootscanning is a phylogenetic method which investigates the phylogenetic relationships of a query DNA sequence by dividing the sequence into segments that overlap, these are called windows. The level of overlap between the windows is also defined. Phylogenetic analysis is then carried out for each window along the genome, scanning the genome. Bootstrap analysis is carried out for each tree that is generated. In traditional bootscanning analysis, such as SimPlot bootscanning analysis which will be described below (see section 2.8.2), the result is shown as a graph of the bootstrap values generated for each window analysed. Bootscanning will show recombination events as a switch in which sequence/sequence group the query sequence is most closely related to. Association Index analysis can be used as a novel bootscanning method.

### 2.8.1 Association Index bootscanning

Association Index bootscanning follows the same principles as described in general for bootscanning above. The difference between this method and other methods is that the result is not a set of bootstrap values, but a set of AI values for the query DNA sequence against predefined sequence groups. The query DNA sequence is divided in windows with a defined overlap and AI analysis with bootstrap replications is carried out for each window. The results are again plotted on a graph, and recombination events will be shown as a switch in the sequence group that the query sequence is closest associated to. In this analysis the scale is 0-1 as described earlier, but now 1 can be taken as meaning strongest association and 0 as no association.

### 2.8.2 SimPlot

The SimPlot program was initially developed to investigate HIV-1 intersubtype recombination. This program is also coupled to the PHYLIP program package. SimPlot can be used in two different ways; to generate similarity plots (which initially was the only function and gave the method its name), or to do bootscanning. Similarity plots investigate the degree of similarity of a query sequence with predefined groups of sequences such as genotypes. Recombination is shown by the query sequence switching which sequence group it is most similar to. This method was used for recombination analysis prior to the advent of bootscanning methods (Salminen et al., 1995). Bootscanning was added to the SimPlot program in 1999. As the results are bootstrap values, a graph is generated with a scale of 0-100%. For each window of the query sequence that is investigated a bootstrap percentage value describes in how many of the generated trees the query sequence falls into a particular group or genotype.

## **Chapter 3**

# **Genotypic Variation Within the Human Hepatitis B Virus Genotypes**

### 3.1 Introduction

Hepatitis B virus (HBV) infection has been detected in most populations throughout the world, as well as in a number of primate species (this will be discussed in more detail in chapter 5). HBV infection is a global health problem, with approximately 1 million deaths from hepatocellular carcinoma and chronic hepatitis estimated to be caused by HBV infection each year (Thomas and Jacyna, 1993). High rates of active infection ranging from 8-15 % are found in South and East Asia, sub-Saharan Africa and amongst indigenous peoples in Central and South America (Andre, 2000). Human HBV can be divided into 8 genotypes based on genetic diversity, differing from each other by 10-13% at the nucleotide level (Figure 3-1). The human genotypes also show a geographically defined distribution pattern (Figure 3-2).

As well as variation between the genotypes, there is also variation within the genotypes. Isolates from the same genotype can show very different virological and clinical characteristics, and the classification of possible subtypes may be important in trying to understand these differences (Miyakawa and Mizokami, 2003). Several reports have described the existence of distinct subgroups within certain genotypes.

One such example is the A' subtype which has been described in HBV isolates from South Africa and Malawi (Bowyer et al., 1997; Kramvis et al., 2002; Owiredu et al., 2001a; Sugauchi et al., 2003c). This subtype does not appear to be associated with a specific disease course (Owiredu et al., 2001b; Sugauchi et al., 2003c) and it is also not known whether it is



the dominant type of genotype A circulating in the southern regions of Africa or whether it is specific to a particular population group or transmission route, although it has now been suggested that this particular subtype, showing a high level of genetic heterogeneity, may be characteristic of the indigenous populations of some of the African countries (Sugauchi et al., 2003c). Genotype A isolates from Europe cluster distinctly from those from southern Africa and those from elsewhere suggesting the existence of other geographically defined HBV variants (Hannoun et al., 2002).

Another genotype that has been reported as existing in different forms is genotype B, one of the predominant Asian genotypes. The different forms in this case are defined as either recombinant or non-recombinant within the core region, with the recombinant form having a core region derived from genotype C variants (Bowyer and Sim, 2000; Sugauchi et al., 2002b). These two variants appear to have distinct geographical distributions, with the non-recombinant strains circulating in Japan, and recombinant strains in Southeast Asia. These subtypes are associated with differences in clinical outcome, with the recombinant genotype B form associated with a more severe disease than the non-recombinant form in Japan. In Japan the most severe liver disease is caused by genotype C (Sugauchi et al., 2002b). The difference in pathogenicity between genotypes B and C has been well documented in many studies (Ding et al., 2002; Kao et al., 2000; Kao et al., 2002; Lindh et al., 1999; Sugauchi et al., 2003a; Sumi et al., 2003). Associations between genotypes and disease is

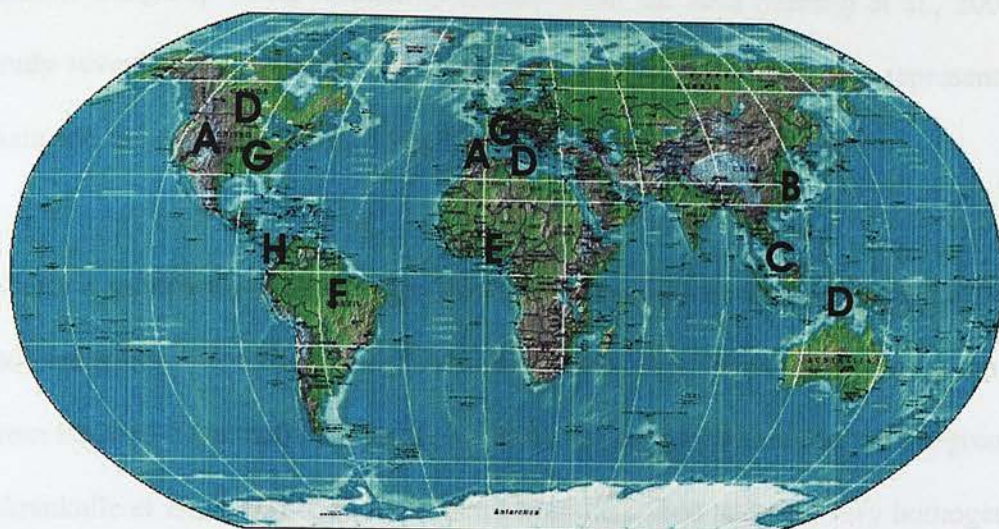


Fig. 3-2. Map of the World showing the distributions of the human HBV genotypes A-H. Genotype A is found predominantly in Europe, the US and Africa, genotypes B and C in East and Southeast Asia, genotype E specifically in Western Africa, genotypes F and H in various population groups in Central and South America. Genotype D is found worldwide, and is also the predominant genotype in the Mediterranean region. Genotype G may have a similar distribution as genotype A or D, having been isolated only from patients in Europe and the US to date.

described in more detail in chapter 1 section 3. Recombination (such as that described for B and C above) between different HBV genotypes will be discussed in more detail in chapter 4.

Apart from existing alongside of genotype B in Asia, genotype C has also been described as existing with at least three subgroups (Alestig et al., 2001a; Alestig et al., 2001b; Huy et al., 2004; Sugauchi et al., 2001), with isolates from Australian aborigines being highly divergent from isolates from other geographical areas (Alestig et al., 2001a; Sugauchi et al., 2001), and

another subgroup being formed by isolates from SE Asia (Alestig et al., 2001b). A recent study revealed the presence of two separate subgroups in Asia, one representing Southeast Asia and the other the Far East and Northern Asia (Huy et al., 2004).

No such groups have been suggested for genotype D, E, and G. Genotype D isolates from isolated tribes of the Nicobar and Andaman Islands have been reported as grouping separately from isolates from mainland India, but were not described as a distinct subgroup or subclade (Arankalle et al., 2003). Genotype E has been described as being very homogeneous (Suzuki et al., 2003) as has genotype G.

Genotype G has so far been reported as consisting of isolates with a high level of homogeneity despite the large distances that separate the two areas where it has so far been identified- France and North America (Kato et al., 2002b; Stuyver et al., 2000; Vieth et al., 2002). It actually appears to be restricted geographically to these two areas, as studies have failed to detect the presence of genotype G in other countries (Kato et al., 2002a; Kato et al., 2002b).

Genotype F has also been reported as existing as a group of geographically defined subsets of strains (Arauz-Ruiz et al., 1997; Arauz-Ruiz et al., 2002; Devesa et al., 2004; Mbayed et al., 1998; Mbayed et al., 2001; Pineiro y Leone et al., 2003; Sanchez et al., 2002). Within these subsets there appear to be high levels of homogeneity within some populations (Nakano et al., 2001) and high levels of heterogeneity in others (Blitz et al., 1998). One genotype F clade representing Central America has now been reported as being defined by a precore mutation associated with the presence of another precore stop codon mutation, although it is not known whether this is a genotype characteristic (Norder et al., 2003). Future research within the field

may reveal similar genetic definitions for other genotype subgroups that may also be related to disease outcome.

Another of the described subgroups was initially tentatively described as a possible new genotype (Arauz-Ruiz et al., 1997; Sanchez et al., 2002), and it has now been identified as such, genotype H (Arauz-Ruiz et al., 2002). As more data becomes available on the prevalence and diversity of genotype H it may turn out that geographically defined subgroups exist within this genotype as well.

The current study aimed to look further at the genotype distributions of HBV within countries in endemic areas, and to investigate further the existence of geographically defined virus variants within the human HBV genotypes. The novel Association Index method will be used to analyse intra-genotypic variation.

## 3.2 Materials and Methods

### 3.2.1 Human serum samples for HBV study

A total of 477 human serum samples were available for screening. 332 of these samples were from indigenous and predominantly rural African communities; 12 from Zaire (now the Democratic Republic of Congo), 51 from Nigeria, 62 from South Africa, 93 from Gambia, 46 from Zimbabwe and 68 from Sudan. The samples were archive samples from previous studies carried out by the Scottish National Blood Transfusion Service (SNBTS). The samples were included in the study as they were from populations in areas where HBV is highly endemic and the contact between humans and other primate species is more likely. The remaining 145 samples were from Papua New Guinea and represented samples taken at time intervals from 38 study subjects. These samples were also representative of indigenous rural populations. Between 1 and 4 samples were available for each study subject. Samples were screened for the presence of HBV DNA as described earlier (chapter 2). Selection criteria for the original samples were indigenous population, further data is not available (Prescott et al., 1999).

### 3.2.2 PreS PCR to investigate primate-like sequences present in human samples

In order to identify any samples that may contain a primate-like virus variant, a short fragment nested PCR was carried out in the PreS region of the HBV genome. This region is phylogenetically informative and polymorphic. Primate HBV and human HBV of genotype D contain a 50bp deletion in this region, so PCR amplification of this region followed by an extended gel electrophoresis can distinguish these types of HBV from all other human genotypes. The first round products from screening PCR 2 (as described in chapter 2) were used to set up an alternative second round PCR with primers 44 and 45. A human control (not genotype D) and a chimpanzee HBV control sample were included. PCR conditions were as described earlier. PCR products were then electrophoresed for 40 minutes at 150 V on a 2%

agarose gel (Ethidium Bromide stained) with a 100 base pair DNA size marker (Promega) and visualised under UV light. A selection of DNA positive samples from Africa was made and sequenced using primers 42 and 43 directly from screening PCR 2 second round amplification products to obtain the PreS region sequence. Phylogenetic trees were then constructed and the genotypes of the samples were determined. DNA positive samples from Papua New Guinea that showed the shorter fragment size were chosen in the first instance for sequencing and were sequenced in the S gene region after this region was determined to be a good region for genotyping (as described in section 3.2.3). Accession numbers for sequences used in the genotyping phylogenetic analyses (PreS and S) can be found in chapter 2.

### 3.2.3 S gene PCR for genotyping

African samples that were sequenced in the PreS region were also sequenced in the S gene region of the genome for genotyping. It was then decided to sequence all human positive samples in the S gene region to get a better idea of the genotype distributions within the countries in this study. Samples were amplified in the S gene region using primers 18 and 31 for first round PCR, and S3 and 26 for second round PCR, with cycling conditions as described earlier. Primer S3 was used for sequencing according to the automated sequencing protocol described earlier (chapter 2). Phylogenetic analysis was then carried out to determine the genotypes of the samples.

### 3.2.4 Genotype-specific phylogenetic analysis of S gene sequences

To investigate any variation within the genotypes, samples were also analysed phylogenetically using larger genotype-specific data sets retrieved from GenBank. Accession numbers for these sequences can be found in appendix A. Mean pairwise distances were also

calculated to enable comparisons of the degree of diversity within the genotypes. This was done using a function of the SIMMONIC sequence analysis package.

### 3.2.5 Association Index analysis

Phylogenetic analysis of S gene sequences using the genotype specific data sets described above appeared to show geographically defined clusters within the trees (see section 3.3.4). To investigate this finding Association Index (AI) analysis was employed. This novel phylogenetic analytical method was initially developed to investigate the segregation of populations of HIV-1 isolates from different body tissues from the same patient (Wang et al., 2001) and is described in more detail in chapter 2. Briefly, the degree of genetic segregation between variants from different locations is scored based on phylogenetic analysis. An association value ( $d$ ) is calculated according to the formula  $d = (n-f)/2^{n-1}$ . In this formula  $f$  is the frequency of the most common sample type in each cluster within the tree, and  $n$  is the number of sequences within the cluster. Random re-sampling of sequences into groups was carried out to provide a control value based on the null hypothesis of no genetic segregation. The Association Index value represents the mean ratio of 100 bootstrap replications of the  $d$  of the test sequences to the controls.

This method is implemented in the SIMMONIC sequence analysis package (see chapter 2). Groups of sequences were labelled according to their geographical origin prior to phylogenetic analysis and calculation of the AI score. The resulting AI values indicate the extent to which these geographically pre-defined groups cluster separately within the phylogenetic tree. The maximum score possible is 1, and this would indicate no segregation according to the pre-defined groupings. A score lower than 0.5, on the other hand, indicates a significant degree of segregation.

### 3.3. Results

#### 3.3.1 Frequency of HBV viraemia

The levels of viraemia detected in the samples for each country were calculated (Table 3-1) based on PCR positive results in both PreS and S gene PCR as described in chapter 2. S gene PCR fragments were 1000 basepairs in length (Figure 3-3) and were used directly for sequencing as described in chapter 2 (see section 3.3.3 for sequencing results).

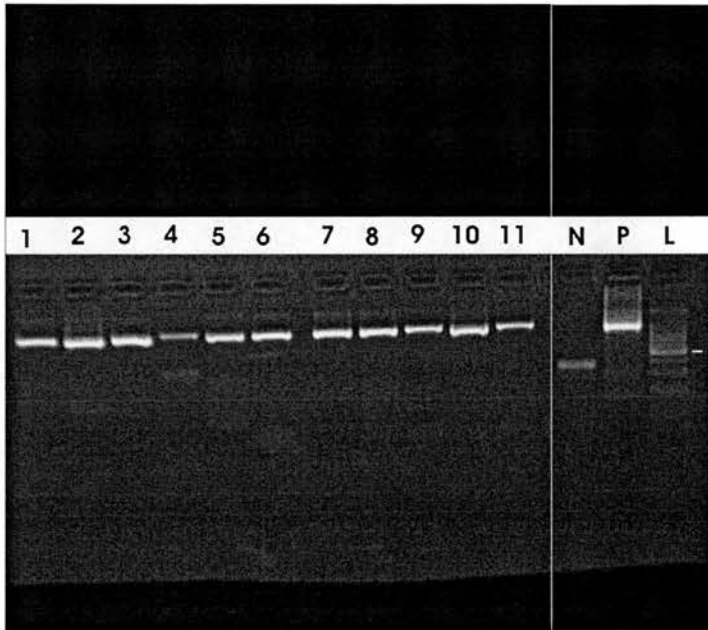


Figure 3-3. Gel image showing an example of S gene PCR products. Lanes 1 to 11 represent samples from Papua New Guinea, lane N is a negative control, lane P is a positive control (size of fragment 1000 bp), and lane L is a molecular size marker. The brightest band seen is 500 bp in length.

#### 3.3.2 PreS PCR to detect possible primate-derived virus variants

African samples were found to contain amplicons of the expected length for human genotypes other than D (Figure 3-4). Some samples from Papua New Guinea were found to be of the shorter length (n = 16). These 16 samples were chosen for S gene amplification and sequencing for genotyping in the first instance, to investigate whether the deletion was due to

genotype D being present or another possibly primate-derived genotype being present. Of the 16 samples that were sequenced 11 were shown to contain genotype D, and the remaining 5 genotype C. No samples from Papua New Guinea were found to contain a genotype other than human genotypes C and D. This suggests that PreS deletions may be more common in genotypes other than genotype D than previously reported, and not exclusively in genotype D and the primate genotypes.

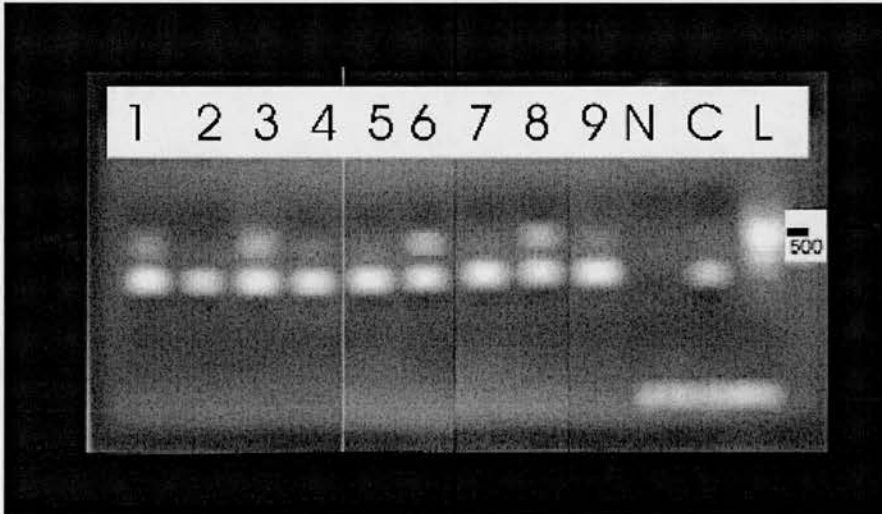


Figure 3-4. Gel image showing an example of the PreS PCR showing size differences. The positive control used in this PCR is chimpanzee derived, and shows the 50 bp deletion (lane C, 150 bp in length). All samples in this PCR are of African origin (lanes 1 to 9, 200 bp in length) and are of the normal length expected for human genotypes other than genotype D. Lane L is a molecular size marker, with the brightest band being 500 bp in length.

### 3.3.3 Genotype distribution

Of the 121 DNA positive samples, a total of 118 positive samples were genotyped. Three samples were only sequenced in the PreS region due to lack of sample material, yielding 115 new S gene region fragments for analysis (Figures 3-5a and 3-5b). Insufficient material was available to allow genotyping of three positive samples.

new S gene region fragments for analysis (Figures 3-5a and 3-5b). Insufficient material was available to allow genotyping of three positive samples.

The majority of samples fall into genotypes A, C, D, and E, as would be expected for samples from these regions. No samples were found to contain human genotypes F, G, and H or any of the primate HBV genotypes. Two samples from South Africa were found to contain genotype B, and when going back to patient information these two patients both had an Asian background (see also Table 3-2 and Figure 3-6).

Samples from Sudan were all found to be negative, which was an unexpected finding. The HBsAg status and e antigen status for these samples were unknown, as the samples had not been previously serology tested for HBV infection.

COUNTRY	NO. PATIENTS	POS. PATIENTS	PREVALENCE (%)
ZAIRE	12	3	25
NIGERIA	51	16	31.37
SOUTH AFRICA	62	6	9.68
THE GAMBIA	93	49	52.69
ZIMBABWE	46	18	39.13
SUDAN	68	0	0
PAPUA NEW GUINEA	38	29	76.32
TOTAL	370	121	32.70

Table 3-1. Prevalence of viraemia detected in the study group. Results are expressed as the number of patients rather than number of samples. As there were multiple samples available from 38 patients from Papua New Guinea, the first sample for each patient that showed a positive result was chosen for further analysis. The number of patients tested therefore is 370, although the actual number of samples available for screening was 477.

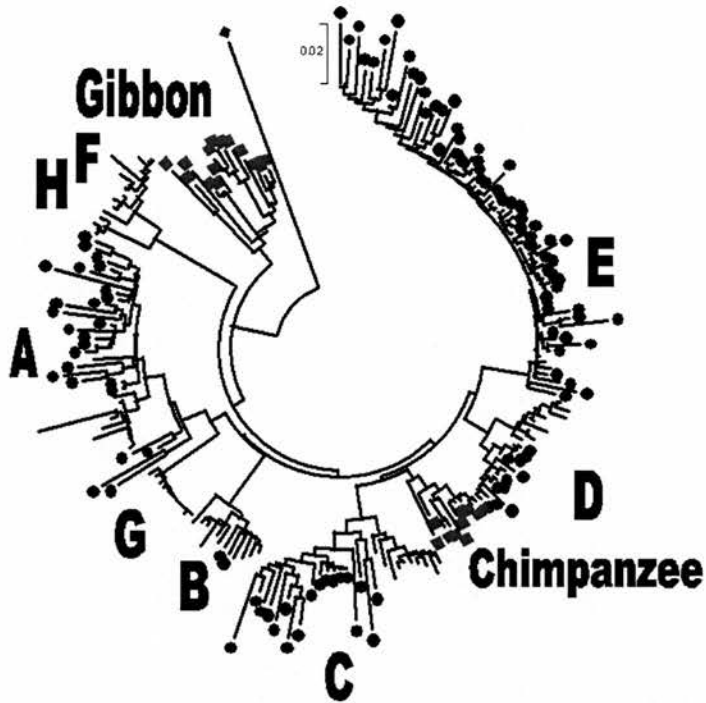


Figure 3-5a. S gene phylogenetic analysis of positive samples for genotyping. Sequences obtained from study subjects are identified with filled in circles. Genotype C and D sequences from this study are all from Papua New Guinea, whereas the A, B and E sequences are of African origin. Primate HBV isolates are indicated with red boxes. The outgroup for the tree is the woolly monkey HBV sequence.

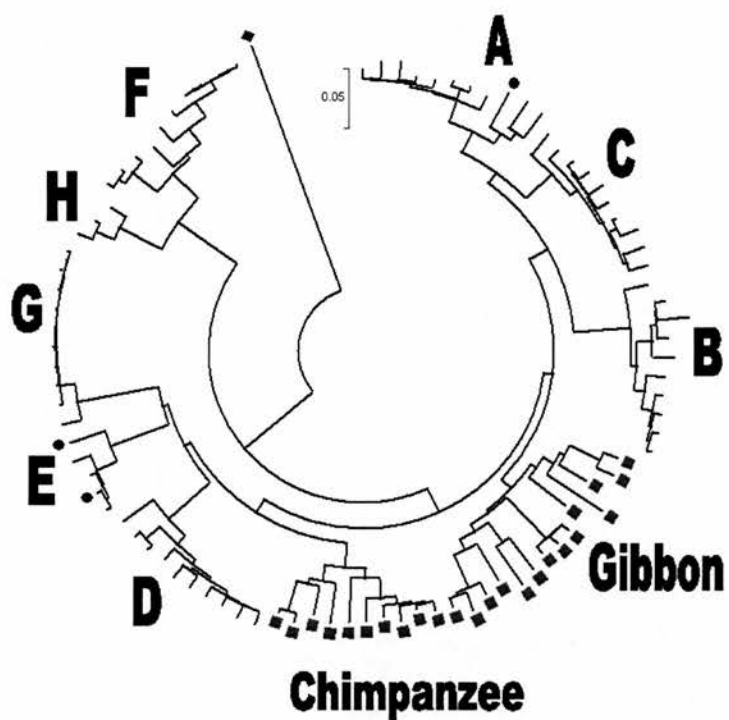


Figure 3-5b. PreS phylogenetic tree for genotype determination of tree DNA positive samples where there was insufficient material to allow S gene sequencing. Sequences from study subjects are identified with filled in circles, all samples are of African origin.

COUNTRY	GENOTYPE					
	A	B	C	D	E	F,G,H
ZAIRE	1	0	0	0	2	0
NIGERIA	1	0	0	0	14	0
SOUTH AFRICA	3	2	0	0	0	0
THE GAMBIA	4	0	0	0	45	0
ZIMBABWE	17	0	0	0	0	0
PAPUA NEW GUINEA	0	0	18	11	0	0
TOTAL	26	2	18	11	61	0

Table 3-2. Genotype distribution of the 118 sequences obtained from the study samples. Zaire, Nigeria and The Gambia have genotype E as the prevalent genotype, which would be expected from their West African location. South Africa and Zimbabwe have genotype A as the prevalent genotype, which corresponds with their more southerly location within the continent of Africa.

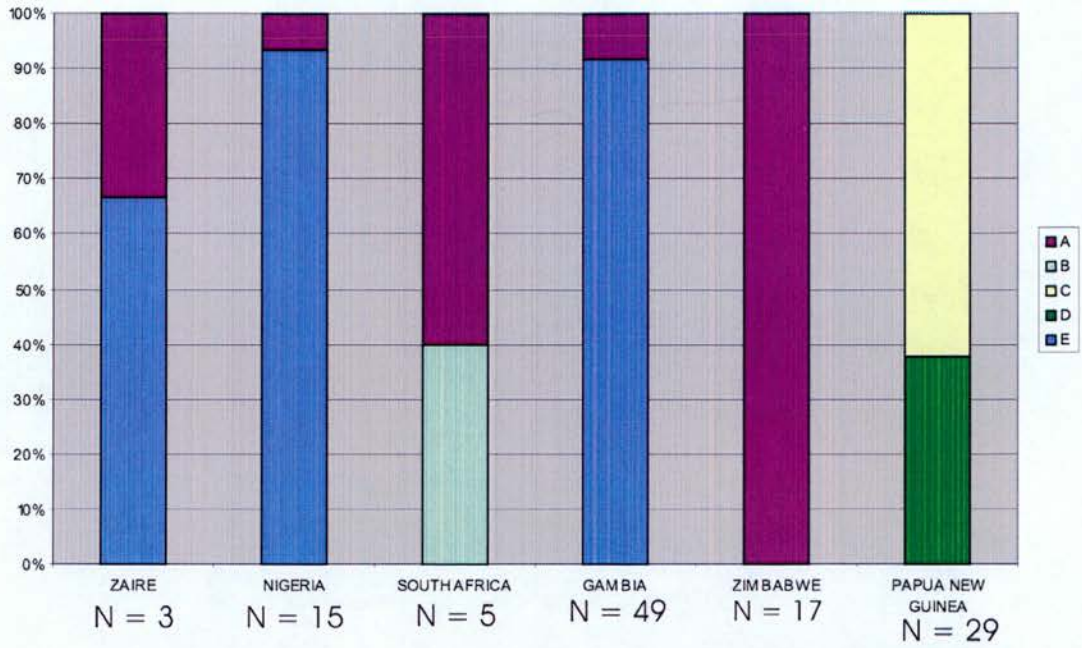


Figure 3-6. Percentage distribution of genotypes in the countries studied.

Genotypes A and E are most prevalent in the African countries, whereas genotypes C and D are found in Papua New Guinea.

#### 3.3.4 Genotype-specific S gene phylogenetic analysis and Association Index analysis

A total of 423 S gene region sequences were downloaded from GenBank and aligned, and where necessary genotyped by phylogenetic analysis. All sequences did not provide genotype information within the GenBank data and therefore needed to be genotyped. Data sets were compiled to represent each of the human genotypes separately. Phylogenetic analysis of the S gene sequences was then carried out to investigate the levels of variation within the genotypes. The woolly monkey sequence was used as an outgroup in all cases. Sequences were selected from GenBank so as to represent as large a number of studies as possible, and as wide a selection of different countries as possible. Even though no samples were detected in this study, genotypes F, G, and H were also analysed in this way, to enable comparisons of the levels of variation seen within all the human genotypes (Figures 3-7 to 3-13). Genotypes F and H are analysed in the same tree due to the lack of available sequences for genotype H, they do however represent separate genotypes and will be discussed as such.

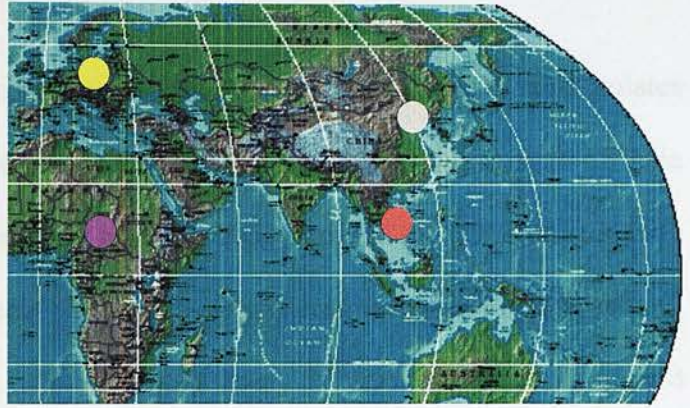
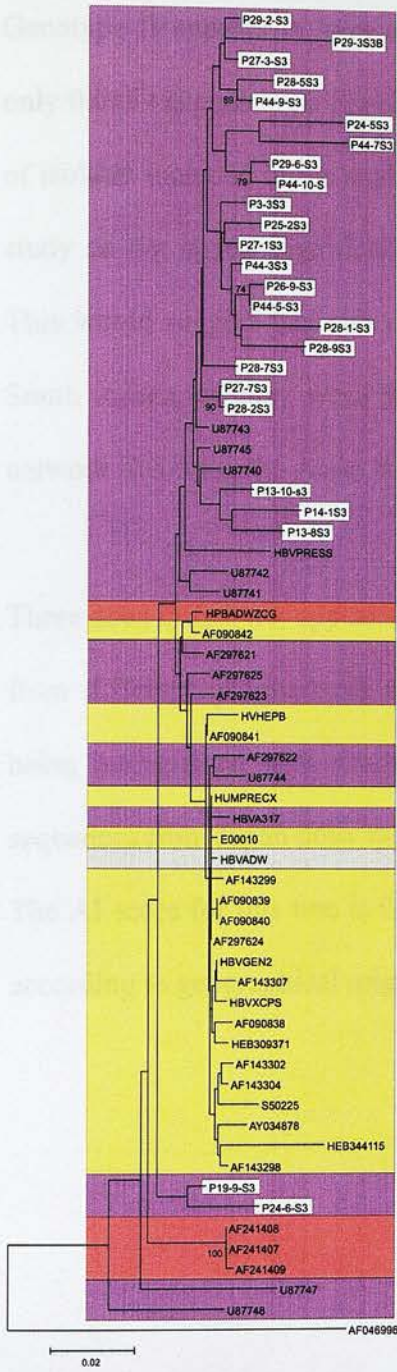
The woolly monkey isolate is the outgroup for all analyses.

The mean pairwise distances were also calculated for each genotype (Table 3-3). Pairwise distances show the level of dissimilarity between sequences, and the higher the distance value, the less related the sequences being compared are. Pairwise distances can be used to estimate the level of diversity seen within a genotype, and therefore as an estimation of the evolutionary history of a particular genotype.

After being labelled both the samples from this study, and the sequences obtained from GenBank, according to the country/region of origin manually, following the data provided either from GenBank or the publication where the sequences were reported, the phylogenetic trees obtained were visually inspected for any clustering of sequences by me. Sequences were then also labelled according to these groupings in the SIMMONIC alignment, and the Association Index analysis was run (Table 3-4). Phylogenetic analysis showing the geographical labelling of HBV sequences for the AI analysis of each genotype were included to allow interpretation of AI values (Figures 3-7 to 3-13).

Genotype A appears to have quite a high level of diversity within the genotype, and this is confirmed by the pairwise distance figure of 0.0309 (Figure 3-7). Samples from this study mainly group together, which may reflect the source population and the transmission networks seen there. The high level of diversity seen among the samples from this study would suggest that the viral evolution is being driven at a fast rate by either host factors or treatment and prevention regimes. The study populations are from endemic areas, and this level of diversity suggests a large pool of circulating HBV variants. S gene analysis distinguishes the A' subgroup, though not so clearly, from European isolates as is shown also by the AI result (see below), as has been suggested before (Sugauchi et al., 2003c). These A' sequences group together to root the main European isolate branch, suggesting a close link between South African and European genotype A variants.

From these data it can be clearly seen that genotype A is mostly found in Europe and southern parts of Africa. There appears to be some degree of segregation between isolates from the different regions. When performing AI analysis on the tree labelled as described the score is 0.2157, indicating a significant degree of segregation according to geographical area of origin. There does not appear to be any separation of HBV strains due to the specific country of origin, with strains from Nigeria, South Africa, Gambia and Zimbabwe being found interspersed with all African isolates within the tree.



Northern Asia is represented by pale grey,  
 South East Asia by pale pink, Europe by pale  
 yellow and Africa by pale purple.

AI = 0.2157

Fig. 3-7. Geographically labelled S gene phylogenetic analysis of HBV genotype A. The samples from this study are highlighted in the white boxes (n = 25). AI analysis suggests segregation of isolates according to geographical origin.

Genotype B appears to have a high level of diversity, but the pairwise distance is only 0.0159 suggesting a lower level than expected (Figure 3-8). The limited number of isolates included in the analysis may be the reason for this. The isolates from this study do not appear significantly different from other isolates within this genotype. This would suggest that either the patients were infected in Asia prior to arrival in South Africa, or they were infected by a strain circulating within a transmission network involving the Asian community in South Africa.

There does in fact not appear to be any separation between genotype B virus isolates from different geographical origins, with strains from the northern parts of Asia being interspersed with strains from southern Asia and Europe. One cluster with sequences from Japan does appear to exist (see sequences marked in Fig. 3-8 above). The AI score for this tree is 0.5205, indicating no significant segregation of isolates according to geographical origin for genotype B.

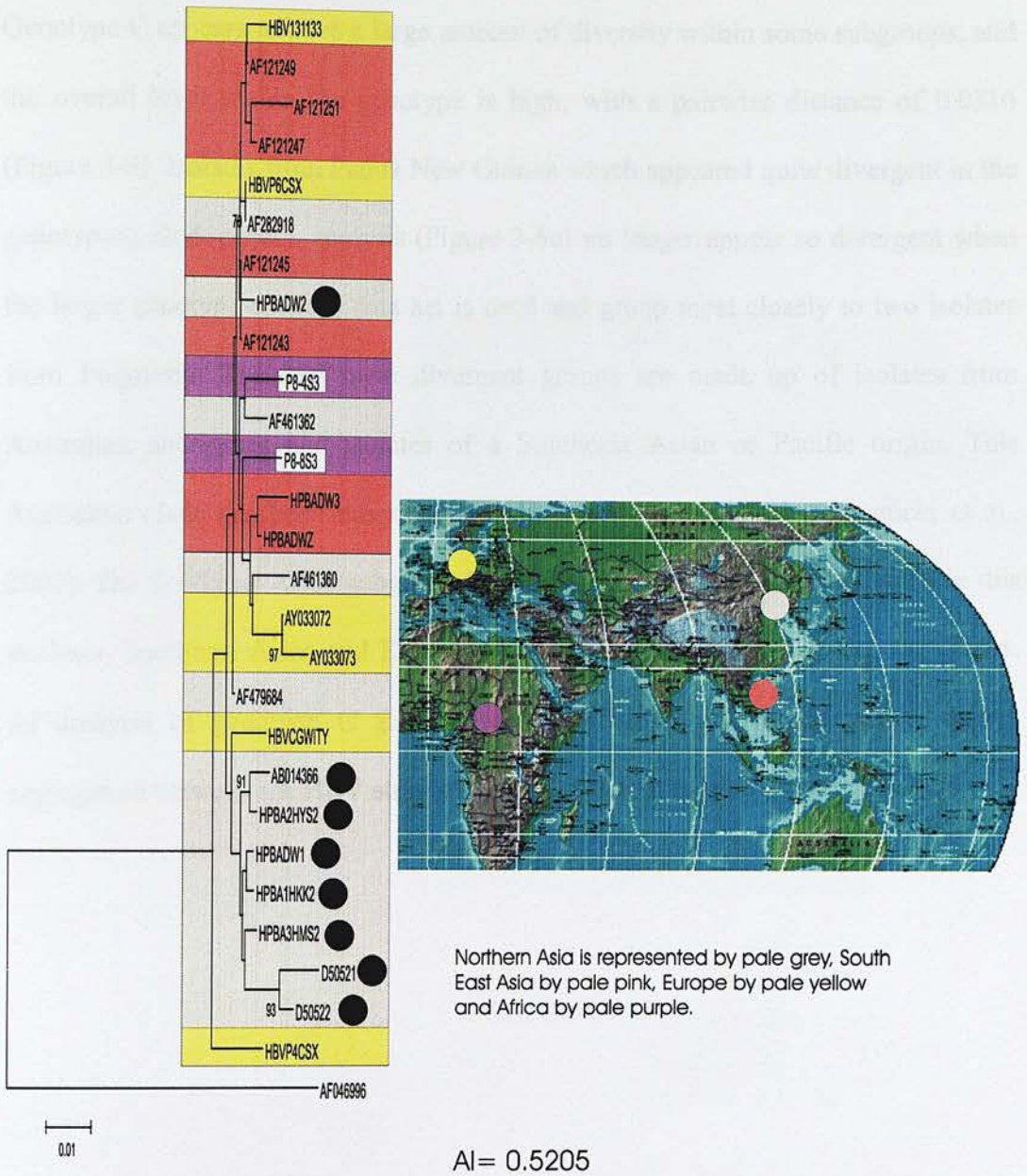
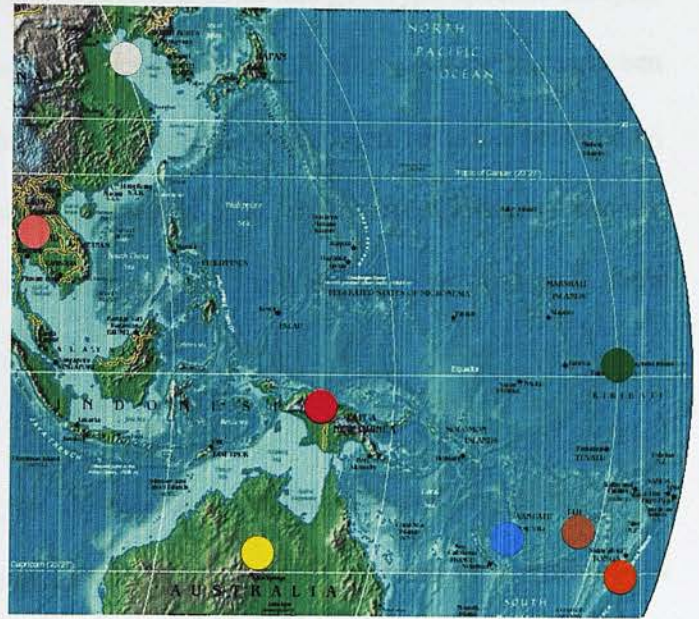
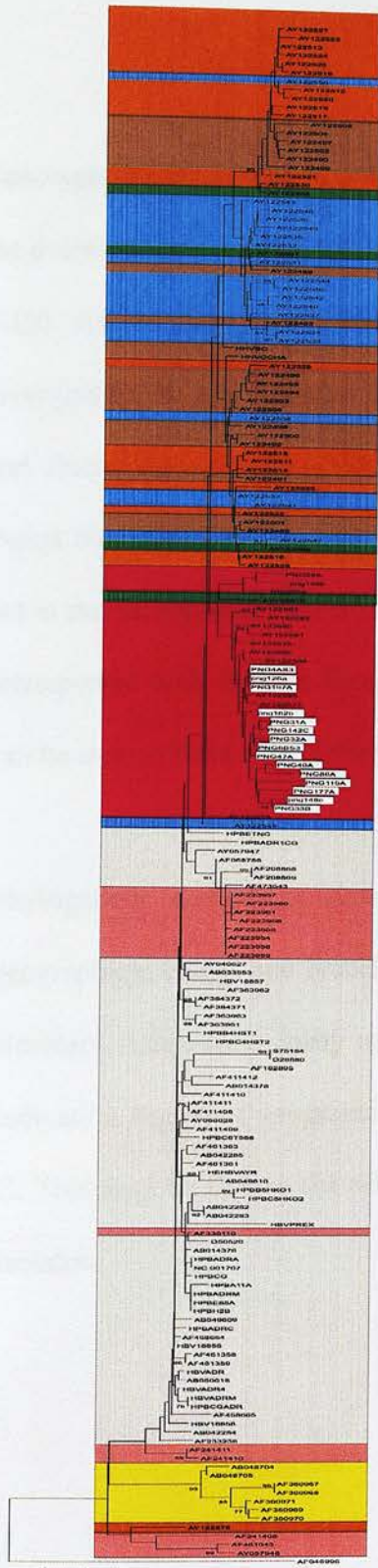


Fig. 3-8. Geographically labelled S gene phylogenetic analysis of HBV genotype B. Samples from this study are highlighted in white boxes (n = 2). Japanese isolates are marked with black filled in circles. AI analysis suggests a lack of segregation according to geographical origin.

Genotype C appears to have a large amount of diversity within some subgroups, and the overall level within the genotype is high, with a pairwise distance of 0.0316 (Figure 3-9). Isolates from Papua New Guinea which appeared quite divergent in the genotyping phylogenetic analysis (Figure 3-5a) no longer appear so divergent when the larger genotype-specific data set is used and group most closely to two isolates from Polynesia. The two most divergent groups are made up of isolates from Australian aborigines and isolates of a Southeast Asian or Pacific origin. This Australian clade has been described before (Alestig et al., 2001a; Sugauchi et al., 2001). The Southeast Asian subgroup (Alestig et al., 2001b) is not as apparent in this analysis. Southeast Asian and European isolates are distributed between the clades. AI analysis of genotype C gives a score of 0.1848, which indicates a strong segregation between the HBV strains from the different regions.



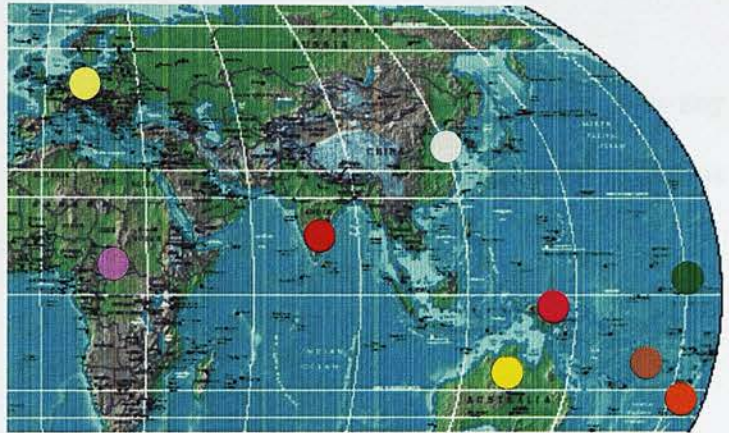
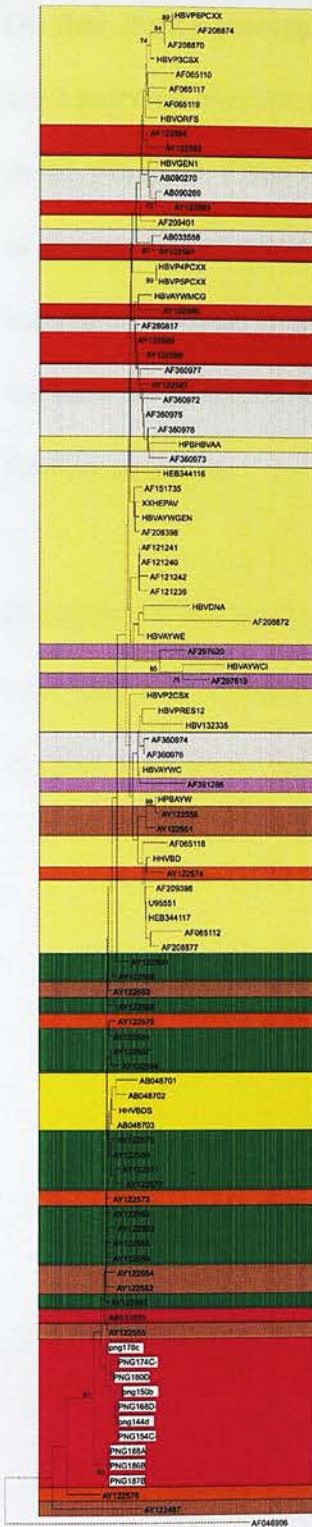
Northern Asia is represented by pale grey, Southern Asia by pale pink, Papua New Guinea by magenta, Fiji by brown, Tonga by orange, Kiribati by green, Vanuatu by blue and Australia by yellow.

AI = 0.1848

Fig. 3-9. Geographically labelled S gene phylogenetic analysis of HBV genotype C. Samples from this study are highlighted in the white boxes (n = 18). AI analysis suggests segregation of isolates according to geographical origin.

Genotype D also appears to contain a fairly high level of diversity within it, although the pairwise distance of 0.0226 indicates a more moderate level of diversity (Figure 3-10). As for genotype C, isolates from Papua New Guinea that appeared highly divergent in the genotyping analysis appear less so in this genotype-specific analysis and cluster closely with isolates from Pacific Islands. A close association between Papua New Guinean isolates and Australian isolates has been reported previously, but is not so apparent in this analysis (Sugauchi et al., 2001). Australian isolates are interspersed with isolates from the Pacific region, and Papua New Guinea isolates can be seen to form two distinct clusters.

Phylogenetic analysis of genotype D sequences results in a similar formation of geographical clades and branching pattern as genotype C. Isolates from Europe and Northern Asia appear fairly mixed. The AI score for this tree is 0.4789. This does indicate a degree of geographical segregation, though not as strong as for genotype C. This may be due to the relationship between the European and Northern Asian isolates.



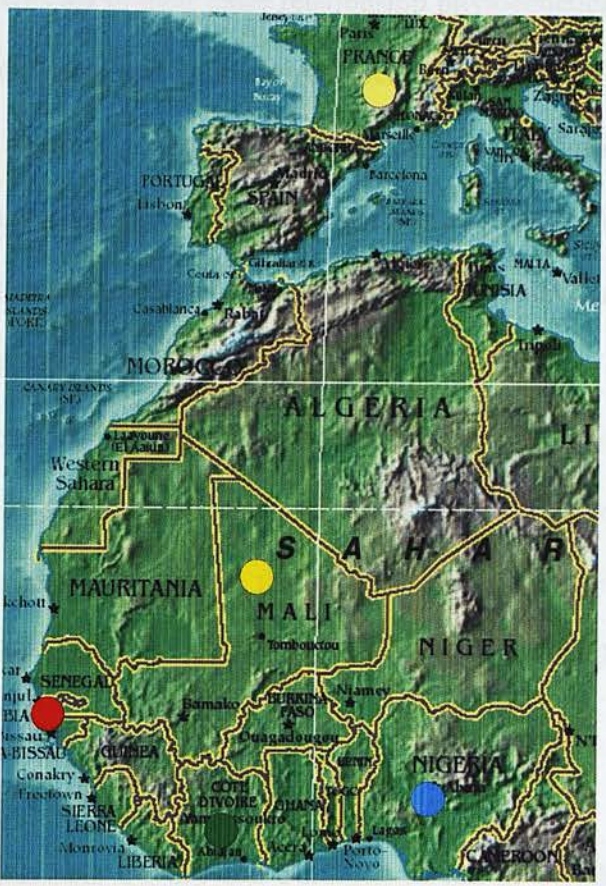
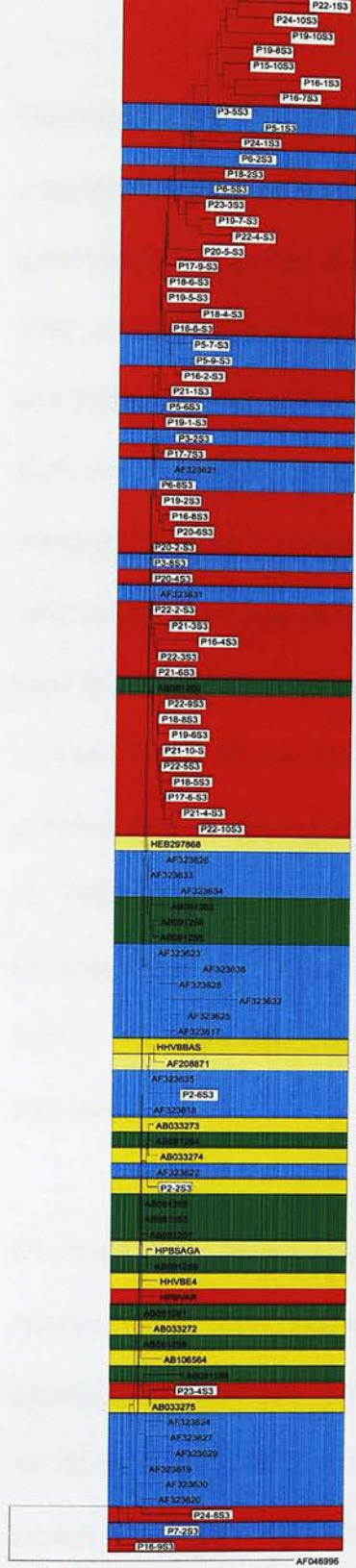
Northern Asia is represented by pale grey, Europe by pale yellow, Africa by pale purple, India by red, Papua New Guinea by bright pink, Australia by bright yellow, Kiribati by green, Fiji by brown and Tonga by orange.

AI = 0.4789

Fig. 3-10. Geographically labelled S gene phylogenetic analysis of HBV genotype D. Samples from this study are highlighted in the white boxes (n = 11). AI analysis suggests segregation according to geographical origin.

On first glance genotype E appears to have less diversity than the other genotypes, but a pairwise distance of 0.0184 suggests a level of diversity similar to that seen for genotypes B, D, F and H (Figure 3-11). This genotype has previously been described as having limited diversity (Suzuki et al., 2003), but this appears not to be the case and may simply be an artefact due to lack of available sequence data until now. From the data set generated for this study it is clear that genotype E is specific to western African countries.

In contrast to the results seen for genotypes C and D, there does not appear to be any segregation of isolates according to the country of origin in genotype E, which is supported by the AI score of 0.6246.



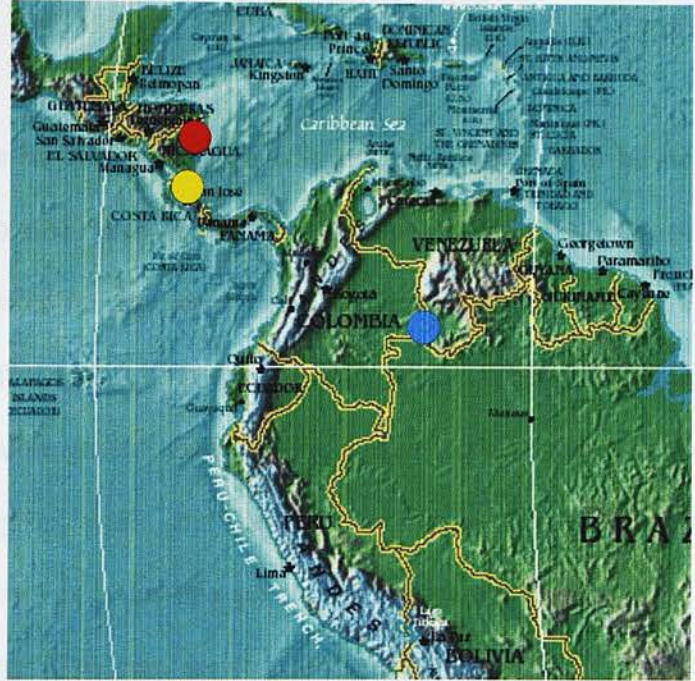
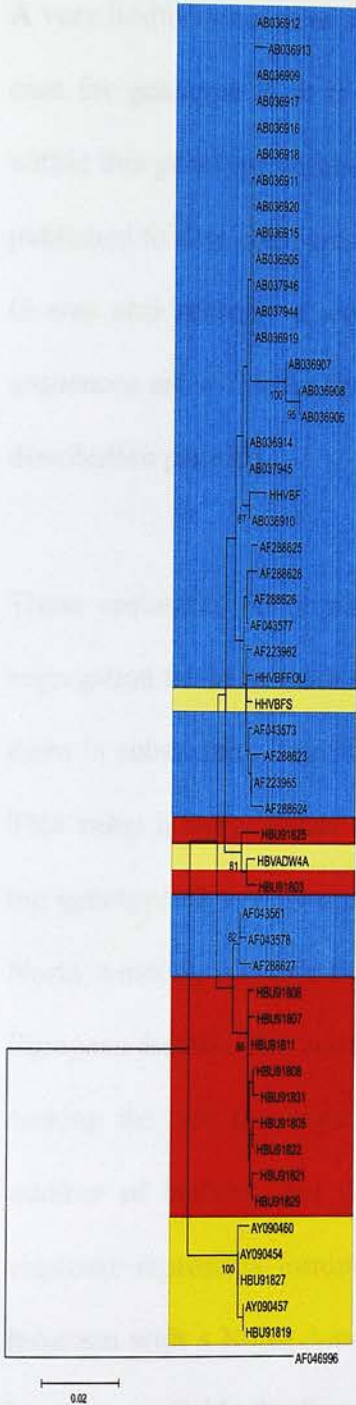
Gambia is represented by red, Nigeria by blue, Ivory Coast by green, Europe by pale yellow and other West African origin by yellow.

$$AI = 0.6246$$

Fig. 3-11. Geographically labelled S gene phylogenetic analysis of HBV genotype E. Samples from this study are highlighted with the white boxes (n = 61). AI analysis suggests there is no segregation according to geographical origin.

Genotypes F and H were analysed phylogenetically together due to the lack of data available for genotype H (Figure 3-12). Prior to its identification as a novel genotype, genotype H was in fact thought to be a subtype of genotype F (Arauz-Ruiz et al., 1997; Arauz-Ruiz et al., 2002; Mbayed et al., 1998; Mbayed et al., 2001; Torbenson and Thomas, 2002), and this is reflected in their close relationship phylogenetically. Both these genotypes are found in Central and Southern America. When carrying out complete genome phylogenetic analysis genotypes F and H are always outliers, whereas as can be seen in the genotyping tree (Fig. 3-5a), when S gene sequences are used they group together with the other human genotypes. The changing relationship between F and H and the other human genotypes depending on the part of the genome used for phylogenetic analysis has been described previously (Naumann et al., 1993). There is limited diversity within both genotypes F and H. The pairwise distance value for F alone is marginally lower than that of H, 0.0132 compared to 0.0021. As more data becomes available for genotype H this analysis should be repeated.

Genotype F has been isolated from both Central and Southern America, and there appears to be some segregation of isolates according to area of origin. When genotype F is analysed, with two groups defined as Central and Southern America, an AI value of 0.4811 is obtained. This suggests that there is some segregation, though not highly significant. The lack of diversity seen between isolates may explain this result, as visually the groupings seem segregated. As more data becomes available for genotype H, this analysis should be carried out to investigate whether geographical variation is seen.



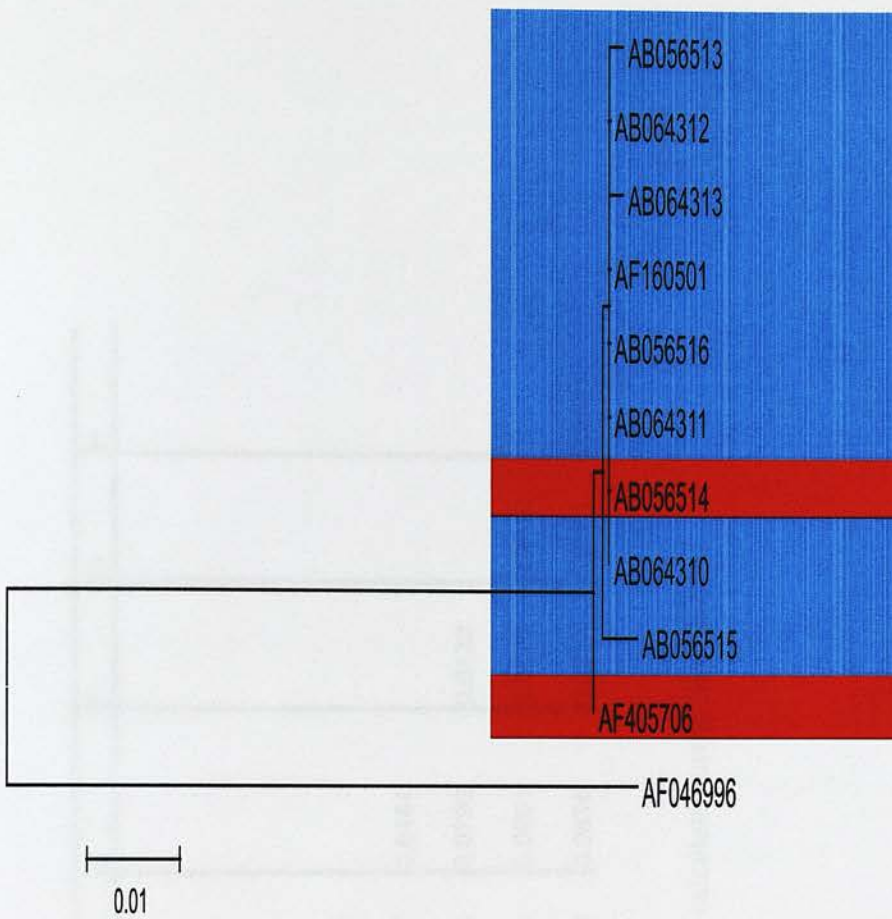
South American genotype F is represented by blue. Central American genotype F by red and genotype H by yellow. Two isolates identified in Europe are represented by pale yellow.

AI = 0.4811

Fig. 3-12. Geographically labelled S gene phylogenetic analysis of HBV genotypes F and H. AI analysis suggests segregation according to geographical origin.

A very limited amount of sequence data is available also for genotype G, so as is the case for genotype H, it is difficult to speculate about the true amount of diversity within this genotype within human populations. The diversity seen within the isolates published to date compares to that of genotype H only. As for genotype H, genotype G was also recently discovered (Stuyver et al., 2000), and a very limited set of sequences are available for analysis making it difficult to analyse the geographical distribution patterns.

These sequences all originate in Europe and the US, and there appears to be no segregation between these two areas even though the geographical distance between them is substantial. The AI value for the genotype G tree is 0.4804 (Figure 3-13). This value is very similar to that seen within genotypes D, and F. When looking at the genotype G tree, two possible groups are seen; one containing all isolates from North America and one European isolate, and the other consisting of one outlying European isolate. The bootstrap values for this genotype G tree are all below 70%, making the tree less significant. The data set for this genotype is limited by the number of isolates, but if this is extrapolated one may argue that the outlying sequence represents European strains and the other European isolate represents an infection with a North American strain. If this is the case, as more genotype G data becomes available the European group may become more defined and the AI value would decrease.



The United States are represented by blue and Europe by red.

AI= 0.4804

Fig. 3-13. Geographically labelled S gene phylogenetic analysis of HBV genotype G.

AI analysis suggests segregation according to geographical origin.

GENOTYPE	GENOTYPE							
	A	B	C	D	E	F	G	H
A	0.0309							
B	0.0574	0.0159						
C	0.0678	0.0650	0.0316					
D	0.0596	0.0594	0.0620	0.0226				
E	0.0684	0.0639	0.0634	0.0527	0.0184			
F	0.0874	0.0778	0.0850	0.0694	0.0793	0.0132		
G	0.0476	0.0602	0.0656	0.0531	0.0601	0.0722	0.0018	
H	0.0872	0.0789	0.0909	0.0720	0.0836	0.0314	0.0751	0.0021

Table 3-3. Mean pairwise distances within and between the genotypes as calculated using SIMMONIC.

GENOTYPE	AI VALUE
A	0.2157
B	0.5205
C	0.1848
D	0.4789
E	0.6246
F	0.4811
G	0.4804

Table 3-4. Association Index analysis results for each of the HBV genotypes.

AI values above 0.5 indicate no significant separation according to the predefined groupings, whereas values below 0.5 do indicate a separation of groupings. Genotype H has not been analysed as there are very few reported isolates to date and they all share a common geographical origin.

### 3.4 Discussion

In this study high rates of HBV DNA positivity were found in all study populations. These high levels may not accurately portray the real level of HBV prevalence within the countries represented in the study, due to for example low numbers of samples available for analysis and the selective nature of the specimens used in the study. The backgrounds of the study groups are unknown, and the possibility therefore arises that they may represent groups of individuals within the general population that are already at a high risk for HBV infection. This would result in an artificially high level of detectable viraemia. The countries represented in this study are however known to be endemic for HBV, so high levels of HBV infection would be expected. Detection of HBV DNA indicates an active infection, but patients who have lost their HBsAg are also HBV DNA positive. Patients with HBsAg expression do not show HBV DNA positivity. This means that the presence of HBV DNA does not identify all carriers of infection.

The distribution of genotypes within the countries was mostly as expected. Genotype A is the most prevalent in the more southerly African nations of Zimbabwe and South Africa and genotype E in the Western African countries of Nigeria and Gambia. Two of the isolates from South Africa were of genotype B, which was not expected. These two samples came from individuals of Asian origin/background, making the result more easily explained. The genotyping results for Papua New Guinea were somewhat unexpected. The presence of genotype C in this population was as expected, whereas the high number of genotype D isolates and absence of genotype B were unexpected. The close proximity of Papua New Guinea to

Southeast Asia suggests a genotype distribution similar to this region would be found.

The finding of genotype D in Papua New Guinea further strengthens the evidence that this is a globally prevalent genotype. No isolates of genotypes F, G, or H were found in any of the study populations. This was as expected, as F and H are found in South and Central America, and G has only been isolated in a small number of European and North American patients to date.

Genotype A shows geographical separation between Europe and Africa (Fig. 3-7) as has been described (Hannoun et al., 2002), but not between different countries in these areas. There is a close relationship between the European isolates and isolates from South Africa (A'), and considering the fact that South Africa was under European rule and domination for a long period of time this is not entirely unexpected.

The separation of European and African isolates is indicative of the HBV transmission networks involved, for example a founder effect. Different geographical regions have circulating virus strains from a common ancestor. Infections appear to be picked up within the region of residence, where HBV is then maintained within the population through for example horizontal and/or vertical transmission within the family setting. This also indicates that the different pools of HBV variants circulating within the two regions share a common evolutionary origin, but are different, having diverged at some point in the past as the carrier populations became more separated.

This type of scenario would be consistent with an HBV origin around the time of the Out of Africa migration around 100 000 years ago. The genotype A ancestor was present in African populations, as they moved into Europe and formed new settlements the virus started to change. A founder effect would mean that there is less variability between the strains circulating within these new population settlements, and slowly over time these geographically separated virus populations become more divergent from each other and the ancestral strain. Considering the long history of European colonization of African nations, it is slightly puzzling that the HBV strains are so separated. Colonization would have provided an opportunity for new virus strains to enter new population groups, and this would be expected to lead to a homogenisation of the viruses found in the native and the colonising populations.

Although genotype A appears to exist in geographically defined subtypes when analysed in the S gene, the previously described A' subtype is not apparent. This subtype has been described following phylogenetic analysis of complete genomes, S gene and Pol gene (Bowyer et al., 1997; Kramvis et al., 2002; Owiredu et al., 2001b), but one previous study also found that small S gene phylogenetic analysis failed to distinguish the group (Sugauchi et al., 2003c). The phylogenetic analysis in the current study was carried out on the small S gene region of the S open reading frame, confirming this previous report (Sugauchi et al., 2003c). As differences can be seen in the geographic cluster/subgroup formation within genotype A further work will be necessary to evaluate the groupings within all the 4 open reading frames of HBV as well as for complete genomes. This data will then be a useful tool when performing epidemiological investigations. Differences in host factors as well as the evolutionary

origin may explain why the level of diversity seen within this genotype is higher than that of other genotypes such as genotype G.

Genotype B is a contrast to genotype A, showing both a lower level of intra-genotype diversity and a lack of geographical segregation of groups of isolates (Fig. 3-9). There is no clear segregation between Southeast Asian and Northern Asian (Japanese) isolates, as is evident from the AI analysis, although one clade of Japanese isolates appears to form (see Fig. 3-9). The difference in pathogenicity of genotype B seen between these two regions is substantial, and the lack of genetic segregation between isolates from the regions makes this difference even more puzzling. Had distinct isolates been found in the two regions one could argue that within the nucleotide changes that separate them lies an important determinant for pathogenicity. The reported prevalence of recombinants with genotype C may play a role in determining these differences (this will be discussed further in chapter 4).

Differences in host factors may also be important in determining the differences in disease outcome in these regions. This type of scenario is more indicative of a more recent origin of the genotype within the affected population. Pathogens new to a population, for example, a virus entering a human population through cross-species transmission, such as Ebola, tends to cause more severe disease in the new host than pathogens which have been circulating for a long time, for example HGV/GBV-C, where both the host and the pathogen have had time to adapt.

Genotype C is another genotype that, like genotype A, shows a high level of diversity within the genotype with a high pairwise distance value (Table 3-3), and this is also reflected in the existence of geographically defined clusters of related isolates which are supported by AI analysis (Fig. 3-10 and Table 3-4). The isolated position of Papua New Guinea and the tribes who live there make the existence of a divergent strain there seem quite likely. This is also reflected in the fact that a similarly divergent subgroup exists in Australian aborigines (Alestig et al., 2001a; Sugauchi et al., 2001). The AI score obtained for the phylogenetic tree (Fig. 3-10) supports previous claims of sub-groups within Asian genotype C (Huy et al., 2004). The high pairwise distance value suggests that this genotype is old, as many differences between strains have accumulated. The existence of geographical subgroups within genotype C suggests some co-evolution, with population based HBV variants being established through a founder effect as described above for genotype A.

The overall level of diversity within genotype D is not as high as for genotype C with a pairwise distance value of 0.0226 (Table 3-3), and the segregation of geographical subgroups not as strong (AI = 0.4789, Table 3-4). The existence of a divergent subgroup of genotype D, like that for genotype C described above, consisting of isolates from Papua New Guinea is not unexpected. Further data is needed from isolated tribes around the world, such as those living in the Nicobar and Andaman Islands (Arankalle et al., 2003). If genetic data were published for HBV isolates from these populations the geographical distribution of genotype D may become clearer. Genotype D is found throughout the world, so it is surprising that there is not more

segregation. One could argue that there are ancestral strains of genotype D such as those from Papua New Guinea, and that the spread of one such strain throughout the world happened relatively recently. Papua New Guinea was inhabited between 40 000 and 60 000 years ago, and isolated tribes still exist there today (Catriona MacFarlane, PhD thesis). This would explain the lower level of diversity and lack of geographical variants.

Genotype E has a low level of diversity compared to the other genotypes, and also shows a lack of geographical subgroups. The limited diversity of this genotype has been described before (Suzuki et al., 2003). The association of genotype E infection with West African countries is clearly demonstrated. There is more diversity within this genotype than previously suggested, but it still appears that the origin of genotype E in its host populations is again more recent. The pairwise distance value for genotype E is 0.0184, suggesting there has been less time available to accumulate differences between strains. This also indicates a rapid spread of infection within West Africa. The lack of segregation of isolates according to the geographical origin (AI = 0.6246) supports this epidemic spread. As more data is made available for this genotype geographical clusters may become more apparent, or the evidence for a recent origin of the genotype may be strengthened. For geographically segregated subgroups of viruses to evolve the populations carrying them need to be isolated. Genotypes are also maintained by separation. Increased frequencies of superinfection with multiple virus types will drive the evolution of the virus towards a lack of genotypes and subtypes through the mechanisms of recombination and mutation. HBV infection is endemic in West Africa which makes superinfection a real

possibility, and therefore, this may be responsible for the lack of diversity seen within genotype E. If this is the case the origin of HBV in West Africa may be more ancient than expected.

Genotypes F and H both show very limited sequence variability, although there does appear to be some geographical segregation of subgroups of isolates. The outlier position of genotypes F and H, this lack of diversity, and the close relationship between them suggests a recent origin from a common ancestor. It has been suggested that South America is the source of HBV infection in the rest of the World as a consequence of European colonisation in the last few hundred years (Bollyky et al., 1997), but the low level of diversity does not agree with this hypothesis (see section 1.5 for more information on this hypothesis). A study recently published analysed a larger data set of sequences from genotype F (Devesa et al., 2004). They showed three distinct geographically defined clades containing isolates from Central America, northern South America and southern South America respectively, as well as two clades containing isolates from Central America and southern South America. This clearly demonstrates that the more data about a genotype becomes available and are analysed, the clearer the picture of the distribution of that genotype in its host populations becomes. If this new data was to be analysed using the Association Index the results may show segregation to a greater extent than that which was obtained in this study.

Genotype G adds further confusion to the problem. It appears to be an Old World genotype of HBV with a very low level of diversity, suggesting a recent origin. The

geographical distance between the only two areas, Europe and The United States of America, where it has been isolated is substantial, yet the genetic distance is not. Modern human travelling habits may be involved in the distribution of this new genotype. As more data becomes available the origin of this genotype may become clearer.

The continuing evolution of HBV appears to be driven by a combination of host, viral and possibly also environmental factors, and discovering the differences between genotypes of different regions will be important for the development of new treatment and prevention strategies, as well as in furthering the understanding of the different disease outcomes seen in HBV infection. Epitope specificities of different genotypes and the same genotype in different regions may differ. This would mean that a novel drug developed using the prevalent HBV type in a particular geographical area may not be as efficient in halting viral replication when used to treat a different genotype or the same genotype in a different part of the world. These differences in epitopes may also help explain why a particular genotype is more pathogenic than another, although host factors such as HLA types need to be taken into consideration here as well.

Within human populations differences can be seen in the severity of disease that can be correlated to the specific genotypes responsible for infection. Differences can also be seen in the severity of disease caused by a specific genotype in different countries (Sugauchi et al., 2002b; Sugauchi et al., 2003b). The geographically defined subgroupings within genotypes may play an important role in explaining these

differences. Molecular characterisation of the complete genomes and amino acid sequences of these variants, together with investigations into any differences in host factors will be important for any future research into HBV pathogenesis and prevention. There is no sequence evidence to date to suggest cross-species transmission from primates to humans. So far there have been no reports of primate-like sequences being isolated from humans, even where they have been in contact with primates (Noppornpanth et al., 2003). This will be discussed further in chapter 5.

AI analysis was used here to investigate the existence of geographically defined subgroups within the human HBV genotypes. This method can also be applied to investigate prevalences of particular variants in certain patient groups, such as those with hepatocellular carcinoma, or to compare the variants isolated from e antigen positive versus e antigen negative individuals. Furthering the application of AI analysis will provide further clues as to the evolutionary pressures put on the HBV genome by a variety of host factors.

The distribution of HBV genotypes globally provides clues about the transmission in the past and origins of this virus in human populations. The understanding of the current epidemiology of HBV infection and its future spread can benefit from this type of information.

# **Chapter 4**

## **Recombination Between HBV**

### **Variants of Different Genotypes**

## 4.1 Introduction

For viruses to evolve, genetic variation needs to be generated. This can occur via a number of mechanisms; for example the accumulation of point mutations, reassortment and recombination. All these may accelerate the rate of evolution of a virus. Mutations of the viral genome may be driven by a positive pressure by the immune system or a treatment regime. This type of pressure may also lead to the selection of recombinant genomes over wild-type genomes during viral replication. Apart from driving evolution, recombination may also influence the interpretation of phylogenetic data and hence the estimation of evolution of a virus. The presence of recombination causes the length of terminal branches, as well as the total branch length, to increase. This will cause the estimation of the time to the closest, most recent ancestor to be underestimated and lead to false evolutionary histories being proposed (Schierup & Hein, 2000). It is therefore very important to be aware of recombination when investigating the evolutionary history of a virus. This section will concentrate on recombination as a mechanism of evolution.

For recombination to take place a number of requirements need to be fulfilled. Firstly, the virus must exist as more than one genotype. Secondly, more than one genotype needs to be present in the population in a specific geographical area. Thirdly, super-infection of one individual with viruses representing different genotypes must be biologically possible and occur. And finally, the replication mechanisms employed by the virus must contain steps at which recombination events are possible.

Recombination between viruses belonging to different genotypes has been documented, for example for HIV-1 (Lole et al., 1999), another parallel that can be

drawn between HIV and HBV infection. A large number of HBV sequences have been published as being recombinants between HBV variants of different genotypes, most notably between A and D, and B and C (Bowyer and Sim, 2000; Fares and Holmes, 2002; Morozov et al., 2000; Sugauchi et al., 2002b; Tran et al., 2003; Yuasa et al., 2000), but also different combinations in fewer cases (Bollyky et al., 1996; Cui et al., 2002; Hannoun et al., 2000b). There are even reports that suggest differences in pathogenicity between HBV variants with or without recombinant elements in their genomes (Sugauchi et al., 2002b; Sugauchi et al., 2003b).

The first here criteria for recombination to take place are fulfilled by HBV; there are 8 different human genotypes, some genotypes can be found in the same areas, for example B and C in Asia and A and D in Europe. Recombination between the particular genotypes mentioned above is as expected the most commonly reported. Mixed genotype infections can be found in certain patient groups, such as intravenous drug users (Chen et al., 2004a; Chen et al., 2004b) and those treated with interferon (Hannoun et al., 2002). The mechanisms involved in HBV recombination events are unknown, but it is clear that is homologous recombination with a specific part of the genome being replaced with that from another genotype. One possibility is that there is a swapping of template during the replication of the viral genome. This is known to occur during the reverse transcription stage of HIV-1 replication (Hu et al., 2003; Negroni & Buc, 2001). HBV also undergoes a reverse transcription stage to produce mature virions, and if more than one template from different genotypes is present, template switching could occur. The integration of HBV DNA fragments into the host genome may also play a role.

The strategies used for investigating the phenomenon of intergenotype recombination in HBV have mainly been based on phylogenetics, and more recently bootscanning methods such as SimPlot (Lole et al., 1999).

Methods such as phylogenetics and bootscanning do have some problems which may lead to the generation of false positive recombination results. Within a viral genome phylogenetically uninformative regions may occur. These are regions where positive selection pressures are affecting the genome. One example of this would be point mutation in known epitopes leading to immune escape mutants, although this is not the immune system driving mutation, but the virus mutating to avoid the immune system. This would lead to the emergence of viruses from different genotypes with genetically highly conserved regions. When performing phylogenetic analysis on such strains, they would appear to belong to one genotype in the region affected, whereas the rest of the genome is phylogenetically most closely related to another genotype. One region of the HBV genome where this may occur is the PreC/C region, where stop codon mutations are known to occur leading to a halt in the production of e antigen.

Another problem that affects bootscanning methods is the existence of sequences that show no affiliation with a specific genotype when analysed phylogenetically, appearing as an outgroup to a genotype rather than a member of it. Bootscanning will include this sequence in the most closely related genotype group, i.e. the one it forms an outgroup for, and this may result in a false recombination result.

In this study a novel bootscanning method was used to re-analyse the previously published recombinant HBV sequences. This method is based on the calculation of association scores for phylogenetic groups, an approach which helps resolve many of

the uncertainties and difficulties of interpretation of results arising from conventional methods such as SimPlot (Wang et al., 2001). The recombination analysis carried out will help to get a better picture of the evolutionary pressure recombination might be putting on HBV, and what effect this might have on the epidemiology and pathogenicity of HBV infection.

## 4.2 Materials and Methods

### 4.2.1 Sequences

The accession numbers of the previously published sequences that were used in this analysis can be found in table 4-1. A number of control sequences were also analysed (table 4-2).

ACCESSION NUMBER	RECOMBINATION BETWEEN	ACCESSION NUMBER	RECOMBINATION BETWEEN
HBVDNA	A/D	AB073824	B/C
HBVP4PCXX	A/D	AB073825	B/C
HBVP5PCXX	A/D	AB073826	B/C
HBVP6PCXX	A/D	AB073827	B/C
HBVAYWC1	A/D	AB073828	B/C
HBVAYWE	A/D	AB073829	B/C
AF297619	A/D	AB073830	B/C
AF297620	A/D	AB073831	B/C
AB031265	B/C	AB073832	B/C
AF121243	B/C	AB073833	B/C
AF121245	B/C	AB073834	B/C
AB100695	B/C	AB073835	B/C
AF100308	B/C	AB073836	B/C
AF100309	B/C	AB073837	B/C
HPBADW2	B/C	AB073839	B/C
HPBADW3	B/C	AB073840	B/C
HPBADWZ	B/C	AB073841	B/C
HBVP4CSX	B/C	HPBADW1	A/B
HBVP6CSX	B/C	AY057948	C/D
HBVCGWITY	B/C	AF241407	A/C
AB073821	B/C	AF241408	A/C
AB073822	B/C	AF241409	A/C
AB073823	B/C		

Table 4-1. Accession numbers and reported genotype recombinations of sequences re-analysed for recombination.

ACCESSION NUMBER	GENOTYPE	ACCESSION NUMBER	GENOTYPE
AY057947	C	AF241410	C
AB031262	C	AF241411	C
AB031266	B	AB073838	B
AB031267	B		

Table 4-2. Accession numbers of sequences used as controls for recombination analysis. Sequences were chosen from studies with reported recombinants.

#### 4.2.2 Association Index bootscanning

The AI method (described in chapter 2) can also be used to look for recombination in HBV genomes. This function of the program utilises the PHYLIP program package (Felsenstein, 1993). Ten representative sequences from each of the HBV genotypes (except genotypes E and H) reported not to be recombinants were selected and used in all recombination analyses (table 4-3). Each genotype was given a different tag, producing the groups required by the software to carry out the analysis. 500bp windows and 100bp overlapping steps were selected for the analysis with 100 bootstrap replications. The resulting data was then plotted onto a graph using Microsoft Excel and compared to the data generated by the SimPlot analysis. A recombination event was determined to have occurred when these two analyses both produced cross-over points in the graphs. Control sequences reported as non-recombinants were analysed using this method. The results can be found in appendix D.

The association index method was also used to scan the HBV genome for the presence of phylogenetically uninformative regions. This was done by making pair wise comparisons of the different genotypes along the entire genome. 500 bp windows were used, with a 100bp overlapping step.

ACCESSION NUMBER	GENOTYPE	ACCESSION NUMBER	GENOTYPE
AF090838	A	AF223965	F
AF090839	A	X69798	F
S50225	A	X75663	F
AF297624	A	X75658	F
AF297625	A	AB056513	G
E00010	A	AB056514	G
AJ344115	A	AB056515	G
L13994	A	AB056516	G
X51970	A	AB064310	G
U87742	A	AB064311	G
AY206391	B	AB064312	G
AF461360	B	AB064313	G
D23677	B	AF160501	G
D23678	B	AF405706	G
AB010292	B	AY090454	H
AF282918	B	AY090457	H
AY033073	B	AY090460	H
D50522	B	AB032432	CHIMPANZEE
HBV131133	B	AB032433	CHIMPANZEE
D23679	B	AF222322	CHIMPANZEE
AB046910	C	AF222323	CHIMPANZEE
AF363961	C	AF242585	CHIMPANZEE
AF233236	C	AF242586	CHIMPANZEE
AF330110	C	AJ131567	GORILLA
AF411408	C	D00220	CHIMPANZEE
AF411412	C	AF498266	CHIMPANZEE
AF461359	C	AF305327	CHIMPANZEE
X75665	C	AY330912	CHIMPANZEE
M12906	C	AY330911	CHIMPANZEE
D00630	C	AB037927	GIBBON
AF121240	D	AB037928	GIBBON
AB090270	D	AY077735	GIBBON
X65257	D	AY077736	GIBBON
Z35716	D	AJ131568	GIBBON
X72702	D	AJ131569	GIBBON
AJ344116	D	AJ131571	GIBBON
U95551	D	AJ131572	GIBBON
AB033559	D	AJ131573	GIBBON
AF280817	D	AJ131574	GIBBON
Y07587	D	AJ131575	GIBBON
AB032431	E	U46935	GIBBON
X75657	E	AF193864	ORANGUTAN
X75664	E	NC_002168	ORANGUTAN
AB036905	F	AY330913	GIBBON
AB036908	F	AY330914	GIBBON
AB036910	F	AY330916	GIBBON
AB036913	F	AY330917	GIBBON
AB036916	F	AF046996	WOOLLY MONKEY
AF223962	F		

Table 4-3. Accession numbers of non-recombinant sequences used in the phylogenetic analysis, AI analysis, and SimPlot analysis.

#### 4.2.3 Bootscanning using SimPlot

The SimPlot program was also developed to investigate HIV-1 intersubtype recombination (Lole et al., 1999). It uses PHYLIP programs to do its calculations, as does the AI method described above. SimPlot bootscanning analyses a query sequence against pre-defined groups of sequences as in AI, and the output is a line graph comparable to that obtainable from the AI results. Sequences were tagged into the groups as defined by genotype in SIMMONIC and then exported in FASTA format. A SimPlot description line was added to the FASTA file as described in the documentation that comes with the program, and this file was then opened up in SimPlot. The bootscanning option was then selected, the query sequence selected, and the parameters selected. To allow comparison between the AI and SimPlot results 500bp windows with 100bp overlapping steps were selected together with neighbour-joining and 100 bootstrap replications. The sequences used for the control groups were those that were used for the AI analysis described above. Control non-recombinant sequences were analysed with this method. Results can be found in appendix D.

#### 4.2.4 Phylogenetic analysis

Example isolates where AI and SimPlot produced different results were analysed phylogenetically as described below.

Sequences that appeared recombinant in the both the AI and SimPlot bootscanning analyses were also analysed phylogenetically. The fragment of the genome up to the cross-over point, the fragment containing the cross-over region, and the fragment of the genome after the cross-over region were selected and phylogenetic analysis

carried out on these. The MEGA2 program was used for the phylogenetic analysis with Jukes-Cantor model, neighbour-joining and 100 bootstrap replications. The sequences used for the phylogenetic analysis of possible recombinants were the same as were used in the AI and SimPlot analyses (table 4-3). The genotype of the sequence was recorded for each segment. The recombinant status of each sequence analysed was then assessed. The suitability of each of the methods used to investigate recombination was then assessed by comparing the results.

## 4.3 Results

### 4.3.1 Phylogenetic segregation of genotypes

A pair wise comparison of the HBV genotypes identifies a major phylogenetically uninformative region (Fig. 4-1). This is found in the region of the HBV genome containing the Pol/X overlap and PreC/C. Genotypes B and C do not segregate fully in the region between bases 1101 and 2400, and genotypes D and E do not segregate fully in the region between bases 1301 and 2600. This implies that these two genotype pairs are under specific selection pressures in this region. PreC/C mutations to evade the immune system may be responsible for this. Genotypes D and E consistently group together in phylogenetic analysis, suggesting a shared evolutionary history, and the fact that their genomes are highly similar in this particular region supports this as well.

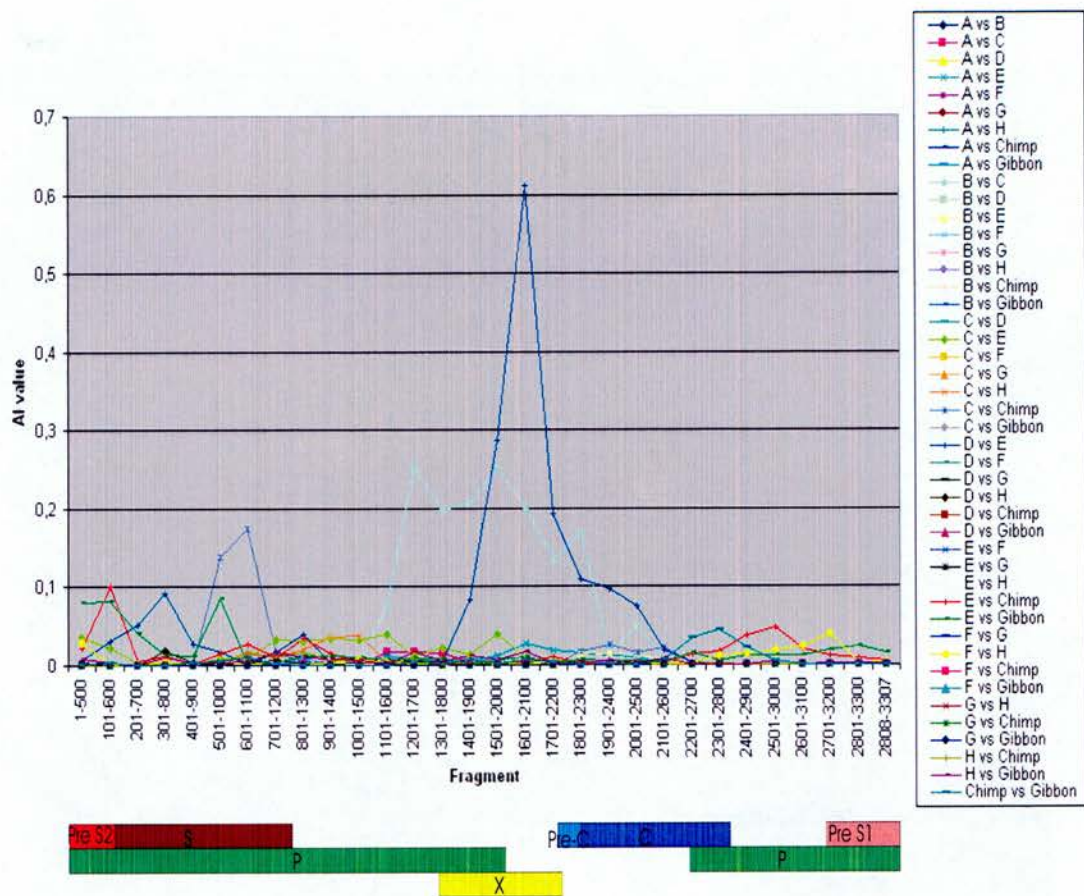


Figure 4-1. Binary AI scan comparisons of the genetic segregation between the human and non-human primate HBV genotypes. A simplified map of the HBV genome is shown below the AI graph to show the segregation across the genome.

#### 4.3.2 Bootscanning and phylogenetics results

Association Index bootscanning results and SimPlot results for each sequence analysed are shown on the same page to allow comparison. A simplified map of the HBV genome has been inserted between the two graphs to enable references to be made to the regions involved in recombination. In these analyses many discrepancies can be seen between the two different methods used. Three control sequences reported to belong to specific genotypes were also tested (Figures 4-2 to 4-4). Of these, the reported genotype B isolate AB073838 (Figure 4-3) appears recombinant using SimPlot, suggesting this method may overestimate recombination. The results of the recombination analyses are shown in tables 4-4 and 4-5.

ACCESSION NUMBER	RECOMBINATION BETWEEN	ANALYSIS METHOD		RECOMBINANT
		AI	SIMPLOT	
HBVDNA	A/D	R	R	Y
HBVP4PCXX	A/D	N	N	
HBVP5PCXX	A/D	N	N	
HBVP6PCXX	A/D	N	R	
HBVAYWC1	A/D	R	R	Y
HBVAYWE	A/D	R	R	Y
AF297619	A/D	N	R	
AF297620	A/D	R	R	Y
AB031265	B/C	R	R	Y
AF121243	B/C	N	R	
AF121245	B/C	N	R	
AB100695	B/C	N	R	
AF100308	B/C	N	R	
AF100309	B/C	N	R	
HPBADW2	B/C	N	R	
HPBADW3	B/C	N	R	
HPBADWZ	B/C	N	R	
HBVP4CSX	B/C	N	R	
HBVP6CSX	B/C	N	R	
HBVCGWITY	B/C	N	R	
AB073821	B/C	N	R	
AB073822	B/C	N	R	
AB073823	B/C	N	R	
AB073824	B/C	N	R	
AB073825	B/C	R	R	Y
AB073826	B/C	N	R	
AB073827	B/C	N	R	
AB073828	B/C	N	R	
AB073829	B/C	N	R	
AB073830	B/C	N	R	
AB073831	B/C	N	R	
AB073832	B/C	N	R	
AB073833	B/C	N	R	
AB073834	B/C	N	R	
AB073835	B/C	N	R	
AB073836	B/C	N	R	
AB073837	B/C	N	R	
AB073839	B/C	N	R	
AB073840	B/C	N	R	
AB073841	B/C	N	R	
HPBADW1	A/B	N	R	
AY057948	C/D	R	R	Y
AF241407	A/C	R	R	Y
AF241408	A/C	R	R	Y
AF241409	A/C	R	R	Y
<b>TOTAL</b>	<b>45</b>	<b>10R</b> <b>35N</b>	<b>43R</b> <b>2N</b>	<b>10</b>

Table 4-4. Results of recombination analysis carried out using AI and SimPlot. N= non-recombinant, R= recombinant, Y= determined recombinant. Highlighted sequences (AF241407, AF241408, and AF241409) are determined recombinants, though not A/C recombinants as reported earlier.

		<b>Association Index</b>	
		<b>Recombinant (R)</b>	<b>Non-recombinant (N)</b>
<b>SimPlot</b>	<b>R</b>	<b>10</b>	<b>33</b>
	<b>N</b>	<b>0</b>	<b>2</b>

Table 4-5. Summary of recombination analyses highlighting the discrepancies between the two methods used in this study. SimPlot overestimates the number of recombinant isolates when compared to Association index analysis.

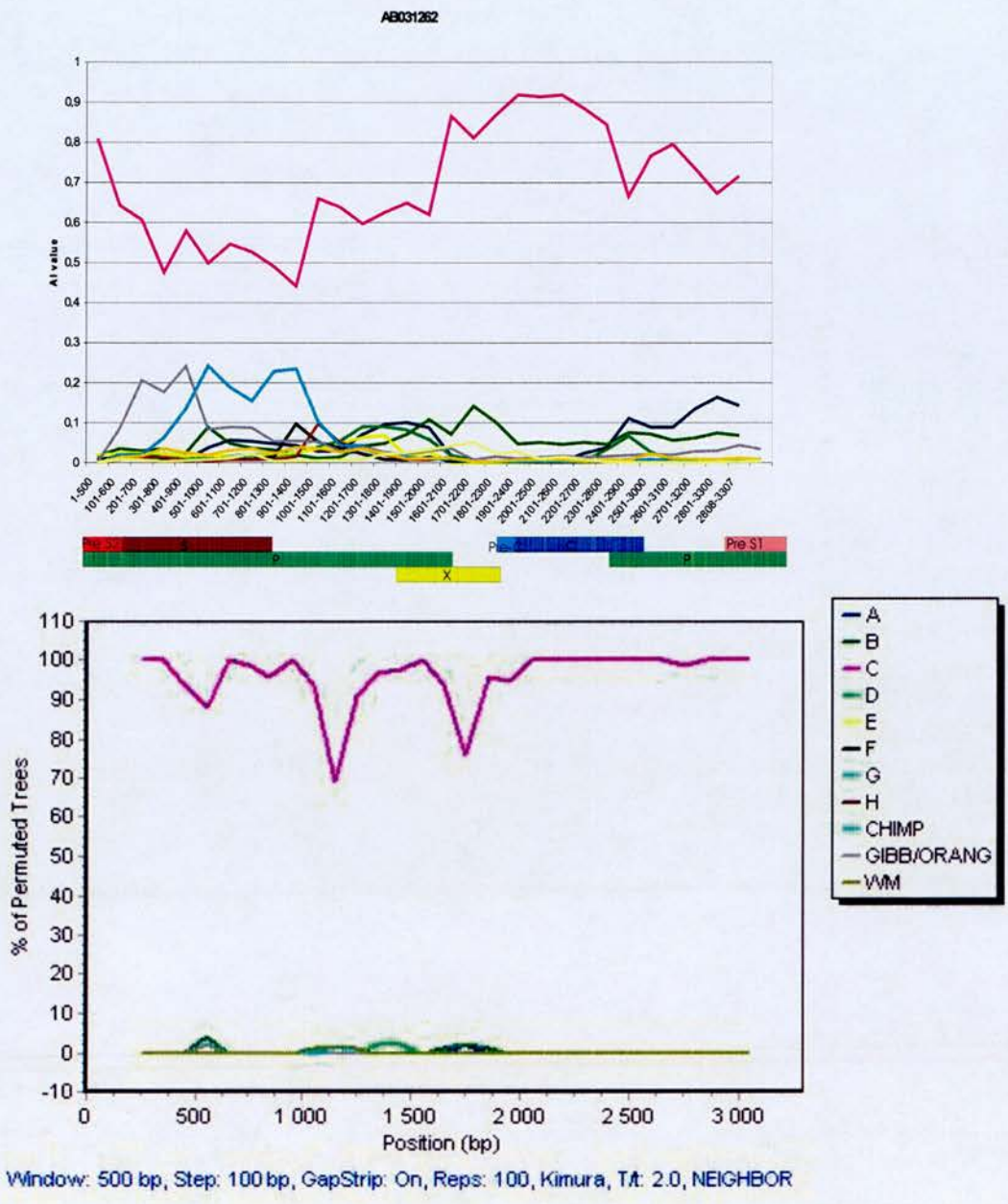


Figure 4-2. Recombination analysis of HBV isolate AB031262 (Yuasa et al., 2000) using the AI (top) and SimPlot (bottom) methods. Isolate was reported to be of genotype C, both analyses supports this.

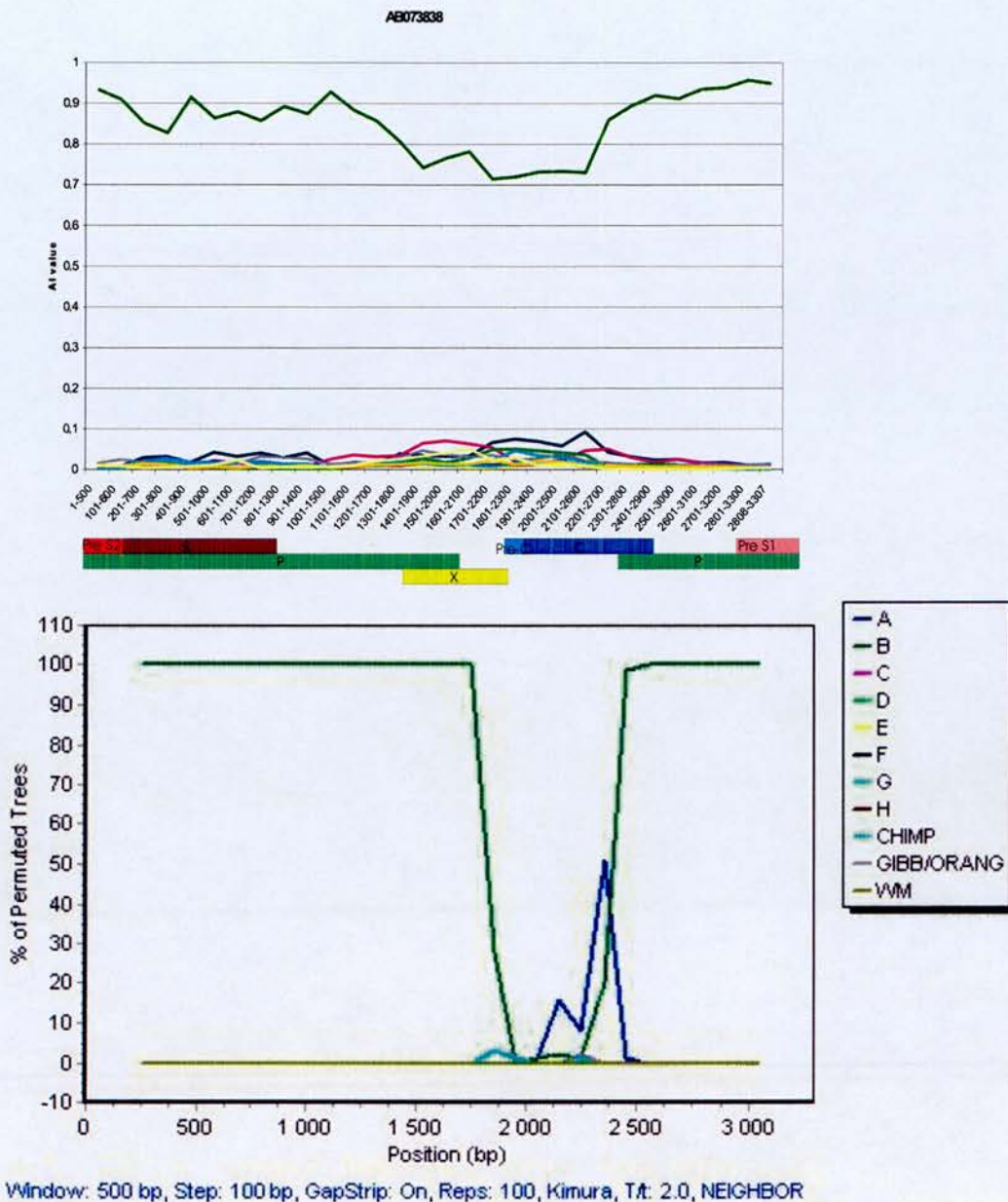


Figure 4-3. Recombination analysis of HBV isolate AB073838 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Isolate was reported to be a genotype B, AI analysis supports this whereas SimPlot shows the isolate to be recombinant in the core region.

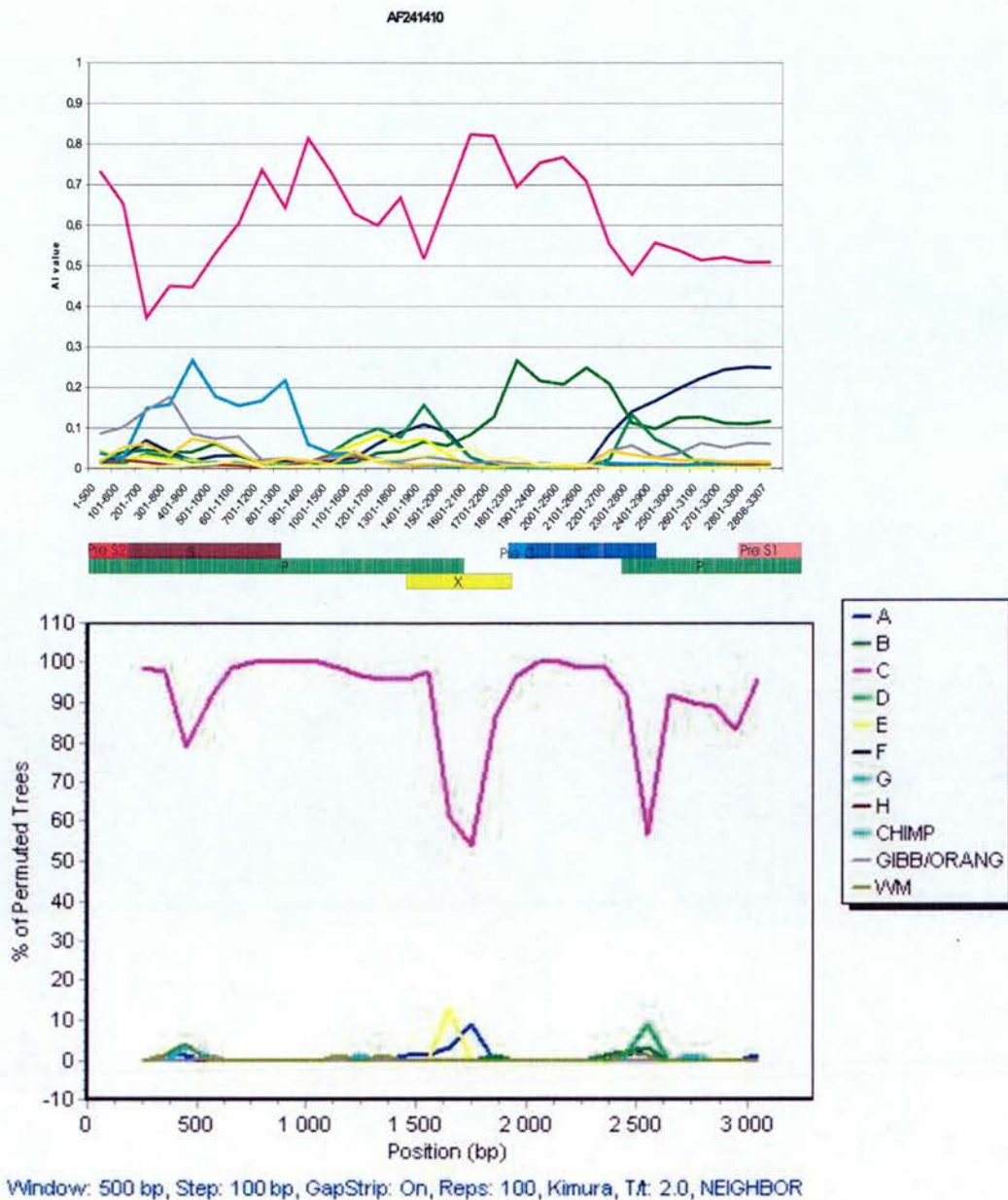


Figure 4-4. Recombination analysis of HBV isolate AF241410 (Hannoun et al., 2000) using the AI (top) and SimPlot (bottom) methods. Isolate was reported to be of genotype C, both analyses support this.

Reported recombinants fall into two main groups; recombinants and non-recombinants. Sequences were deemed to be recombinant when both AI scanning and SimPlot bootscanning showed the sequences to be recombinant. Graphical depictions of the analyses of 3 of the 10 isolates found to be recombinant can be found below (Figures 4-5, 4-7, 4-9), and the remaining 7 can be found in appendix B. 7 of the 10 recombinant isolates were confirmed to be recombinants between the genotypes they were initially reported to be recombinants of. 3 isolates showed different results. Isolates AF241407, AF241408, AF241409 (Hannoun et al., 2000b) were reported to be recombinants between genotypes A and C. The analyses carried out for this study shows that these isolates are recombinants between genotype C and a possible new undiscovered genotype of HBV, showing characteristics both of genotype A and G. Genotype G had not been discovered when these isolates were published, and they would have appeared recombinant with genotype A in SimPlot analysis. SimPlot analysis carried out with genotype G included, as was done in all analyses for this study, shows these isolates to be closer to genotype G than A. Phylogenetic analysis of recombinant isolates was also carried out. The analyses of isolates HBVAYWC1, AB073825, and AF241408 can be seen below (Figures 4-6, 4-8, 4-10). The remaining phylogenetic analyses can be found in Appendix C.

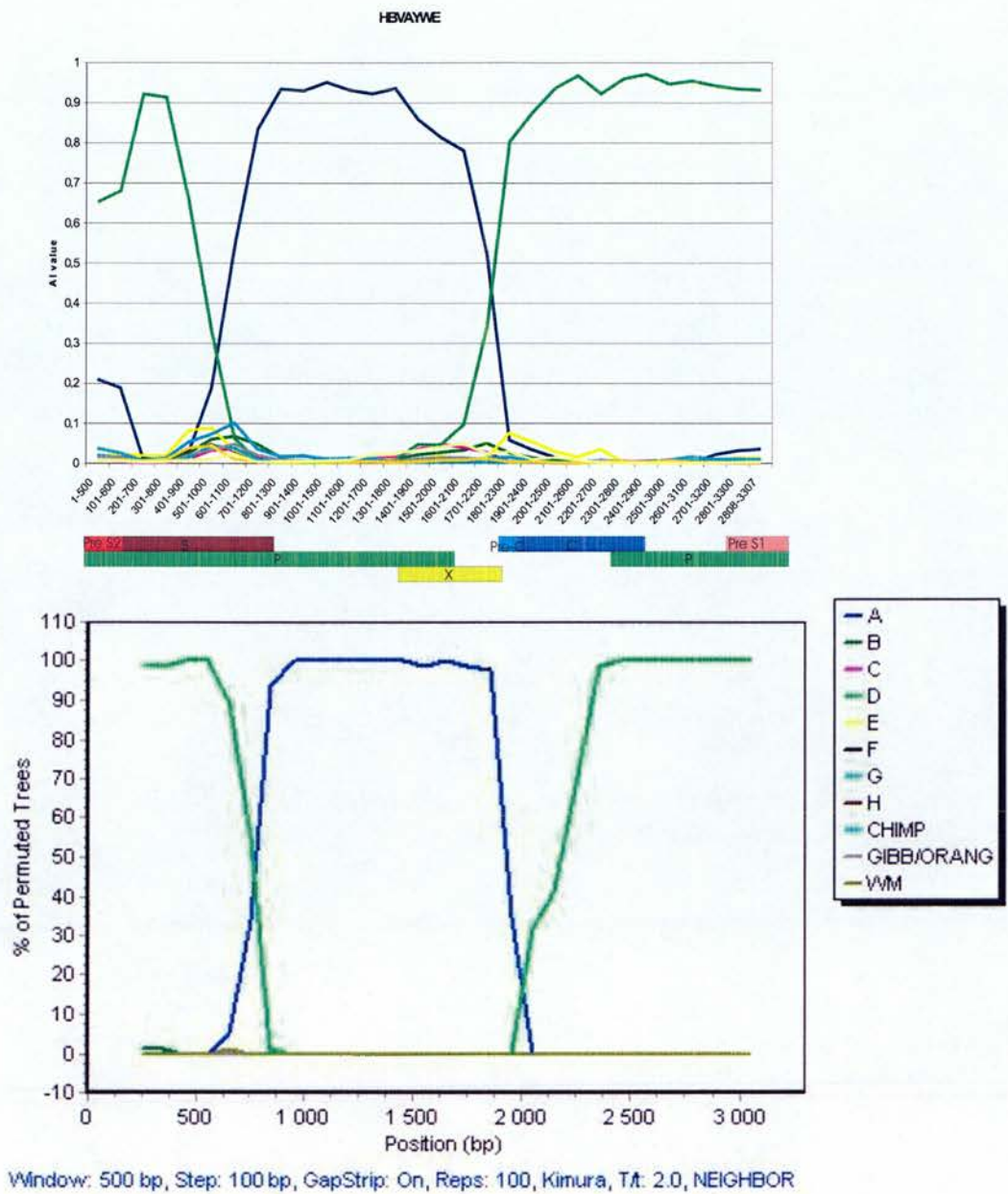


Figure 4-5. Recombination analysis of HBV isolate HBVAYWC1(Bowyer et al., 2000) using the AI (top) and SimPlot (bottom) methods. Interpretation: recombinant using both methods.



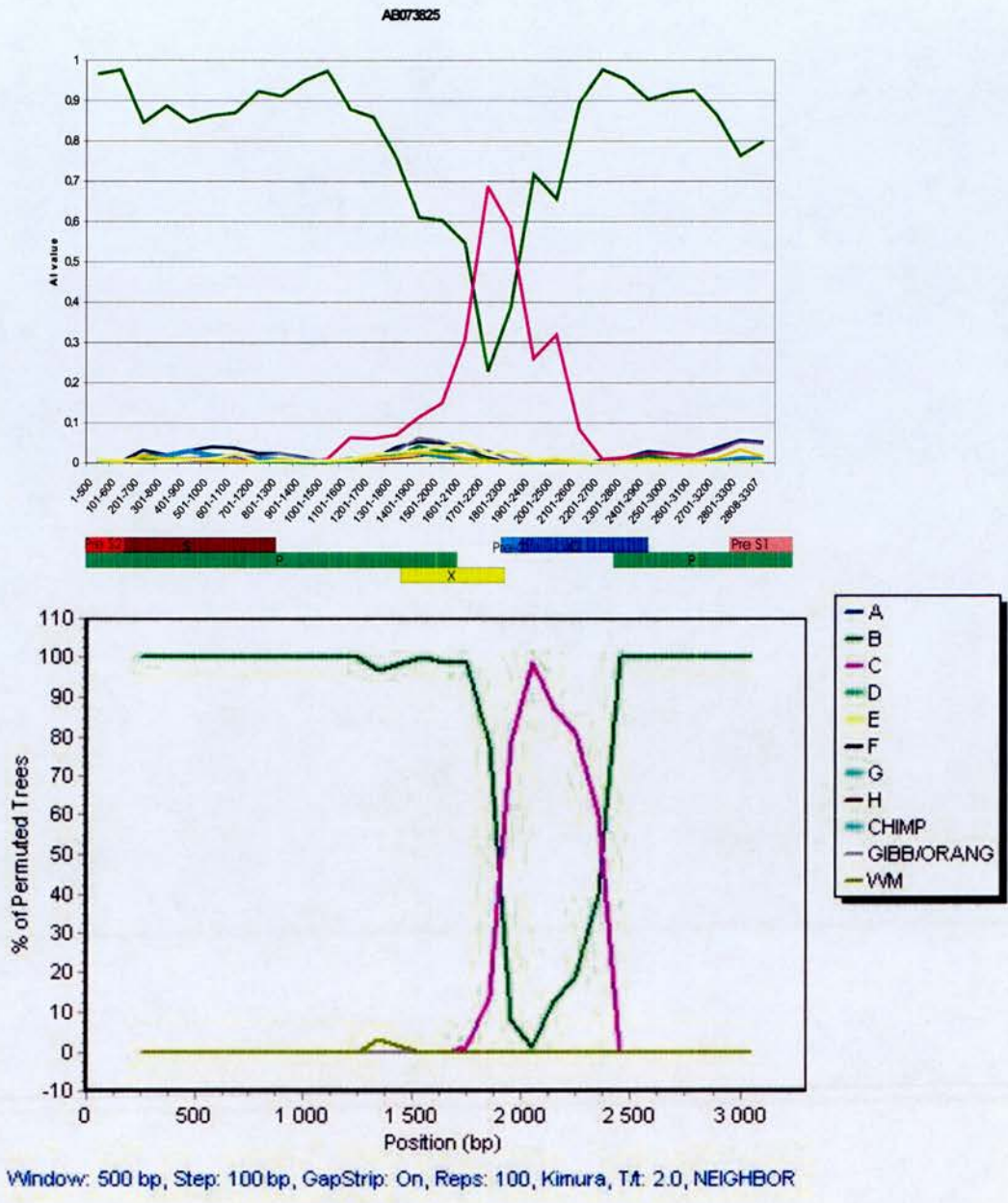


Figure 4-7. Recombination analysis of HBV isolate AB073825 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: recombinant using both methods.

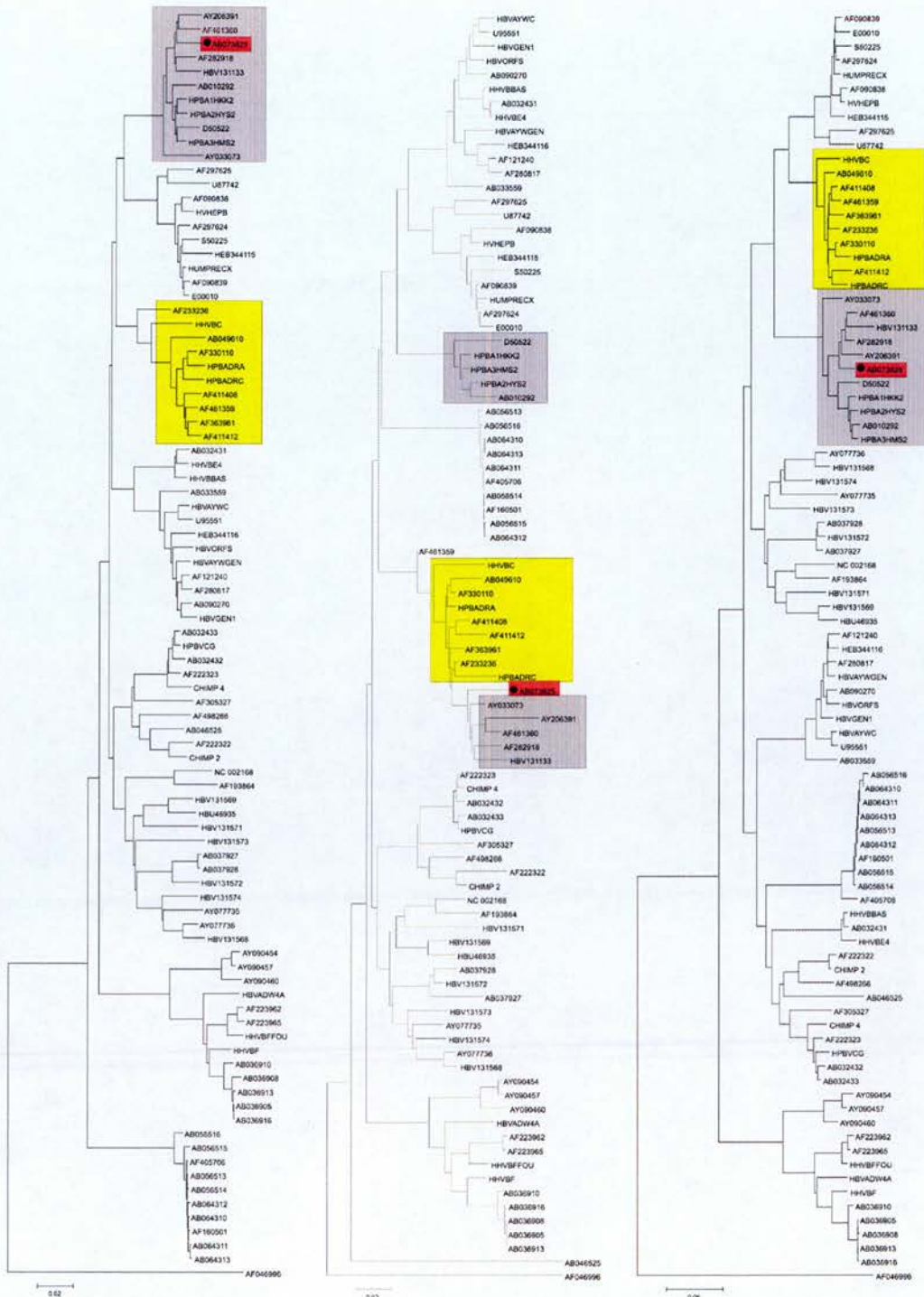


Figure 4-8. Phylogenetic analysis of HBV isolate AB073825, reported to be a B/C recombinant. Analysis does not fully support recombinant result from AI and SimPlot analysis. Phylogenetic analysis of the recombinant fragment shows this to coincide with the uninformative region, and the isolate does not group with either genotype B or C.

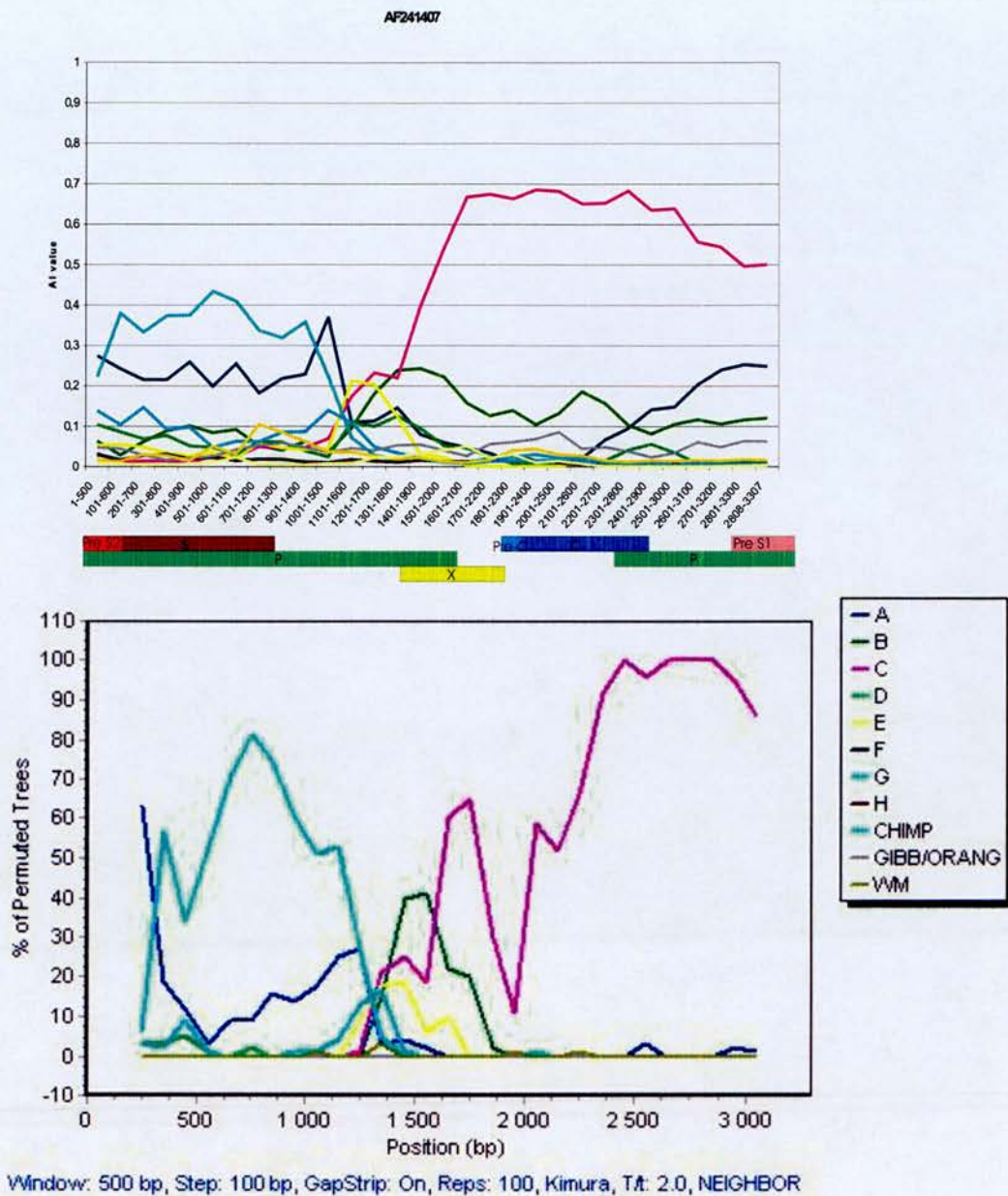


Figure 4-9. Recombination analysis of HBV isolate AF241408 (Hannoun et al., 2000) using the AI (top) and SimPlot (bottom) methods. Interpretation: recombinant using both methods.

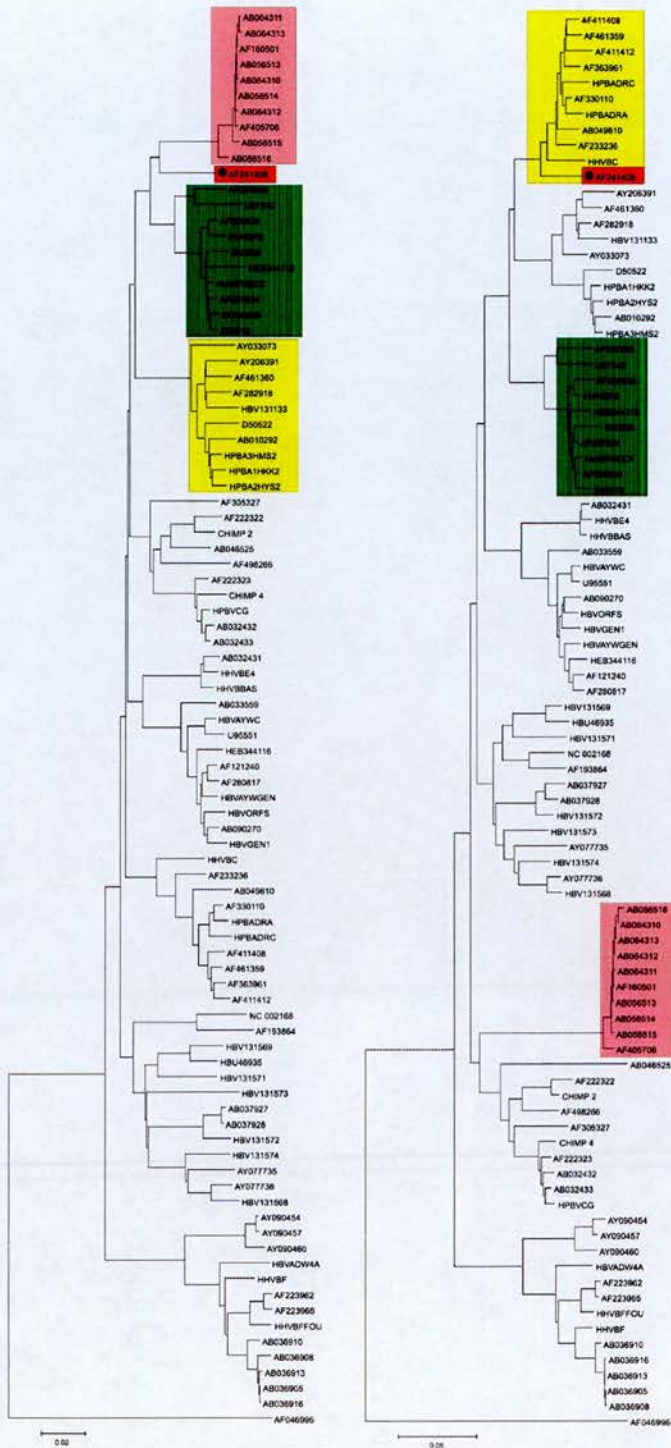


Figure 4-10. Phylogenetic analysis of HBV isolate AF241408, reported to be an A/C recombinant. Analysis supports recombinant result from AI and SimPlot analysis, and shows the isolate does not group with any specific genotype in the recombinant region.

Non-recombinant isolates were those that showed no recombination in the AI analysis and recombination in SimPlot analysis (43 isolates) or no recombination in SimPlot analysis (2 isolates; HBVP4PCXX, HBVP5PCXX). These two isolates were identified using an author-written program (Bowyer & Sim, 2000) and phylogenetic analysis. The problems regarding false positive recombination identification described earlier explain why these isolates were reported as recombinant, and have now been found to be non-recombinant.

HBV isolates HBVP4PCXX and HBVP5PCXX were reported to be A/D recombinants, however, both AI and SimPlot analysis showed this not to be the case. Phylogenetic analysis supports this result. The region involved in the recombination event coincides with the uninformative region described earlier, where genotypes B and C, and D and E fail to segregate. In the tree it can be seen that genotype D is divided into two branches, separated by the genotype E isolates. Isolate HBVP4PCXX can be found on the one of these two genotype D branches that outgroups the other genotype D and E isolates. This could lead to the sequence being judged to be more closely related to the genotype A isolates.

Apart from the two isolates described above, one of the reported genotype A and D recombinants (HBVP6PCXX) was found to be non-recombinant. Isolate HBVP6PCXX has a divergent X region in the AI analysis, and this is consistent with the phylogenetically uninformative region identified in Fig. 4-1.

The majority (30/32) reported genotype B and C recombinants were found to be non-recombinants in the AI analysis. The region seen to be involved in the recombination event from the SimPlot analysis can be found inside the phylogenetically uninformative region described in Fig. 4-1, where genotypes B and C fail to segregate.

The only reported genotype A and B recombinant was found to be a non-recombinant, and again the proposed region of recombination falls within the uninformative region described above. Three examples of discrepant results are shown below (Figures 4-11, 4-13, 4-15). Remaining discrepant results can be found in Appendix D.

Phylogenetic trees were constructed for selected examples where SimPlot and AI showed discrepant results. Trees were constructed to show the region of the genome preceding the reported recombination event, the region containing the recombinant part of the genome, and the region following the recombination region.

Reported genotype B/genotype C recombinants were found to be non-recombinants in the majority of cases. Phylogenetic analysis of selected examples of these recombinants supports these findings (Figures 4-12, 4-14, 4-16). The regions of the genome reported to be recombinant falls within the phylogenetically uninformative region described in Fig. 4-1, where genotype B and C fail to segregate. The phylogenetic trees show this quite clearly. Half of the genotype B isolates included in this analysis group on a separate branch, whereas the other isolates form a branch within the genotype C clade.

The only genotype A/B recombinant reported to date (HPBADW1) was also found to be non-recombinant using the AI analysis. The phylogenetic analysis supports this

finding. In the tree the isolate groups with the five genotype B isolates that group separately from the genotype B/C branch in this region. This smaller genotype B group falls as an outgroup to genotypes A, D and E in this region, and this could explain why SimPlot shows HPBADW1 as belonging to genotype A in this region.

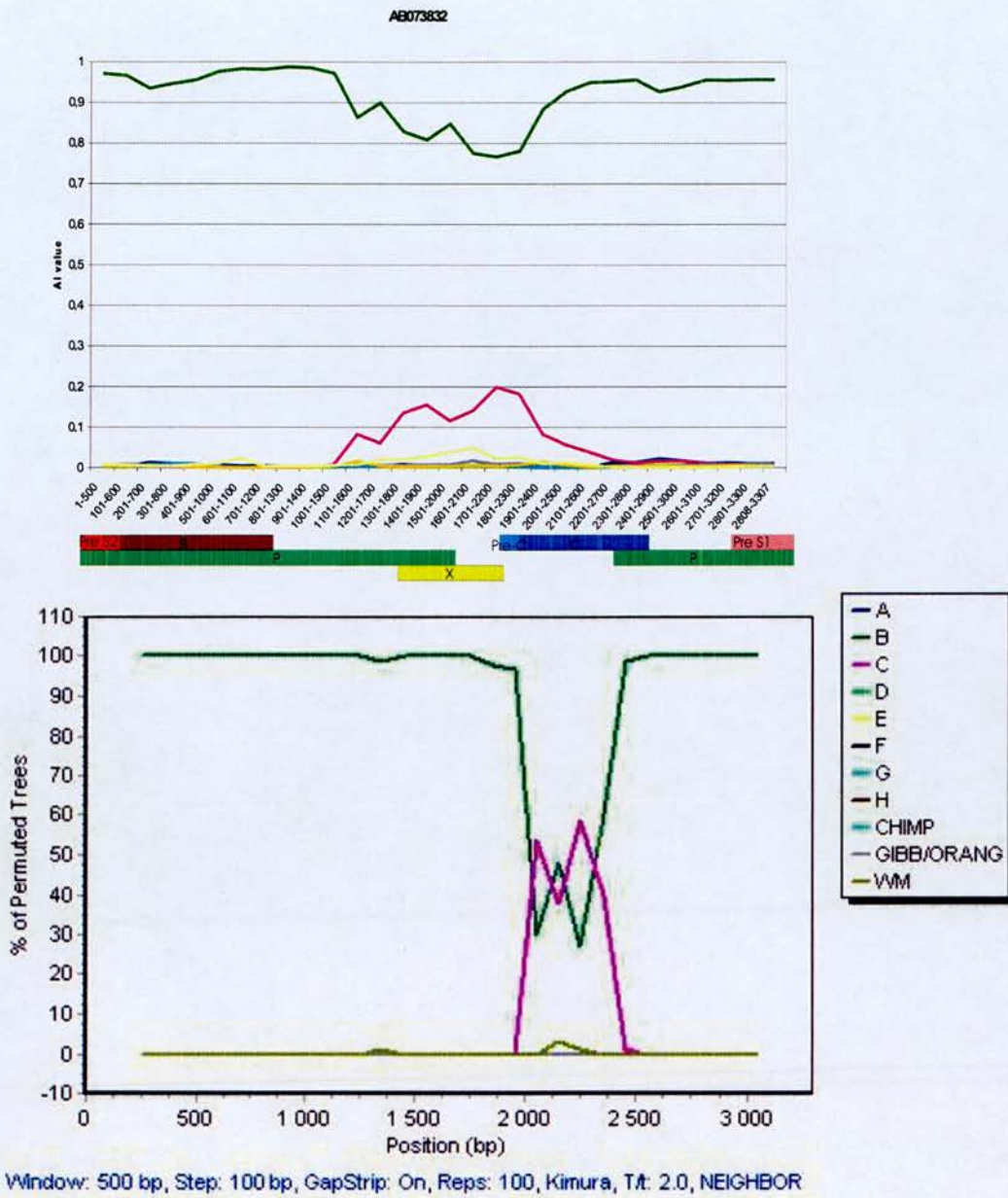


Figure 4-11. Recombination analysis of HBV isolate AB073832 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

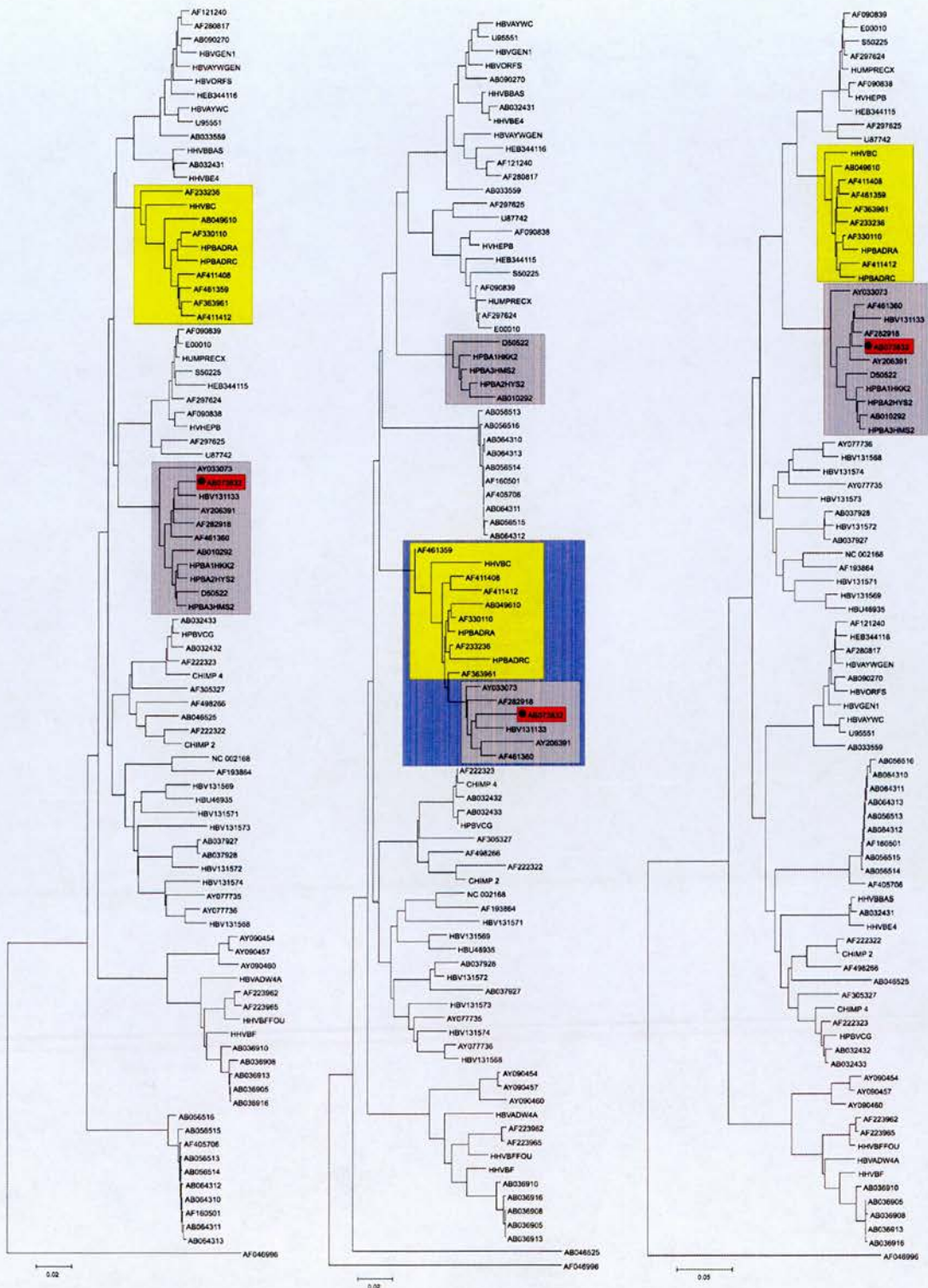


Figure 4-12. Phylogenetic analysis of HBV isolate AB073832, reported to be a B/C recombinant. Genotype B is highlighted in the grey box, genotype C in the yellow box, and the isolate in the red box. The lilac box shows the failure of genotypes B and C to segregate in this region of the genome. Phylogenetic analysis supports the AI non-recombinant result.

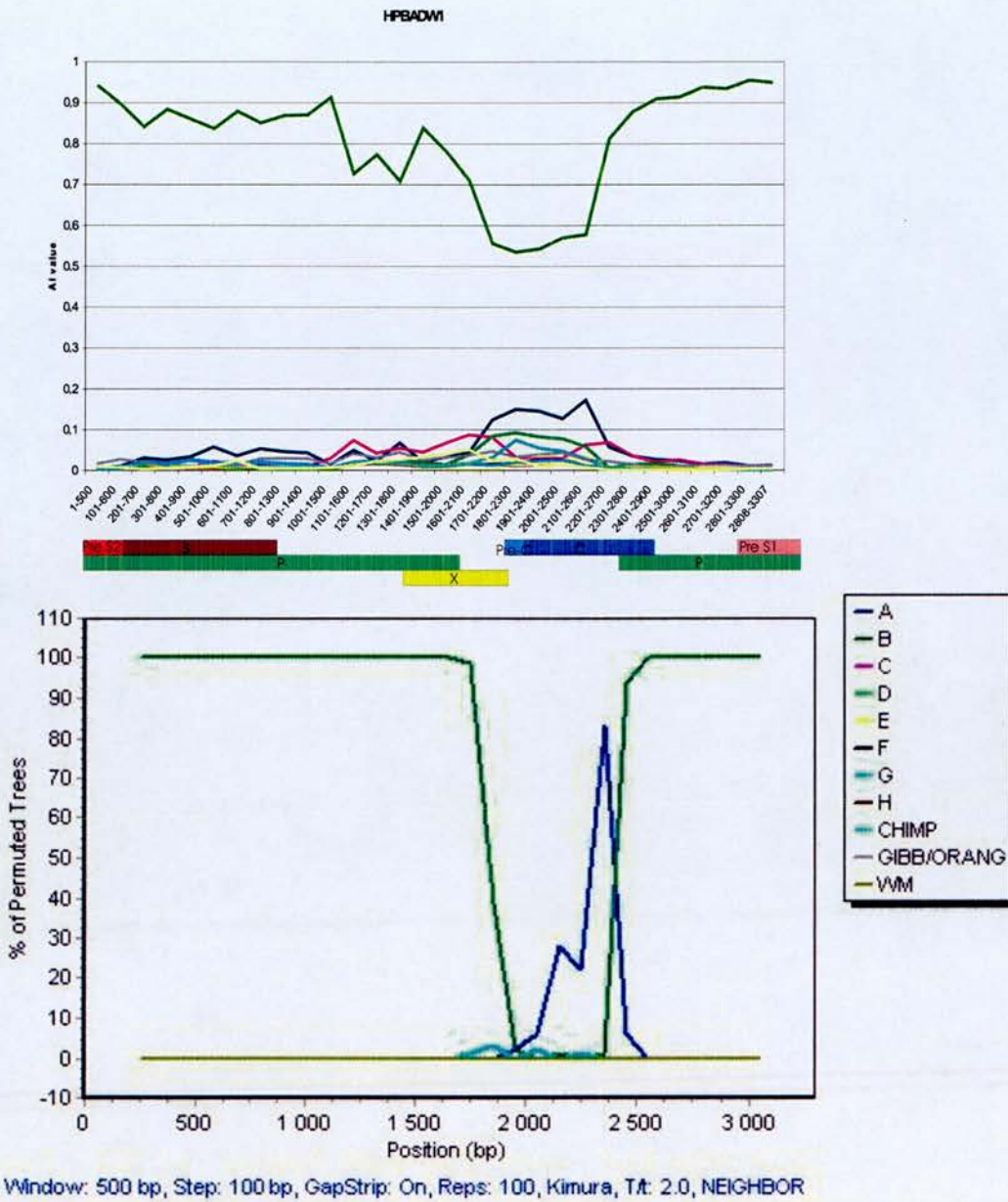


Figure 4-13. Recombination analysis of HBV isolate HPBADW1 (Bollyky et al., 1996) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant with a divergent core region, SimPlot – recombinant in the core region.

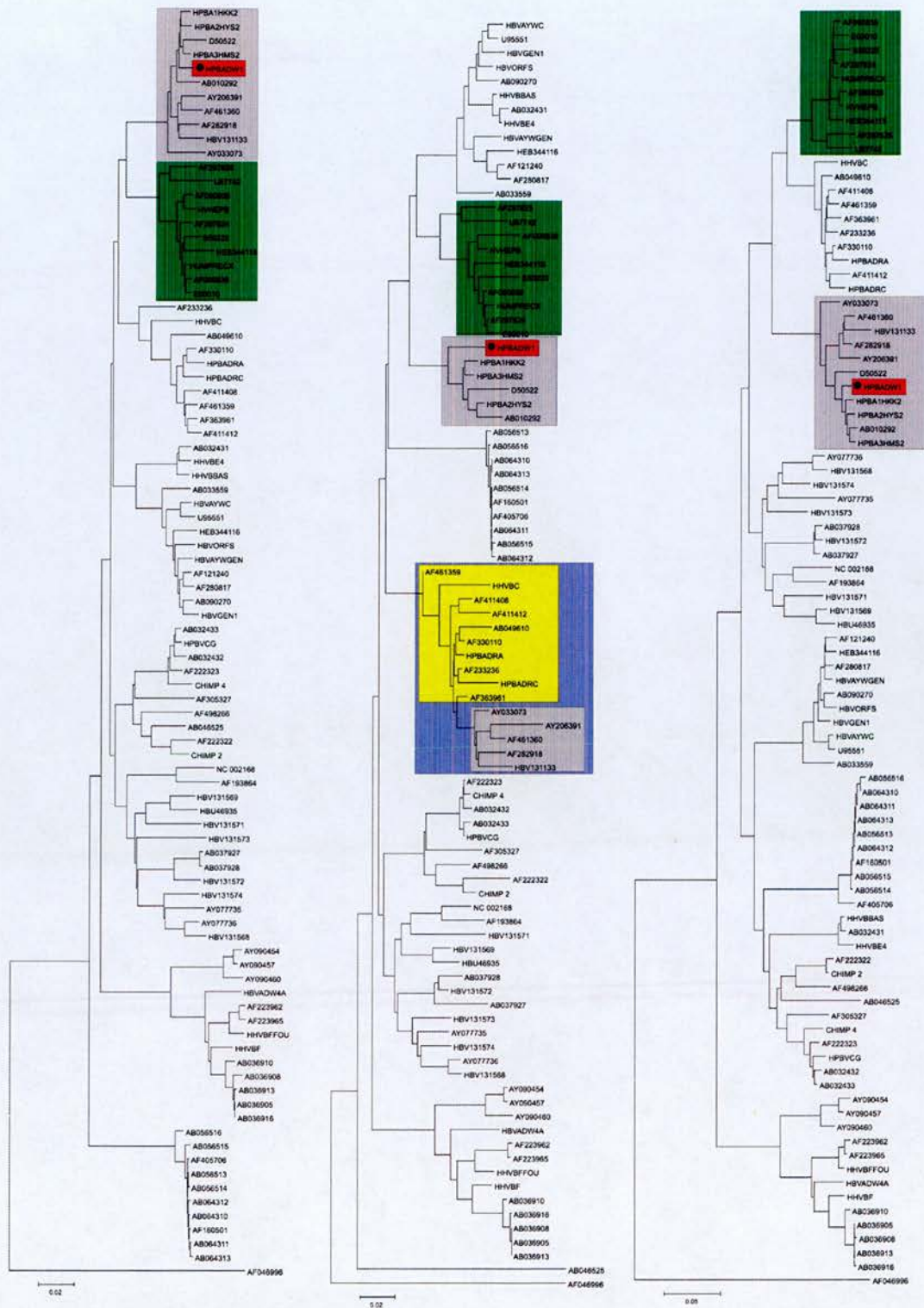


Figure 4-14. Phylogenetic analysis of HBV isolate HPBADW1, reported to be an A/B recombinant. Genotype A is highlighted in the green box, genotype B in the grey box, genotype C in the yellow box, and the isolate in the red box. The lilac box shows the failure of genotypes B and C to segregate in this region of the genome. Phylogenetic analysis supports the AI non-recombinant result.

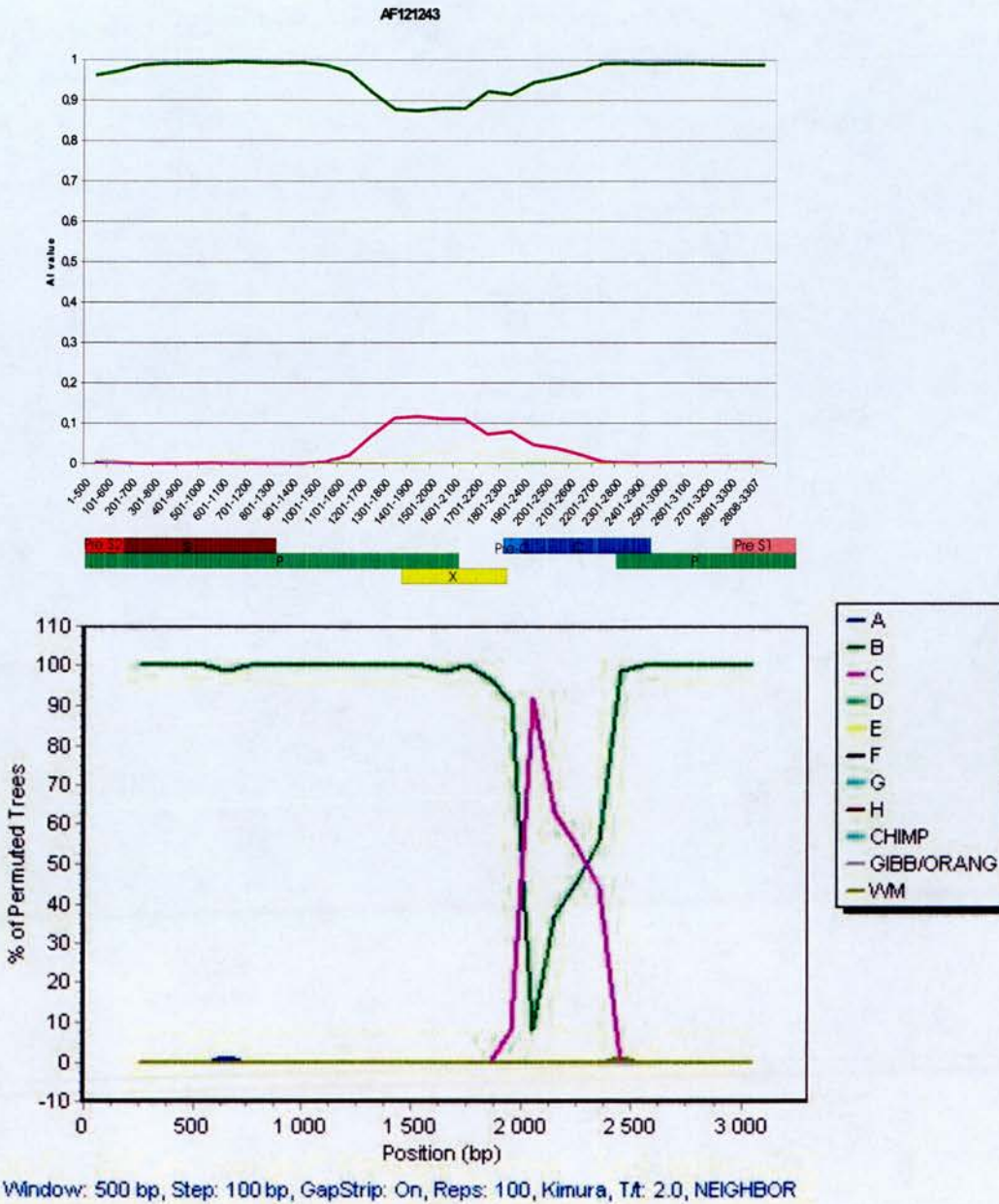


Figure 4-15. Recombination analysis of HBV isolate AF121243 (Hannoun et al., 2000) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

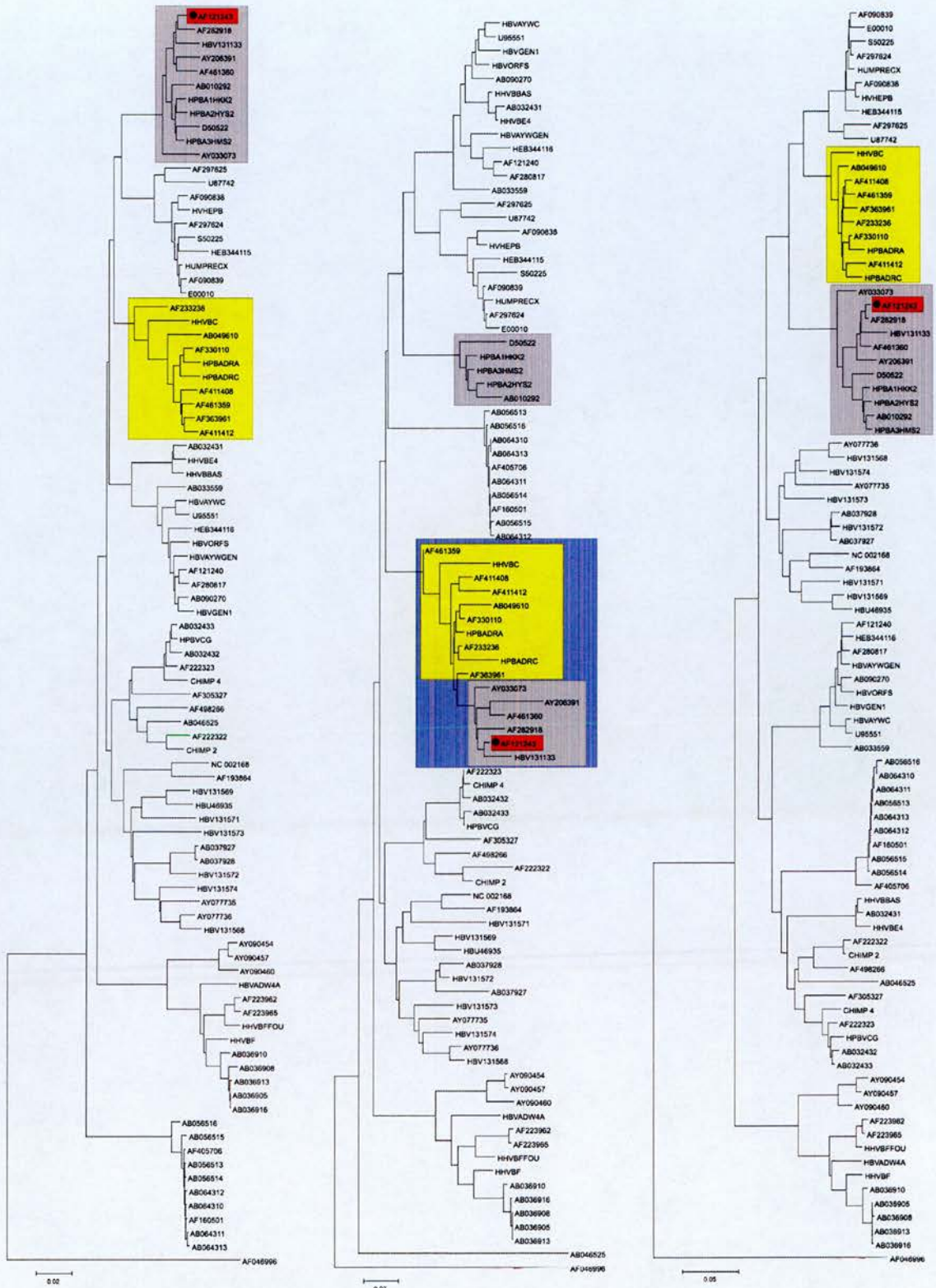


Figure 4-16. Phylogenetic analysis of HBV isolate AF121243, reported to be a B/C recombinant. Genotype B is highlighted in the grey box, genotype C in the yellow box and the isolate in the red box. The lilac box shows the failure of genotypes B and C to segregate in this region of the genome. Phylogenetic analysis supports the AI non-recombinant result.

#### 4.4 Discussion

Recombination does appear to occur in HBV infection although maybe not as frequently as previously reported. Using recombination positive results for both the AI and SimPlot methods (as described in chapter 2) as the criteria for calling an isolate a recombinant, 10 recombinant HBV genomes, out of the 45 reported genomes that were re-analysed, were identified. A/D recombination seems to be more common than any other (4/10 isolates), but as more sequence data becomes published from studies throughout the world this will have to be reviewed.

SimPlot identified a total of 43 recombinants, making the discrepancy between the AI and SimPlot methods very large. AI scanning of the HBV genotypes in pair wise combinations identified a phylogenetically uninformative region, where in particular genotypes B and C, and D and E respectively, fail to segregate. The majority of the isolates identified as recombinants using SimPlot only (30/33 isolates) were reported to be B/C recombinants. In each case the region of the genome reported to be involved in recombination corresponds with this uninformative region. The way SimPlot calculates the relatedness of an isolate to the genotypes leads to the isolates being deemed to be part of the genotype it groups closest to, even though the sequence actually falls outside this group in the phylogenetic analysis.

Phylogenetic analysis confirmed the findings of the AI analysis, both in the identification of recombinants and non-recombinants. Phylogenetic analysis on its own requires interpretation from the researcher, and a sequence forming an outgroup may be taken as belonging to the genotype it groups closest to, as is the case for SimPlot analysis as well, although it should be judged as having no affiliation with any specific genotype.

A/D recombinants and the C/D recombinant have much larger regions of overlap than the B/C recombinants. This suggests that the recombination events have taken place over an extended period of time, as both these genotypes are found widely throughout the world, and a longer time would be required to accumulate the different recombinations seen (Fares and Holmes, 2002). The B/C breakpoints are also all found in non-overlapping regions of the genome, and the fragments involved are always smaller, involving around 500 bp long fragments. Fares (2002) found all B/C recombinants to have originated from one single recent recombination event, due to the close relatedness between these isolates in phylogenetic analysis. A/D recombination points can be found in both overlapping and non-overlapping regions of the genome, and the isolates do not appear to be related, suggesting a longer time-span since the different recombination events that gave rise to these isolates (Fares and Holmes, 2002).

If A/D recombinants represent older recombination events this could explain why they seem to be more frequent than recombination events between other genotypes. A longer period of time will mean that the recombined parts of the genome have had time to become stabilised, and any recombination events affecting the replicative fitness and infectivity of the virus will have been removed from circulating virus populations. If recombination occurs in a region of the genome that has constraints on it, such as the Pol gene (Mizokami et al., 1997) the new genome may be unable to replicate and therefore these will not be produced and potentially spread to new hosts. The genotype B isolates with the reported genotype C recombination, shown in this study to be a divergent core region, have been described as being more pathogenic than non-recombinant genotype B in southeast Asia (Sugauchi et al., 2002b). If they have a recent origin as Fares suggests (Fares and Holmes, 2002) their genomes may not be as stable with breakpoints within epitope regions or other regions important for the replication and infectivity of the virus, and their new genomes may be responsible for this

difference in disease. Genotype B is most prevalent in Southeast Asia, whereas genotype C is more frequent in for example Japan. The disease caused by genotype C in Japan is more severe than genotype C infection in Southeast Asia. Population factors as well as viral factors must be considered when investigating such differences.

Recombination between HBV isolates from different genotypes does occur. The genotypes involved in these recombination events reflect the genotypes in circulation in the countries of origin. A/D recombinants all originate in Europe, where these two are the two most prevalent genotypes, and B/C recombinants originate in Asia. The fact that C/D recombinants are found in Tibet confirms the worldwide distribution seen for genotype D. The existence of some highly divergent HBV strains in Vietnam that do appear recombinant, but do not seem to be recombinant with any of the known human or non-human primate genotypes, hints at the existence of further genotypes of HBV. These genotypes may be infecting humans or other species of primate, in which case cross-species transmission as an evolutionary mechanism for HBV must be considered. Recombination has to be considered an evolutionary force when considering HBV infection. It appears to be occurring currently, based on the B/C recombination events, but has also occurred further back in the evolutionary history of hepatitis B virus. Recombination may be responsible for the creation of the genotypes seen today.

Differences in disease severity may be associated with these recombinant genomes, and needs to be investigated further. Differences in host factors also need to be taken into consideration. Recombination has to be considered as one of the mechanisms of HBV evolution that may become more important in the future. If cross-species transmission of HBV is possible, then

recombination between native and new HBV genomes could create new viruses capable of causing much more severe disease and making treatment and prevention more difficult.

Recombination appears to be an important factor in the evolution of HBV, and this means it is very important that the methods used to investigate the extent of recombination are adequate. The SimPlot method appears to overestimate the amount of recombination, by forcing sequences with low affiliation into the closest genotype. Phylogenetics on its own also appears to be a poor method of choice when looking for recombination, although this may be due to interpretation. Many genomes have crossover points in regions of the viral genome that are uninformative, and may therefore give false positive results. Phylogenetics is most accurate when it is performed over multiple regions, with overlap between them. Otherwise any uninformative regions may go unnoticed. The AI method eliminates the two major problems associated with using SimPlot and/or phylogenetics as recombination detection methods. Sequences with low affiliations will not be artificially assigned a genotype, but will be shown as having no strong association with any particular genotype. Phylogenetically uninformative regions are also identified when using the AI method. All the above points mean that studies such as that by Fares & Holmes (2000) significantly overestimate the amount of recombination that occurs for HBV. This also means that any conclusions drawn regarding the evolutionary history have to be treated with caution. A combination of methods such as AI and phylogenetics will be more sensitive in detecting recombination than either method on its own.

## **Chapter 5**

# **Survey of HBV Prevalence in Non- human Primates**

The infection frequency and phylogenetic data presented in this chapter were published in 2003 (Starkman et al., 2003). This manuscript can be found in appendix F.

## 5.1 Introduction

Since the discovery of HBV, primates have been involved in research into the infectivity, pathogenicity, and disease association of the virus, and also in vaccine trials. In the early days of HBV research chimpanzees were widely used, but even as early as 1969 there were concerns about the use of chimpanzees in research, due to the threat of extinction facing this species (Deinhardt, 1976). Morality issues of using our closest genetic cousin for experimentation also have to be taken into consideration.

### 5.1.1 History of HBV infection in primates

Primates have been screened for HBV before entering research centres, and so far only the great and lesser apes (i.e. chimpanzees, gorillas, orang-utans and gibbons) have been found to carry the virus (Deinhardt, 1976; Eichberg and Kalter, 1980). Most of this screening was carried out to ensure that animals were not already infected with HBV as they were to be used, amongst other things, in infection susceptibility experiments. These experiments also showed that only the great or lesser apes were susceptible to infection in the laboratory environment, although there were some early reports of vervet monkeys and rhesus monkeys becoming infected (Deinhardt, 1976). These results have not been reproduced, partly because it has become less and less acceptable to carry out these types of experiments.

As early as the late 1960's and early 1970's, scientists were realising that it was unsustainable to use chimpanzees in HBV research due to the threat of extinction. The only other primate that seemed susceptible to infection was the gibbon, another family of species that is unsustainable to use, as it is also under threat of extinction. More recently, baboons (*Papio spp.*) and macaques (*Macaca spp.*) have been evaluated as possible primate models for HBV research. Baboon studies from 1996 (Michaels et al., 1996) concluded that baboons were in

fact resistant to HBV infection, whereas a study from 2000 (Kedda et al., 2000) showed that chacma baboons (*Papio ursinus orientalis*) were susceptible to HBV infection and developed a chronic carrier state. Further experiments with chacma baboons have confirmed these findings (Baptista et al., 2003). Infectivity studies using macaques also indicate the possible use of this primate species in HBV research (Gheit et al., 2002). This would be advantageous, as this type of small primate is relatively easy to breed and keep in captivity.

It was always assumed that any HBV infected primates had picked up their HBV infections from humans involved in the chain of handling before they were tested. Primates captured in the wild were often given injections of pooled human sera, as a protection against human diseases, prior to their transport to research facilities around the world. It is possible that this practice was in fact responsible for some of the infections detected in the primates. The tests of the time did not look at the genetic sequence of the virus, but only looked for serological positivity. It is therefore possible that at least some of the infections detected were due to natural infections, with virus variants closely enough related to the human HBV viruses to be cross-reactive in serological tests. It is now known that the primate HBV variants share the same serological determinants as the human genotypes and are closely related to them. A more detailed description of each of the primate HBV variants detected to date is found below.

#### 5.1.2 History of HBV infection in chimpanzees and gorillas

The first major report of hepatitis in chimpanzees came out in 1978, after an outbreak at London Zoo (Zuckerman et al., 1978). At the time it was thought that the infection was of human origin, but later when the entire genome of the isolate from one of the chimpanzees involved in the outbreak was sequenced in 1988 (Vaudin et al., 1988), it became clear that this

virus was closely related to, but distinct from human HBV. Since this first isolate was published several more isolates have been reported. In 2000 three papers were published in a short period of time that described more isolates (Hu et al., 2000; MacDonald et al., 2000; Takahashi et al., 2000), and it became clear that chimpanzee HBV was a distinct genotype of HBV specific for this species. In 2001 a further two papers were published that concentrated on chimpanzee HBV, and they showed that HBV infection is indigenous to the ape species, with the possible existence of sub-species specific sub-genotypes within the ape HBV clades (Hu et al., 2001; Takahashi et al., 2001).

One paper in 2000 (Grethe et al., 2000) also reported a chimpanzee isolate, although this isolate was most closely related to gibbon HBV (see sections 5.1.6 and 5.3.2), and the first isolate from a western lowland gorilla (*Gorilla gorilla gorilla*). The gorilla isolate was very closely related to chimpanzee HBV. A recently published study of HBV prevalence in non-human primates in Gabon and the Congo (Makuwa et al., 2003) reported a high frequency of infection in western lowland gorillas, but unfortunately no sequence data was published. This sequence data will be very important in the future for determining whether there is a separate gorilla genotype of HBV, or whether gorillas are infected with a variant of chimpanzee HBV. The papers from 2000 and 2001 were beginning to speculate on the virus-host relationships of HBV and the possible origins and evolutionary history of the virus. Hu (2001) suggested that there were subspecies types of HBV within the chimpanzee HBV clade, and used mitochondrial data to back up his claims. In 2002 there was a report published where researchers had isolated an HBV strain from *Pan troglodytes schweinfurthii*, which meant that HBV was present in all the four subspecies of the common chimpanzee (Vartanian et al., 2002). This isolate branches off separately in the phylogenetic tree, supporting the existence of subspecies-specific HBV strains as suggested by Hu (Hu et al., 2001). To elucidate further

the exact relationships between chimpanzee HBV and the subspecies it would be helpful if isolates were genetically characterised in the bonobo (*Pan paniscus*), eastern lowland gorilla (*G. g. graueri*), and the mountain gorilla (*G. g. berengeti*).

In this study I have screened samples from common chimpanzees for the presence of HBV to further elucidate the chimpanzee HBV clade.

### 5.1.3 History of HBV infection in gibbons

Although seropositivity of gibbons for HBV was reported in the early years of HBV research it was not until 1996 that HBV DNA was isolated from a gibbon and the complete genome was sequenced (Norder et al., 1996). There had been an earlier report of a PreS gene sequence from a gibbon (Mimms et al., 1993), but gibbon HBV was not fully characterised until 1996. After the publication of this sequence it became clear that this was another species-specific HBV variant. A large number of gibbon sequences have now been published (Aiba et al., 2003; Grethe et al., 2000; Lanford et al., 2000; Mimms et al., 1993; Noppornpanth et al., 2003; Norder et al., 1996), and as is the case for chimpanzee HBV, it has become clear that there is a separate and distinct gibbon HBV genotype.

In this study I have screened samples from a variety of gibbon species for the presence of HBV DNA, and have investigated the possible presence of subtypes within the gibbon clade.

### 5.1.4 History of HBV infection in orang-utans

The complete genome of an orang-utan HBV isolate (accession number NC\_002168, orang-utan Somad) was first reported in 1999 (Warren et al., 1999), and orang-utan HBV has since been characterised further (Verschoor et al., 2001). Orang-utan HBV is closely related to gibbon HBV, but appears to be separate from it, in contrast to the gorilla/chimpanzee HBV

relationship. Large numbers of orang-utans have been screened for HBV as part of rehabilitation and release schemes, and there appears to be a high frequency of infection, both active and past (Warren et al., 1998). However, only a limited number of isolates have been sequenced and published.

In this study I screened orang-utan samples for the presence of HBV DNA in an attempt to increase the amount of sequence data available for this species and evaluate the orang-utan HBV clade.

#### 5.1.5 HBV in other primates

Apart from the reports of HBV in vervet monkeys (Deinhardt, 1976) there have been few reports of HBV infection in other primate species. There has been only one report of a single *Cercopithecus aethiops* monkey showing signs of infection (Heckel et al., 2001), and the detection of hepadnavirus-like particles in two ruffed lemurs (*Varecia variegata variegata*) (Worley and Stalis, 2002), but no sequences have been published. No Old World primates apart from the great and lesser apes have been shown to harbour HBV infection with naturally occurring species-specific virus variants, neither in older studies (Deinhardt, 1976; Eichberg and Kalter, 1980; Michaels et al., 1996) nor in more recent ones (Makuwa et al., 2003).

New World monkeys have also been screened for the presence of HBV, and to date there is only one report of one species being positive- the woolly monkey (*Lagothrix lagotricha*) (Lanford et al., 1998). This will be discussed further later. In this study I screened a large number of serum/plasma samples from Old World monkeys in an attempt to isolate an HBV-like virus.

As part of the study HBV screening data for primates already published was compiled to investigate the overall frequencies of infection in primates.

#### 5.1.6 Cross-species transmission of HBV

As mentioned earlier, it is possible to infect apes with human HBV variants, and this used to be thought to be the explanation for the presence of HBV infection in captured primates. The detection of human genotype E in a chimpanzee (accession number AB032431) is most likely explained by this type of scenario. The detection of a gibbon-like HBV variant in a chimpanzee (Grethe et al., 2000) on the other hand suggests the possibility of HBV crossing the species barrier among non-human primates in the other direction, from primate to human. Since we now know that there exist in nature a number of nonhuman primate HBV variants, the possibility arises of cross-species transmission from the natural hosts to humans.

This scenario seems especially likely in parts of Africa, where bush meat is gaining in popularity again and is posing a threat to the world's primate populations. It is also a concern that must be considered in the medical field, where xenotransplantation using primate organs is being investigated as a possible solution for the organ shortages experienced worldwide. It has already been shown that infections can cross the species barrier through this type of procedure (Michaels et al., 1996).

Baboon livers have been used in xenotransplantations, as these primates have been considered resistant to HBV infection (Lanford et al., 1995). As it has now been shown that at least one species of baboon, the chacma baboon (Baptista et al., 2003; Kedda et al., 2000), can be infected with HBV the suitability of baboons for this type of medical procedure is doubtful, although screening for HBV in donor animals would eliminate the risk. As mentioned above

(see section 5.1.5), the ability of these primates to support human HBV replication may indicate the existence of a natural baboon HBV variant. In one sample taken from a baboon liver xenograft given to an HBV patient HBV DNA was detected in one assay (Lanford et al., 1995). The most likely source of this DNA is extrahepatic HBV DNA of patient origin entering the liver via the cardiovascular system, but the possibility remains that it may be of primate donor origin. Transmission of infectious baboon cytomegalovirus through xenotransplantation has been documented (Michaels et al., 2001), so zoonotic transmission from baboon to human is a real possibility.

There has been one report of possible gorilla to human transmission of HBV in a Zoo (Linnemann, Jr. et al., 1984), although this was not backed up by any sequence data from either the gorilla or the keeper involved. Large scale screening of primate handlers working with gibbons has also failed to produce any evidence for primate to human transmission (Noppornpanth et al., 2003). Nevertheless, cross-species transmission and subsequent evolution within new hosts has been proposed as a possible scenario for the origin of human HBV infections in modern populations (as discussed in more detail in chapter one, section 1.5.2).

In this study I combined human and primate HBV sequence data in a search for evidence of past cross-species transmission events, in an attempt to shed more light on the evolutionary origin and history of the hepatitis B viruses infecting primates today.

## 5.2 Materials and methods

### 5.2.1 Primate serum and plasma samples

A total of 137 primate serum samples were available for screening. 71 were samples from drills (*Mandrillus leucophaeus*) residing at a rescue centre in Nigeria, 48 of which were wild-born and 23 captive-born, all previously tested HBsAg negative. 24 samples were obtained from the following species; 8 drills, 2 mandrills (*Mandrillus sphinx*), 4 cherry-capped mangabeys (*Cercopithecus torquatus*) and 7 chimpanzees (*Pan troglodytes spp.*), kept at the Limbe wildlife centre in Cameroon, all of which were wild-born. The drills, mandrills and mangabeys were all HBsAg negative, 2 chimpanzees were HBsAg positive, 3 were negative and 2 were unknown. One sample from a known HBsAg positive chimpanzee from a rescue centre in Nigeria and 3 samples from mona monkeys (*Cercopithecus mona*), 2 samples from putty-nosed monkeys (*C. nictitans*), and 3 samples from red-eared monkeys (*C. erythrotis*) all known to be HBsAg negative were also available. HBsAg positivity normally indicates a carrier of HBV, although it is known that in humans chronic carriers can lose their HBsAg expression and still show the presence of HBV DNA. Therefore all animals previously tested for HBsAg and found to be negative were also screened for the presence of HBV DNA.

20 samples from various gibbon species (*Hylobates spp.*, *Nomascus spp.*), 4 of which were known to be HBsAg positive and 16 negative, and 14 Bornean orang-utan (*Pongo pygmaeus pygmaeus*) samples, all HBsAg negative were sent from the Pintung rescue centre in Taiwan. These animals were all wild-born, and the veterinarian Dr John Lewis took samples on site. Samples were stored at 4°C prior to shipping to the UK according to CITES regulations, and then stored at -25°C prior to analysis. One male and one female lar gibbon (*Hylobates lar*) both captive born and from the Welsh Mountain Zoo were also included. These samples were provided by Dr Andrew Greenwood. They were HBsAg tested at the City Hospital,

Edinburgh, and the female was found to be positive whereas the male was negative. Samples were extracted and screened for the presence of HBV DNA as described in chapter 2 sections 1, 2, and 3.

### 5.2.2 Screening, sequencing and phylogenetics

Serum or plasma samples from the primates were extracted as described in chapter 2. The DNA was then screened for the presence of HBV DNA using two separate nested PCR assays as described earlier in chapter 2. Positive samples were subjected to further PCR amplification to obtain overlapping fragments representing the entire genome of the HBV isolate. Primers used for this sequencing are described in chapter 2. Complete sequences were then aligned with all previously published primate HBV isolates as well as representative sequences from all the human genotypes of HBV. These sequences were then subjected to phylogenetic analysis, using both the neighbour-joining method in MEGA2 and the maximum likelihood method (carried out by Edward Holmes in Oxford).

### 5.2.3 Screening of Old World monkey samples with highly conserved primers

To investigate whether Old World monkey species may be carrying a highly divergent HBV variant that is undetectable using current serological and molecular screening strategies a new PCR was developed. A region of the HBV genome which is highly conserved between all the human, primate and rodent HBV variants was identified. A new set of primers was selected to amplify this region of the HBV genome. Primers 21, 22, 23, and 24 (see chapter 2 for primer sequences) were chosen. An alignment showing the segment of the S gene region and the primer sites can be found below (Figure 5-2). These sets of primers were shown in control experiments to be equal in sensitivity when compared with the two other sets of primers used

in screening. PCR buffers and thermal cycling conditions were as described earlier in chapter 2 section 2.

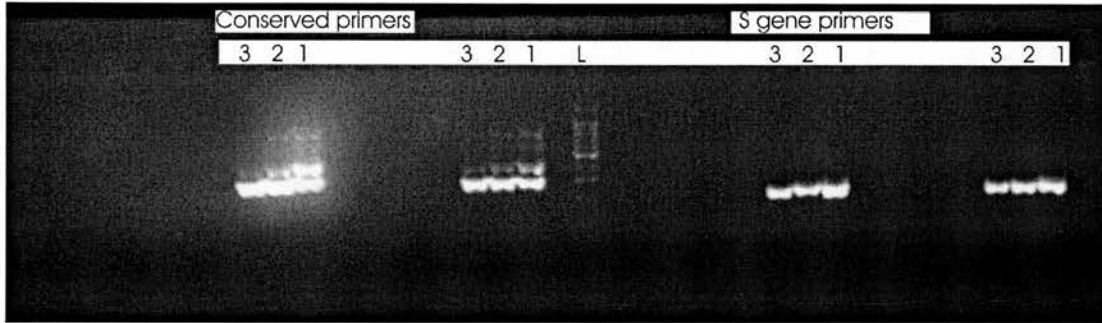


Figure 5-1. Gel image showing a comparison of the S gene PCR (chapter 2) and the conserved PCR. Lane 1 in each case is undiluted human HBV DNA positive control DNA, lane 2 is a 1/100 dilution of the control DNA and lane 3 is a 1/1000 dilution of the control DNA. Each set of primers was tested in duplicate. Both sets of primers failed to amplify a 1/10000 dilution of control DNA. Lane L is a molecular size marker, with the brightest band having a size of 500bp. The fragment amplified by the S gene primers is 200bp long and the fragment amplified by the conserved primers is 300 bp long.

Primer	Primer 21	Primer 22	Primer 23	Primer 24
H-A	537 GACTT CTCCTCAATTT CTCAGGGG	311 GATGTGG CTCGGGGCTT TTAGTAT	638 GATACCT CACCGGAT CCGGCA	659 GCGAAGAG AACCGATTA ATGCA
H-B	.....	.....	.....	.....
H-C	.....	.....	.....	.....
H-D	.....	.....	.....	.....
H-E	.....	.....	.....	.....
H-F	.....	.....	.....	.....
H-G	.....	.....	.....	.....
H-H	.....	.....	.....	.....
CH	.....	.....	.....	.....
CB	.....	.....	.....	.....
CC	.....	.....	.....	.....
CD	.....	.....	.....	.....
CE	.....	.....	.....	.....
CF	.....	.....	.....	.....
CG	.....	.....	.....	.....
CH	.....	.....	.....	.....
CI	.....	.....	.....	.....
CJ	.....	.....	.....	.....
CK	.....	.....	.....	.....
CL	.....	.....	.....	.....
CM	.....	.....	.....	.....
CN	.....	.....	.....	.....
CO	.....	.....	.....	.....
CP	.....	.....	.....	.....
CQ	.....	.....	.....	.....
CR	.....	.....	.....	.....
CS	.....	.....	.....	.....
CT	.....	.....	.....	.....
CU	.....	.....	.....	.....
CV	.....	.....	.....	.....
CW	.....	.....	.....	.....
CX	.....	.....	.....	.....
CY	.....	.....	.....	.....
CZ	.....	.....	.....	.....
CA	.....	.....	.....	.....
CB	.....	.....	.....	.....
CC	.....	.....	.....	.....
CD	.....	.....	.....	.....
CE	.....	.....	.....	.....
CF	.....	.....	.....	.....
CG	.....	.....	.....	.....
CH	.....	.....	.....	.....
CI	.....	.....	.....	.....
CJ	.....	.....	.....	.....
CK	.....	.....	.....	.....
CL	.....	.....	.....	.....
CM	.....	.....	.....	.....
CN	.....	.....	.....	.....
CO	.....	.....	.....	.....
CP	.....	.....	.....	.....
CQ	.....	.....	.....	.....
CR	.....	.....	.....	.....
CS	.....	.....	.....	.....
CT	.....	.....	.....	.....
CU	.....	.....	.....	.....
CV	.....	.....	.....	.....
CW	.....	.....	.....	.....
CX	.....	.....	.....	.....
CY	.....	.....	.....	.....
CZ	.....	.....	.....	.....
CA	.....	.....	.....	.....
CB	.....	.....	.....	.....
CC	.....	.....	.....	.....
CD	.....	.....	.....	.....
CE	.....	.....	.....	.....
CF	.....	.....	.....	.....
CG	.....	.....	.....	.....
CH	.....	.....	.....	.....
CI	.....	.....	.....	.....
CJ	.....	.....	.....	.....
CK	.....	.....	.....	.....
CL	.....	.....	.....	.....
CM	.....	.....	.....	.....
CN	.....	.....	.....	.....
CO	.....	.....	.....	.....
CP	.....	.....	.....	.....
CQ	.....	.....	.....	.....
CR	.....	.....	.....	.....
CS	.....	.....	.....	.....
CT	.....	.....	.....	.....
CU	.....	.....	.....	.....
CV	.....	.....	.....	.....
CW	.....	.....	.....	.....
CX	.....	.....	.....	.....
CY	.....	.....	.....	.....
CZ	.....	.....	.....	.....
CA	.....	.....	.....	.....
CB	.....	.....	.....	.....
CC	.....	.....	.....	.....
CD	.....	.....	.....	.....
CE	.....	.....	.....	.....
CF	.....	.....	.....	.....
CG	.....	.....	.....	.....
CH	.....	.....	.....	.....
CI	.....	.....	.....	.....
CJ	.....	.....	.....	.....
CK	.....	.....	.....	.....
CL	.....	.....	.....	.....
CM	.....	.....	.....	.....
CN	.....	.....	.....	.....
CO	.....	.....	.....	.....
CP	.....	.....	.....	.....
CQ	.....	.....	.....	.....
CR	.....	.....	.....	.....
CS	.....	.....	.....	.....
CT	.....	.....	.....	.....
CU	.....	.....	.....	.....
CV	.....	.....	.....	.....
CW	.....	.....	.....	.....
CX	.....	.....	.....	.....
CY	.....	.....	.....	.....
CZ	.....	.....	.....	.....
CA	.....	.....	.....	.....
CB	.....	.....	.....	.....
CC	.....	.....	.....	.....
CD	.....	.....	.....	.....
CE	.....	.....	.....	.....
CF	.....	.....	.....	.....
CG	.....	.....	.....	.....
CH	.....	.....	.....	.....
CI	.....	.....	.....	.....
CJ	.....	.....	.....	.....
CK	.....	.....	.....	.....
CL	.....	.....	.....	.....
CM	.....	.....	.....	.....
CN	.....	.....	.....	.....
CO	.....	.....	.....	.....
CP	.....	.....	.....	.....
CQ	.....	.....	.....	.....
CR	.....	.....	.....	.....
CS	.....	.....	.....	.....
CT	.....	.....	.....	.....
CU	.....	.....	.....	.....
CV	.....	.....	.....	.....
CW	.....	.....	.....	.....
CX	.....	.....	.....	.....
CY	.....	.....	.....	.....
CZ	.....	.....	.....	.....
CA	.....	.....	.....	.....
CB	.....	.....	.....	.....
CC	.....	.....	.....	.....
CD	.....	.....	.....	.....
CE	.....	.....	.....	.....
CF	.....	.....	.....	.....
CG	.....	.....	.....	.....
CH	.....	.....	.....	.....
CI	.....	.....	.....	.....
CJ	.....	.....	.....	.....
CK	.....	.....	.....	.....
CL	.....	.....	.....	.....
CM	.....	.....	.....	.....
CN	.....	.....	.....	.....
CO	.....	.....	.....	.....
CP	.....	.....	.....	.....
CQ	.....	.....	.....	.....
CR	.....	.....	.....	.....
CS	.....	.....	.....	.....
CT	.....	.....	.....	.....
CU	.....	.....	.....	.....
CV	.....	.....	.....	.....
CW	.....	.....	.....	.....
CX	.....	.....	.....	.....
CY	.....	.....	.....	.....
CZ	.....	.....	.....	.....
CA	.....	.....	.....	.....
CB	.....	.....	.....	.....
CC	.....	.....	.....	.....
CD	.....	.....	.....	.....
CE	.....	.....	.....	.....
CF	.....	.....	.....	.....
CG	.....	.....	.....	.....
CH	.....	.....	.....	.....
CI	.....	.....	.....	.....
CJ	.....	.....	.....	.....
CK	.....	.....	.....	.....
CL	.....	.....	.....	.....
CM	.....	.....	.....	.....
CN	.....	.....	.....	.....
CO	.....	.....	.....	.....
CP	.....	.....	.....	.....
CQ	.....	.....	.....	.....
CR	.....	.....	.....	.....
CS	.....	.....	.....	.....
CT	.....	.....	.....	.....
CU	.....	.....	.....	.....
CV	.....	.....	.....	.....
CW	.....	.....	.....	.....
CX	.....	.....	.....	.....
CY	.....	.....	.....	.....
CZ	.....	.....	.....	.....
CA	.....	.....	.....	.....
CB	.....	.....	.....	.....
CC	.....	.....	.....	.....
CD	.....	.....	.....	.....
CE	.....	.....	.....	.....
CF	.....	.....	.....	.....
CG	.....	.....	.....	.....
CH	.....	.....	.....	.....
CI	.....	.....	.....	.....
CJ	.....	.....	.....	.....
CK	.....	.....	.....	.....
CL	.....	.....	.....	.....
CM	.....	.....	.....	.....
CN	.....	.....	.....	.....
CO	.....	.....	.....	.....
CP	.....	.....	.....	.....
CQ	.....	.....	.....	.....
CR	.....	.....	.....	.....
CS	.....	.....	.....	.....
CT	.....	.....	.....	.....
CU	.....	.....	.....	.....
CV	.....	.....	.....	.....
CW	.....	.....	.....	.....
CX	.....	.....	.....	.....
CY	.....	.....	.....	.....
CZ	.....	.....	.....	.....
CA	.....	.....	.....	.....
CB	.....	.....	.....	.....
CC	.....	.....	.....	.....
CD	.....	.....	.....	.....
CE	.....	.....	.....	.....
CF	.....	.....	.....	.....
CG	.....	.....	.....	.....
CH	.....	.....	.....	.....
CI	.....	.....	.....	.....
CJ	.....	.....	.....	.....
CK	.....	.....	.....	.....
CL	.....	.....	.....	.....
CM	.....	.....	.....	.....
CN	.....	.....	.....	.....
CO	.....	.....	.....	.....
CP	.....	.....	.....	.....
CQ	.....	.....	.....	.....
CR	.....	.....	.....	.....
CS	.....	.....	.....	.....
CT	.....	.....	.....	.....
CU	.....	.....	.....	.....
CV	.....	.....	.....	.....
CW	.....	.....	.....	.....
CX	.....	.....	.....	.....
CY	.....	.....	.....	.....
CZ	.....	.....	.....	.....
CA	.....	.....	.....	.....
CB	.....	.....	.....	.....
CC	.....	.....	.....	.....
CD	.....	.....	.....	.....
CE	.....	.....	.....	.....
CF	.....	.....	.....	.....
CG	.....	.....	.....	.....
CH	.....	.....	.....	.....
CI	.....	.....	.....	.....
CJ	.....	.....	.....	.....
CK	.....	.....	.....	.....
CL	.....	.....	.....	.....
CM	.....	.....	.....	.....
CN	.....	.....	.....	.....
CO	.....	.....	.....	.....
CP	.....	.....	.....	.....
CQ	.....	.....	.....	.....
CR	.....	.....	.....	.....
CS	.....	.....	.....	.....
CT	.....	.....	.....	.....
CU	.....	.....	.....	.....
CV	.....	.....	.....	.....
CW	.....	.....	.....	.....
CX	.....	.....	.....	.....
CY	.....	.....	.....	.....
CZ	.....	.....	.....	.....
CA	.....	.....	.....	.....
CB	.....	.....	.....	.....
CC	.....	.....	.....	.....
CD	.....	.....	.....	.....
CE	.....	.....	.....	.....
CF	.....	.....	.....	.....
CG	.....	.....	.....	.....
CH	.....	.....	.....	.....
CI	.....	.....	.....	.....
CJ	.....	.....	.....	.....
CK	.....	.....	.....	.....
CL	.....	.....	.....	.....
CM	.....	.....	.....	.....
CN	.....	.....	.....	.....
CO	.....	.....	.....	.....
CP	.....	.....	.....	.....
CQ	.....	.....	.....	.....
CR	.....	.....	.....	.....
CS	.....	.....	.....	.....
CT	.....	.....	.....	.....
CU	.....	.....	.....	.....
CV	.....	.....	.....	.....
CW	.....	.....	.....	.....

#### 5.2.4 Mitochondrial 12S region PCR and sequencing of primate samples

Since no species information was available for the chimpanzees and some of the gibbon samples it was decided to try to determine the species using mitochondrial data to enable more precise comparison of HBV and species phylogenies. Without this type of information it is difficult to make any hypotheses regarding the origins and evolutionary history of HBV in primates and the relationships between the viruses and their hosts. A search was done on GenBank for mitochondrial sequence data. Different regions of the mitochondrion were chosen and phylogenetic analysis carried out to evaluate the potential use of the regions for the purpose of determining species. The 12S region was chosen, as the phylogenetic tree obtained using this data showed a close congruence to the evolutionary tree of primates as accepted today (Catriona MacFarlane, PhD thesis).

Primers 12S-S (5'-CCATAAACAMAYAGGYTTGGTCC-3'; positions 641 to 664) and 12S-AS (5'-CAGGGTTTGCTGAAGATGGCGGTATATA-3'; positions 1270 to 1298) were designed using the sequence data from GenBank (Catriona MacFarlane, PhD thesis). These sequences were then used in the phylogenetic analysis of the 12S sequences. Accession numbers can be found in this tree (Figure 5-5).

A single round PCR was set up using 2 µl of extracted sample DNA in a total reaction volume of 50 µl. The PCR buffer was made up as described earlier for the screening of the samples for the presence of HBV. The PCR was carried out using the following conditions; 40 cycles of 94°C for 60 sec, 55°C for 60 sec and 72°C for 60 sec, followed by a final extension step at 72°C for 7 minutes and a holding step at 20°C for 10 minutes. 15µl of product was loaded onto a 2% agarose gel (stained with ethidium bromide) and electrophoresed for 40 minutes at

150V. A 100 bp marker was run alongside the products to check the fragment size. PCR fragments were then visualised under UV light and a picture was taken.

Fragments were then sequenced using primer 12S-S and the BigDye automated sequencing protocol as described earlier. Sequencing data was then imported into the SIMMONIC sequence analysis program and aligned to the GenBank sequences used to design the primers. Phylogenetic trees were then constructed using MEGA2 as described earlier. An attempt was then made to identify the species of the primate samples for which this information was unavailable.

#### 5.2.5 Association Index analysis

Association Index (AI) analysis was carried out as described earlier in chapter 3. This analysis was done in the first instance to investigate in which parts of the HBV genome the human and primate genotypes segregate into the groups seen in phylogenetic analyses (see Fig. 3-1). The analysis was carried out across the genome using 250bp sizes fragment windows, and moving along the genome in 100 bp overlapping steps. The results were then plotted into a barchart using Microsoft Excel. AI analysis was also carried out to investigate groupings seen within the primate clades using the same method as used for investigating geographical subgroupings within human HBV genotypes (see chapter 3 section 2.4). Chimpanzee isolates, gibbon isolates and orang-utan isolates were tagged separately in the analysis. Genotype F and the woolly monkey HBV sequence were omitted from the analysis due to the lesser relatedness of these to the remaining genotypes.

### 5.2.6 Recombination analysis

To investigate the existence of recombinant HBV genomes within the primate genotypes in the search for evidence of cross-species transmission events two methods were used. The AI method can be used to look for recombination in HBV genomes as described in chapter 4. 600 bp windows moving along the genome in 20 bp steps were used in the first instance, and then 500bp windows with 100bp overlapping steps were used after this was shown to provide the same results (see chapter 4, page 119). SimPlot analysis was then carried out using the same sequence groups, window and step sizes (Lole et al., 1999).

## 5.3 Results

### 5.3.1 Frequency of HBV viraemia in primates

The frequency of HBV viraemia in primates was calculated both for samples screened in this study, and for previously published results in combination with the results of this study.

#### 5.3.1.1 Frequency of viraemia seen in study samples

A total of 7 out of 136 primates tested positive for HBV DNA, all of which were within the ape group. No old World monkey species tested positive with the two HBV screening PCR's described in chapter 2. When these samples were re-screened using highly conserved primers for part of the S gene region able to pick up rodent HBV variants as well as all the known human and primate HBV variants, no positive samples were found either (Table 5-1). Two chimpanzees were found to be positive for HBV DNA by PreS and S gene PCR.

Chimpanzee 1, Louisa, was HBsAg tested in 1997 and found to be negative. Louisa is of the *Pan troglodytes troglodytes* subspecies of chimpanzee (this was shown by amplification and phylogenetic analysis of mitochondrial DNA, see section 5.2.5), and was captured and then confiscated in south-western Cameroon. Prior to confiscation she was kept by local people. She is kept with other chimpanzees at the Limbe wildlife centre in Cameroon. Since confiscation she has had no contact with primates of other species, but the exact details of how she was kept prior to confiscation are unavailable.

The second chimpanzee is Osang who belongs to the *vellerosus* subspecies of chimpanzee (see section 5.2.5). He was donated to a conservation project in the south-eastern part of Nigeria on the 16<sup>th</sup> of January 1996. He was HBsAg tested on arrival and was found to be positive. Osang was also kept by local people prior to his arrival at the project. He is kept with other chimpanzees at the centre.

Five gibbons were found to be HBV DNA positive by PCR. Of these 4 were wild-caught and one was captive born. The captive born animal is a female named Tamang (*Hylobates lar*), and as mentioned before she was HBsAg tested at the City Hospital in Edinburgh and found to be positive. She was born in 1989 at Cricket St Thomas and moved to the Welsh Mountain Zoo in 1995. She was born to a captive born father (Paignton Zoo) with known wild-caught parents from Malaysia and Thailand, and a captive born mother (Stuttgart Zoo) with unknown parental origins. At the Welsh Zoo she is kept with a male, Jake who was born at the Zoo and is both HBsAg and HBV DNA negative.

The remaining 4 positive gibbons consist of Crazy Woman (*H. agilis*) confiscated in Taiwan 06-09-1993, TB Black (species not identified) confiscated in Taiwan 17-05-1996, Happy (*Nomascus gabriellae*) confiscated in Taiwan 08-10-1996, and Wendy (*N. gabriellae*) confiscated in Taiwan 27-05-1999. Their histories prior to confiscation are unknown, as are their origins. Their origins can be estimated from their species denotation. They have been kept on their own since confiscation. All four were tested and found to be HBsAg positive.

The species of the chimpanzees was determined by mitochondrial sequencing, whereas the sanctuary provided most species names for the gibbons. Mitochondrial sequencing was carried out on gibbon samples as well to provide species denotations for those animals for which this information was unavailable.

### 5.3.1.2 Frequency of viraemia seen in all primates

Looking first at the results for the ape species, all three species groups of great ape, as well as the lesser ape gibbon have prevalences over 15%, with an overall prevalence of infection of 18.7% (Table 5-2). This figure compares to that found in human populations in endemic areas.

Moving on to look at the Old World monkeys, a large number of different species have been screened. So far none have been found to contain a HBV-like virus using current serological and molecular screening techniques. Even when using highly conserved primers in a new PCR strategy, no HBV-like virus is detected in these species of primate (Table 5-3). Apart from the captive woolly monkey, no New World primate species have been found to harbour HBV either (Table 5-4). The woolly monkey HBV isolate will be discussed in more detail later.

<u>SPECIES</u>	<u>NO.</u>	<u>PCR</u>	<u>sAg</u>	<u>ANTI-CORE</u>	<u>TOTAL POS.</u>	<u>% POS.</u>
<i>Pan troglodytes spp.</i>	8	2	3	2	2	25
<i>Hylobates spp.</i>	22	5	5	0	5	22.7
<i>Pongo pygmaeus pygmaeus</i>	14	0	0	1	0	0
<i>Mandrillus leucophaeus</i>	78	0	0	0	0	0
<i>M. sphinx</i>	2	0	0	0	0	0
<i>Cercopithecus torquatus</i>	4	0	0	0	0	0
<i>C. mona</i>	3	0	0	0	0	0
<i>C. nictitans</i>	2	0	0	0	0	0
<i>C. erythrotis</i>	3	0	0	0	0	0
<b>TOTAL</b>	<b>136</b>	<b>7</b>	<b>8</b>	<b>3</b>	<b>7</b>	<b>5.15</b>
<b>TOTAL APES</b>	<b>44</b>	<b>7</b>	<b>8</b>	<b>3</b>	<b>7</b>	<b>15.91</b>
<b>TOTAL OWM</b>	<b>92</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

Table 5-1. Frequency of HBV infection in study samples. High prevalence rates are seen in chimpanzees and gibbons, whereas all other apes and Old World monkeys screened were found to be negative. All chimpanzee, gibbon and orang-utan samples included in this study were tested for HBsAg and anti-HBc. Those animals found to be PCR positive for HBV DNA were also HBsAg positive.

SPECIES	TOTAL PCR	HBsAg	ACTIVE INFECTION (%)	ANTI-HBc	REFERENCES FOR DATA NOT FROM THIS STUDY
<i>Gorilla gorilla</i> spp.	10/46	5/27	10/46 (21.7%)	3/12	(Grethe et al., 2000; Zuckerman et al., 1978; Linnemann et al., 1984; Thornton et al., 2001; Worley and Stalis, 2002; Makuwa et al., 2003)
<i>Pan troglodytes</i> spp.	75/162	52/620	101/641 (15.8%)	48/130	(Takahashi et al., 2000; Takahashi et al., 2001; Hu et al., 2000; Hu et al., 2001; MacDonald et al., 2000; Grethe et al., 2000; Zuckerman et al., 1978; Ogata et al., 1993; Vaudin et al., 1988; Deinhardt, 1976; Eichberg and Kalter, 1980; Vartanian et al., 2002; Makuwa et al., 2003)
<i>Hylobates</i> spp.	53/177	35/191	55/213 (25.8%)	48/130	(Grethe et al., 2000; Norder et al., 1996; Lanford et al., 2000; Vaudin et al., 1988; Deinhardt, 1976; Thornton et al., 2001; Noppornpanth et al., 2003)
<i>Pongo pygmaeus</i> spp.	32/104	58/141	58/297 (19.5%)	1/14	(Warren et al., 1999; Davis et al., 2000; Vaudin et al., 1988; Deinhardt, 1976)
<b>TOTAL APES</b>	<b>170/489</b>	<b>150/979</b>	<b>224/1197 (18.7%)</b>	<b>100/286</b>	

Table 5-2. Combined levels of HBV infection in apes. An overall prevalence of 18.7 % in ape species compares to rates for human populations in endemic areas.

SPECIES	TOTAL PCR	HBsAg	ACTIVE INFECTION (%)	ANTI-HBc	REFERENCES FOR DATA NOT FROM THIS STUDY
<i>Papio</i> spp.	0/4	0/168	0	0/103	(Deinhardt, 1976; Michaels et al., 1996; Eichberg and Kalter, 1980)
<i>Mandrillus leophaeus</i>	0/78	0/78	0	0/78	
<i>M. sphinx</i>	0/96	0/2	0	0/2	(Makuwa et al., 2003)
<i>Cebus albifrons</i>	-	0/10	0	-	(Eichberg and Kalter, 1980)
<i>Cercopithecus aetiops</i>	-	0/19	0	-	(Eichberg and Kalter, 1980)
<i>C. torquatus</i>	0/24	0/4	0	0/4	(Makuwa et al., 2003)
<i>C. mona</i>	0/3	0/3	0	0/3	
<i>C. nictitans</i>	0/24	0/2	0	0/2	(Makuwa et al., 2003)
<i>C. erythrotis</i>	0/3	0/3	0	0/3	
<i>C. cephus</i>	0/28	-	-	-	(Makuwa et al., 2003)
<i>C.h. solatus</i>	0/6	-	-	-	(Makuwa et al., 2003)
<i>C. neglectus</i>	0/2	-	-	-	(Makuwa et al., 2003)
<i>C. pogonias</i>	0/2	-	-	-	(Makuwa et al., 2003)
<i>C. albigena</i>	0/6	-	-	-	(Makuwa et al., 2003)
<i>Miopithecus talapoin</i>	0/5	-	-	-	(Makuwa et al., 2003)
<b>TOTAL</b>	<b>0/281</b>	<b>0/289</b>	<b>0</b>	<b>0/195</b>	

Table 5-3. Combined levels of HBV infection in Old World Monkeys. No Old World monkeys have been found to be HBV positive.

SPECIES	TOTAL PCR	HBsAg	ACTIVE INFECTION (%)	ANTI-HBc	REFERENCES FOR DATA NOT FROM THIS STUDY
<i>Lagothrix lagothericha</i>	10/35	7/31	10/33 (30.3%)	-	(Lanford et al., 1998)
<i>Saimiri sciureus</i>	-	0/20	0	-	(Eichberg and Kalter, 1980)
<i>Callithrix jacchus</i>	-	0/6	0	-	(Eichberg and Kalter, 1980)
<i>Saguinus oedipus</i>	-	0/12	0	-	(Eichberg and Kalter, 1980)
<b>TOTAL</b>	<b>10/35</b>	<b>7/69</b>	<b>10/33 (30.3%)</b>		

Table 5-4. Combined levels of HBV infection in New World Monkeys. Only woolly monkeys have to date been found to carry HBV.

### 5.3.2 Phylogenetic analysis of primate HBV isolates

Following complete genome sequencing of the 7 HBV positive primate samples phylogenetic analysis was carried out (Figure 5-3). All previously published primate isolates were included in the analysis (n = 25), as well as 10 representative sequences from each of the human genotypes A-H as described earlier (accession numbers for all sequences used in this analysis can be found in chapter 2). The complete genome analysis splits the resulting tree into three sections; South America (both primate and human), all other human genotypes, and all the remaining primate genotypes (see Fig. 5-3).

Complete genome analysis shows that the primate isolates from this study fall into the previously described primate clades (Fig. 5-3). The gibbon/orang-utan clade is made up of four main branches. The first of these is made up of the two orang-utan isolates and two of the gibbon isolates from this study (Crazy Woman and TB Black). One of these is an agilis gibbon (Crazy Woman), and the other is of unknown species (TB Black). This unknown gibbon species most likely belongs to the *Hylobates* subgenus (see section 5.3.3) using mitochondrial data, although it was not possible to determine the exact species using this method. One lar gibbon (Tamang) from this study can be found on a branch with other isolates from lar gibbons and the chimpanzee isolate mentioned earlier (accession number HBV131575, see section 5.1.6), and a *gabriellae* gibbon (Wendy) can be found on another branch which has mainly concolor gibbons within it. Another *gabriellae* gibbon (Happy) is an outlier. The final branch contains isolates from *pileatus* gibbons.

Within the chimpanzee clade one divergent isolate can be seen; AB046525 belonging to the *troglodytes* subspecies (Fig. 5-3). The other isolates are separated onto two main branches, with the exception of AB032431 which groups within human genotype E. One branch

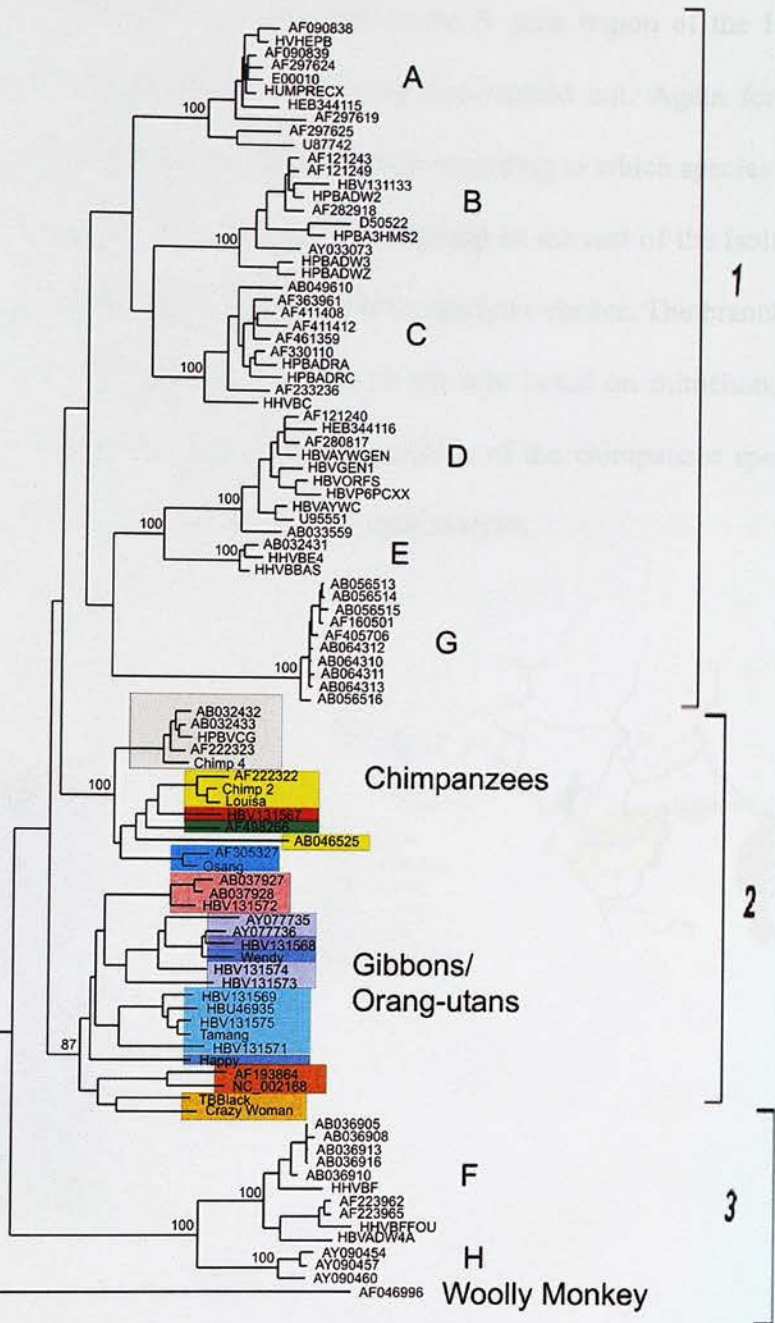
contains isolates of the verus subspecies, and the other isolates from the vellerosus, troglodytes and schweinfurthii subspecies, as well as the single gorilla isolate. Vellerosus isolates, including Osang from this study, branch off first within this clade, followed by the schweinfurthii isolate, the gorilla isolate and then the troglodytes isolates including Louisa from this study. There again appears to be separation of isolates depending on the host species of primate. This species separation of HBV variants resembles the evolutionary relationship between the species (Fig. 5- 7b), but does not mirror it exactly. This is partly due to the inlier position of the gorilla isolate. HBV from the vellerosus subspecies is also more closely related to the troglodytes/schweinfurthii clade than to that made up of verus isolates.

One isolate from a chimpanzee originally reported to be of the verus subspecies (Chimp 2 from MacDonald et al., 2000) is found in the troglodytes cluster. Serum from this chimpanzee was available for testing, so mitochondrial PCR and sequencing was carried out. This showed that this chimpanzee was most likely of the troglodytes species and not a verus as first reported. This information has subsequently been verified by the sanctuary where the chimpanzee is now being kept. In this analysis, apart from the isolates grouping according to which species they were derived from, there is also an East/West split within the African primates based on the habitat ranges. The geographical distributions of gorillas and chimpanzees are described further in section 5.3.4.

The main human clade splits into two as well, one containing genotypes A, B, and C, the other containing genotypes D, E, and G. The overall branching order within the complete genome tree resembles quite closely the evolutionary tree of the primates (Fig. 5-7a), with gibbon and orang-utan branching off first, followed by chimpanzee and gorilla, and then

human genotypes A-E and G, with the New World woolly monkey isolate forming the outgroup. Only the outlier position of human genotypes F and H does not fit.

- *Gorilla gorilla gorilla*
- *P. t. schweinfurthii*
- *P. t. vellerosus*
- *P. t. troglodytes*
- *P. t. verus*
- *N. concolor*
- *N. leucogenys*
- *N. gabriellae*
- *H. pileatus*
- *H. agilis/H. moloch*
- *H. lar*
- *Pongo pygmaeus*



— 0.01 substitutions/site

Figure 5-3. Phylogenetic analysis of complete genomes of HBV isolated from study primates. This analysis splits the tree into three sections; 1- all human genotypes except F and H, 2- all primate genotypes, and 3- all South American genotypes, human and primate.

A number of further isolates for primates are available in the S gene region of the HBV genome only, so phylogenetic analysis of this region was also carried out. Again for the chimpanzee clade there is distinct separation of HBV variants according to which species was the host species (Fig. 5-4). The gorilla isolate is now an outgroup to the rest of the isolates, and the schweinfurthii isolate is most closely related to the troglodytes cluster. The branching order of the chimpanzee S gene phylogenetic tree mirrors the tree based on mitochondrial DNA data (Fig. 5-7b) very closely. The geographical separation of the chimpanzee species into Central/Eastern and Western areas is also reflected in this analysis.

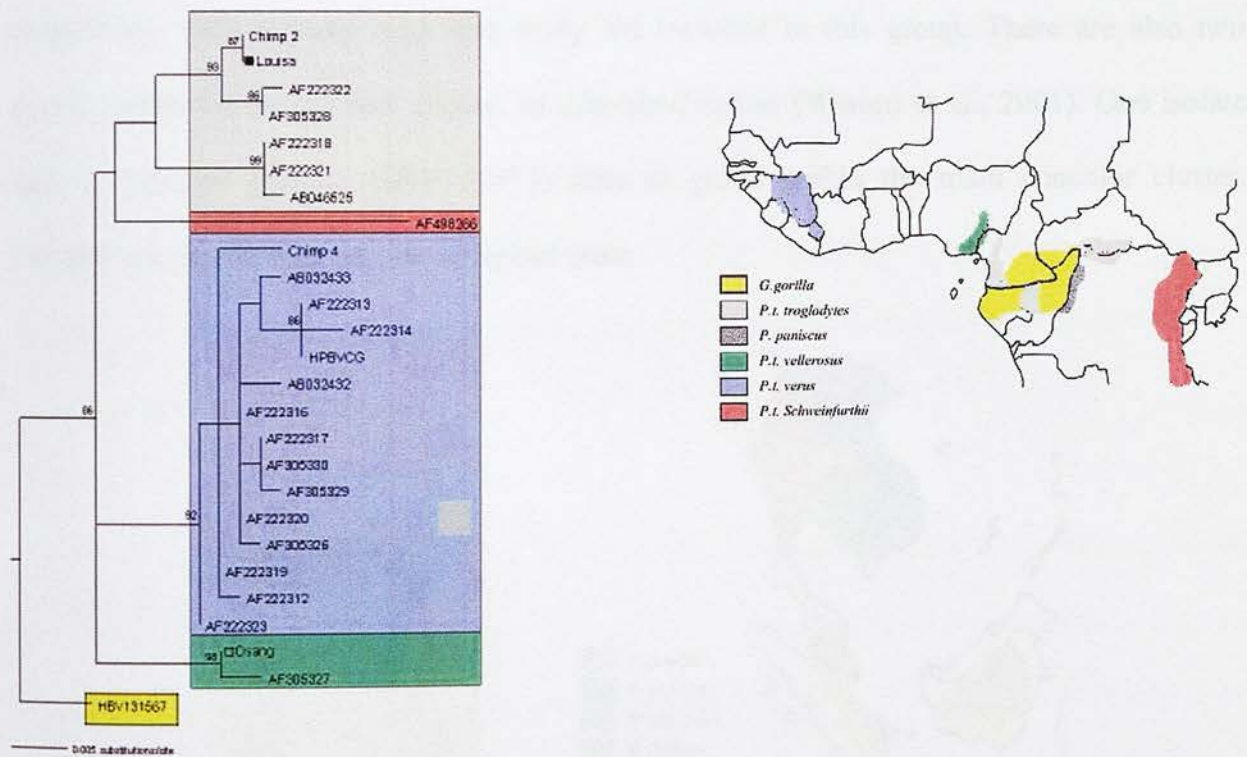


Figure 5-4. S gene phylogenetic analysis of chimpanzee HBV isolates, and the geographical distribution of species within Africa. Species specific clusters can be seen. These are supported by an AI analysis value of 0.0074 (Table 5-5). These clusters also segregate according to the geographical origin of the species (AI = 0.0010, Table 5-5).

Within the gibbon clade there also again appears to be separation according to species (Fig. 5-5), although again this does not reflect the phylogenetic relationship between the species, as isolates from the *Nomascus* subgenus can be found on a branch separating two clades with species-specific isolates from the *Hylobates* subgenus. One isolate is highly divergent in this region (Happy from this study). One interesting finding from this analysis that also would not be predicted from the evolutionary tree of gibbons is the very close relationship between the orang-utan isolates and gibbon isolates from species with habitat ranges that overlap with the orang-utans. Two isolates from this study are included in this group. There are also two groups within the orang-utan cluster, as described earlier (Warren et al., 2001). One isolate from a concolor gibbon (HBV131573) fails to group within the main concolor cluster. Possible reasons for this will be discussed later.

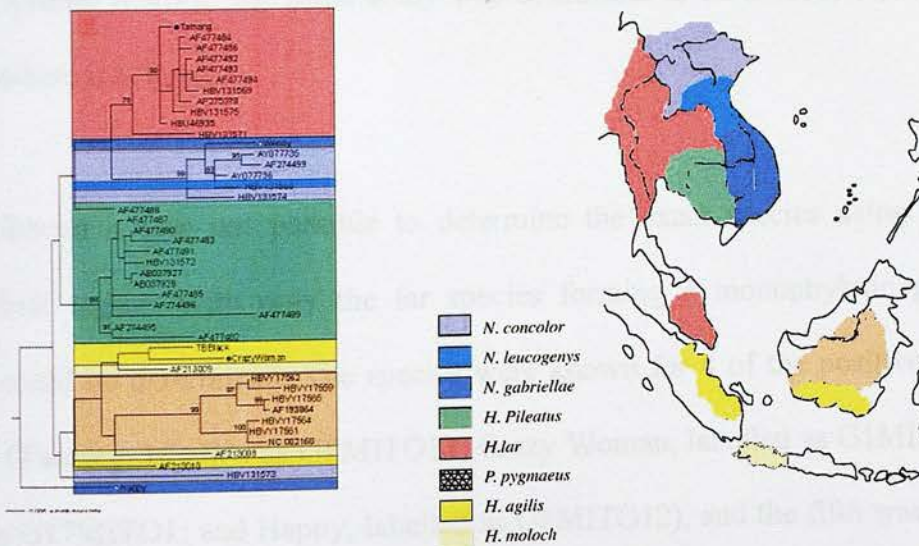


Figure 5-5. S gene phylogenetic analysis of gibbon and orang-utan HBV isolates, and the geographical distribution of these species within Southeast Asia. Some species specific clusters are seen, and these are supported by an AI value of 0.1774 (Table 5-5). Geographical segregation of species and HBV variants is more strongly supported (AI = 0.0349, Table 5-5).

### 5.3.3 Mitochondrial PCR and sequencing of primate samples

Mitochondrial 12S region PCR and sequencing was carried out in an attempt to determine the subspecies of primate samples for which this information was unavailable (Fig. 5-6). For the chimpanzees it was possible to differentiate between troglodytes, verus and vellerosus subspecies using sequence data from this region. Troglodytes is the central African subspecies of chimpanzee. Most of the animals screened in this study originated in Cameroon, and as expected most were of the troglodytes subspecies. One of the positive chimpanzees from this study was found to be of the vellerosus subspecies (Osang, originating in Nigeria), and the other was of the troglodytes subspecies (Louisa, a chimpanzee from Cameroon, labelled as L21 in the mitochondrial DNA phylogenetic tree). This analysis also showed that chimp 2 (L18) (MacDonald et al., 2000) was in fact not of the verus subspecies as claimed in the publication, but of the troglodytes subspecies. This has since been confirmed by the sanctuary. Chimp 4 from that same study was confirmed to be a verus chimpanzee by the sanctuary where it is kept.

For the gibbons it was not possible to determine the exact species using this particular mitochondrial region, with only the lar species forming a monophyletic group, but the subgenus could be determined. The species were known for 4 of the positive gibbons from this study (Tamang, labelled as GFMITO12; Crazy Woman, labelled as G1MITO12; Wendy, labelled as G17MITO1; and Happy, labelled as G8MITO12), and the fifth was determined to be of the hylobates subgenus (TB Black, labelled 6MITO12). A phylogenetic tree showing the species for all chimpanzee, gibbon and orang-utan samples and a selection of samples representing other species is shown in figure 5-4. This tree was not used to determine the species phylogenies used in the analysis of the relationship between the HBV variants and

their hosts. Previously published data with suggested gibbon phylogenies were used for that purpose.

It is not surprising that it was not possible to determine the exact species of the gibbons using this particular region of the mitochondria, as there is currently major controversy regarding the species phylogenies within the gibbon species group. Different genomic and mitochondrial regions have been investigated as potential species phylogeny determinants, but so far there is no agreement about which way is the best (Roos and Geissmann, 2001). A lot of species determination is still done in the field using morphology, fur colour and song pattern, as well as habitat. The gibbon species are all found in distinct habitats, with small overlapping regions in some cases. There is some interbreeding between species in the wild, but this phenomenon is much more common in captivity. Hybrids in the wild do not seem to form stable family connections and do not tend to produce any offspring (Nowak, ed., 1999).

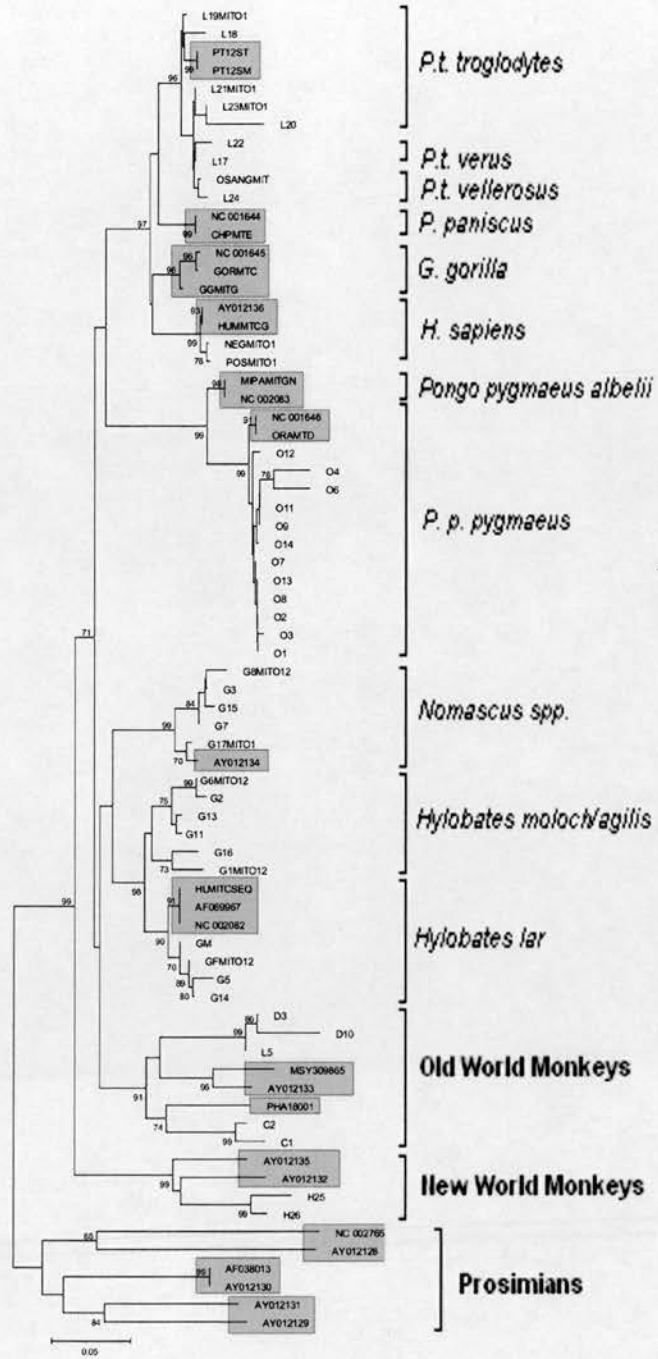


Figure 5-6. Phylogenetic analysis of mitochondrial 12S sequences obtained from study samples. Samples downloaded from GenBank are highlighted in the yellow boxes. Two human controls were also sequenced as part of the study. These are NEGMITO1 (serum negative for HBV) and POSMITO1 (serum positive for HBV). They were included as an internal contamination check, as they were also used as the positive and negative controls for HBV PCR screening.

#### 5.3.4 Geographical distribution of primates in Africa and Asia and HBV variants

As can be seen in the map in figure 5-5, African primates have geographically separated habitats. The only species with overlapping habitats are found within the central and eastern parts of Africa. The gorilla exists as three subspecies. *Gorilla gorilla gorilla* is also known as the western lowland gorilla, and this habitat is also shared with the troglodytes subspecies of chimpanzee. In the East there is also an overlap of habitat, this time between the schweinfurthii subspecies of chimpanzee and the two eastern species of gorilla; *G. g. graueri* (the eastern lowland gorilla) and *G. g. berengei* (the mountain gorilla). If we relate this to the species phylogeny of the African apes (Fig. 5-7b) we see that the East/West split of habitats is reflected in the genetic differences within both gorilla and chimpanzee subspecies populations. If we then relate this to the HBV trees from section 5.3.2 we see that the geographic distribution of the primates explains the HBV variant separation very well, even though the species phylogeny also fits the data.

The gibbon species are much more overlapping in their habitats, which can be seen from the map in figure 5-6. Not all species are shown on this map, only the ones from which HBV has been isolated have been selected. The species distribution of the subgenera is described below. Species where HBV has not been isolated are also described, to give a better picture of the distribution and interactions of these primates in the wild. Gibbons are divided into four subgenera; *nomascus*, *hylobates*, *bunopithecus* and *symphalangus*. These are determined by differences in the numbers of chromosomes.

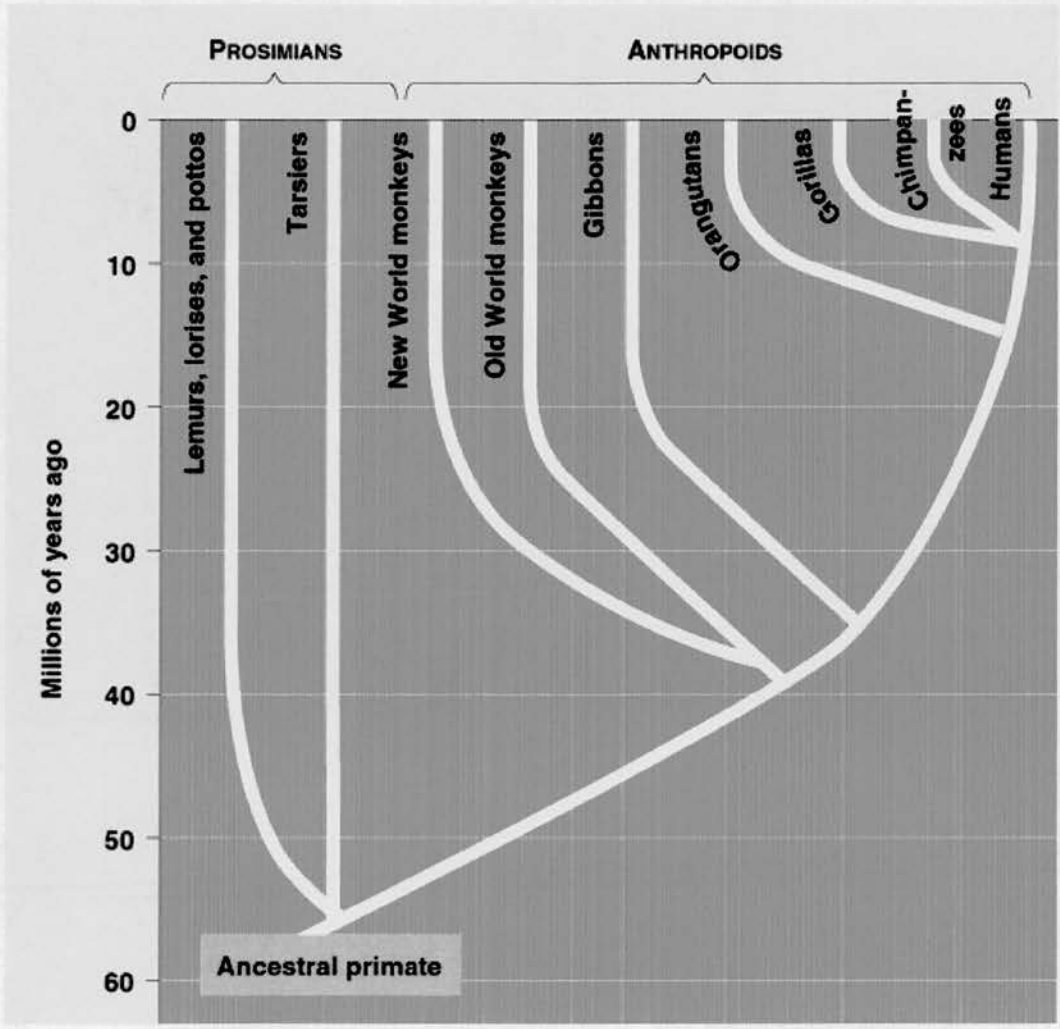
Subgenus *nomascus* is made up of three species; *concolor*, *leucogenys* and *gabriellae*. *Concolor* gibbons can be found in an area from south-eastern China to north-western Laos and northern Vietnam, as well as Hainan. *Leucogenys* can be found in southern Yunnan and throughout Central Laos and Central Vietnam. *Gabriellae* can be found in southern Laos,

through eastern Cambodia and southern and central Vietnam. Subgenus *hylobates* is made up of six species; *lar*, *pileatus*, *agilis*, *moloch*, *muelleri*, and *klossii*. *Lar* gibbons can be found in southern Yunnan through eastern and southern Burma, Thailand, eastern and southern mainland Malaysia and northern Sumatra. *Pilaetus* gibbons can be found in southeastern Thailand, south-western Laos and Cambodia. *Agilis* gibbons can be found in southern peninsular Thailand, north-western mainland Malaysia, southern Sumatra and south-western Borneo. *Moloch* gibbons are found only on Java and *muelleri* gibbons in all regions of Borneo except the southwest. *Klossii* gibbons can be found on a number of the Mentawai Islands off western Sumatra. Subgenus *bunopithecus* is made up of only one species, the hoolock gibbon, which can be found in areas of eastern Bangladesh, Burma and the border region of Yunnan. Subgenus *symphalangus* is made up of the *syndactylus* species, or siamang, which is found in mainland Malaysia and Sumatra. Orang-utans exist as two separate species, *Pongo pygmaeus pygmaeus* which can be found on Borneo, and *P. p. albelii* which is found on Sumatra.

If we relate this information to the species phylogeny of the Asian primates (Fig. 5-8) we see that the geographical distribution of the species reflects the genetic differences. A rough line can be drawn placing *hylobates* species to the west and *nomascus* species to the east. If we then relate this to the trees from section 5.3.2, we again see that there is a better correlation between the geographical distributions of the species (though not necessarily the subgenus) and the HBV variants carried, than between the species phylogeny and HBV variant.

Figure 5-7. Phylogenetic trees showing the ancestral relationships between the major primate groups (A), and the African ape species (B).

A



© 1999 Addison Wesley Longman, Inc.

Figure 5-7a was taken from [http://www.mun.ca/biology/scarr/Phylogeny\\_of\\_Primata.htm](http://www.mun.ca/biology/scarr/Phylogeny_of_Primata.htm),

B

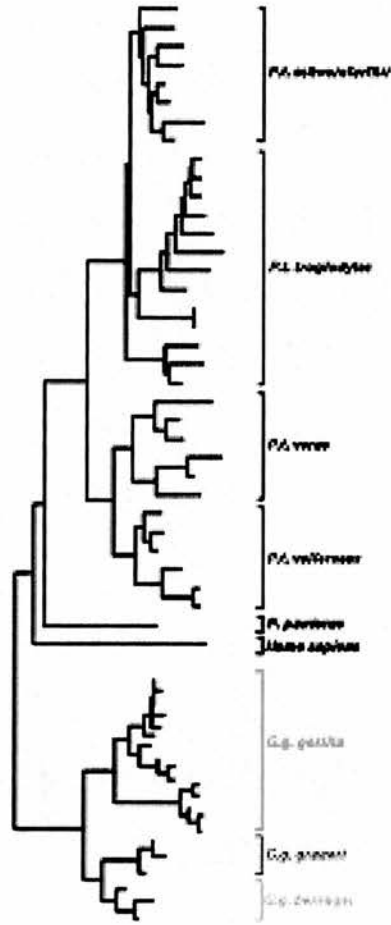


Figure 5-7b is an adapted composite of mitochondrial DNA data published by Gao et al (1999) and Hofreiter et al (2003).

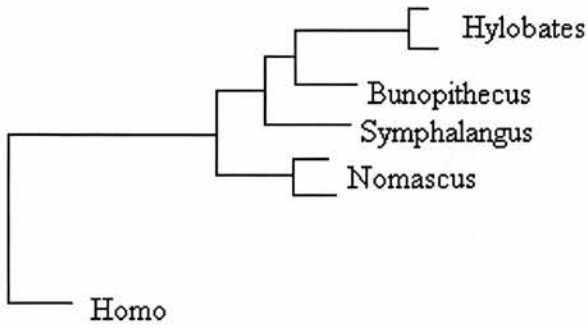


Figure 5-8. Phylogenetic tree showing the ancestral relationships between the gibbon species. This tree is adapted from a maximum likelihood tree based on mitochondrial control region DNA sequence data published by Roos & Geissmann (2001). This tree is rooted with a human sequence although the orang-utan is the next closest relative, as can be seen in figure 5-7a.

### 5.3.5 Association Index analysis of primate HBV genomes

AI analysis was used to investigate the groupings observed within the chimpanzee and gibbon/orang-utan clades respectively (Table 5-5).

As this analysis shows, there is strong support for both geographical and species determined HBV variant distribution within the chimpanzee/gorilla clade. This is seen both for complete genomes and the S gene region. The geographical determinant has a stronger support than the species determinant. The difference between the two is largest when analysing the complete genome. Within the gibbon/orang-utan clade the geographical separation of HBV types is most strongly supported, again particularly when looking at the complete genome.

	Complete genome Chimpanzee clade	Gibbon clade	S gene Chimpanzee clade	Gibbon clade
Species	0.1006	0.3248	0.0074	0.1774
Geography	0.0166	0.0043	0.0010	0.0349

Table 5-5. AI results for the geographical and species separation of HBV variants infecting the different species of primates. Species denotes AI analysis of groups of sequences defined by the species from which they were isolated, whereas geography denotes the arrangement of sequences into groups depending on the geographical origin of the host-species. The geographical ranges were defined as East and West for the chimpanzee clade analysis and North and South for the gibbon clade analysis.

### 5.3.6 Recombination analysis

Primate samples were investigated for the presence of recombinant sequences using two methods, AI bootscanning and SimPlot bootscanning. Control sequences for gibbon, orang-utan, gorilla and chimpanzee HBV were tested in the first instance to provide profiles for comparison. Samples from this study and divergent variants published by other groups were then analysed. At first gibbon and orang-utan were tested as separate groups, but after analysis of the orang-utan isolate gave a puzzling result with low association to both gibbon and orang-utan HBV (Figure 5-9B), they were also tested as one group (Figure 5-10). As can be seen from the control analysis using the AI method, combining the gibbon and orang-utan isolates into one group does not affect the control analysis of the gibbon isolate, whereas the result for the orang-utan is improved. This suggests that these two groups of isolates are very closely related, like two members of the same genotype.

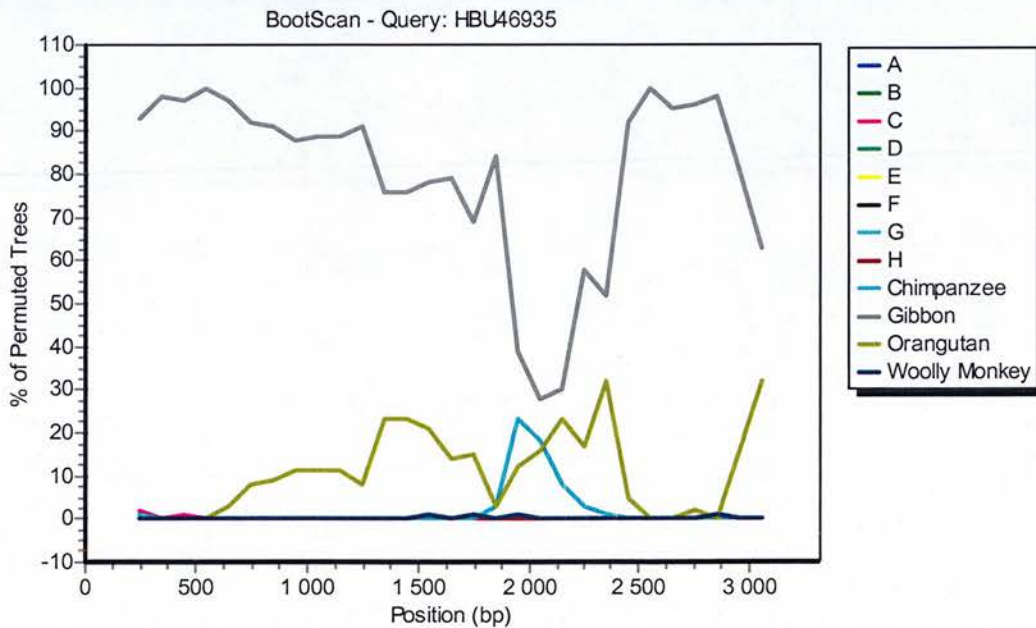
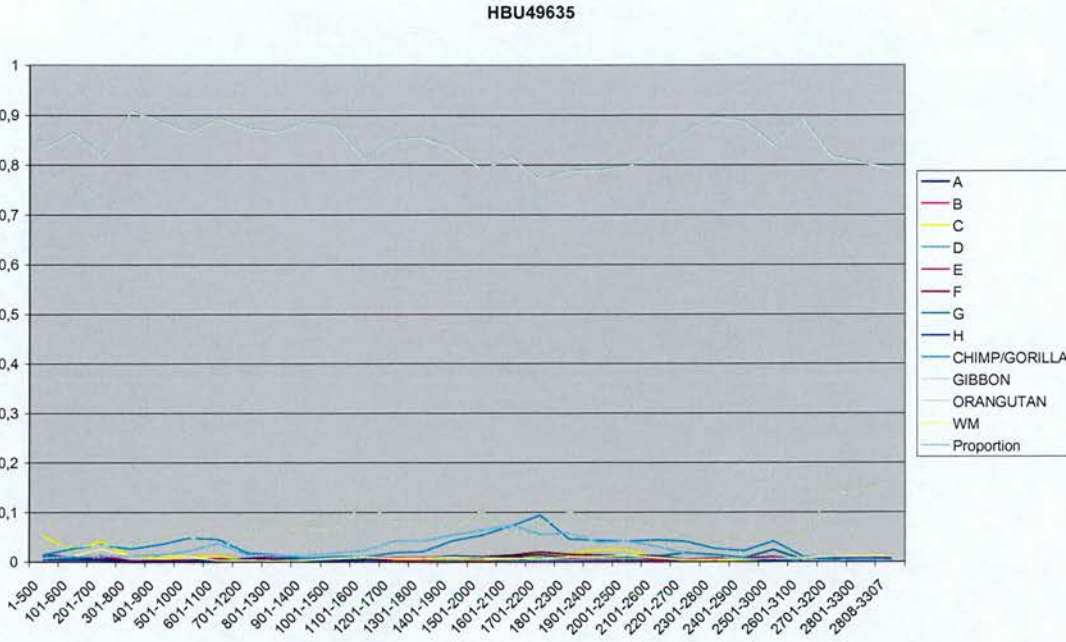
The SimPlot results for the gibbon isolate shows a greater degree of relatedness to the combined group than to the separate gibbon group. The SimPlot results for the orang-utan isolate is the opposite. It was decided to use the combined group in all further analyses. In the gibbon and orang-utan analyses a combined gorilla and chimpanzee group is used, after data showed these groups to be closely related also (see appendix E). Only the analysis of one lar gibbon from this study (Tamang) is shown here (Fig. 5-11). The remainder of the analyses can be found in appendix E.

The gorilla isolate showed a high degree of association with the chimpanzee isolates, except for part of the S gene region (bases 1-1000). A chimpanzee control sequence was also tested using both a combined gorilla and chimpanzee group and gorilla and chimpanzee separate.

There is no significant difference between the two analyses using either the AI method or SimPlot, so it was decided to use the combined group for all further analysis.

Figure 5-9. AI and SimPlot analysis of gibbon (A) and orang-utan (B) control sequences, using separate groups for these species.

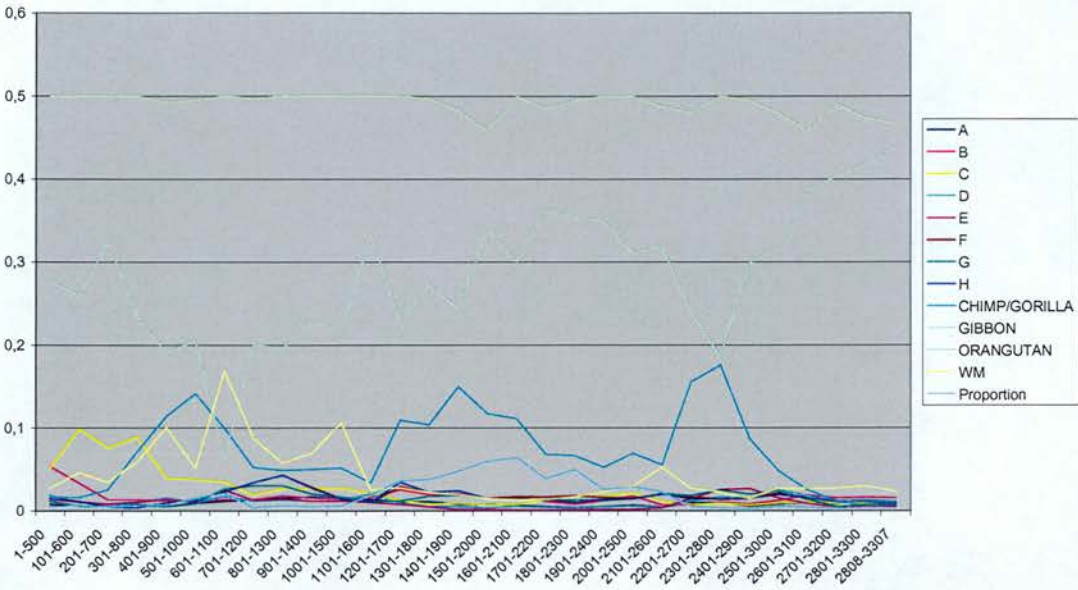
A



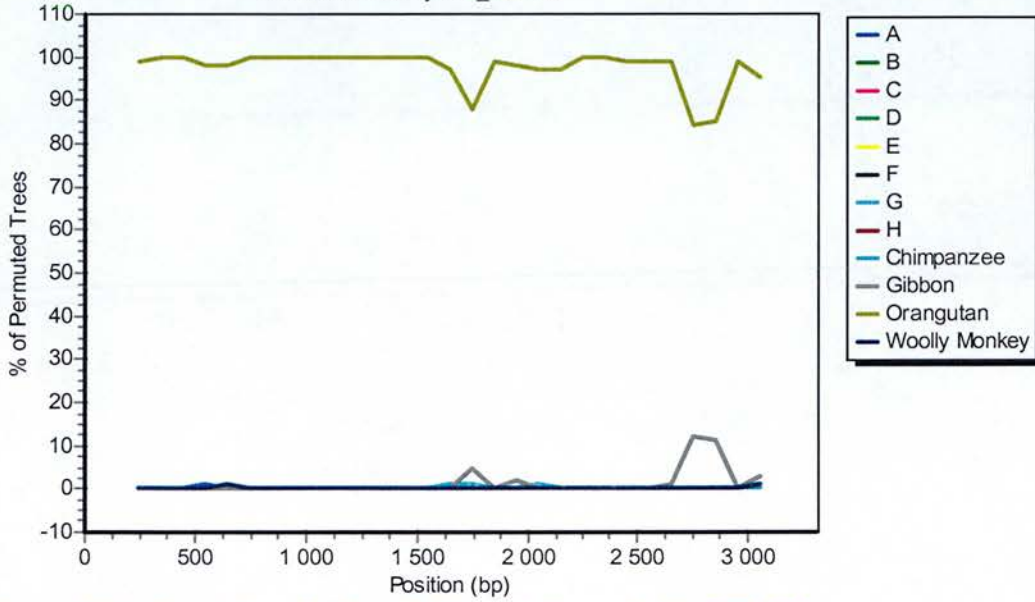
Window : 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

B

NC\_002168



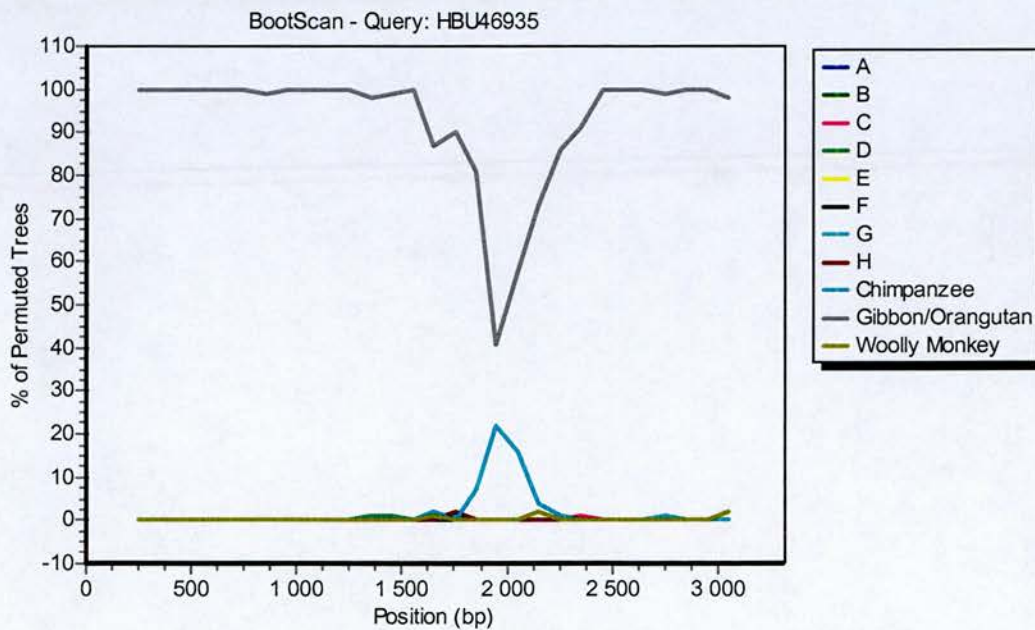
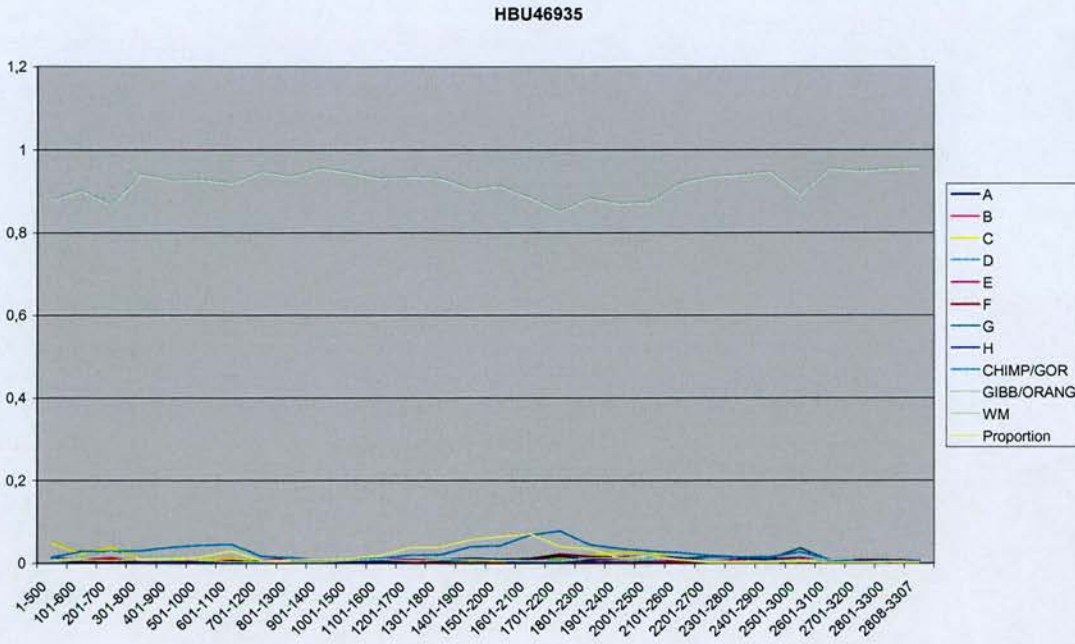
BootScan - Query: NC\_002168



Window : 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

Figure 5-10. AI and SimPlot analysis of gibbon (A) and orang-utan (B) control sequences, with these species together as one group.

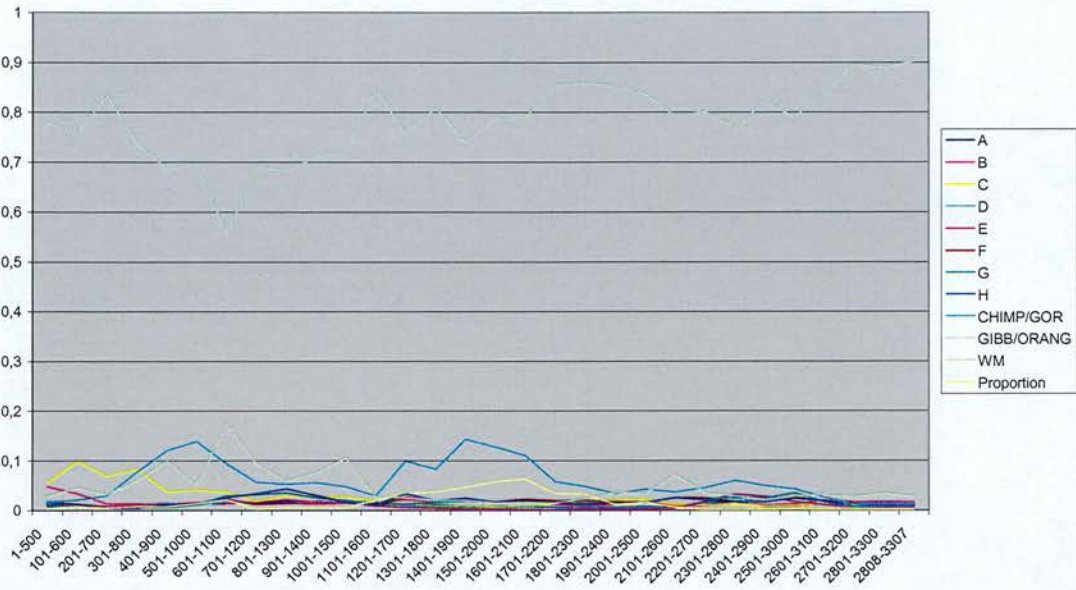
A



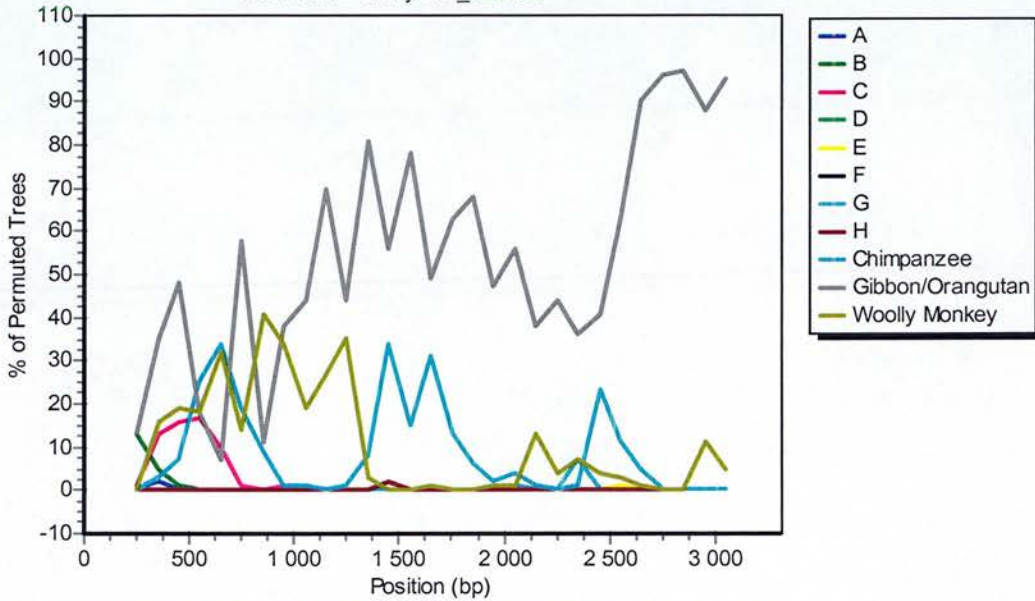
Window : 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

B

NC\_002168



BootScan - Query: NC\_002168



Window: 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

In the AI analysis the isolates from chimpanzees Osang and Louisa both show a strong association with the chimpanzee/gorilla group, with profiles resembling that of the control HPBVCG. SimPlot shows this as well, but does also show a larger amount of variation within the S gene region for both isolates, although more so for the Louisa isolate. Neither isolate shows any evidence of possible recombination.

The isolate from gibbon Tamang shows a great resemblance in analysis profile with the control sequence HBU46935, both in the AI and SimPlot analyses, and shows no signs of recombination. There are no areas of the genome that appear divergent from previous data. The SimPlot analysis does identify a region within the core gene with a higher degree of variation. This region coincides with the region of the HBV genome where the support for the human/primate split is weaker.

The isolate from gibbon Wendy again shows a high degree of agreement between the two analyses. SimPlot analysis identifies an increase in sequence variation within the S gene, but this cannot be seen in the AI analysis. No evidence to suggest recombination is seen.

The isolate from gibbon TB Black shows much more divergence from the control group, especially in the S gene and core gene. These areas are identified in both analyses. There is a lot of discrepancy between the results of the two analyses, with AI showing the isolate as divergent and SimPlot as a recombinant with human genotype C in the S gene and chimpanzee HBV in the core gene.

The isolate from gibbon Crazy Woman also shows a degree of divergence in the core gene that is identified in both analyses. The remainder of the genome shows a high degree of relatedness to other gibbon/orang-utan isolates in the AI analysis. SimPlot identifies a second

area of divergence in the S gene. Although divergent, this isolate is not shown to be a recombinant in the SimPlot analysis or the AI analysis.

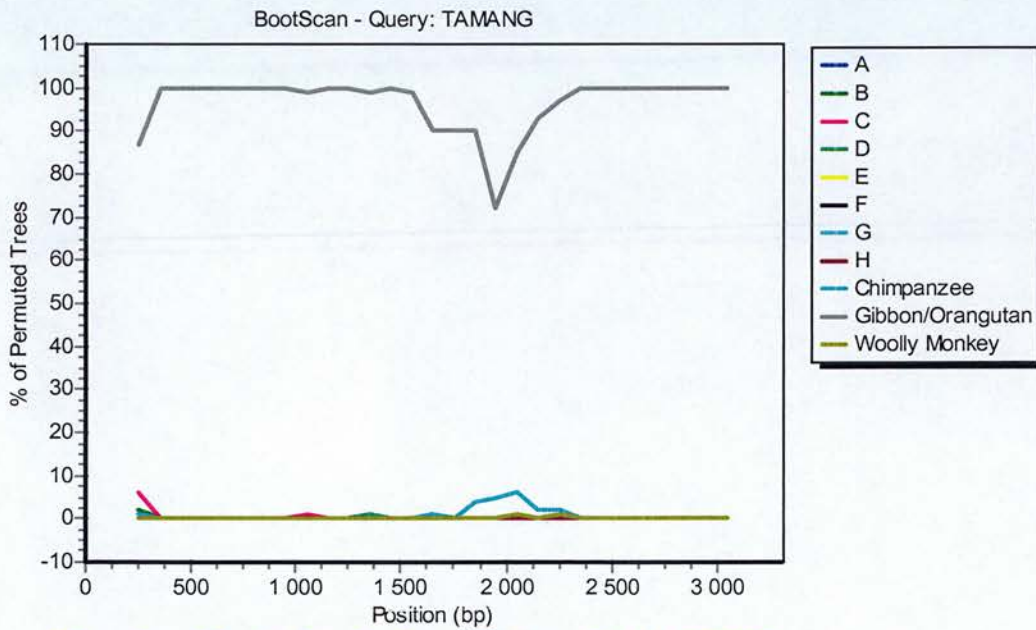
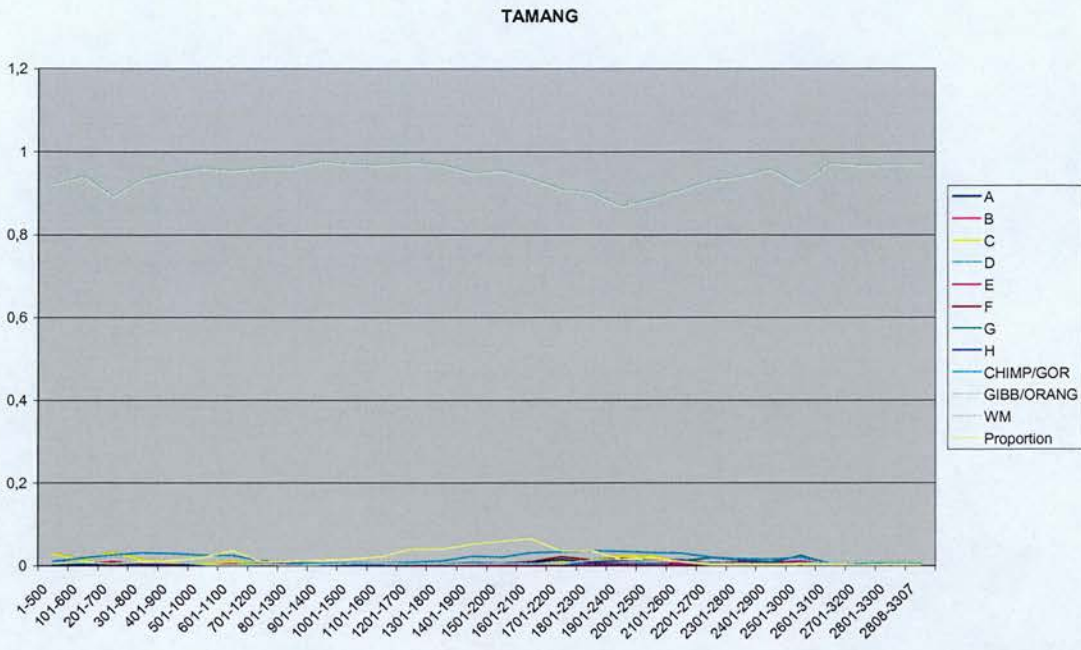
The isolate from gibbon Happy again shows a higher degree of divergence, especially in the S gene. SimPlot again identifies a second area of divergence, again in the core gene, that is not seen in the AI analysis, but does not show the isolate as a recombinant either.

Isolate AB032431 is a chimpanzee isolate that groups with human genotype E (see Fig. 5-3). Both AI and SimPlot analysis show the sequence to be very closely related to genotype E over the entire genome, with no signs of recombination with HBV of any other genotype.

Isolate HBV131575 is a chimpanzee isolate that groups with the gibbon isolates. AI and SimPlot analysis both show a high degree of relatedness to the other gibbon/orang-utan isolates. SimPlot does however identify one region of increased divergence in the core gene not shown by the AI data. Neither analysis suggests a recombination event.

Isolate AB046525 is a chimpanzee isolate from a troglodytes subspecies animal. It was reported as having a divergent core gene (Takahashi et al., 2001). This divergence is clearly shown in both AI and SimPlot analysis. AI analysis does not suggest recombination with any of the known genotypes, although there is a small region where the closest relationship is with gibbon HBV. SimPlot also fails to show a distinct recombination event, although there is one small region most closely related to the woolly monkey isolate and another with genotype G. None of these regions show strong enough association to suggest recombination in either the AI or the SimPlot analysis.

Figure 5-11. A and SimPlot analysis of the complete genome sequence of HBV isolated from gibbon Tamang.



Window : 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

#### 5.4 Discussion and conclusions

When the results of this study are added to a number of previous studies that report screening data for primates it becomes clear that HBV infection is prevalent in a very restricted range of primate species. The frequency of infection in chimpanzees and gibbons found in this study are high, at 22.7 and 25% respectively. Other studies have found similarly high frequencies in chimpanzees, gorillas, gibbons and orang-utans (Grethe et al., 2000; Lanford et al., 2000; Makuwa et al., 2003; Noppornpanth et al., 2003; Warren et al., 1998; Warren et al., 1999). When data from multiple studies are combined an overall frequency of active infection of 18.7% for all ape species is obtained, and those for the individual species are all over 15%. Human populations in endemic areas have active infection rates that compare to these figures. The fact that HBV appears to be endemic in other ape species apart from humans is not the only parallel that can be drawn between these infections.

Transmission patterns may also be the same. In humans vertical transmission is the main mode of transmission in Asia, whereas in Africa horizontal transmission predominates. Evidence for both these modes of transmission has been reported for gibbons (Grethe et al., 2000; Lanford et al., 2000; Noppornpanth et al., 2003). No such data have been published for chimpanzees or gorillas, as most frequently the animals sampled have not been kept in large groups/family groups, but rather represent orphaned and confiscated unrelated animals. However, since such similarities exist between human and gibbon HBV transmission, one may speculate that a similar situation will occur in African apes as well.

This study failed to identify the presence of a HBV-like virus in any of the Old World monkey species screened. A recent study showed identical results following screening of a large number of animals representing 9 different species (Makuwa et al., 2003). To date HBV

negative screening results have been published for 15 different species. This evidence suggests that these species do not carry HBV. One cannot, however, definitely conclude the absence of HBV in these monkeys. There are a large number of monkey species in the Old World, and the 15 tested represent only a fraction of these. One explanation for the lack of HBV in Old World monkeys so far is that we simply haven't found the carrier species yet.

Another explanation for this lack of HBV detection could be that these species actually do carry HBV, only the genomes are so divergent they cannot be detected using current serological and molecular screening techniques, possibly even more divergent than the woolly monkey isolate (Lanford et al., 1998). To investigate this possibility a new PCR assay was developed, using primers specific for the S gene region of the viral genome and conserved enough to enable detection of HBV from the rodent species. This alternative screening strategy failed to identify any HBV positive animals among the Old World monkeys in this study. This strengthens the evidence that suggests that apes are the only carriers of HBV. Further screening of more different species is necessary before any further conclusions can be made regarding the presence or absence of HBV in Old World monkeys. The fact that rodents and birds carry HBV-like viruses suggests the existence of a monkey virus.

HBV infection in New World primates is another issue that requires resolution. To date 10 woolly monkeys, all captives, are the only examples of HBV infection in primates other than apes (Lanford et al., 1998). Only four species have so far been screened for HBV, so the same argument that applies to Old World monkeys has to be applied. There may be new genotypes of HBV still to be discovered, with unknown primate carriers. Also, all positive animals are captives, so there is no evidence to support the existence of woolly monkey HBV in the wild. It is a possibility that the infection could have arisen in captivity through cross-species

transmission from another unidentified primate species. Woolly monkeys do seem to suffer liver disease and mortality due to this infection, although the frequency of this is not known, as many deaths due to liver disease were assigned as HBV deaths without performing any tests (Lanford et al., 1998).

Phylogenetically the isolates from non-human primates fall into species-specific genotypes as described by previous studies (Grethe et al., 2000; Hu et al., 2001; Takahashi et al., 2001; Verschoor et al., 2001). There also appear to exist sub-types within these genotypes, determined partly by the sub-species of primate which is the carrier. This is supported by mitochondrial data both from this study and others (Hu et al., 2001). The close relationship between the gorilla isolate and those from chimpanzees is confirmed. An analogous relationship emerges in Southeast Asia, where gibbon and orang-utan isolates cluster together, which contradicts the previous claims that orang-utans are infected with a specific genotype of HBV (Verschoor et al., 2001; Warren et al., 1999). There appears to be some separation of isolates according to the species of origin, although this does not mirror the phylogenetic relationship of the species (Fig. 5-8). If a North/South split is made according to the geographical range of the habitats of the species this corresponds well with the branches within this complete genome tree. More information about the geography of gibbons and orang-utans is found in section 5.3.4.

There is a lack of diversity between the orang-utan isolates, and this coupled with the fact that this study failed to detect HBV in 14 orang-utans suggests that HBV may not be so widespread in this species. This is highly indicative of a more recent origin and cross-species transmission, as in both Africa and Asia there are regions of habitat overlap between species infected with HBV. Chimpanzees and gorillas share habitat ranges in Africa, and gibbon species and orang-utans share habitats in Southeast Asia..

As has been suggested recently (Fares and Holmes, 2002) humans may be the vector responsible for this spread. No other primate moves over the distances that humans are able to, although this has dramatically increased in the last few hundred years or so. Humans are still however, the most widespread primate in the world as well as the most prevalent, and they do come into contact with other primate species regularly. Using the estimates rate of sequence change of  $2.1 \times 10^{-5}$  nucleotide changes/site/year, the divergence of the HBV variants infecting the different subspecies of chimpanzee can be dated at 1600 to 2300 years ago, which agrees with the 6000 years ago divergence date for the hominid HBV viruses suggested by Fares (2002).

Although HBV appears to exist as sub-species variants within the different primate HBV genotypes, the HBV phylogeny and the host phylogeny do not always agree. The relationship between the African apes and the HBV isolates that they carry is closer than that seen in Asia, both when analysing complete genomes and S gene sequences. In Asia there appears to be a greater concordance between the actual geographical origin of the host species and the HBV variant rather than the species. This geographical separation is seen also in Africa. To investigate the sub-clades seen in chimpanzee and gibbon HBV the Association Index analysis was employed. For the African apes both species separation and geographical separation are strongly supported, whereas in Asia the geographical separation is most strongly supported. This existence of geographical variants is another parallel between human and non-human primate HBV, but is also a parallel that can be drawn with avian and mammalian HBV (Pult et al., 2001b; Testut et al., 1996). The importance of such geographical variants still needs to be investigated.

Recombination between HBV isolates from different genotypes is seen in human HBV infection, but not in the other primate HBV genotypes. This may be due to the fact that humans from different geographical areas, and humans in general, come in contact with other carriers more frequently than other species of primates do. Due to human expansion throughout Asia and Africa the habitats for the other apes have been decimated, and where once there were large regions of overlap between habitats there are now ever decreasing such regions. Humans are known to be able to become superinfected with more than one genotype of HBV. This has not been seen in any of the HBV infected primates detected to date. If primates are able to become superinfected, there is the possibility that recombinants could arise. The fact that we have not seen any recombinant sequences may also be due to the rarity of this event. Recombinants are rare in human populations, and far fewer primates have been HBV tested than humans. The fact that chimpanzees, gorillas, gibbons and orang-utans are all on the endangered species list also means the likelihood of contact between different species, even though there is some habitat overlap, is also decreasing. More extensive sampling of primates for HBV research may turn up recombinant genomes.

Cross-species transmission certainly does appear to be possible. The infectivity of human HBV in other species of primate has been well documented. In this study we have confirmed one case of human genotype E infection in a chimpanzee (Takahashi et al., 2000) as a non-recombinant virus of human origin. The absence of human genotypes in infected primates screened in this study and others support this. If humans and primates shared genotypes, or did so in the past, one would expect to find evidence for this in sequences isolated from these primates. Cross-species transmission between other species of primate also appears possible, as the gibbon-like HBV isolate from a chimpanzee (Grethe et al., 2000) shows. This animal was kept with gibbons in a Zoo, and the recombination analysis shows that this isolate is only

of gibbon origin, with no chimpanzee HBV segments. From this same study (Grethe et al., 2000) an isolate from a concolor gibbon was among those published. In the phylogenetic analysis carried out as part of this project this isolate (HBV131573) did not group with isolates from other concolor gibbons. This animal was also a captive animal, and may have caught HBV from another gibbon species whilst in captivity. The close relationship between gorilla and chimpanzee isolates, and gibbon and orang-utan isolates also suggests cross-species transmission. The existence of a highly divergent chimpanzee isolate (AB046525) (Takahashi et al., 2001) with a core gene that is not closely related to any HBV genotypes isolated to date is also highly suggestive of cross-species transmission, although this could also be due to the existence of another chimpanzee genotype, AB046525 being the first example of this. This isolate hints at the possible existence in nature of highly divergent HBV isolates, in as yet unidentified primate hosts. No HBV-like viruses have so far been reported from Old World rodents, and therefore these species must also be considered as a possible source of this infection.

Recent cross-species transmissions between species is one possible explanation for the distribution of the HBV variants of the non-human primates. Another possible explanation is that the genotypes seen today arose during the evolution of the primate species. The rate of sequence change estimated for HBV does not agree with this. The chimpanzee subspecies started diverging around 1.5 million years ago (Morin et al., 1994), yet the HBV variants infecting them can be calculated to have diverged between 1600 and 2200 years ago. However, the constraints that exist for the evolution of the HBV genome (Mizokami et al., 1997) may prevent the fixation of many mutations into the genome. However, when non-coding regions of the HBV genome are used to calculate divergence low rates are found, and a more recent divergence date of 6000 years for hominoid HBV is obtained (Fares and

Holmes, 2002). The close relationship between the gibbon and orang-utan HBV isolates, and a potentially analogous relationship between chimpanzee and gorilla HBV, also do not fit into this model. Another trait difficult to explain is the number of different genotypes that infect humans, although genotypes A-E and G have a monophyletic origin, the position of F and H is not accounted for. If a population age for modern humans is taken to be 150 000 years (Stringer, 2002) and the divergence of chimpanzees as 6 million years ago there is a discrepancy between the level of diversity seen for the HBV isolates. There is less diversity within the chimpanzee clade than within the human HBV group, even though there has been much more time to evolve. The sizes of chimpanzee populations are vastly smaller than those seen for humans, placing a constraint on the number of different isolates that are able to circulate within these populations. This in turn will lead to a decreased immune pressure on the HBV genome to evolve.

To get a truer picture of the HBV carrier rate in wild populations of primates new strategies for sampling must be developed. Non-invasive sampling of faecal matter and hair from nest sites are the best way to sample a large number of individuals without disruption. This type of sampling has proved effective for SIV researchers (Santiago et al., 2003a; Santiago et al., 2003b). Recently a novel assay for HBV DNA in faecal samples was developed (Makuwa et al., 2003), but only one positive sample was identified out of 36 tested. As a high rate of active infection was detected using plasma samples this may not be a sensitive enough assay. More sequence data is needed from gorillas and orang-utans to further evaluate the relationship between HBV carried by these species and HBV carried by their neighbouring species. The screening of further species for the presence of HBV, such as bonobos and the two eastern species of gorilla, as well as more monkey species would also be very helpful in

the search for the origin of HBV. The development of more sensitive non-invasive assays like that described above are needed to facilitate this research.

A recent origin of HBV in the different primates with frequent cross-species transmission events is proposed, but the role of humans in this needs to be investigated.

# **Chapter 6**

## **Discussion and Conclusions**

## 6.1 Frequency of viraemia and epidemiology

HBV infection appears to be endemic in both human and non-human primate populations. Rates of infection of 15% or more can be found in human populations, and current primate data suggests this level of infection in the ape species as well. The level of diversity seen within the chimpanzee and gibbon HBV genotypes is also indicative of endemicity. The distribution of genotypes in human populations was as expected with genotypes A and E in Africa and C and D in Papua New Guinea. The ape isolates found in this study also fell within the expected genotypes for each species. The apparent restriction of HBV to apes is surprising, considering the existence of HBV viruses in rodents and birds. Recombination analysis and conventional phylogenetic analysis failed to identify any evidence for the presence of primate-like viruses in humans or human viruses in any of the primates. Subgroups exist within the non-human primate genotypes, and when looking at chimpanzee HBV for example, these groups can be compared to the different genotypes that exist for human HBV.

The transmission networks responsible for maintaining HBV infection within a host population can be either horizontal or vertical, and both these types have been demonstrated for gibbons in captivity (Noppornpanth et al., 2003). In Asia the predominant route of infection is from mother-to-child, whereas in Africa horizontal, or contact, transmission plays a larger part. Until this type of epidemiological data is published for chimpanzees we can only assume that transmission follows the same routes as has been demonstrated for gibbons and humans. The development of sensitive non-invasive screening methods is needed to obtain this type of data. Similar studies on gibbons in the wild would also provide valuable information. This type of sampling will be able to provide a better picture of the actual rates of infection in the wild populations of primates, and will be able to identify any new viruses

carried by ape species from which none have been reported to date, such as bonobos and the East African gorilla species. As humans continue to expand their villages, towns and cities further, the habitats of other primates are disappearing. The collection of epidemiological data for infections occurring naturally in these primates is very important. As human and primate habitats get closer the chance of encounters increases. Cross-species transmission of primate viruses to humans have been documented, and as bushmeat is gaining in popularity in Africa the risks to human health have to be evaluated (Brooks et al., 2002; Lerche et al., 2001; Peeters et al., 2002).

## 6.2 Geographical distribution of genotypes

A further parallel between human and non-human primate HBV is the existence of geographically defined clusters of related isolates. In chimpanzees and gorillas these also correspond very well with the actual species or subspecies that is infected. The groups seen within the primate genotypes can be compared to the different human genotypes. Human genotypes A to E and G form a single monophyletic group within the phylogenetic tree, and the distribution of the genotypes is geographically defined. The outlying position of the genotype F clade is difficult to explain. There is also geographical distribution of variants within human genotypes A, C, and D, and evidence has recently been published suggesting a similar scenario for genotype F in Central and South America (Devesa et al., 2004). The geographically defined subgroups may play a role in differences in infectivity and pathogenicity of HBV variants in different populations. Molecular characterization of these variants will be important in future investigations of HBV pathogenicity and treatment.

These geographical variants are also analogous to the variants seen in the non-human primate genotypes. The geographical origin appears to be more important than species in determining which HBV variant is found infecting populations. The fact that the human genotypes show geographical distribution supports this. The geographical variants within the human genotypes suggest that transmission networks are localised. Differences in host factors may help explain why this occurs. The geographical distribution of the HBV genotypes infecting humans and other apes holds clues as to the origins of these viruses. This will be discussed later.

The AI analysis carried out in these investigations will provide a helpful tool for researchers looking to compare differences between patient groups. Certain types of isolates may be more commonly associated with acute infection, whereas other types may be more associated with a chronic carrier state, e antigen expressors may have different types than those infected individuals that do not express e antigen. When carrying out these analyses one must remember that the genotype of the virus variant may not be directly responsible for the phenotype seen in the infected individual, host factors play an important role and must be considered. AI analysis can be applied to multivariate data sets to compare both host and viral genotypes with the infection phenotype.

### 6.3 Recombination within HBV genomes

A recombinant HBV genome is one which contains distinct genetic elements with the characteristics of different and distinct genotypes. For recombination to occur, several criteria have to be met. Multiple genotypes must exist, superinfection with more than one genotype within the same individual must be possible and occurring, and a mechanism for recombination must exist. For HBV recombination may take place during replication of the viral genome, by template swapping.

Recombination between HBV genomes from different genotypes does occur, within the human genotypes. Little evidence has been published to date to suggest this occurs also for non-human primate genotypes. This may simply be due to the relative lack of available sequence data for this HBV group in comparison to human HBV. One chimpanzee isolate does appear to be recombinant (Takahashi et al., 2001), but the genotype responsible for donating that part the core gene remains to be identified. For human HBV recombination occurs most often between genotypes A and D, and B and C. Considering the geographical distributions of these genotypes these combinations are not unexpected. The human genotypes split into three main groups when analysed phylogenetically (Fig. 5-3). One containing the South American genotypes F and H, one containing genotypes D, E, and G, and one containing A, B, and C. The close relationship seen between genotype F and H is also seen between genotypes B and C, and between D and E. This suggests three ancestral human HBV viruses. The recombination seen between B and C has been proposed as a recent event (Fares and Holmes, 2002), but considering the phylogenetic relationship

between these two genotypes it may also be a remnant of a common ancestor. This could also explain the A and D recombinants.

Recombinants exist that, like the chimpanzee isolate described above, contain sequence elements that do not belong to any of the known genotypes (Hannoun et al., 2000b) although initial analyses identified these segments as genotype A. This suggests that recombinant sequences exist that have not been correctly identified, which may confuse the genotyping of isolates and lead to false identification of recombinants. The close relationship seen between chimpanzee and gorilla HBV, and gibbon and orang-utan HBV suggests recombination may have played a role in the evolution of these genotypes, in combination with cross-species transmission. HBV may have initially existed as totally distinct species-specific variants, but through cross-species transmission and recombination these have evolved into a larger, less species-specific genotype infecting the African apes. When genotype G was first reported it was thought to be a recombinant genotype (Stuyver et al., 2000). This all suggests that recombination may be playing a large and complex role in the evolution of HBV by creating further variability among circulating virus populations. Further research needs to be carried out to clarify the mechanisms involved in recombination and the impact this has on the pathology of HBV infection.

#### 6.4 Origins and evolution of HBV within primates, human and non-human

Each of the primate genotypes has a single origin, and a similar observation can be made for human genotypes A-E and G, whereas genotypes F and H have a separate origin to all other HBV variants. The relationship between F and H is similar to that between B and C, and D and E. If the “Out of Africa” hypothesis (Magnius and Norder, 1995; Norder et al., 1994) is applied to this data F and H would represent the oldest HBV variants. If these genotypes existed at the time of migration out of Africa and were spread as far as the New World, the geographical isolation of that region until around 400 years ago could explain why they are so divergent. The more continuous movement of humans through Europe, Africa and Asia, due to the continuous landmass making access to different areas easier, would give more opportunity for diversification into further genotypes. Founder effects following migration of populations to new geographically separated areas would lead to a smaller number of variants circulating among susceptible individuals. This would mean that they would evolve independently of other virus populations in other areas. Cross-species transmission and recombination between viral genomes could then have given rise to new genotypes. This seems to fit the hypothesis, but it fails to explain why gibbon and chimpanzee HBV variants are more closely related to each other than they are to the human genotypes that can be found in the same geographical areas.

If the “New World origin” hypothesis (Bollyky et al., 1997) is applied to the data it does not fit. The amount of diversity within each genotype is too great to have been generated in such a short time-span, and the geographical distribution of the

genotypes, and subtypes within the genotypes, is very difficult to explain if this was the case. Genotypes F and H are restricted in their distribution to mainly native populations in Central and South America. Genotypes F and H have not been reported in any other countries, and this study failed to identify any in the countries surveyed. Other genotypes can be found in the Americas, but F and H are prevalent, especially in native populations (Devesa et al., 2004; Pineiro y Leone et al., 2003). No data published to date supports this hypothesis for the origin of HBV.

Cross-species transmission does appear to have played a part in the emergence of the primate genotypes in Africa and Southeast Asia, but there is no evidence to date to suggest cross-species transmission between humans and other primate species. There is, however, no evidence to support the hypothesis of co-speciation/co-evolution as proposed previously (MacDonald et al., 2000). The relationship between the ape genotypes is closer than the relationship between genotypes from the same geographical areas, i.e. between genotypes B and C and gibbon HBV in Asia, and genotype E and chimpanzee HBV in Africa. If the position of genotypes F and H were not as outliers, but as part of the monophyletic group that contains all the other human genotypes, coevolution would be a possibility, as gibbon HBV appears to have diverged before chimpanzee HBV, and the human genotypes were the last to diverge. This agrees with the speciation of the primates. Genotypes F and H do remain outliers however, making this scenario unlikely.

The data from this study has provided evidence for the correlation between the geographical origin and sequence divergence. In non-human primates this seems to

cut across the species-barriers, with African apes sharing a common genotype and in Asia overlapping habitat ranges between gibbons and orang-utans result in the clustering of HBV isolates. A scenario is proposed for the relatively recent spread of HBV among the African and Southeast Asian primates, with cross-species transmission between animals in overlapping habitats. This accounts much better for the geographical association, rather than species association, of HBV genotypes in these primates than previous co-speciation hypotheses (MacDonald et al., 2000; Magnusius and Norder, 1995; Norder et al., 1994). This agrees most closely with the recent origin proposed by Fares (Fares and Holmes, 2002), and the role of humans in the dissemination of HBV throughout the world needs to be more fully evaluated.

## 6.5 Conclusions

Several different hypotheses have been proposed for the origin of HBV, but reconstructing the evolutionary history of human and non-human primate HBV has proven difficult. The evolution of the viral genome is subjected to constraints on the level of nucleotide change that can occur without a detrimental effect on virus replication, recombination between different genotypes is occurring, and the genotypes show geographical distribution which in humans also is associated with differences in pathology. This study, like previous studies, has demonstrated the existence of HBV variants which group interspersed with the human genotypes. No ape isolates have been identified that are as divergent as the human genotypes F and H or the woolly monkey isolate. One chimpanzee isolate which has a highly divergent core sequence (Takahashi et al., 2001) hints at the existence of such divergent variants, but the species that is its host remains to be identified. Further work is necessary to identify the original species distribution of these divergent variants, and identifying them will help resolve the evolutionary history, origin, and past epidemiology of HBV in primates, including humans. There is also no evidence of cross-species transmission of HBV between humans and other primate species in their natural habitats. These findings deepen the mystery of HBV origins and evolution in humans, and provide a context for ongoing studies of HBV biological variability and genotype-associated differences in pathogenicity and outcomes of infection.

# **Chapter 7**

## **Bibliography**

Aiba, N, Nishimura, H, Arakawa, Y., and Abe, K (2003). Complete nucleotide sequence and phylogenetic analyses of hepatitis B virus isolated from two pileated gibbons. *Virus Genes* **27**, 219-226.

Alestig, E, Hannoun, C., Horal, P., and Lindh, M (2001). Phylogenetic origin of hepatitis B virus strains with precore C-1858 variant. *Journal of Clinical Microbiology* **39**, 3200-3203.

Alestig, E, Hannoun, C., Horal, P., and Lindh, M (2001). Hepatitis B virus genotypes in Mongols and Australian Aborigines). *Archives of Virology* **146**, 2321-2329.

Andre, F. (2000). Hepatitis B epidemiology in Asia, the Middle East and Africa. *Vaccine* **18 Suppl 1**, S20-S22.

Angus, P, Vaughan, R, Xiong, S, Yang, H, Delaney, W, Gibbs, C, Brosgart, C, Colledge, D, Edwards, R, Ayres, A, Bartholomeusz, A, and Locarnini, S (2003). Resistance to adefovir dipivoxil therapy associated with the selection of a novel mutation in the HBV polymerase. *Gastroenterology* **125**, 292-297.

Arankalle, V. A, Murhekar, K. M, Gandhe, S. S, Murhekar, M. V, Ramdasi, A. Y, Padbidri, V., and Sehgal, S. C (2003). Hepatitis B virus: predominance of genotype D in primitive tribes of the Andaman and Nicobar islands, India (1989-1999). *Journal of General Virology* **84**, 1915-1920.

Arauz-Ruiz, P., Norder, H., Visona, K. A., and Magnus, L. O. (1997). Molecular epidemiology of hepatitis B virus in Central America reflected in the genetic variability of the small S gene. *Journal of Infectious Diseases* **176**, 851-858.

Arauz-Ruiz, P., Norder, H., Visona, K. A., and Magnus, L. O. (1997). Genotype F prevails in HBV infected patients of Hispanic origin in Central America and may carry the precore stop mutant. *Journal of Medical Virology* **51**, 305-312.

Arauz-Ruiz, P., Norder, H., Robertson, B. H., and Magnus, L. O. (2002). Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. *Journal of General Virology* **83**, 2059-2073.

Baptista, M, Kramvis, A, Jammeh, S, Naicker, J, Galpin, JS, and Kew MC (2003). Follow up of infection of chacma baboons with inoculum containing a and non-a genotypes of hepatitis B virus. *World Journal of Gastroenterology* **9**, 731-735.

Bläckberg, J and Kidd-Ljunggren, K (2003). Mutations within the hepatitis B virus genome among chronic hepatitis B patients with hepatocellular carcinoma. *Journal of Medical Virology* **71**, 18-23.

Blitz, L., Pujol, F. H., Swenson, P. D., Porto, L., Atencio, R., Araujo, M., Costa, L., Monsalve, D. C., Torres, J. R., Fields, H. A., Lambert, S., Van Geyt, C., Norder, H., Magnus, L. O., Echevarria, J. M., and Stuyver, L. (1998). Antigenic diversity of hepatitis B virus strains of genotype F in Amerindians and other population groups from Venezuela. *Journal of Clinical Microbiology* **36**, 648-651.

Blow, J. A, Turell, M. J, Walker, E. D, and Silverman, A (2002). L. Post-bloodmeal diuretic shedding of hepatitis B virus by mosquitoes (Diptera: Culicidae). *Journal of Medical Entomology* **39**, 605-612.

Blumberg, B. S, Larouze, B, London, W. T, Werner, B, Hesser, J. E, Millman, I, Saimot, G, and Payet, M (1975). The relation of infection with hepatitis B agent to primary hepatic carcinoma. *American Journal of Pathology* **81**, 669-682.

Blumberg, B. S, Hesser, J. E, Economidou, I, Hadziyannis, S., Gioustozi, A, Heliakis, P, and Livadas, D (1975). The variety of responses within a community to infection with Australia (hepatitis B) antigen. *Developments in Biological Standardization* **30**, 270-283.

Blumberg, B. S (2004). Hepatitis B; the hunt for a killer virus. U.S.A, Princeton University Press.

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.

Bollyky, P. L., Rambaut, A., Grassly, N., Carman, W. F., and Holmes, E. C (1997). Hepatitis B virus has a new world evolutionary origin. *Hepatology* **26**, 765.

Bollyky, P. L. and Holmes, E. C. (1999). Reconstructing the complex evolutionary history of hepatitis B virus. *Journal of Molecular Evolution* **49**, 130-141.

Botta, A, Lu, M, Zhen, X, Kemper, T, and Roggendorf, M (2000). Naturally occurring woodchuck hepatitis virus (WHV) deletion mutants in chronically WHV-infected woodchucks. *Virology* **277**, 226-234.

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**, 1719-1729.

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.

Breiner, K. M, Urban, S, and Schaller, H (1998). Carboxypeptidase D (gp180), a golgi-resident protein, functions in attachment and entry of avian hepatitis B viruses. *Journal of Virology* **72**, 8098-8104.

Brind, A, Jiang, J, Samuel, D, Gigou, M, Feray, C, Brechot, C., and Kremsdorf, D (1997). Evidence for selection of hepatitis B mutants after liver transplantation through peripheral blood mononuclear cell infection. *Journal of Hepatology* **26**, 228-235.

Brooks, J. I, Rud, E. W, Pilon, R. G, Smith, J. M, Switzer, W. M, and Sandstrom, P. A. (2002). Cross-species retroviral transmission from macaques to human beings. *The Lancet* **360**, 387-388.

Chang, D, Sugimoto, C, Wang, M, Tsai, R-T, and Yogo, Y (1999). JC virus genotypes in a Taiwan aboriginal tribe (Bunun): implications for its population history. *Archives of Virology* **144**, 1081-1090.

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.

Chang, S. F, Netter, H. J, Hildt, E, Schuster, R, Schaefer, S., Hsu, Y-C, Rang, A, and Will, H (2001). Duck hepatitis B virus expresses a regulatory HBx-like protein from a hidden open reading frame. *Journal of Virology* **75**, 161-170.

Charrel, R. N., de Micco, P., and de Lamballerie, X (1999). Phylogenetic analysis of GB viruses A and C: evidence for cospeciation between virus isolates and their primate hosts. *Journal of General Virology* **80**, 2329-2335.

Chen, B-F, Kao, J-H, Liu, C-J, Chen, D-S, and Chen, P-J (2004). Genotypic dominance and novel recombinations in HBV genotype B and C coinfecting intravenous drug users. *Journal of Medical Virology* **73**, 13-22.

Chen, B-F, Chen, P-J., Jow, G-M, Sablon, E, Liu, C-J, Chen, D-S, and Kao, J-H (2004). High prevalence of mixed genotype infections in hepatitis B virus infected intravenous drug users. *Journal of Medical Virology* **74**, 536-542.

Chen, H. S, Kew, M. C., Hornbuckle, W. E, Tennant, B. C, Cote, P. J, Gerin, J. L, Purcell, R. H., and Miller, R. H (1992). The precore gene of the woodchuck hepatitis virus genome is not essential for viral replication in the natural host. *Journal of Virology* **66**, 5682-5684.

Chironna, M., Germinario, C., Lopalco, P. L., Quarto, M., and Barbuti, S. (2000). HBV, HCV and HDV infections in Albanian refugees in Southern Italy (Apulia region). *Epidemiology and Infection* **125**, 163-167.

Chu, C-j, Hussain, M, and Lok, A. S. F (2002). Hepatitis B virus genotype B is associated with earlier HBeAg seroconversion compared with hepatitis B virus genotype C. *Gastroenterology* **122**, 1756-1762.

Cooper, A, Paran, N, and Shaul, Y (2003). The earliest steps in hepatitis B virus infection. *Biochimica et Biophysica Acta* **1614**, 89-96.

Cova, L., Lambert, V, Chevallier, A, Hantz, O., Fourel, I, Jaquet, C, Pichoud, C., Boulay, J, Chomel, B, Vitvitski, L., and et al (1986). Evidence for the presence of duck hepatitis B virus in wild migrating ducks. *Journal of General Virology* **67**, 537-547.

Cui, C., Shi, J., Hui, L., Xi, H., Zhuoma, Quni, Tsedan, and Hu, G. (2002). The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. *Journal of General Virology* **83**, 2773-2777.

Dane, D. S, Cameron, C. H, and Briggs, M (1970). Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *Lancet* **1**, 695-698.

Davis, H. L., Suparto, I. I., Weeratna, R. R., Jumintarto, Iskandriati, D. D., Chamzah, S. S., Ma'ruf, A. A., Nente, C. C., Pawitri, D. D., Krieg, A. M., Heriyanto, Smits, W., and Sajuthi, D. D. (2000). CpG DNA overcomes hyporesponsiveness to hepatitis B vaccine in orangutans. *Vaccine* **18**, 1920-1924.

Deinhardt, F. (1976). Hepatitis in primates. *Advances in Virus Research* **20**, 113-157.

Devesa, M, Rodriguez, C, Leon, G, Liprandi, F., and Pujol, F. H (2004). Clade analysis and surface antigen polymorphism of hepatitis B virus American genotypes. *Journal of Medical Virology* **72**, 377-384.

Diao, J, Slaney, D. M, and Michalak, T. I (2003). Modulation of the outcome and severity of hepadnaviral hepatitis in woodchucks by antibodies to hepatic asialoglycoprotein receptor. *Hepatology* **38**, 629-638.

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

Ding, X., Mizokami, M, Ge, X, Orito, E., Iino, S, Ueda, R, and Nakanishi, M (2002). Different hepatitis B virus genotype distributions among asymptomatic carriers and patients with liver diseases in Nanning, southern China. *Hepatology Research* **22**, 37-44.

Ding, X., Park, Y. N., Taltavull, T. C., Thung, S. N., Jin, X, Jin, Y, Trung, N. S, Edamoto, Y., Sata, T, and Abe, K (2003). Geographic characterization of hepatitis B virus infections, genotyping of hepatitis B virus, and p53 mutation in hepatocellular carcinoma analyzed by in situ detection of viral genomes from carcinoma tissues: comparison among six different countries. *Japanese Journal of Infectious Diseases* **56**, 12-18.

Ding, X., Gu, H, Zhong, Z-H, Zilong, X, Tran, H. T-T, Iwaki, Y, Li, T-C, Sata, T, and Abe, K (2003). Molecular epidemiology of hepatitis viruses and genotypic distribution of hepatitis B and C viruses in Harbin, China. *Japanese Journal of Infectious Diseases* **56**, 19-22.

Eichberg, J. W. and Kalter, S. S. (1980). Hepatitis A and B: serologic survey of human and nonhuman primate sera. *Laboratory Animal Science* **30**, 541-543.

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

Feitelson, M. A, Millman, I, Halbherr, T, Simmons, H, and Blumberg, B. S (1986). A newly identified hepatitis B type virus in tree squirrels. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 2233-2237.

Felsenstein, J. (1993). "PHYLIP Inference Package version 3.5." Department of Genetics, University of Washington, Seattle.

Feng, G, Yue, L, White, A. T, Pappas, P. G, Barchue, J, Greene, B. M, Sharp, P. M, Shaw, G. M, and Hahn, B. H (1992). Human infection by genetically diverse SIVsm-related HIV-2 in West Africa. *Nature* **358**, 495-499.

Flodgren, E., Bengtsson, S., Knutsson, M., Strebkova, E. A., Kidd, A. H., Alexeyev, O. A., and Kidd-Ljunggren, K. (2000). Recent high incidence of fulminant hepatitis in Samara, Russia: molecular analysis of prevailing hepatitis B and D virus strains. *Journal of Clinical Microbiology* **38**, 3311-3316.

Fortuin, M., Chotard, J, Jack, A. D, Maine, N. P, Mendy, M, Hall, A. J, Inskip, H. M, George, M. O, and Whittle, H. C (1993). Efficacy of hepatitis B virus vaccine in the Gambian expanded programme on immunisation. *The Lancet* **341**, 1129-1131.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M., and Hahn, B. H. (1999). Origins of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* **397**, 436-441.

Georges-Courbot, M. C, Lu, C. Y, Makuwa, M, Telfer, P, Onanga, R, Dubreuil, G, Chen, Z, Smith, S. M, Georges, A, Gao, F, Hahn, B. H, and Marx, P. A (1998). Natural infection of a household pet red-capped mangabey (*Cercopithecus torquatus torquatus*) with a new simian immunodeficiency virus. *Journal of Virology* **72**, 600-608.

Gheit, T, Sekkat, S, Cova, L., Chevallier, M., Petit, M. A, Hantz, O., Lesenechal, M, Benslimane, A, Trepo, C., and Chemin, I (2002). Experimental transfection of *Macaca sylvanus* with cloned human hepatitis B virus. *Journal of General Virology* **83**, 1645-1649.

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.

Grgacic, E. V. L (2002). Identification of structural determinants of the first transmembrane domain of the small envelope protein of duck Hepatitis B virus essential for particle morphogenesis. *Journal of General Virology* **83**, 1635-1644.

Guo, J, Kitamura, T, Ebihara, H, Sugimoto, C, Kunitake, T, Takehisa, J., Qun Na, Y, Al-Ahdal, M. N, Hallin, A, Kawabe, K, Taguchi, F, and Yogo, Y (1996). Geographical distribution of the human polyomavirus JC virus types A and B and isolation of a new type from Ghana. *Journal of General Virology* **77**, 919-927.

Hannoun, C., Horal, P., and Lindh, M. (2000). Long-term mutation rates in the hepatitis B virus genome. *Journal of General Virology* **81**, 75-83.

Hannoun, C., Norder, H., and Lindh, M. (2000). An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam. *Journal of General Virology* **81**, 2267-2272.

Hannoun, C., Krogsgaard, K., Horal, P., and Lindh, M. (2002). Genotype mixtures of hepatitis B virus in patients treated with interferon. *Journal of Infectious Diseases* **186**, 752-759.

Hattwell, J. N and Sharp, P. M (2000). Evolution of human polyomavirus JC. *Journal of General Virology* **81**, 1191-1200.

Heckel, J. O., Rietschel, W., and Hufert, F. T (2001). Prevalence of hepatitis B virus infections in nonhuman primates. *Journal of Medical Primatology* **30**, 14-19.

Hino, O, Kitagawa, T, Koike, K, Kobayashi, M., Hara, M, Mori, W, Nakashima, T, Hattori, N, and Sugano, H (1984). Detection of hepatitis B virus DNA in hepatocellular carcinomas in Japan. *Hepatology* **4**, 90-95.

Hofreiter, M, Siedel, H, Van Neer, W, and Vigilant, L (2003). Mitochondrial DNA sequence from an enigmatic gorilla population (*Gorilla gorilla uellensis*). *American Journal of Physical Anthropology* **121**, 361-368.

Hollinger, F. B. Hepatitis B virus. Fields, B. N, Knipe, D. M, Howley, P. M, et al, and Eds. *Field's Virology* (3rd edition), 2739-2807. 1996. Philadelphia, Lippincott-Raven Publishers.

Hu, W. S, Rhodes, T, Dang, Q, and Pathak, V (2003). Retroviral recombination: review of genetic analyses. *Frontiers in Bioscience* **8**, 143-155.

Hu, X., Margolis, H. S., Purcell, R. H., Ebert, J., and Robertson, B. H. (2000). Identification of hepatitis B virus indigenous to chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America*. **97**, 1661-1664.

Hu, X., Javadian, A., Gagneux, P., and Robertson, B. H. (2001). Paired chimpanzee hepatitis B virus (ChHBV) and mtDNA sequences suggest different ChHBV genetic variants are found in geographically distinct chimpanzee subspecies. *Virus Research* **79**, 103-108.

Huy, T. T-T, Ushijima, H, Quang, V. X, Win, K. M., Leungrojanakul, P, Kikuchi, K, Sata, T, and Abe, K (2004). Genotype C of hepatitis B virus can be classified into at least two subgroups. *Journal of General Virology* **85**, 283-292.

Hwang, G-Y, Huang, C-J, Lin, C-Y, and Wu, C-C (2003). Dominant mutations of hepatitis B virus variants in hepatoma accumulate in B-cell and T-cell epitopes of the HBx antigen. *Virus Research* **92**, 157-164.

Iavarone, M, Trabut, J-B, Delpuech, O, Carnot, F, Colombo, M., Kremsdorf, D, Brechot, C., and Thiers, V (2003). Characterisation of hepatitis B virus X protein mutants in tumour and non-tumour liver cells using laser capture microdissection. *Journal of Hepatology* **39**, 253-261.

Ishikawa, T and Ganem, D (1995). The pre-S domain of the large viral envelope protein determines host range in avian hepatitis B viruses. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 6259-6263.

Jilbert, A. R and Kotlarski, I (2000). Immune response to duck hepatitis B virus infection. *Developmental and Comparative Immunology* **24**, 285-302.

Kane, A, Lloyd, J, Zaffran, M, Simonsen, L, and Kane, M (1999). Transmission of hepatitis B, hepatitis C and human immunodeficiency viruses through unsafe injections in the developing world: model-based regional estimates. *Bulletin of the World Health Organization* **77**, 801-807.

Kao, J-H, Wu, N-H, Chen, P-J., Lai, M-Y, and Chen, D-S (2000). Hepatitis B virus genotypes and the response to interferon therapy. *Journal of Hepatology* **33**, 998-1002.

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.

Kao, J. H., Chen, P. J., Lai, M. Y., and Chen, D. S (2002). Genotypes and clinical phenotypes of hepatitis B virus in patients with chronic hepatitis B virus infection. *Journal of Clinical Microbiology* **40**, 1207-1209.

Kato, A, Kitamura, T, Sugimoto, C, Ogawa, Y, Nakazato, K, Nagashima, K, Hall, W. W, Kawabe, K, and Yogo, Y (1997). Lack of evidence for the transmission of JC polyomavirus between human populations. *Archives of Virology* **142**, 875-882.

Kato, H., Orito, E., Sugauchi, F., Ueda, R., Gish, R. G., Usuda, S., Miyakawa, Y., and Mizokami, M. (2001). Determination of hepatitis B virus genotype G by polymerase chain reaction with hemi-nested primers. *Journal of Virological Methods* **98**, 153-159.

Kato, H., Orito, E., Gish, R. G., Sugauchi, F., Suzuki, S., Ueda, R., Miyakawa, Y., and Mizokami, M. (2002). Characteristics of hepatitis B virus isolates of genotype G and their phylogenetic differences from the other six genotypes (A through F). *Journal of Virology* **76**, 6131-6137.

Kato, H., Orito, E., Gish, R. G., Bzowej, N., Newsom, M., Sugauchi, F., Suzuki, S., Ueda, R., Miyakawa, Y., and Mizokami, M. (2002). Hepatitis B e antigen in sera from individuals infected with hepatitis B virus of genotype G. *Hepatology* **35**, 922-929.

Kato, H., Ruzibakiev, R., Yuldasheva, N., Hegay, T., Kurbanov, F., Achundjanov, B., Tuichiev, L., Usuda, S., Ueda, R., and Mizokami, M. (2002). Hepatitis B virus genotypes in Uzbekistan and validity of two different systems for genotyping. *Journal of Medical Virology* **67**, 477-483.

Kawsar, M. and Goh, B. T. (2002). Hepatitis B virus infection among Chinese residents in the United Kingdom. *Sexually Transmitted Diseases* **78**, 166-168.

Kedda, M. A., Kramvis, A., Kew, M. C., Lecatsas, G., Paterson, A. C., Aspinall, S., Stark, J. H., De Klerk, W. A., and Gridelli, B. (2000). Susceptibility of chacma baboons (*Papio ursinus orientalis*) to infection by hepatitis B virus. *Transplantation* **69**, 1429-1434.

Kew, M. C., Miller, R. H., Chen, H. S., Tennant, B. C., and Purcell, R. H. (1993). Mutant woodchuck hepatitis virus genomes from virions resemble rearranged hepadnaviral integrants in hepatocellular carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 10211-10215.

Kidd-Ljunggren, K., Courouce, A. M., Oberg, M., and Kidd, A. H. (1994). Genetic conservation within subtypes in the hepatitis B virus pre-S2 region. *Journal of General Virology* **75**, 1485-1490.

Kidd-Ljunggren, K., Oberg, M., and Kidd, A. H. (1995). The hepatitis B virus X gene: analysis of functional domain variation and gene phylogeny using multiple sequences. *Journal of General Virology* **76**, 2119-2130.

Kirk, G. D., Camus-Randon, A.-M., Mendy, M., Goedert, J. J., Merle, P., Trepo, C., Brechot, C., Hainaut, P., and Montesano, R. (2000). Ser-249 p53 mutations in plasma DNA of patients with hepatocellular carcinoma from the Gambia. *Journal of the National Cancer Institute* **92**, 148-153.

Koike, K., Tsutsumi, T., Fujie, H., Shintani, Y., and Kyoji, M. (2002). Molecular mechanism of viral hepatocarcinogenesis. *Oncology* **62 Suppl 1**, 29-37.

Koschel, M., Thomssen, R., and Bruss, V. (1999). Extensive mutagenesis of the hepatitis B virus core gene and mapping of mutations that allow capsid formation. *Journal of Virology* **73**, 2153-2160.

Kramvis, A., Weitzmann, L., Owiredu, W. K., and Kew, M. C. (2002). Analysis of the complete genome of subgroup A' hepatitis B virus isolates from South Africa. *Journal of General Virology* **83**, 835-839.

Kumar, S., Tamura, K., Jakobsen, I. B., and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics*. **17**, 1244-1245.

Lanford, R. E., Michaels, M. G., Chavez, D., Brasky, K., Fung, J., and Starzl, T. E (1995). Persistence of extrahepatic hepatitis B virus DNA in the absence of detectable hepatic replication in patients with baboon liver transplants. *Journal of Medical Virology* **46**, 207-212.

Lanford, R. E., Chavez, D., Brasky, K. M., Burns, R. B. 3., and Rico-Hesse, R. (1998). Isolation of a hepadnavirus from the woolly monkey, a New World primate. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 5757-5761.

Lanford, R. E., Chavez, D., Rico-Hesse, R., and Mootnick, A. (2000). Hepadnavirus infection in captive gibbons. *Journal of Virology* **74**, 2955-2959.

Langer, B. C. A, Frösner, G. G, and von Brunn, A (1997). Epidemiological study of viral hepatitis types A, B, C, D and E among Inuit's in West Greenland. *Journal of Viral Hepatitis* **4**, 339-349.

Lee, M-O, Choi, Y-H, Shin, E-C, Kang, H-J, Kim, Y-M, Jeong, S-Y, Seong, J. K, Yu, D-Y, Cho, H, Park, J. H, and Kim, S. J (2002). Hepatitis Bi virus X protein induced expression of interleukin 18 (IL-18): a potential mechanism for liver injury caused by hepatitis B virus (HBV) infection. *Journal of Hepatology* **37**, 380-386.

Lerche, N. W, Switzer, W. M, Yee, J. L, Shanmugam, V, Rosenthal, A. N, Chapman, L. E, Folks, T. M, and Heneine, W (2001). Evidence of infection with simian type D retrovirus in persons occupationally exposed to nonhuman primates. *Journal of Virology* **75**, 1783-1789.

Lin, B and Anderson, D. A (2000). A vestigial X open reading frame in duck hepatitis B virus. *Intervirology* **43**, 185-190.

Lindh, M., Andersson, A-S, and Gusdal, A (1997). Genotypes, nt 1858 variants, and geographic origin of hepatitis B virus- large-scale analysis using a new genotyping method. *The Journal of Infectious Diseases* **175**, 1285-1293.

Lindh, M., Gonzales, J. E, Norkrans, G., and Horal, P (1998). Genotyping of hepatitis B virus by restriction pattern analysis of a pre-S amplicon. *Journal of Virological Methods* **72**, 163-174.

Lindh, M., Hannoun, C., Dhillon, A. P, Norkrans, G., and Horal, P (1999). Core promoter mutations and genotypes in relation to viral replication and liver damage in East Asian hepatitis B virus carriers. *The Journal of Infectious Diseases* **179**, 775-782.

Linnemann, C. C., Jr., Kramer, L. W., and Askey, P. A. (1984). Familial clustering of hepatitis B infections in gorillas. *American Journal of Epidemiology* **119**, 424-430.

Liu, C. J., Kao, J. H., Chen, P. J., Lai, M. Y., and Chen, D. S. (2002). Molecular epidemiology of hepatitis B viral serotypes and genotypes in Taiwan. *Journal of Biomedical Science* **9**, 166-170.

- Lole, K. S, Bollinger, R. C, Paranjape, R. S, Gadkari, D, Kulkarni, S. S, Novak, N. G, Ingersoll, R, Sheppard, H. W, and Ray, S. C (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology* **73**, 152-160.
- Lu, M, Hilken, G, Kruppenbacher, J, Kemper, T, Schirmbeck, R, Reiman, J, and Roggendorf, M (1999). Immunization of woodchucks with plasmids expressing woodchuck hepatitis virus (WHV) core antigen and surface antigen suppresses WHV infection. *Journal of Virology* **73**, 281-289.
- MacDonald, D. M., Holmes, E. C., Lewis, J. C., 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.
- Magnius, L. O. and Norder, H. (1995). Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. *Intervirology* **38**, 24-34.
- Makuwa, M, Souquiere, S, Telfer, P, Leroy, E, Bourry, O, Rouquet, P, Clifford, S, Wickings, E. J, Roques, P, and Simon, F (2003). Occurrence of hepatitis viruses in wild-born non-human primates: a 3 year (1998-2001) epidemiological survey in Gabon. *Journal of Medical Primatology* **32**, 307-314.
- 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 Academy of Sciences of the United States of America* **77**, 2941-2945.
- 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.
- Mbayed, V. A., Lopez, J. L., Telenta, P. F., Palacios, G., Badia, I., Ferro, A., Galoppo, C., and Campos, R. (1998). Distribution of hepatitis B virus genotypes in two different pediatric populations from Argentina. *Journal of Clinical Microbiology* **36**, 3362-3365.
- Mbayed, V. A., Barbini, L., Lopez, J. L., and Campos, R. H. (2001). Phylogenetic analysis of the hepatitis B virus (HBV) genotype F including Argentine isolates. *Archives of Virology* **146**, 1803-1810.
- Michaels, M. G., Lanford, R., Demetris, A. J., Chavez, D., Brasky, K., Fung, J., and Starzl, T. E. (1996). Lack of susceptibility of baboons to infection with hepatitis B virus. *Transplantation* **61**, 350-351.
- Michaels, M. G., Jenkins, F. J, StGeorge, K, Nalesnik, M. A, Starzl, T. E, and Rinaldo, C. R. Jr (2001). Detection of infectious baboon cytomegalovirus after baboon-to-human liver xenotransplantation. *Journal of Virology* **75**, 2825-2828.
- Mimms, L. T., Solomon, L. R, Ebert, J. W., and Fields, H (1993). Unique preS sequence in a gibbon-derived hepatitis B virus variant. *Biochemical and Biophysical Research Communications* **195**, 186-191.
- Miyakawa, Y and Mizokami, M (2003). Classifying hepatitis B virus genotypes. *Intervirology* **46**, 329-338.

Mizokami, M, Nakano, T, Orito, E., Tanaka, Y., Sakugawa, H, Mukaide, M, and Robertson, B. H (1999). Hepatitis B virus genotype assignment using restriction fragment length polymorphism patterns. *FEBS Letters* **450**, 66-71.

Mizokami, M., Orito, E., Ohba, K., Ieko, K., Lau, J. Y., and Gojobori, T. (1997). Constrained evolution with respect to gene overlap of hepatitis B virus. *Journal of Molecular Evolution* **44 Suppl 1**, S83-90.

Morin, P. A., Moore, J. J, Chakraborty, R, Jin, L, Goodall, J, and Woodruff, D. S (1994). Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* **265**, 1193-1201.

Morozov, V, Pisavera, M, and Groudinin, M (2000). Homologous recombination between different genotypes of hepatitis B virus. *Gene* **26**, 55-65.

Murhekar, M. V, Murhekar, K. M, Arankalle, V. A, and Sehgal, S. C (2002.). Epidemiology of hepatitis B virus infection among the Nicobarese- a mongoloid tribe of the Andaman and Nicobar Islands, India. *Epidemiology and Infection* **128**, 465-471.

Naito, H, Hayashi, S., and Abe, K (2001). Rapid and specific genotyping system for hepatitis B virus corresponding to six major genotypes by PCR using type-specific primers. *Journal of Clinical Microbiology* **39**, 362-364.

Nakai, K, Win, K. M., Oo, S. S, Arakawa, Y., and Abe, K (2001). Molecular characteristic-based epidemiology of hepatitis B, C, and E viruses and GB virus C/hepatitis G virus in Myanmar. *Journal of Clinical Microbiology* **39**, 1536-1539.

Nakamoto, Y, Guidotti, L. G, Kuhlen, C. V, Fowler, P, and Chisari, F. V (1998). Immune pathogenesis of hepatocellular carcinoma. *Journal of Experimental Medicine* **188**, 341-350.

Nakano, T, Lu, L, Hu, X, Mizokami, M, Orito, E., Shapiro, C. N., Hadler, S. C, and Robertson, B. H (2001). Characterization of hepatitis B virus genotypes among Yuupa Indians in Venezuela. *Journal of General Virology* **82**, 359-365.

Nakayoshi, T, Maeshiro, T, Nakayoshi, T, Nakasone, H, Sakugawa, H, Kinjo, F, Orito, E., and Mizokami, M (2003). Difference in prognosis between patients infected with hepatitis B virus with genotype B and those with genotype C in the Okinawa Islands: a prospective study. *Journal of Medical Virology* **70**, 350-354.

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. *Journal of General Virology* **74**, 1627-1632.

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, Arauz-Ruiz, P., Blitz, L., Pujol, F. H., Echevarria, J. M., and Magnius, L. O (2003). The T(1858) variant predisposing to the precore stop mutation correlates with one of two major genotype F hepatitis B virus clades. *Journal of General Virology* **84**, 2083-2087.

Norder, H., Courouce, A. M., and Magnius, L. O. (1992). Molecular basis of hepatitis B virus serotype variations within the four major subtypes. *Journal of General Virology* **73**, 3141-3145.

Norder, H., Hammas, B., Lofdahl, S., Courouce, 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**, 1201-1208.

Norder, H., Hammas, B., Lee, S. D., Bile, K., Courouce, A. M., Mushahwar, I. K., and Magnius, L. O. (1993). Genetic relatedness of hepatitis B viral strains of diverse geographical origin and natural variations in the primary structure of the surface antigen. *Journal of General Virology* **74**, 1341-1348.

Norder, H., Courouce, A. M., and Magnius, L. O. (1993). Complete nucleotide sequences of six hepatitis B viral genomes encoding the surface antigen subtypes ayw4, adw4q-, and adrq- and their phylogenetic classification. *Archives of Virology* **8 Suppl**, 189-199.

Norder, H., Courouce, 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.

Nowak, R. M ed. Walker's Primates of the World. 1999. United States of America, Johns Hopkins University Press.

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.

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 Academy of Sciences of the United States of America* **86**, 7059-7062.

Orito, E., Ichida, T, Sakugawa, H, Sata, M, Horiike, N, Hino, K, Okita, K, Okanoue, T, Iino, S, Tanaka, E, Suzuki, K, Watanabe, H., Hige, S, and Mizokami, M (2001). Geographic distribution of hepatitis B virus (HBV) genotype in patients with chronic HBV infection in Japan. *Hepatology* **34**, 590-594.

Owiredu, W. K. B. A, Kramvis, A, and Kew MC (2001). Molecular analysis of hepatitis B virus genomes isolated from black African patients with fulminant hepatitis B. *Journal of Medical Virology* **65**, 485-492.

Owiredu, W. K. B. A, Kramvis, A, and Kew MC (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.

Paran, N, Cooper, A, and Shaul, Y (2003). Interaction of hepatitis B virus with cells. *Reviews in Medical Virology* **13**, 137-143.

Parekh, S, Zoulim, F., Ahn, S. H, Tsai, A, Li, J, Kawai, S, Khan, N, Trepo, C., Wands, J, and Tong, S (2003). Genome replication, virion secretion, and e antigen expression of naturally occurring hepatitis B virus core promoter mutants. *Journal of Virology* **77**, 6601-6612.

Park, S. G, Kim, Y, Park, E, Ryu, H. M, and Jung, G (2003). Fidelity of hepatitis B virus polymerase. *European Journal of Biochemistry* **270**, 2929-2936.

Peeters, M., Courgnaud, V, Abela, B, Auzel, P, Pourrut, X, Bibollet-Ruche, F, Loul, S, Liegeois, F, Butel, C, Koulagna, D, Mpoudi-Ngole, E, Shaw, G. M, Hahn, B. H, and Delaporte, E (2002). Risk to humans health from a plethora of simian immunodeficiency viruses in primate bushmeat. *Emerging Infectious Diseases* **8**, 451-457.

Pineiro y Leone, F. G, Mbayed, V. A., and Campos, R. H (2003). Evolutionary history of Hepatitis B virus genotype F: an in-depth analysis of Argentine isolates. *Virus Genes* **27**, 103-110.

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.

Prescott, L. E., MacDonald, D. M., Davidson, F, Mokili, J, Pritchard, D. I, Arnot, D. E, Riley, E. M, Greenwood, B. M, Hamid, S, Saeed, A. A, McClure, M. O., Smith, D. B., and Simmonds, P (1999). Sequence diversity of TT virus in geographically dispersed human populations. *Journal of General Virology* **80**, 1751-1758.

Prince, A. M., Whalen, R, and Brotman, B (1997). Successful nucleic acid based immunization of newborn chimpanzees against hepatitis B virus. *Vaccine* **15**, 916-919.

Pult, I, Abbott, N, Zhang, Y-Y, and Summers, J (2001). Frequency of spontaneous mutations in an avian hepadnavirus infection. *Journal of Virology* **75**, 9623-9632.

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.

Quintero, A., Martinez, D., Alarcon, D. N., Costagliola, A., Urbina, L., Gonzalez, N., Liprandi, F., Castro, D. G., and Pujol, F. H. (2002). Molecular epidemiology of hepatitis B virus in Afro-Venezuelan populations. *Archives of Virology* **147**, 1829-1836.

Roche, B, Feray, C, Gigou, M, Roque-Afonso, A. M, Arulnaden, J. L, Delvart, V, Dussaix, E, Guettier, C, Bismuth, H, and Samuel, D (2003). HBV DNA persistence 10 years after liver transplantation despite successful anti-HBS passive immunoprophylaxis. *Hepatology* **38**, 86-95.

Roos, C and Geissmann, T (2001). Molecular phylogeny of the major hylobatid divisions. *Molecular Phylogenetics and Evolution* **19**, 486-494.

Ryu, C. J, Cho, D-Y, Gripon, P, Kim, H. S, Guguen-Guillouzo, C, and Hong, H. J (2000). An 80-kilodalton protein that binds to the Pre-S1 domain of hepatitis B virus. *Journal of Virology* **74**, 110-116.

Sahu, G. K, Tai, P-C, Chatterjee, S. B, Lin, M-H, Tennant, B, Gerin, J, and Shih, C (2002.). Out-of-frame versus in-frame core internal deletion variants of human and woodchuck hepatitis B viruses. *Virology* **292**, 35-43.

Salemi, M, Strimmer, K, Hall, W. W, Duffy, M, Delaporte, E, Mboup, S, Peeters, M., and Vandamme, A-M (2001). Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution. *The FASEB Journal* **15**, 276-278.

Sanchez, L. V, Maldonado, M, Bastidas-Ramirez, B. E, Norder, H, and Panduro, A (2002). Genotypes and S-gene variability of Mexican hepatitis B virus strains. *Journal of Medical Virology* **68**, 24-32.

Santiago, M. L, Lukasik, M, Kamenya, S, Li, Y, Bibollet-Ruche, F, Bailes, E., Muller, M. N., Emery, M, Goldenberg, D. A, Lwanga, J. S, Ayouba, A, Nerrienet, E, McClure, H. M, Heeney, J. L., Watts, D. P, Pusey, A. E, Collins, D. A, Wrangham, R. W., Goodall, J, Brookfield, J. F, Sharp, P. M, Shaw, G. M, and Hahn, B. H (2003). Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (*Pan troglodytes schweinfurthii*). *Journal of Virology* **77**, 7545-7562.

Santiago, M. L, Bibollet-Ruche, F, Bailes, E., Kamenya, S, Muller, M. N., Lukasik, M, Pusey, A. E, Collins, D. A, Wrangham, R. W., Goodall, J, Shaw, G. M, Sharp, P. M, and Hahn, B. H (2003). Amplification of a complete simian immunodeficiency virus genome from fecal RNA of a wild chimpanzee. *Journal of Virology* **77**, 2233-2242.

Sánchez-Tapias, J. M, Costa, J., Mas, A, Bruguera, M, and Rodés, J (2002). Influence of hepatitis B virus genotype on the long-term outcome of chronic hepatitis B in Western patients. *Gastroenterology* **123**, 1848-1856.

Schierup, M. H and Hein, J (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**, 870-891.

Schlicht, H. J and Schaller, H (1989). The secretory protein of human hepatitis B virus is expressed on the cell surface. *Journal of Virology* **63**, 5399-5404.

Schuster, R, Hildt, E, Chang, S. F, Terradillos, O, Pollicino, T., Lanford, R., Gerlich, W. H., Will, H, and Schaefer, S (2002). Conserved transactivating and pro-apoptotic functions of hepadnaviral X protein in ortho- and avihepadnaviruses. *Oncogene* **21**, 6606-6613.

Seeger, C and Mason, W. S (2000). Hepatitis B virus biology. *Microbiology and Molecular Biology Reviews* **64**, 51-68.

Simmonds, P. (2001). 2000 Fleming Lecture. The origin and evolution of hepatitis viruses in humans. *Journal of General Virology* **82**, 693-712.

Simmonds, P. (2001). Reconstructing the origins of human hepatitis viruses. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **356**, 1013-26.

Sirisena, N. D, Njoku, M. O, Idoko, J. A, Isamade, E, Barau, C, Jelve, D, Zamani, A, and Otowo, S (2002). Carriage rate of hepatitis-B surface antigen (HBsAg) in an urban community in Jos, Plateau State, Nigeria. *Nigerian Postgraduate Medicine Journal* **9**, 7-10.

Souquiere, S, Bibollet-Ruche, F, Robertson, D. L., Makuwa, M, Apetrei, C, Onanga, R, Kornfeld, C, Plantier, J-C, Gao, F, Abernethy, K, White, L. J. T, Karesh, W, Telfer, P, Wickings, E. J, Mauclore, P, Marx, P. A, Barre-Sinoussi, F, Hahn, B. H, Muller-Trutwin, M. C, and Simon, F (2001.). Wild Mandrillus sphinx are carriers of two types of lentivirus. *Journal of Virology* **75**, 7086-7096.

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. E, MacDonald, D. M., Lewis, J. C. M, Holmes, E. C., and Simmonds, P (2003). Geographic and species association of hepatitis B virus genotypes in non-human primates. *Virology* **314**, 381-393.

Stewart, C-B and Disotell, T. R (1998). Primate evolution - in and out of Africa. *Current Biology* **8**, R582-R588.

Stringer, C (2002). Modern human origins: progress and prospects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **357**, 563-579.

Stroffolini, T, Mele, A, Tosti, M. E, Gallo, G, Balocchini, E, Ragni, P, Santonastasi, F, Marzolini, A, Ciccozzi, M, and Moiraghi, A (2000). The impact of the hepatitis B mass immunisation campaign on the incidence and risk factors of acute hepatitis B in Italy. *Journal of Hepatology* **33**, 980-985.

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

Sugauchi, F., Mizokami, M., Orito, E., Ohno, T., Kato, H., Suzuki, S., Kimura, Y., Ueda, R., Butterworth, L. A., and Cooksley, W. G. (2001). A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. *Journal of General Virology* **82**, 883-892.

Sugauchi, F., Chutaputti, A, Orito, E., Kato, H., Suzuki, S, Ueda, R, and Mizokami, M (2002). Hepatitis B virus genotypes and clinical manifestation among hepatitis B carriers in Thailand. *Journal of Gastroenterology and Hepatology* **17**, 671-676.

Sugauchi, F., Orito, E., Ichida, T, Kato, H., Sakugawa, H, Kakumu, S, Ishida, T, Chutaputti, A, Lai, C-L., Ueda, R, Miyakawa, Y, and Mizokami, M (2002). Hepatitis B virus of genotype B with or without recombination with genotype C over the precore region plus the core gene. *Journal of Virology* **76**, 5985-5992.

Sugauchi, F., Orito, E., Kato, H., Suzuki, S, Kawakita, S, Sakamoto, Y, Fukushima, K, Akiba, T, Yoshihara, N, Ueda, R, and Mizokami, M (2003). Genotype, serotype, and phylogenetic characterization of the complete genome sequence of hepatitis B virus isolates from Malawian chronic carriers of the virus. *Journal of Medical Virology* **69**, 33-40.

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

Sugauchi, F., Orito, E., Ichida, T., Kato, H., Sakugawa, H., Kakumu, S., Ishida, T., Chutaputti, A., Lai, C-L., Gish, R. G., Ueda, R., Miyakawa, Y., and Mizokami, M (2003). Epidemiologic and virologic characteristics of hepatitis B virus genotype B having the recombination with genotype C. *Gastroenterology* **124**, 925-932.

Sugimoto, C, Kitamura, T, Guo, J, Al-Ahdal, M. N, Shchelkumov, S. N, Otova, B, Ondrejka, P, Chollet, J-Y, El-Safi, S, Ettayebi, M, Gre'senguet, G, Kocagoz, T, Chaiyarasamee, S, Thant, K. Z, Thein, S, Moe, K, Kobayashi, N, Taguchi, F, and Yogo, Y (1997). Typing of urinary JC virus DNA offers a novel means of tracing human migrations. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 9191-9196.

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, 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 Academy of Sciences of the United States of America* **75**, 4533-4537.

Suzuki, S, Sugauchi, F., Orito, E., Kato, H., Usuda, S., Siransy, L, Arita, I, Sakamoto, Y, Yoshihara, N, El-Gohary, A, Ueda, R, and Mizokami, M (2003). Distribution of hepatitis B virus (HBV) genotypes among HBV carriers in the Cote d'Ivoire: complete genome sequence and phylogenetic relatedness of HBV genotype E. *Journal of Medical Virology* **69**, 459-465.

Tai, P-C, Banik, D, Lin, G-I, Pai, S, Pai, K, Lin, M-H, Yuoh, G, Che, S, Hsu, S. H, Chen, T-C, Kuo, T-T, Lee, C-S, Yang, C-S, and Shih, C (1997). Novel and frequent mutations of hepatitis B virus coincide with a major histocompatibility complex class-I restricted T-cell epitope of the surface antigen. *Journal of Virology* **71**, 4852-4856.

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.

Takahashi, K., Mishiro, S., and Prince, A. M. (2001). Novel hepatitis B virus strain from a chimpanzee of Central Africa (*Pan troglodytes troglodytes*) with an unusual antigenicity of the core protein. *Intervirology* **44**, 321-326.

Tanaka, J. (2000). Hepatitis B epidemiology in Latin America. *Vaccine* **18 Suppl 1**, S17-S19.

Tang, H and McLachlan (2002). Avian and mammalian hepadnaviruses have distinct transcription factor requirements for viral replication. *Journal of Virology* **76**, 7468-7472.

Tennant, B. C and Gerin, J. L (2001). The woodchuck model of hepatitis B virus infection. *Institute for Laboratory Animal Research Journal* **42**, 89-102.

Teo, C. G (1992). The virology and serology of hepatitis: an overview. *CDR Review* **2**, R109-R113.

Testut, P, Renard, C-A, Terradillos, O, Vitvistsky-Trepo, L, Tekaia, F, Degott, C, Blake, J, Boyer, B, and Buenida, M. A (1996). A new hepadnavirus endemic in arctic ground squirrels in Alaska. *Journal of Virology* **70**, 4210-4219.

Thomas, H. C and Jacyna, M. R. Hepatitis B virus: pathogenesis and treatment of chronic infection. Zuckerman, A. J., Thomas, H. C., and (Eds.). *Viral Hepatitis*, 185-207. 1993. Edinburgh, Churchill Livingstone.

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

Torbenson, M and Thomas, D. L (2002). Occult hepatitis B. *The Lancet Infectious Diseases* **2**, 479-486.

Tran, H. T-T, Ushijima, H, Ngoc, T. T, Ha, L. D, Hayashi, S., Sata, T, and Abe, K (2003). Recombination of genotypes B and C in hepatitis B virus isolated from a Vietnamese patient with fulminant hepatitis. *Japanese Journal of Infectious Diseases* **56**, 35-37.

Tu, H, Bonura, C, Giannini, C, Mouly, H, Soussan, P, Kew, M, Paterlini-Brechot, P, Brechot, C., and Kremsdorf, D (2001). Biological impact of natural COOH-terminal deletions of hepatitis B virus X protein in hepatocellular carcinoma tissues. *Cancer Research* **61**, 7803-7810.

Tyler, G. V, Summers, J, and Synder, R. L (1981). Woodchuck hepatitis virus in natural woodchuck populations. *Journal of Wildlife Diseases* **17**, 297-301.

Urban, S, Breiner, K. M, Fehler, F, Klingmuller, U, and Schaller, H (1998). Avian hepatitis B virus infection is initiated by the interaction of a distinct Pre-S subdomain with the cellular receptor gp180. *Journal of Virology* **72**, 8089-8097.

Vartanian, J. P., Pineau, P., Henry, M., Hamilton, W. D., Muller, M. N., Wrangham, R. W., and Wain-Hobson, S. (2002). Identification of a hepatitis B virus genome in wild chimpanzees (*Pan troglodytes schweinfurthi*) from East Africa indicates a wide geographical dispersion among equatorial African primates. *Journal of Virology* **76**, 11155-11158.

Vaudin, M., Wolstenholme, A. J., Tsiquaye, K. N., Zuckerman, A. J., and Harrison, T. J. (1988). The complete nucleotide sequence of the genome of a hepatitis B virus isolated from a naturally infected chimpanzee. *Journal of General Virology* **69**, 1383-1389.

Verschoor, E. J., Warren, K. S., Langenhuijzen, S., Heriyanto, Swan, R. A., and Heeney, J. L. (2001). Analysis of two genomic variants of orang-utan hepadnavirus and their relationship to other primate hepatitis B-like viruses. *Journal of General Virology* **82**, 893-897.

Vieth, S, Manegold, C, Drosten, C, Nippraschk, T, and Gunther, S (2002). Sequence and phylogenetic analysis of hepatitis B virus genotype G isolated in Germany. *Virus Genes* **24**, 153-156.

Vryheid, R. E, Yu, E. S, Mehta, K. M, and McGhee, J (2001). The declining prevalence of hepatitis B virus infection among Asian and Pacific Islander children. *Asian American Pacific Islands Journal of Health* **9**, 162-178.

Wang, T. H., Donaldson, Y. K., Brettle, R. P., Bell, J. E., and Simmonds, P (2001). Identification of shared populations of HIV-1 infecting microglia and tissue macrophages outside the central nervous system. *Journal of Virology* **75**, 11686-11699.

Warren, K. S., Niphuis, H, Heriyanto, Verschoor, E. J., Swan, R. A., and Heeney, J. L (1998). Seroprevalence of specific viral infections in confiscated orangutans (*Pongo pygmaeus*). *Journal of Medical Primatology* **27**, 33-37.

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.

Warren, K. S., Verschoor, E. J., Langenhuijzen, S., Heriyanto, Swan, R. A., Vigilant, L, and Heeney, J. L (2001). Speciation and intrasubspecific variation of Bornean orangutans, *Pongo pygmaeus pygmaeus*. *Molecular Biology and Evolution* **18**, 472-480.

Westland, C, Delaney IV, 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 dipivoxil. *Gastroenterology* **125**, 107-116.

Whittle, H. C., Inskip, H, Hall, A. J, Mendy, M, Downes, R, and Hoare, S (1991). Vaccination against hepatitis B and protection against chronic viral carriage in The Gambia. *The Lancet* **337**, 747-750.

Worley, M. B and Stalis, I. H (2002). Detection of virus-like particles in the liver of black and white ruffed lemurs with hepatitis. *Journal of Wildlife Diseases* **38**, 258-265.

Yuasa, R, Takahashi, K., Dien, B. V, Bihn, N. H, Morishita, T, Sato, K, Yamamoto, N, Isomura, S, Yoshioka, T, Mishiro, S., and Kakumu, S (2000). Properties of hepatitis B virus genome recovered from Vietnamese patients with fulminant hepatitis in comparison with those of acute hepatitis. *Journal of Medical Virology* **61**, 23-28.

Yuen, M-F, Sablon, E, Wong, D. K-H, Yuan, H-J, Wong, B. C-Y, Chan, A. O-O, and Lai, C-L (2003). Role of hepatitis B virus genotypes in chronic hepatitis B exacerbation. *Clinical Infectious Diseases* **37**, 593-597.

Zhang, Z, Torii, N, Hu, Z, Jacob, J, and Liang, T. J (2001). X-deficient woodchuck hepatitis virus mutants behave like attenuated viruses and induce protective immunity in vivo. *The Journal of Clinical Investigation* **108**, 1523-1531.

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**

**Accession Numbers Used for**

**Phylogenetic Analysis of S Gene**

**Sequences**

## **GENOTYPE A**

HBVADW, HVHEPB, S50225, E00010, HUMPRECX, HPBADWZCG, HBVA317, HBVXCPS, HBVGEN2, HEB344115, AF297625, AF297624, AF297623, AF297622, AF297621, AF090842, AF090838, AF090839, AF090840, AF090841, AF143298, AF143299, AF143302, AF143304, AF143307, AY034878, U87742, U87740, U87741, U87743, U87744, U87745, U87747, U87748, HEB309371, HBVPRESS

## **GENOTYPE B**

AB014366, HPBA1HKK2, HPBA2HYS2, HPBA3HMS2, D50521, D50522, AF79684, AY033072, AY033073, HBV131133, AF121251, AF121249, AF121247, AF282918, AF461360, AF461362

## **GENOTYPE C**

HPBETNC, HBV18858, HHVBC, HPBADRC, HPBCG, HPBE8A, HPBH2B, HPBB4HST1, HPBB5HKO1, HPBC5HKO2, HPBC6T588, HPBA11A, HPBADRA, HPBADR1CG, S75184, HBVADR, HBVADR4, HEHBVAYR, HBVADRM, HBVPREX, HHVCCHA, AF461358, AF233236, NC\_001707, AF473543, AF461043, AF458665, AF458664, AY066028, AF384372, AY057947, AF411408-AF411412, AF182805, AF330110, AF223961, AY040627, AF384371, AF363961-AF363963, AB048704, AB049610, AB049609, AB050018, HBV18856, D28880, AF461359, AF461361, AF461363, AB033553, AB042282-AB042285, AB048705, AF208868, AF208869, HPBC4HST2, HPBCGADR, AB014378, D50520, HBV18857, AF068756, AB014376, AF223954-AF223960, AF360967-AF360971

## **GENOTYPE D**

HBVGEN1, HPBAYW, HPBHBVAA, U95551, XXHEPAV, HBVAYWMCG, HBVAYWC, HBVORFS, HBVPRES12, HBVP2CSX, HBVP3CSX, HBVAYWGEN, AF151735, AF121239-AF121242, HBV132335, HEB344116, HEB344117, AB048701-AB048703, AB033558, AB033559, AF065110, AF065112, AF065117-AF065119, AF208870, AF208872, AF208874, AF208877, AF209396, AF209398, AF209401, AF280817, AF391286, HHVBDS, HHVBD, AF360972-AF360978, AB090269, AB090270

## **GENOTYPE E**

HHVBBAS, HHVBE4, AF323617-AF323636, AF208817, HEB297868, HPBVAR, HPBSAGA, AB091255-AB09126, AB033272-AB033275, AB166564

## **GENOTYPE F**

HBVADW4A, HHVBF, HHVBFFOU, AB036905-AB036920, AF223962, AF223965, AB037944-AB037946, AF043561, AF043573, AF043577, AF043578, HHVBFS, AF288623-AF288628, HBU91803, HBU91805, HBU91806, HBU91807, HBU91808, HBU91811, HBU91821, HBU91822, HBU91825, HBU91829, HBU91831

## **GENOTYPE G**

AF160501, AF405706, AB064310-AB064313, AB056513-AB056516

## **GENOTYPE H**

AY090457, AY090460, AY090454, HBU91819, HBU91827

# **Appendix B**

## **AI and SimPlot Analysis of Recombinant HBV Genomes**

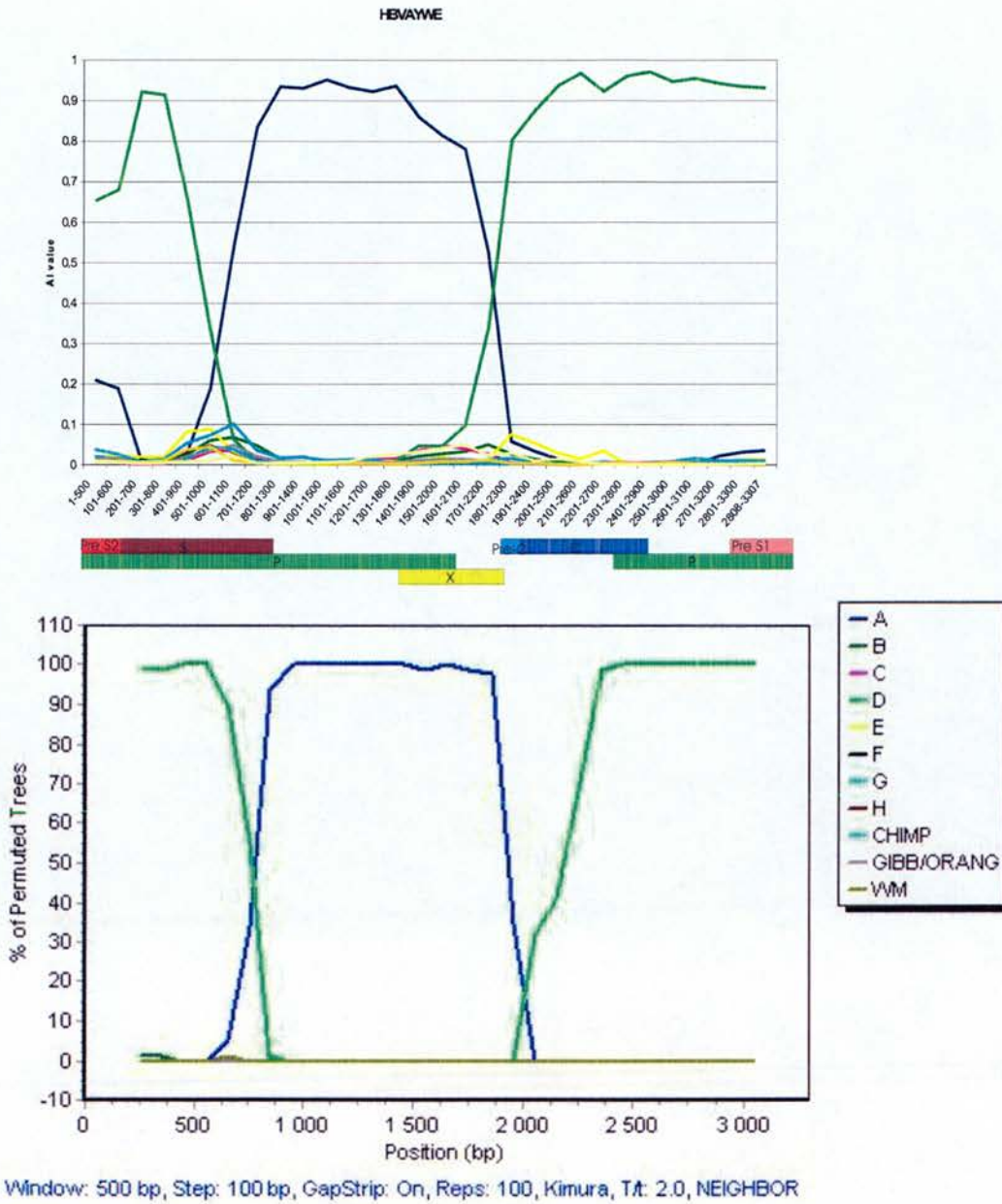


Figure B-1. Recombination analysis of HBV isolate HBVADWE () using the AI (top) and SimPlot (bottom) methods. Interpretation: recombinant using both methods.

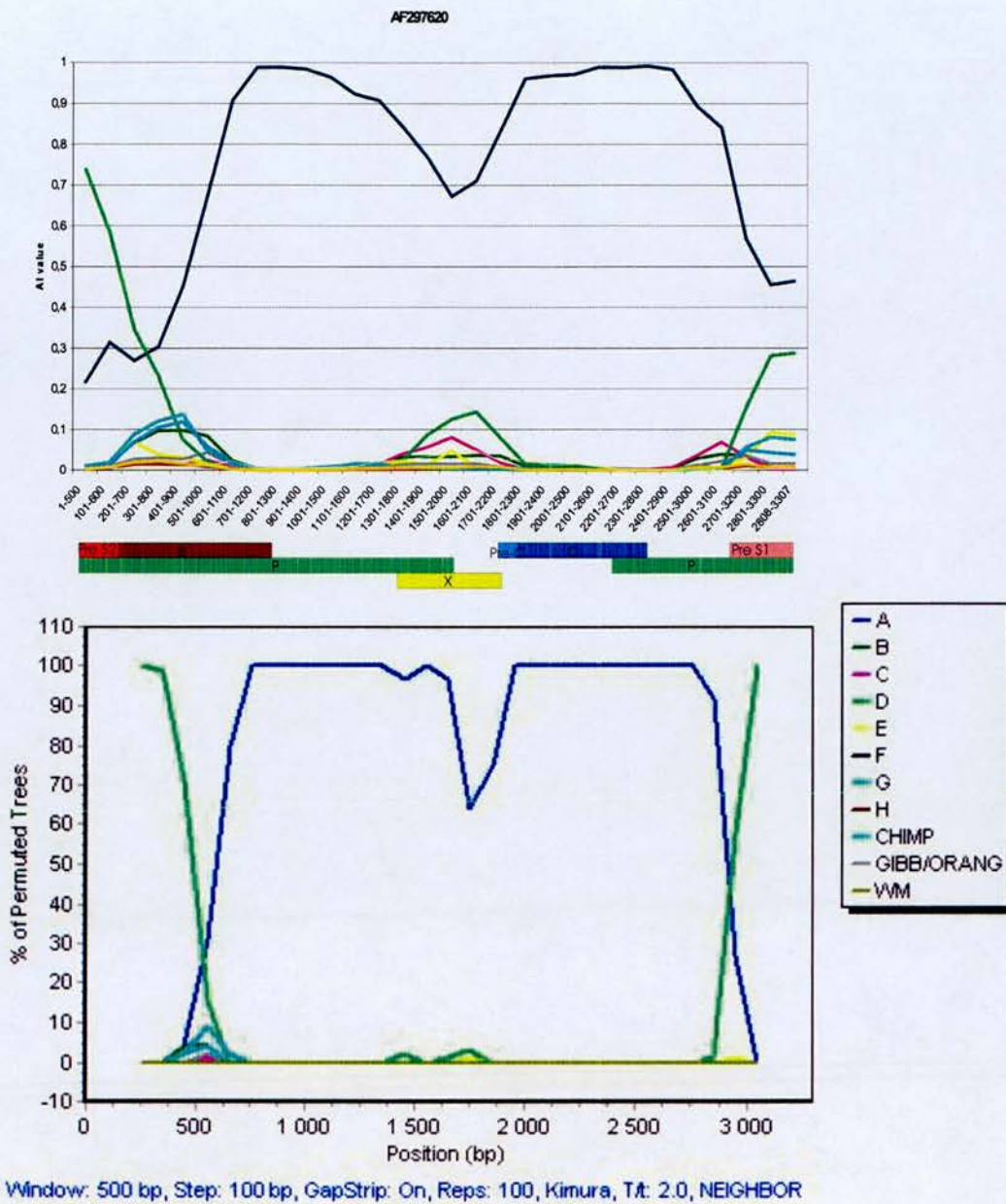


Figure B-2. Recombination analysis of HBV isolate AF297620 ( ) using the AI (top) and SimPlot (bottom) methods. Interpretation: recombinant using both methods.

AF241408

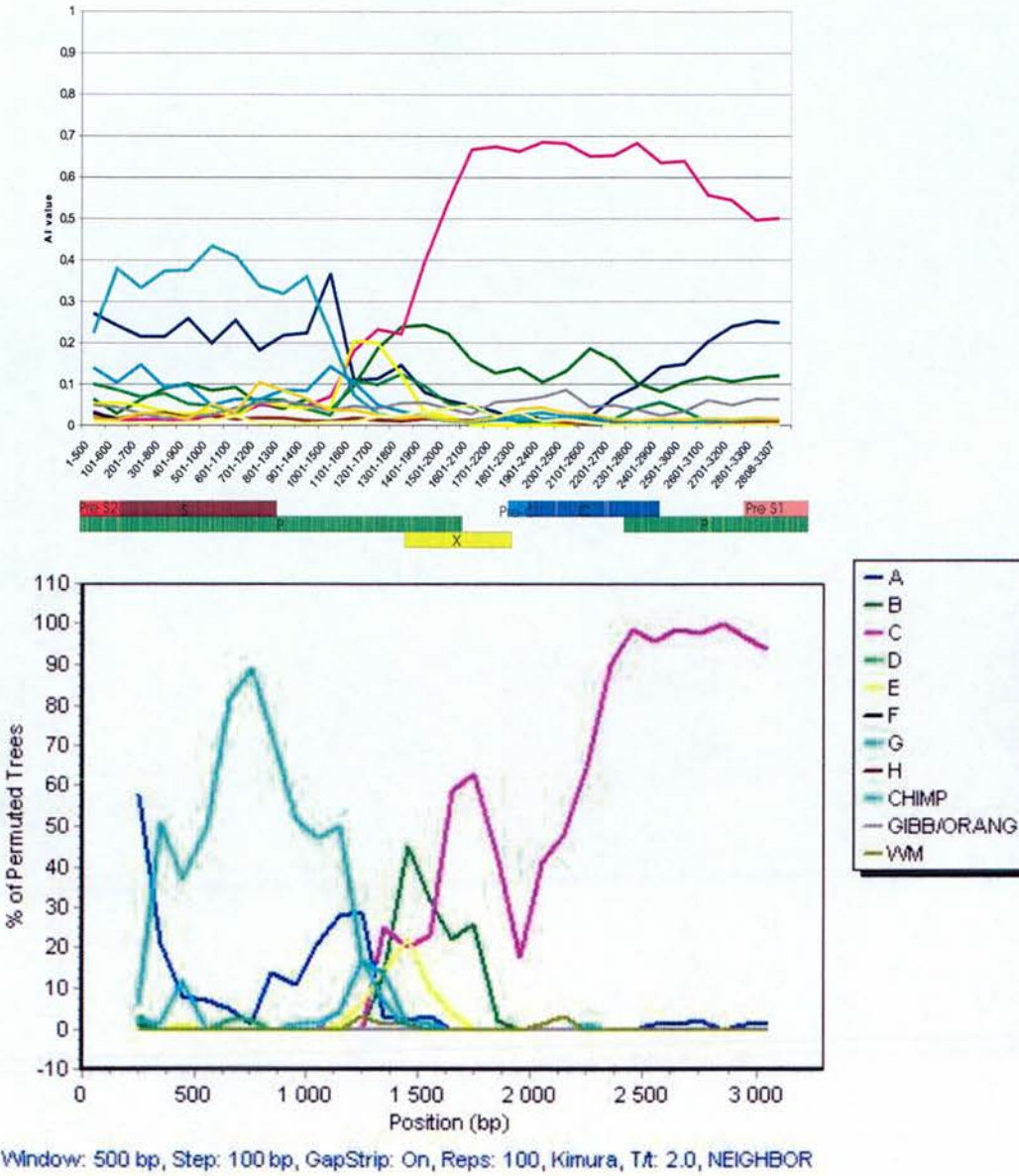
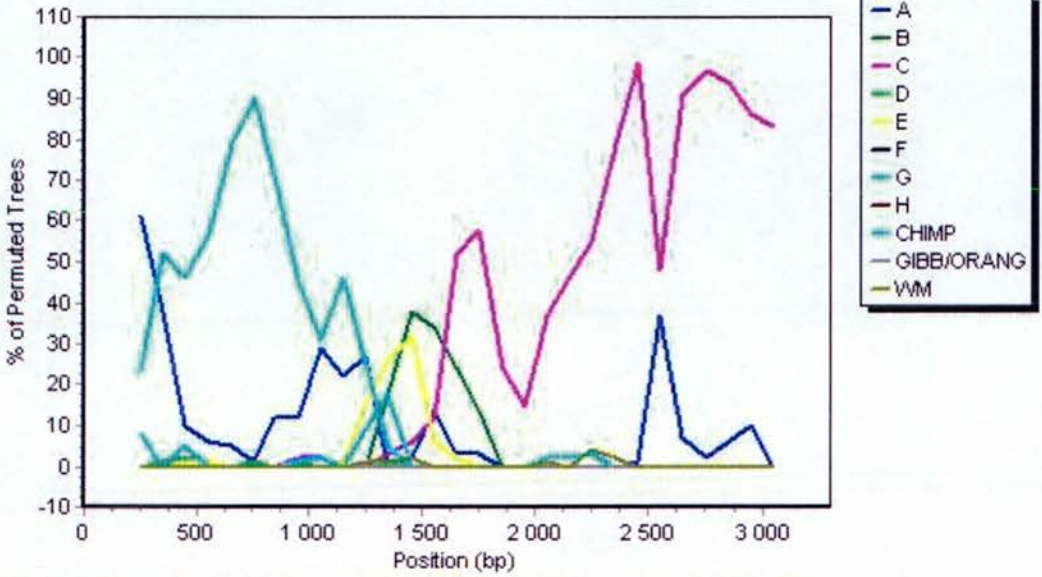
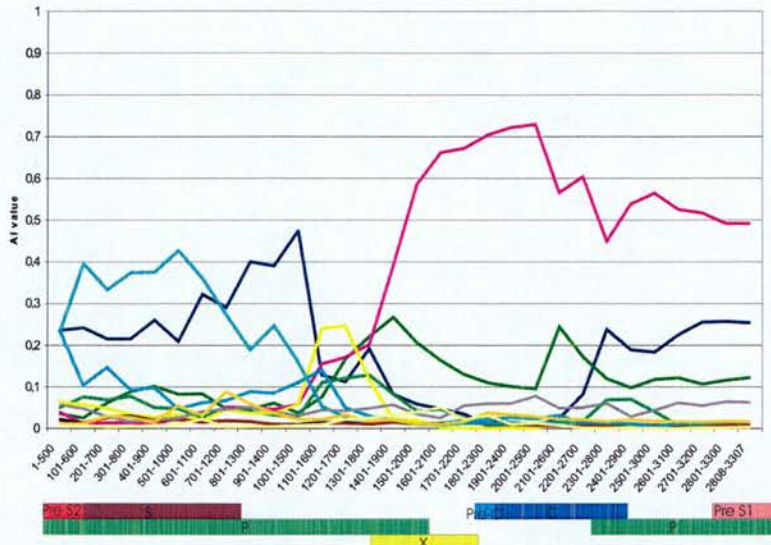


Figure B-3. Recombination analysis of HBV isolate AF241407 ( ) using the AI (top) and SimPlot (bottom) methods. Interpretation: recombinant using both methods.

AF241409



Window: 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, Tt: 2.0, NEIGHBOR

Figure B-4. Recombination analysis of HBV isolate AF241409 () using the AI (top) and SimPlot (bottom) methods. Interpretation: recombinant using both methods.

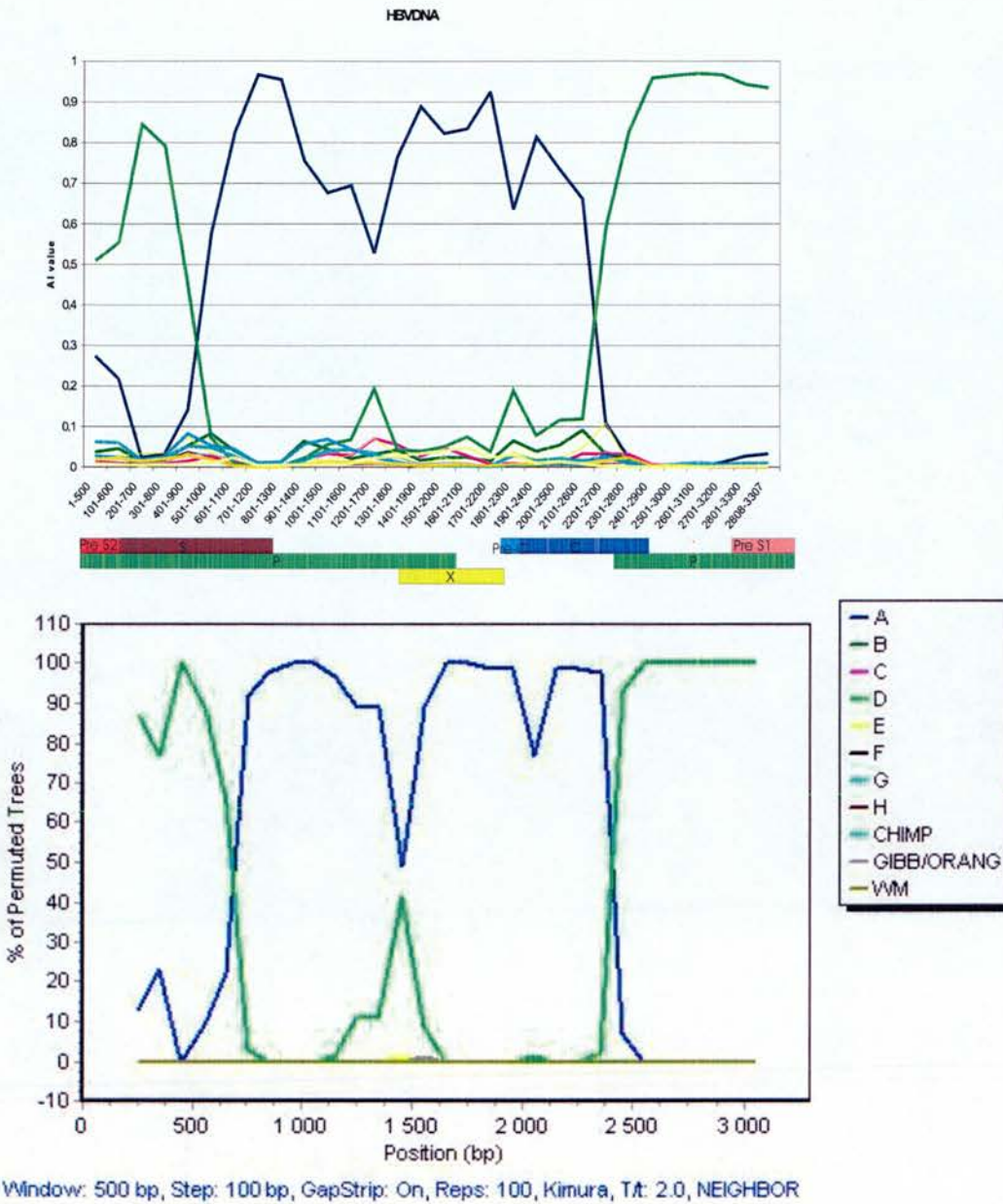


Figure B-5. Recombination analysis of HBV isolate HBVDNA () using the AI (top) and SimPlot (bottom) methods. Interpretation: recombinant using both methods.

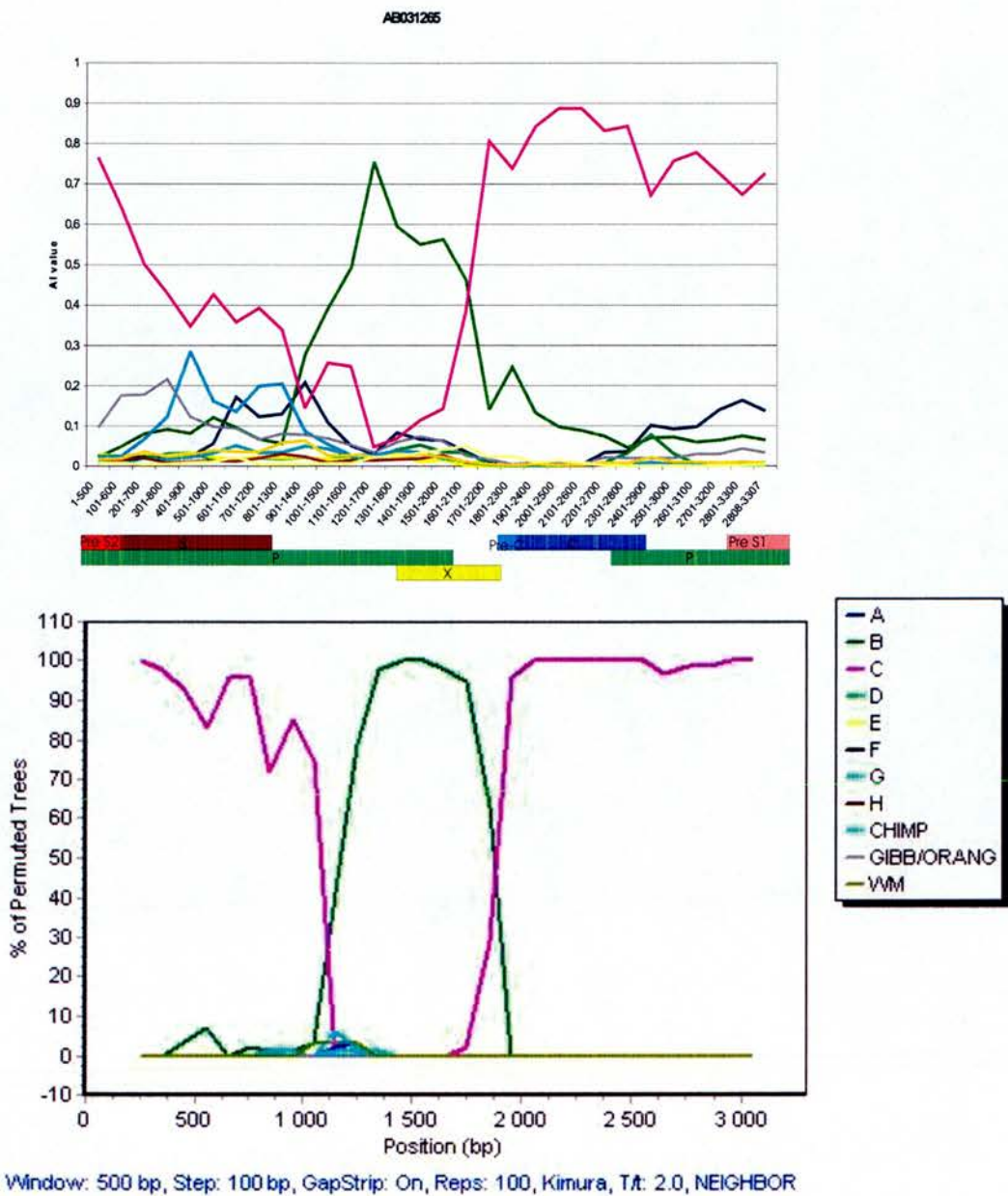


Figure B-6. Recombination analysis of HBV isolate AB031265 ( ) using the AI (top) and SimPlot (bottom) methods. Interpretation: recombinant using both methods.

AY057948

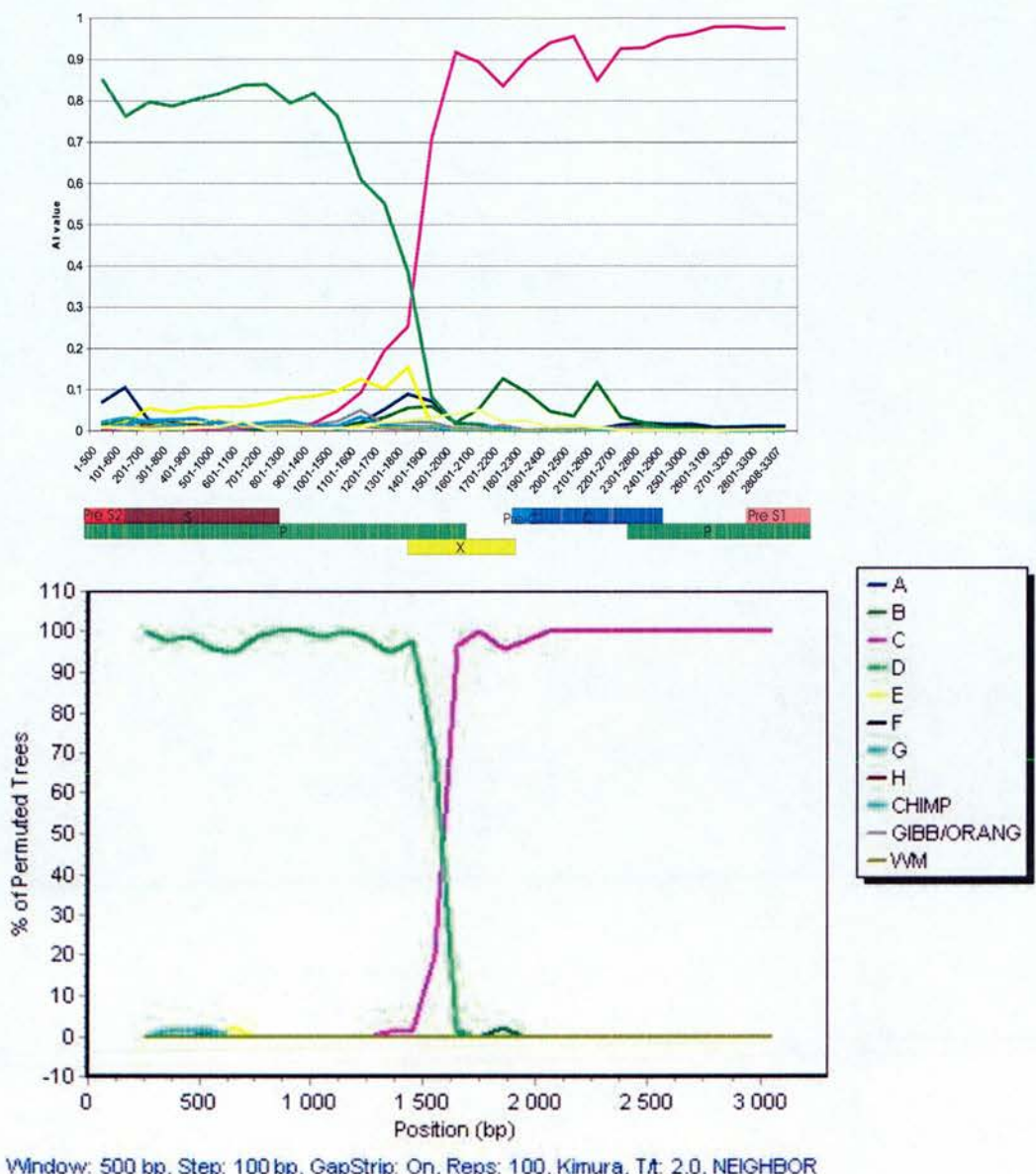


Figure B-7. Recombination analysis of HBV isolate AY057948 ( ) using the AI (top) and SimPlot (bottom) methods. Interpretation: recombinant using both methods.

**Appendix C**

**Phylogenetic Analysis of**

**Recombinant HBV Genomes**

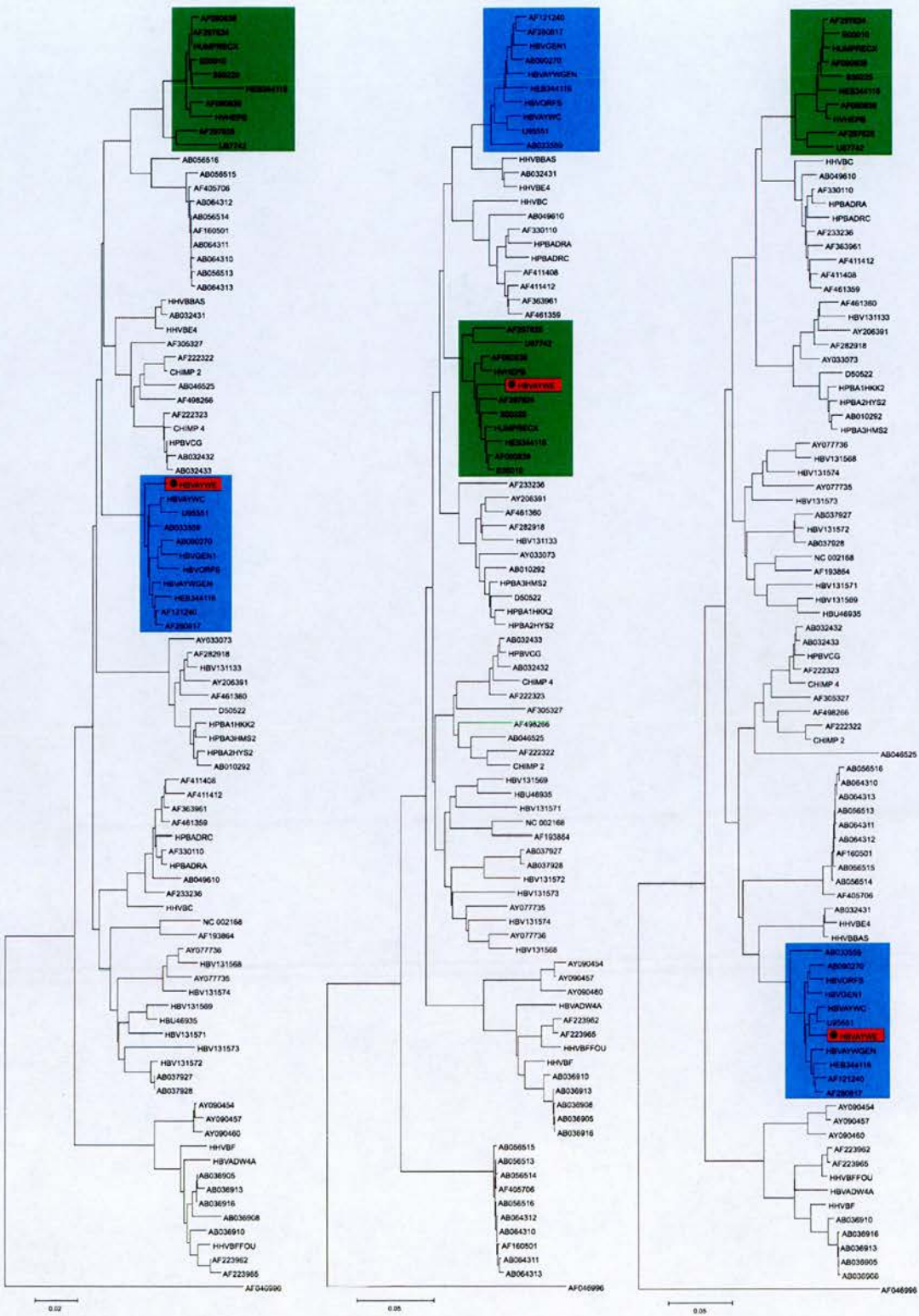


Figure C-1. Phylogenetic analysis of HBV isolate HBVAYWE, reported to be an A/D recombinant. Analysis supports the recombinant result from AI and SimPlot analysis.

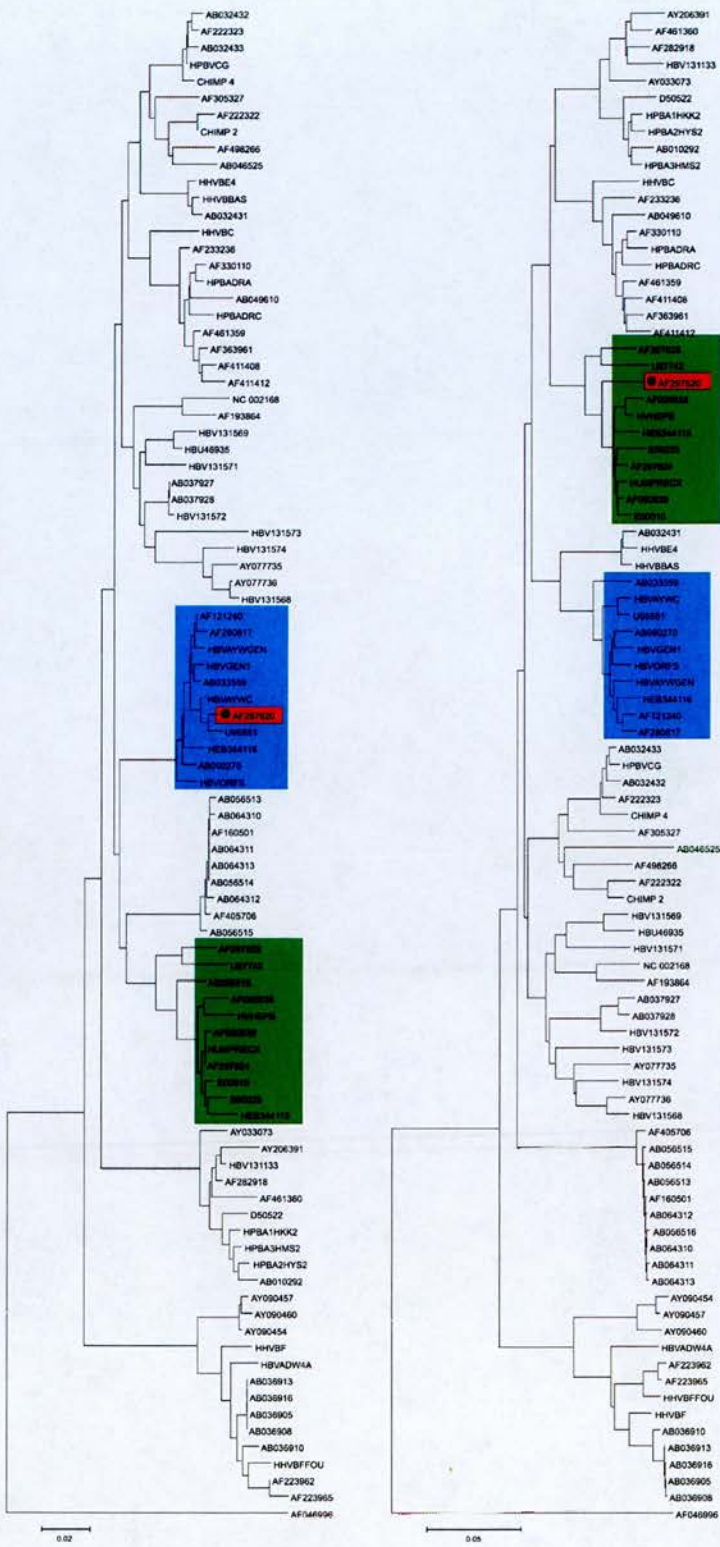


Figure C-2. Phylogenetic analysis of HBV isolate AF297620, reported to be an A/D recombinant. Analysis supports recombinant result from AI and SimPlot analysis.

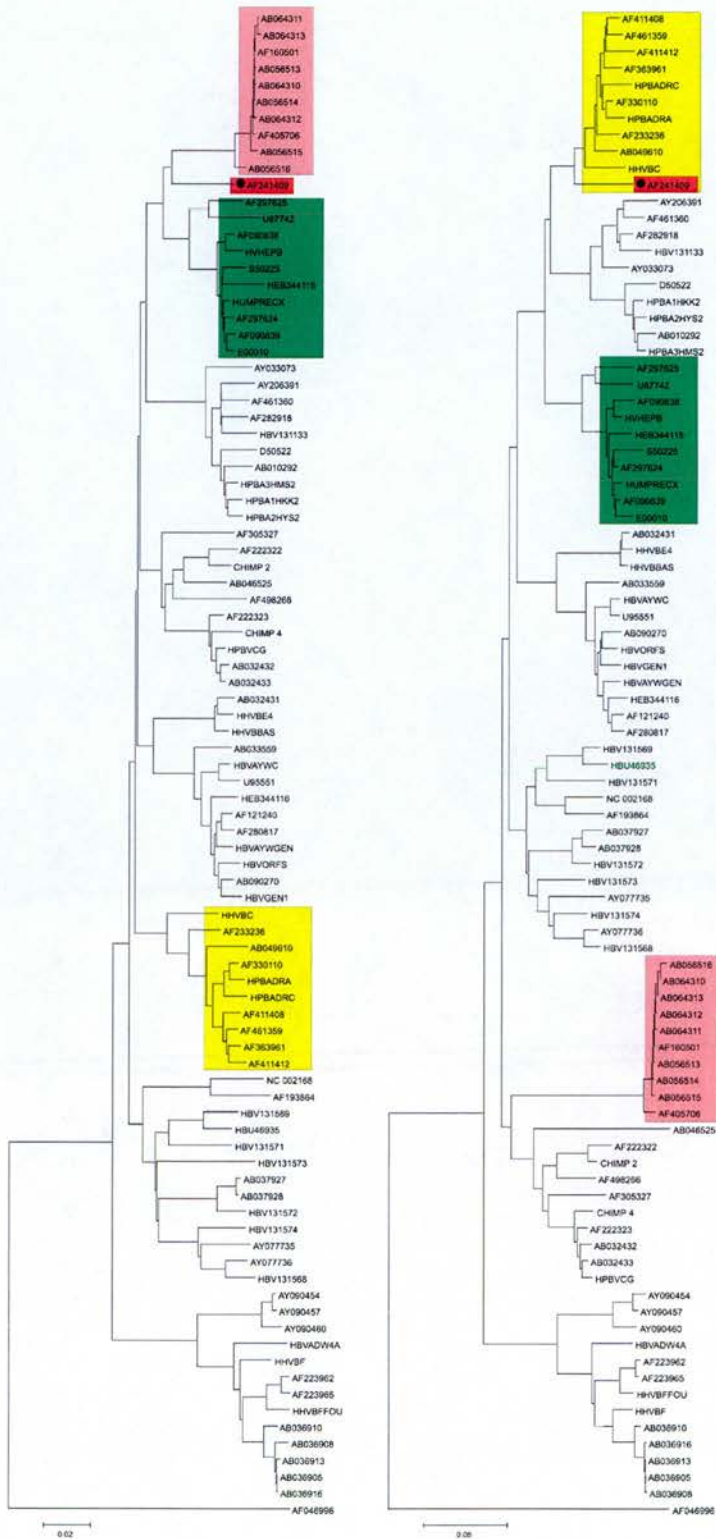


Figure C-3. Phylogenetic analysis of HBV isolate AF241409, reported to be an A/C recombinant. Analysis supports recombinant result from AI and SimPlot analysis, and shows that the isolate does not group with any specific genotype in the recombinant fragment.

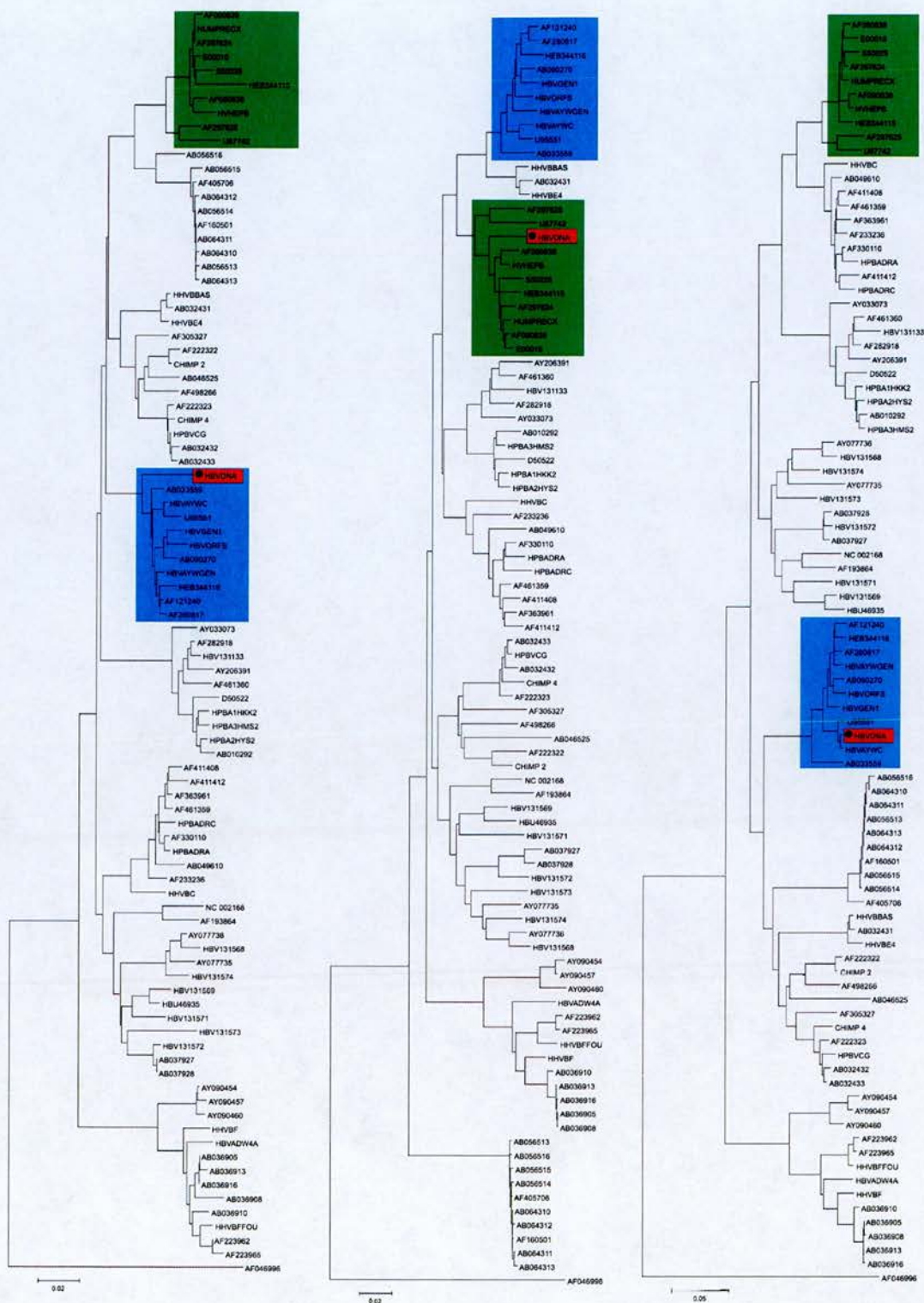


Figure C-4. Phylogenetic analysis of HBV isolate HBVDNA, reported to be an A/D recombinant. Genotype A is highlighted in the green box, genotype D in the blue box, and the isolate in the red box. Phylogenetic analysis supports the AI and SimPlot recombinant results.

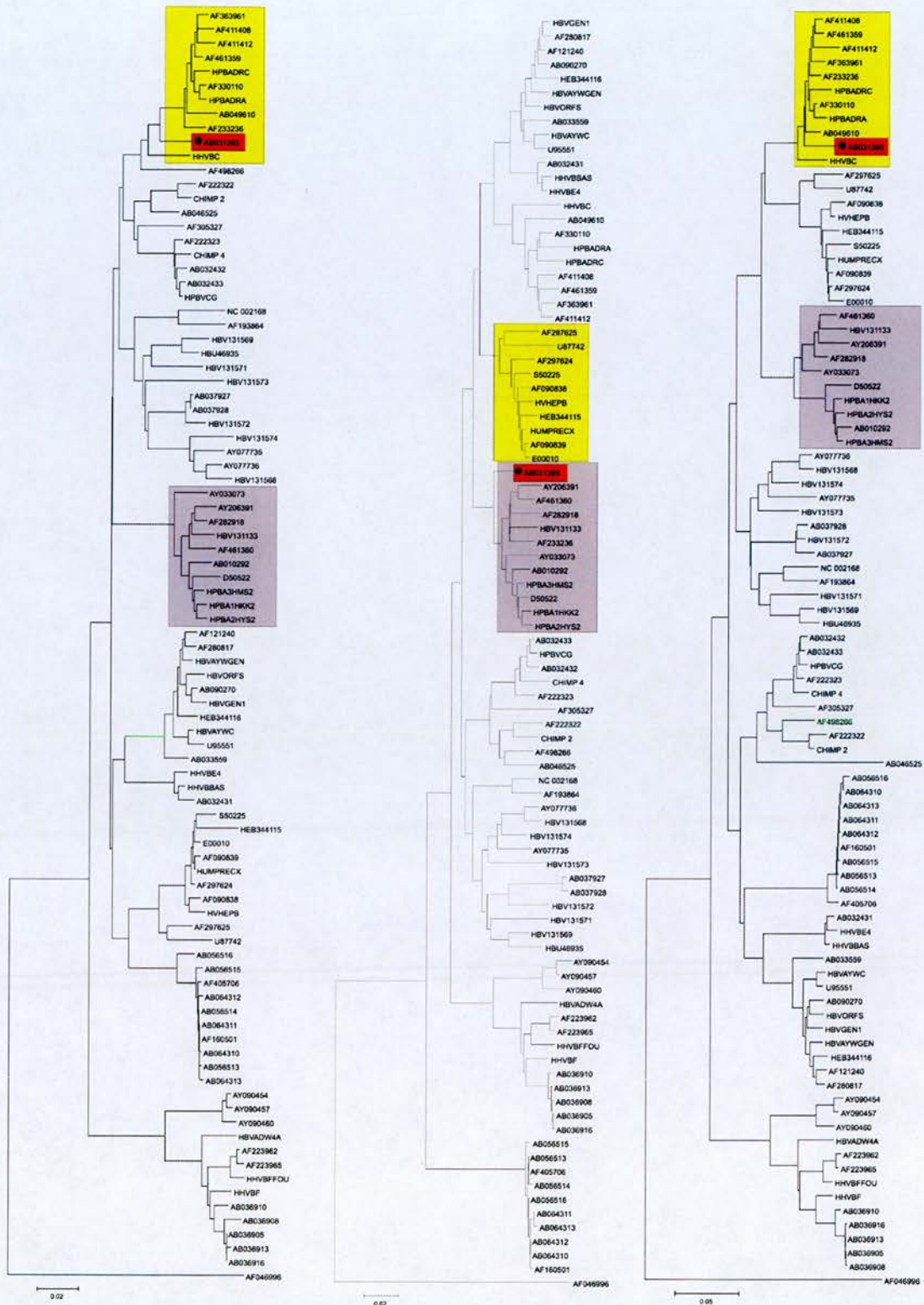


Figure C-5. Phylogenetic analysis of HBV isolate AB031265, reported to be a B/C recombinant. Genotype B is highlighted in the grey box, genotype C in the yellow box, and the isolate in the red box. Phylogenetic analysis supports the AI and SimPlot recombinant results.

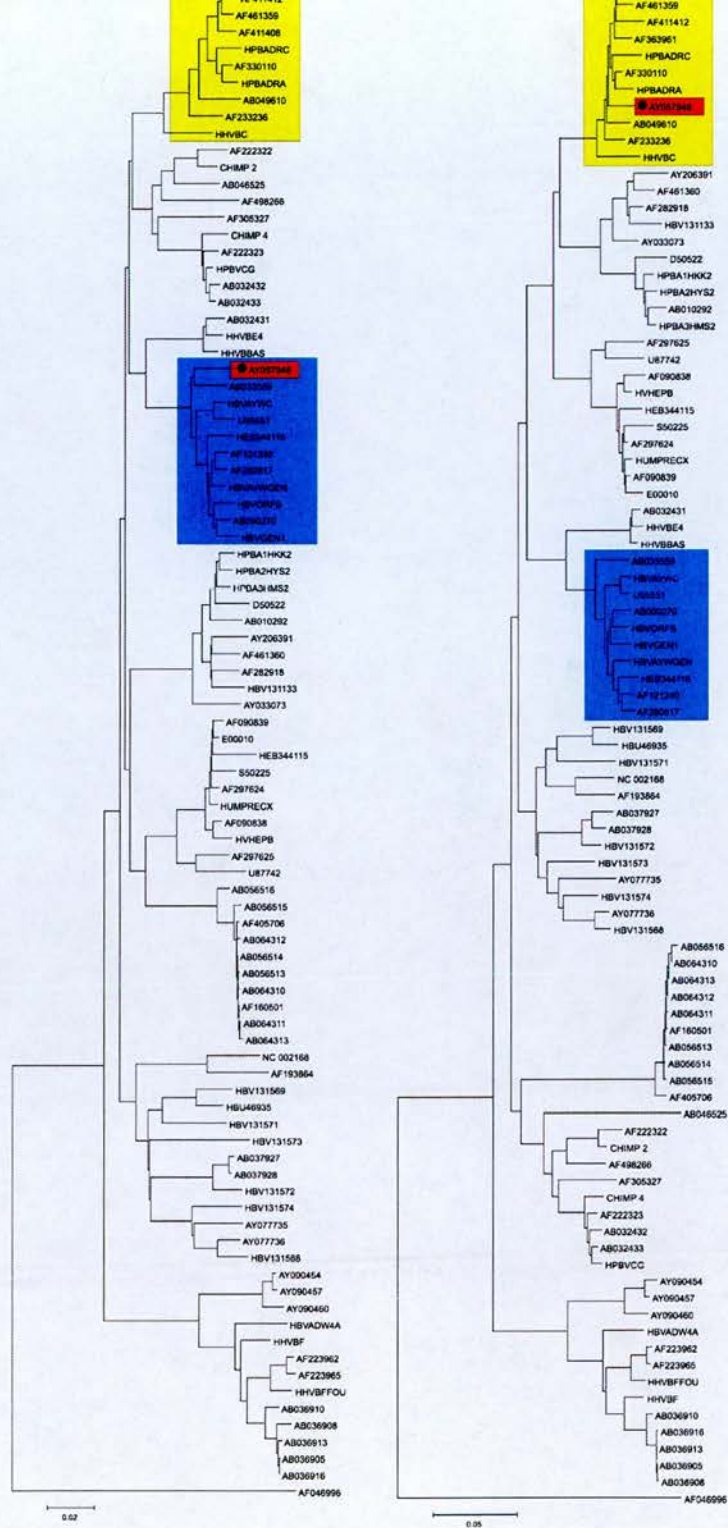


Figure C-6. Phylogenetic analysis of HBV isolate AY057948, reported to be a C/D recombinant. Genotype C is highlighted in the yellow box, genotype D in the blue box, and the isolate in the red box. Phylogenetic analysis supports the AI and SimPlot recombinant results.

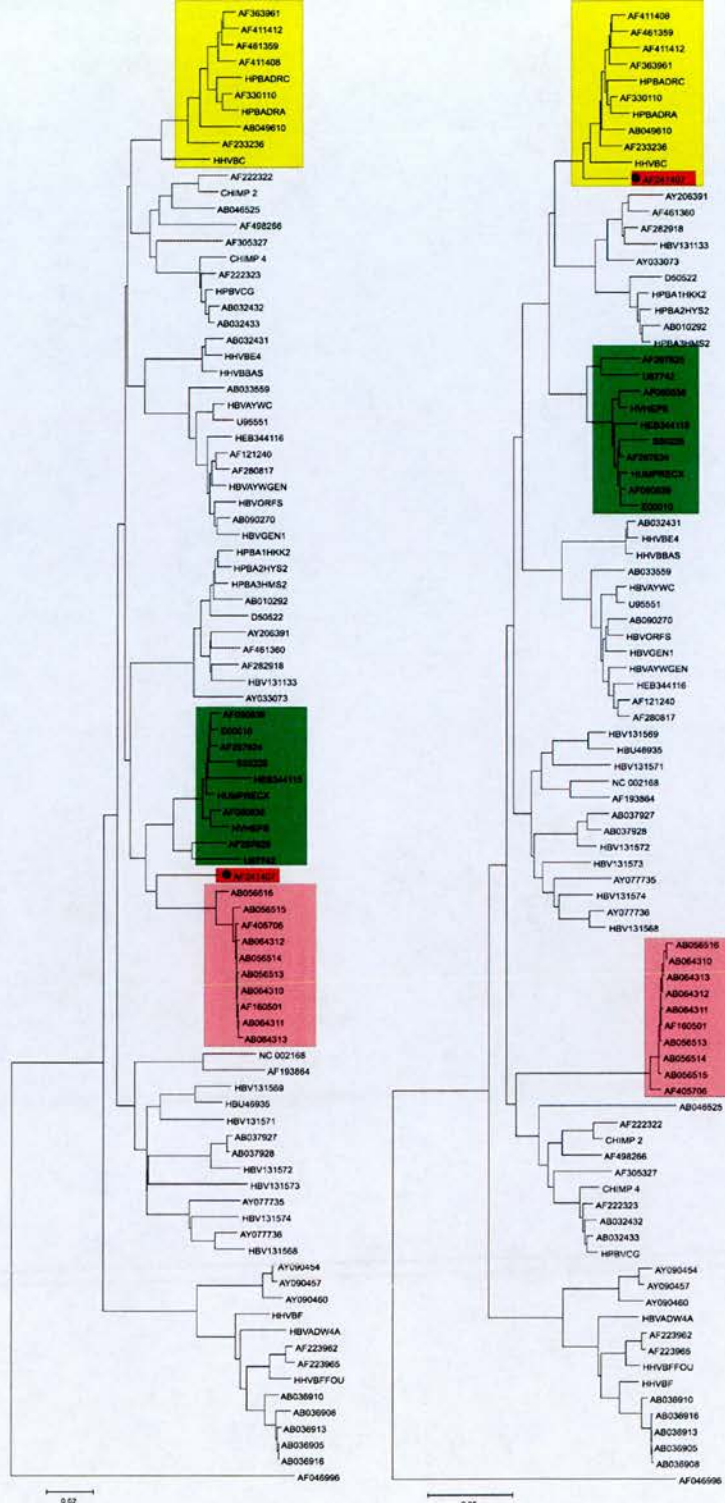


Figure C-7. Phylogenetic analysis of HBV isolate AF241407, reported to be an A/C recombinant. Genotype A is highlighted in the green box, genotype C in the yellow box, genotype G in the pink box and the isolate in the red box. Phylogenetic analysis supports the AI and SimPlot findings that this is not an A/C recombinant, but may be a recombinant with an unknown genotype related to genotypes A and G.

# **Appendix D**

## **Recombination Analysis of Reported Recombinants Producing Discrepant Results for the Two Methods Used**

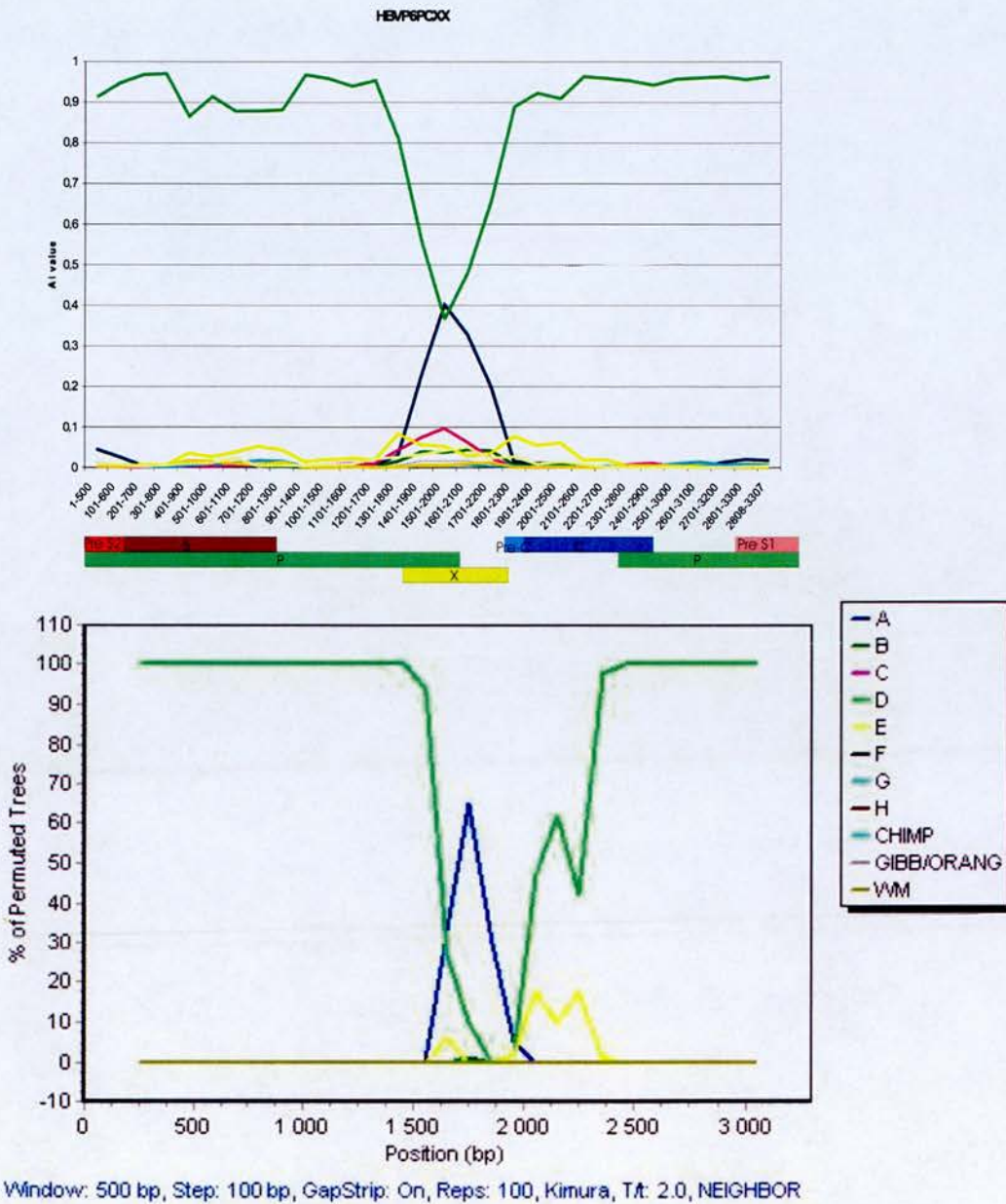


Figure D-1. Recombination analysis of HBV isolate HBVP6PCXX (Bowyer, 2000) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant with a divergent core region, SimPlot – recombinant in the core region.

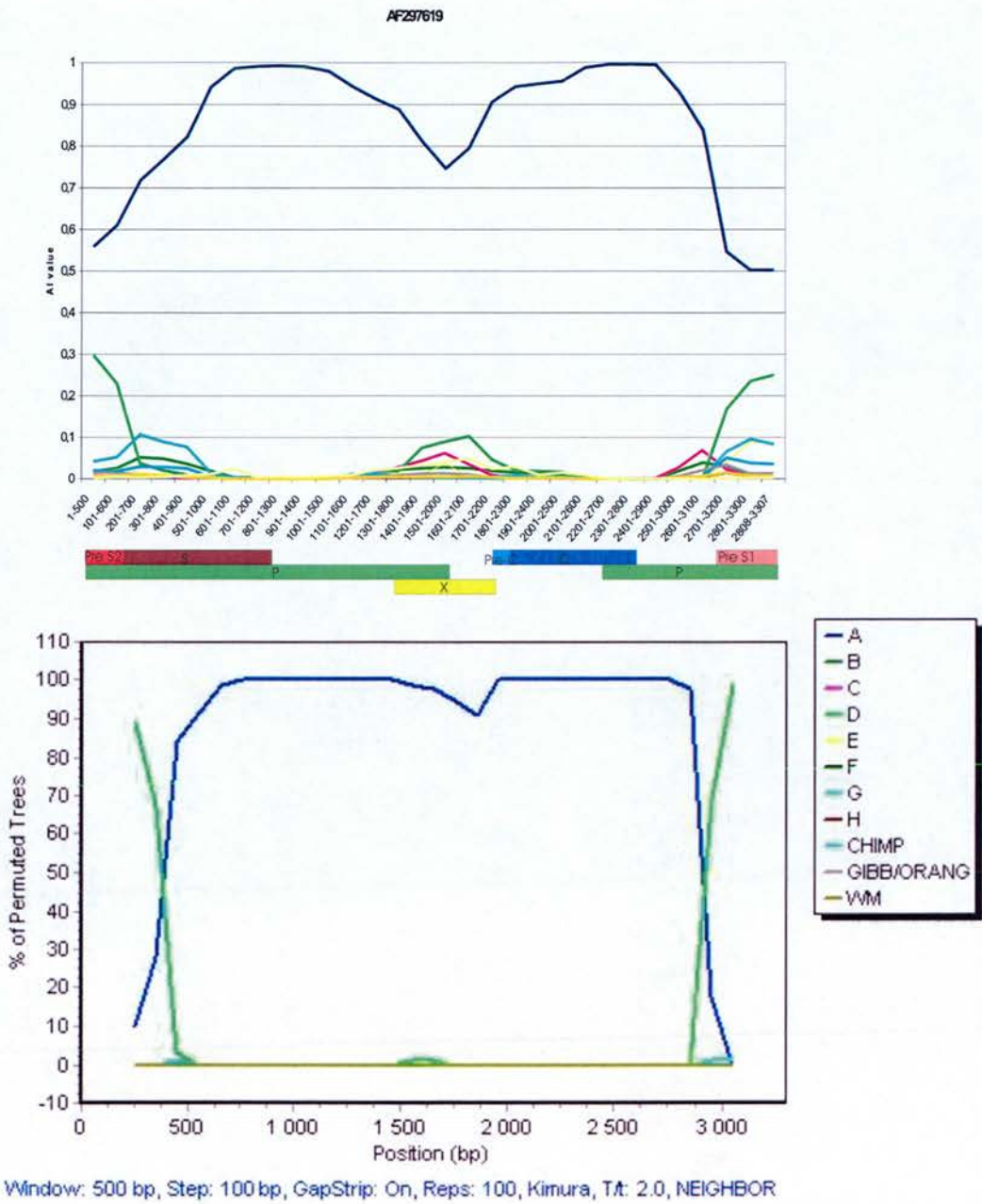


Figure D-2. Recombination analysis of HBV isolate AF297619 (Owiredu et al., 2001) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the PreS1/PreS2 region.

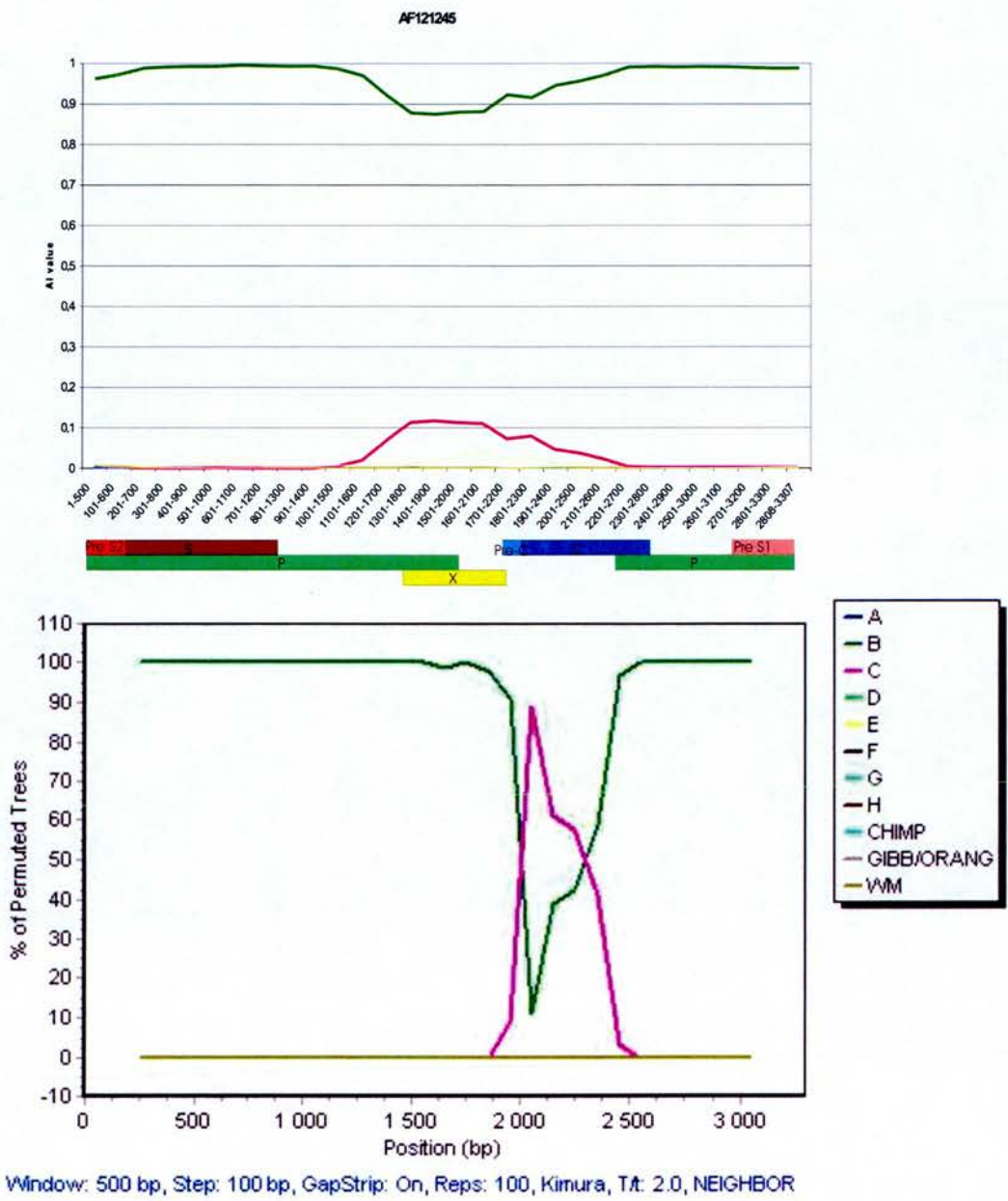


Figure D-3. Recombination analysis of HBV isolate AF121245 (Hannoun et al., 2000) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

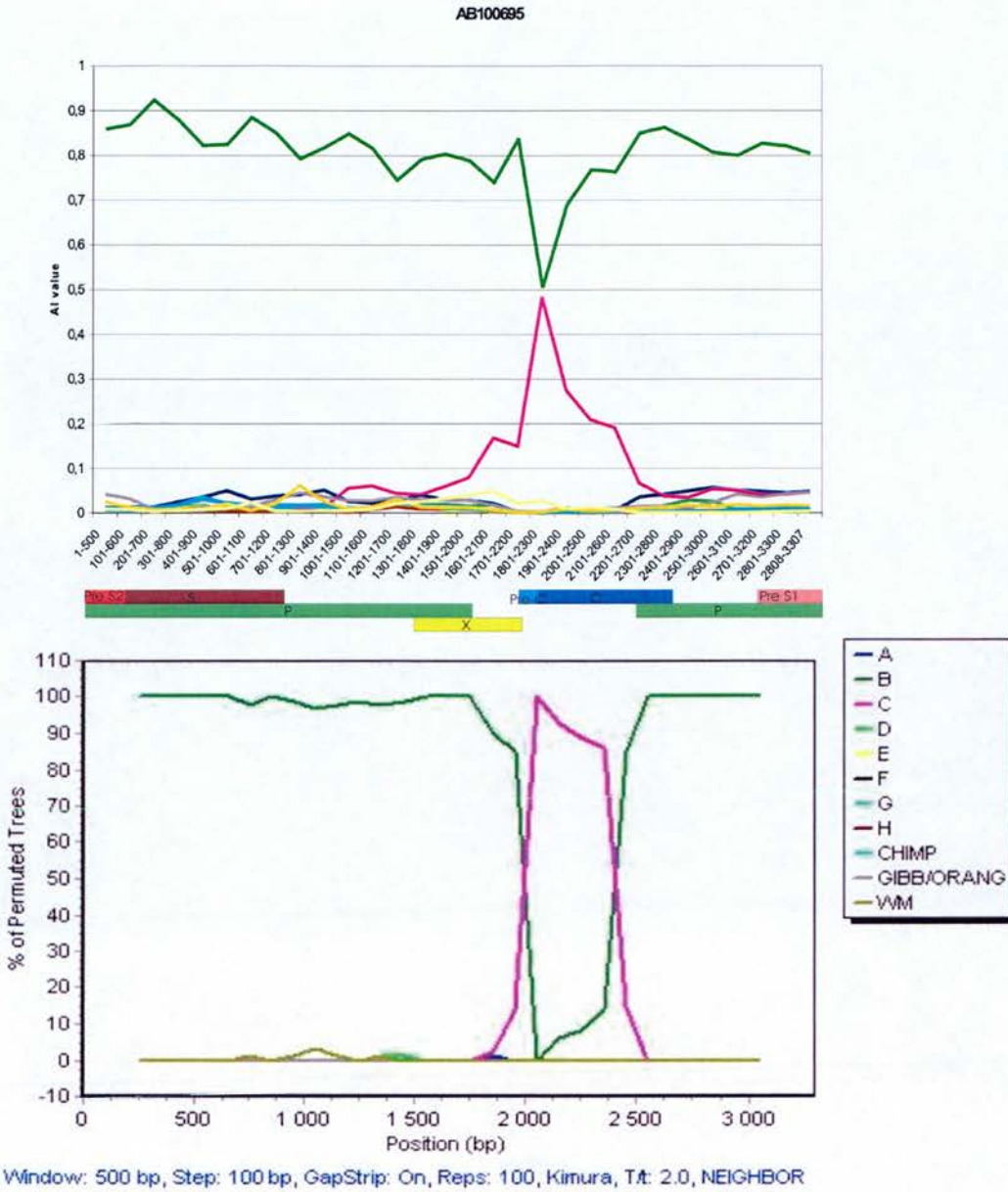


Figure D-4. Recombination analysis of HBV isolate AB100695 (Tran et al., 2003) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant with a divergent core region, SimPlot – recombinant in the core region.

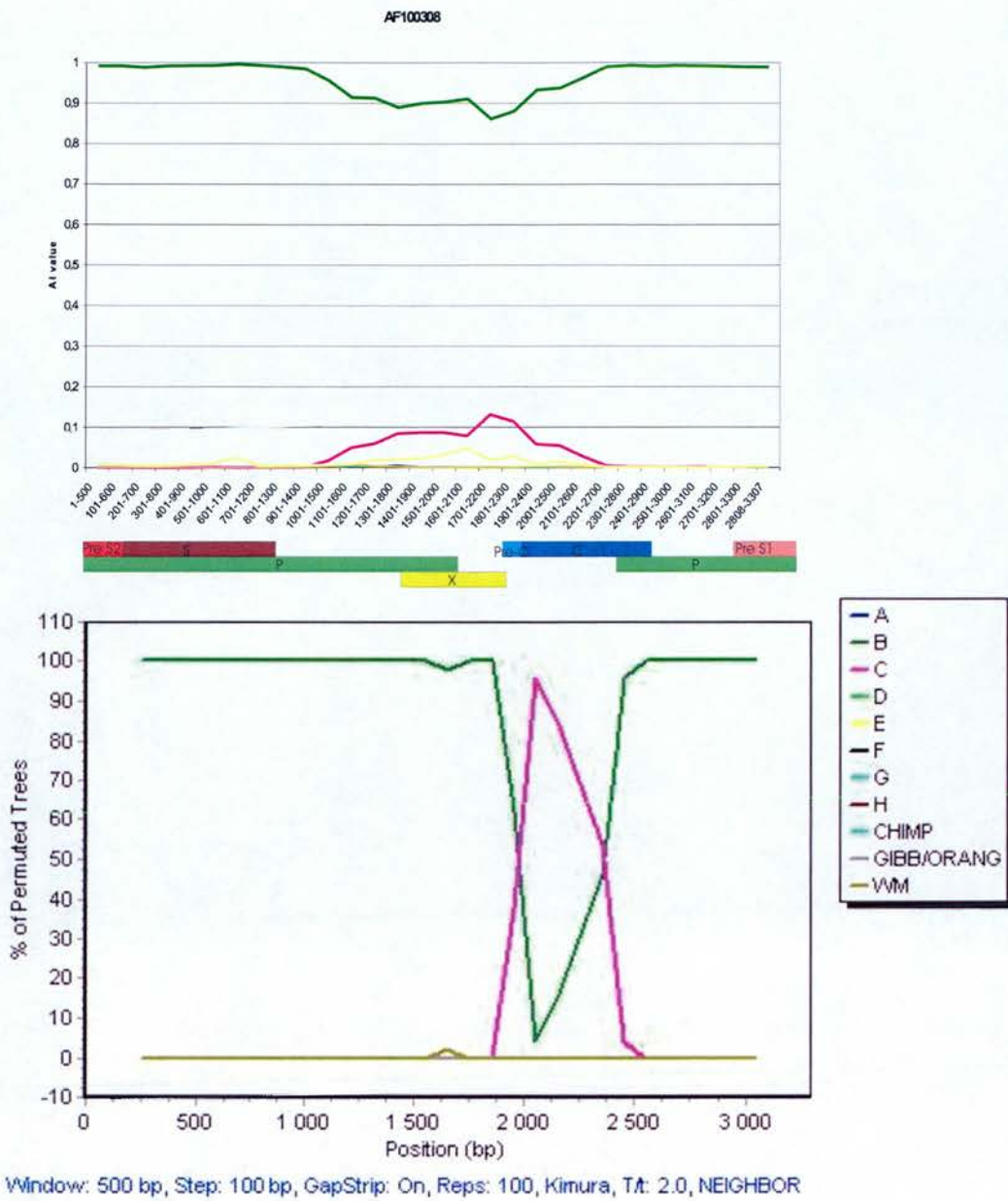


Figure D-5. Recombination analysis of HBV isolate AF100308 (Fares & Holmes, 2002) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

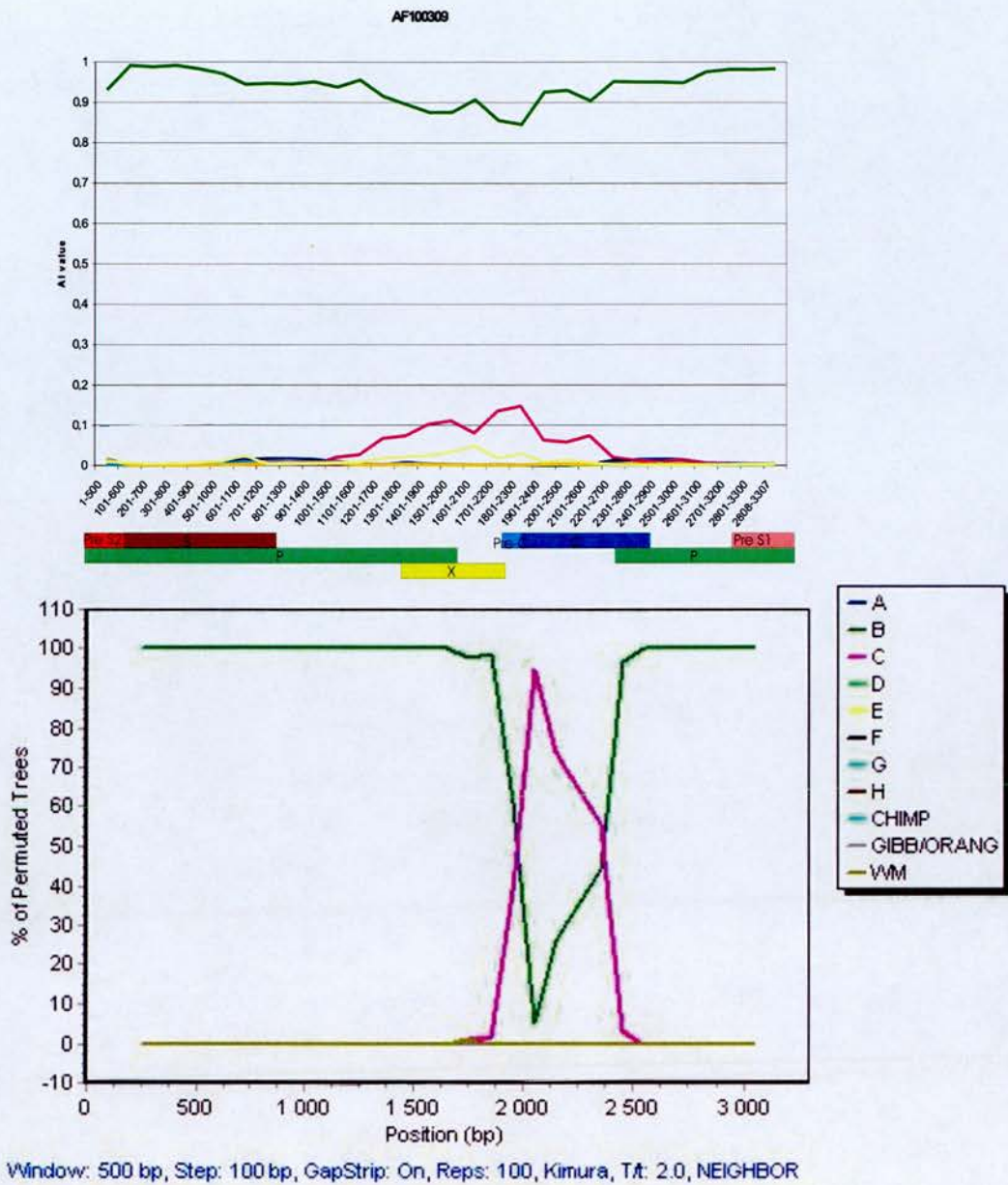


Figure D-6. Recombination analysis of HBV isolate AF100309 (Fares & Holmes, 2002) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

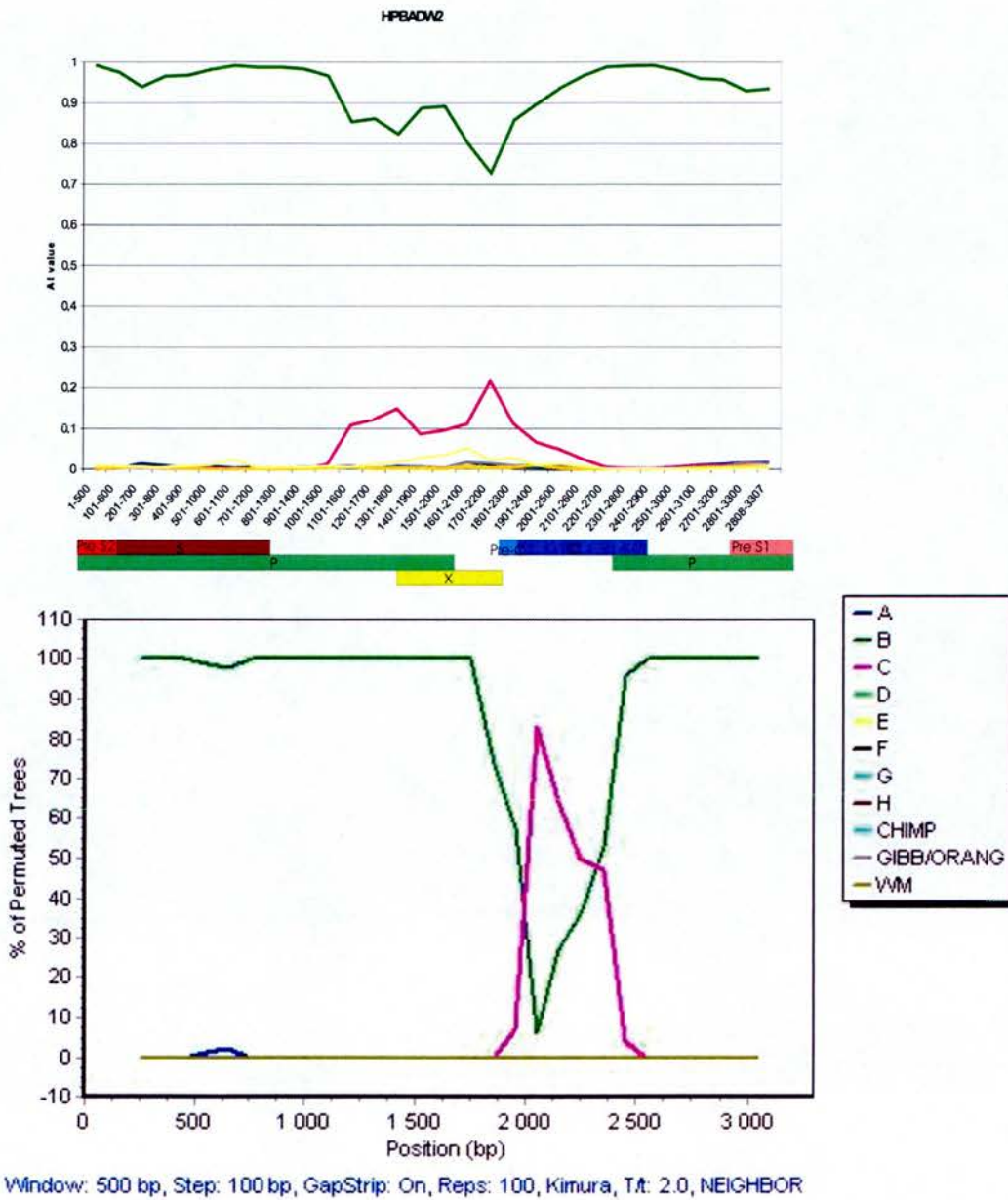


Figure D-7. Recombination analysis of HBV isolate HPBADW2 (Bowyer, 2000) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

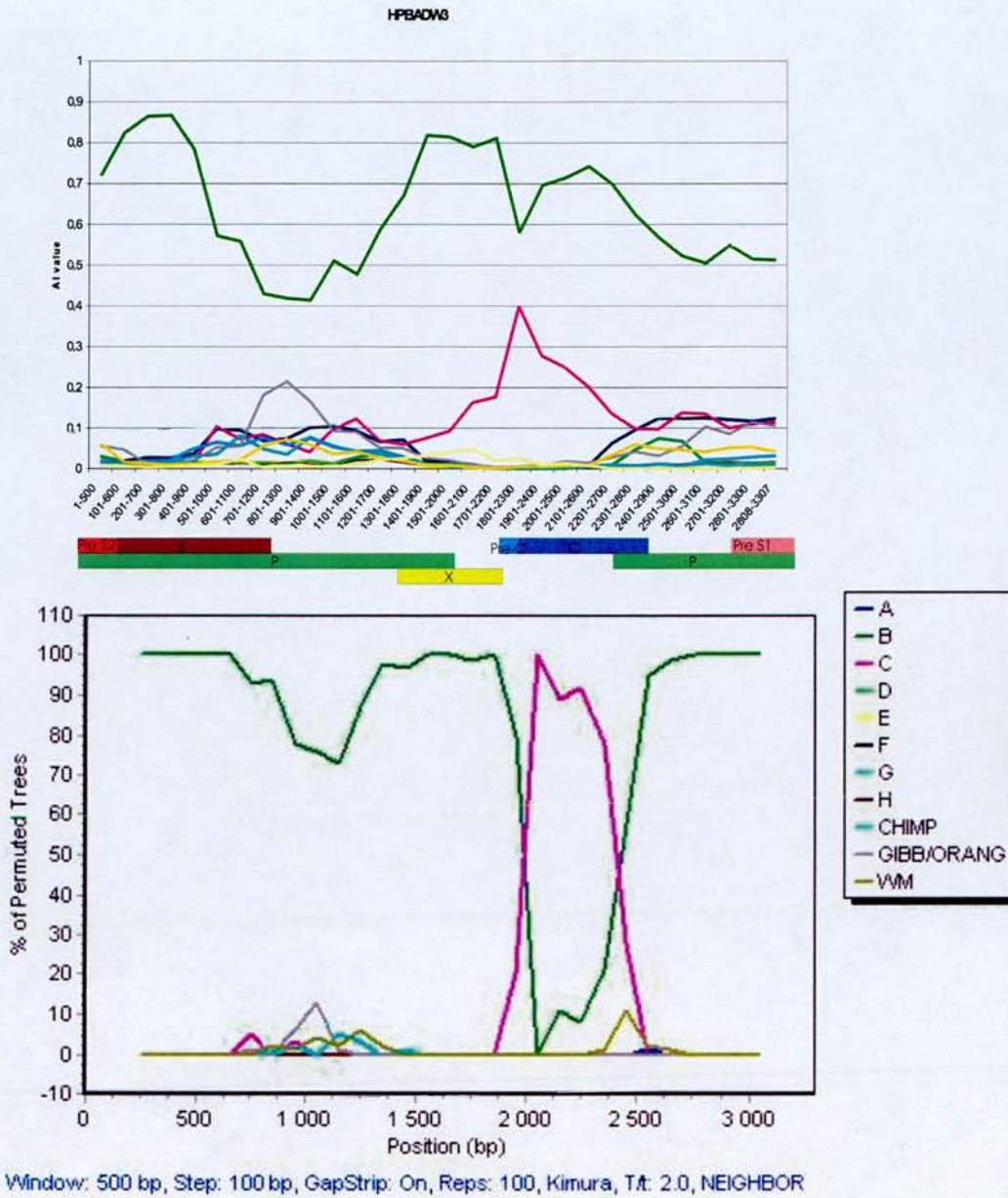


Figure D-8. Recombination analysis of HBV isolate HPBADW3 (Norder et al., 2003) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant divergent across the entire genome, SimPlot – recombinant in the core region with a divergent fragment in the single-coding region of the P ORF.

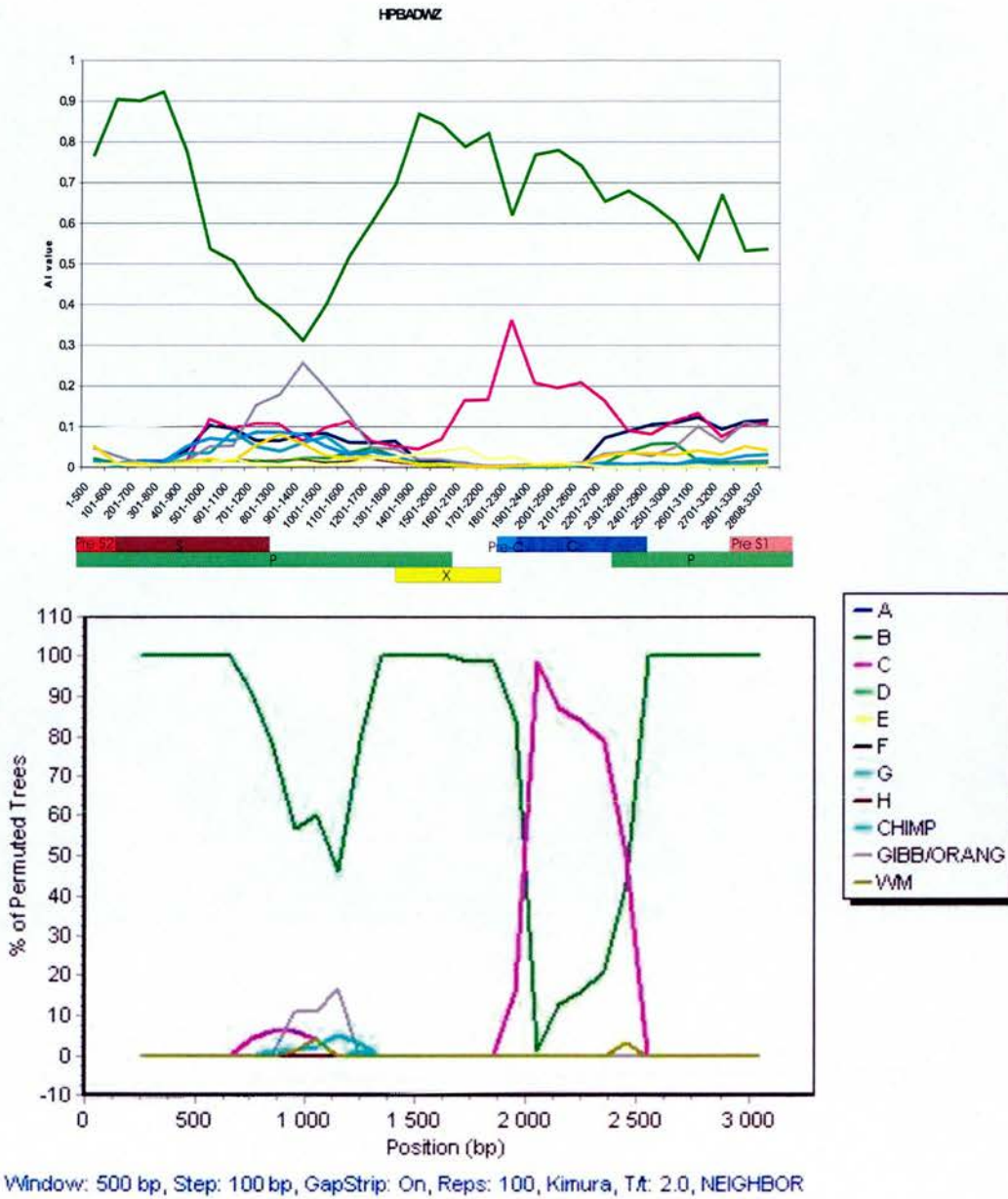


Figure D-9. Recombination analysis of HBV isolate HPBADWZ (Bowyer, 2000) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant divergent across the entire genome, particularly the region spanning the S gene and Pol overlap, SimPlot – recombinant in the core region with a divergent fragment in the single-coding part of the P ORF.

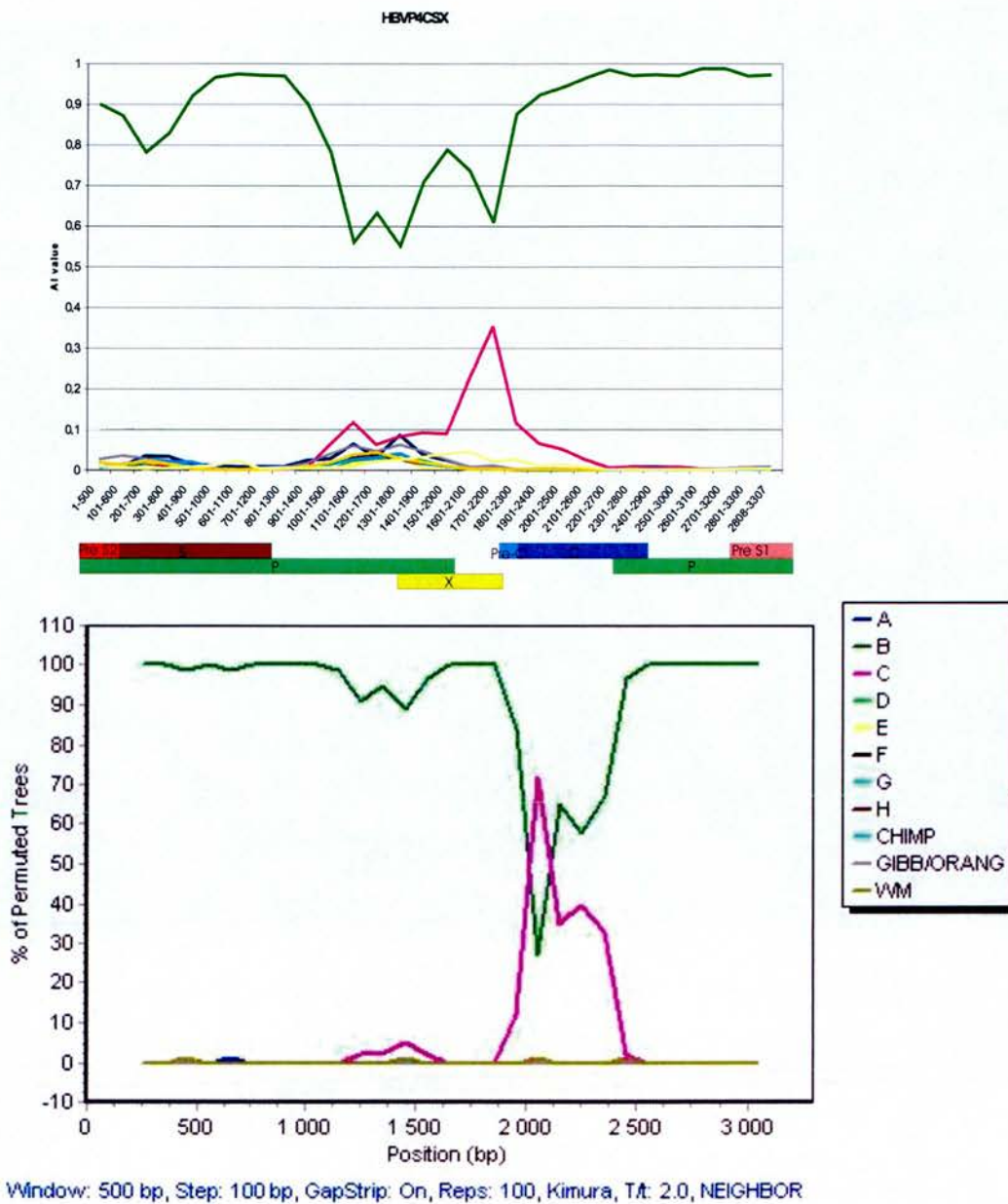


Figure D-10. Recombination analysis of HBV isolate HBVP4CSX (Bowyer, 2000) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant with a divergent P/X region, SimPlot – recombinant in the core region.

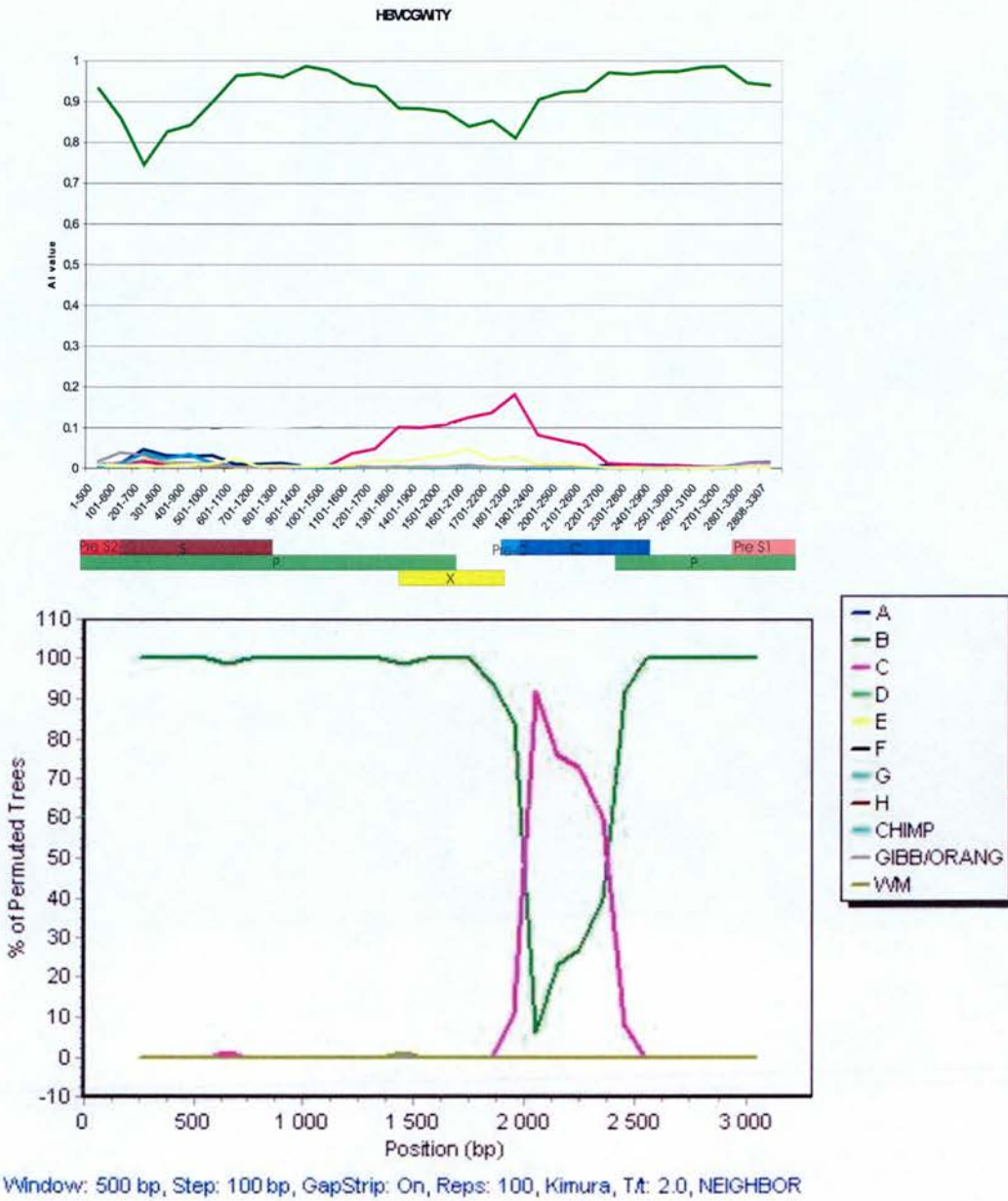


Figure D-11. Recombination analysis of HBV isolate HBVCWGITY (Bowyer, 2000) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

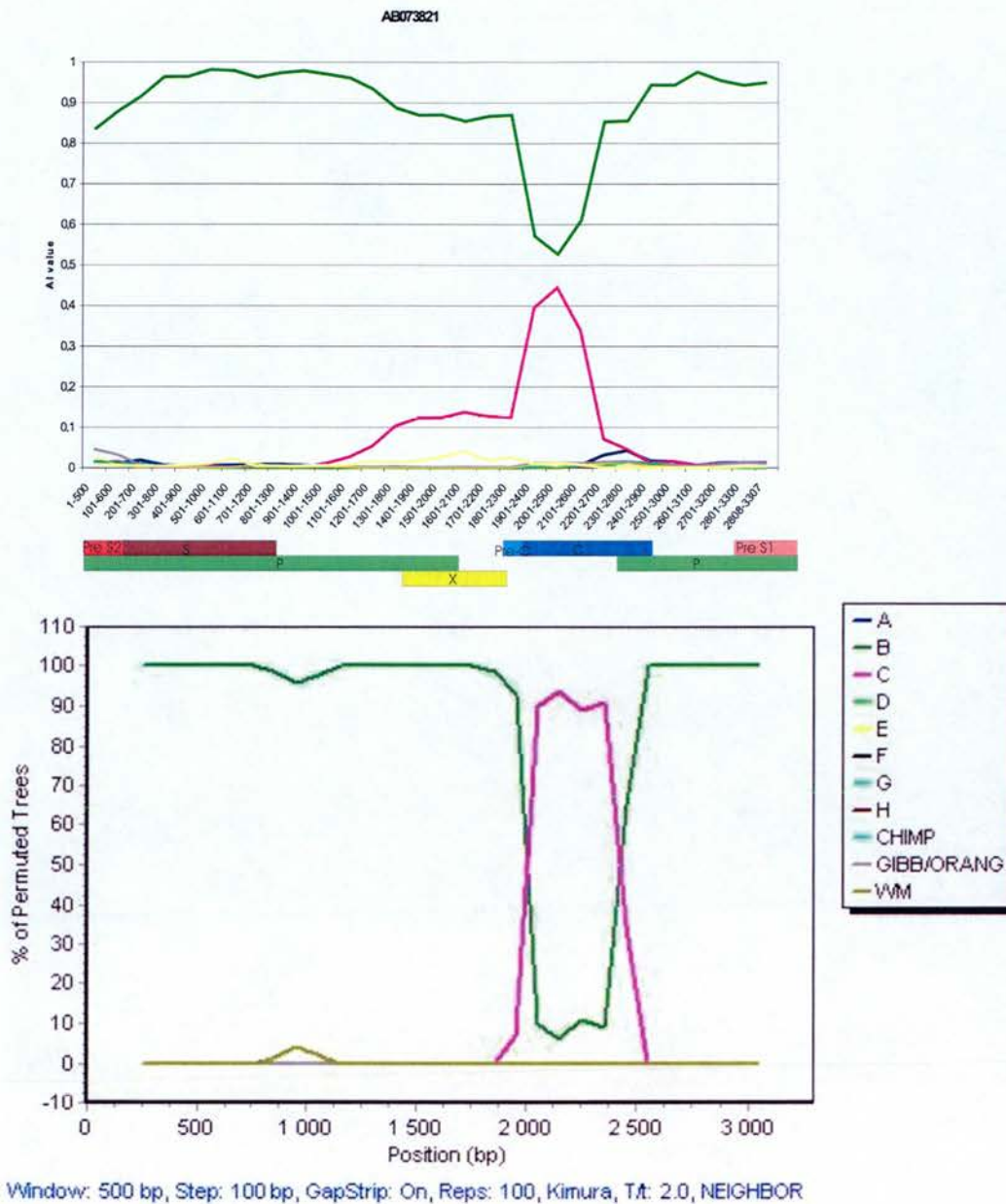


Figure D-12. Recombination analysis of HBV isolate AB073821 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant with a divergent core region, SimPlot – recombinant in the core region.

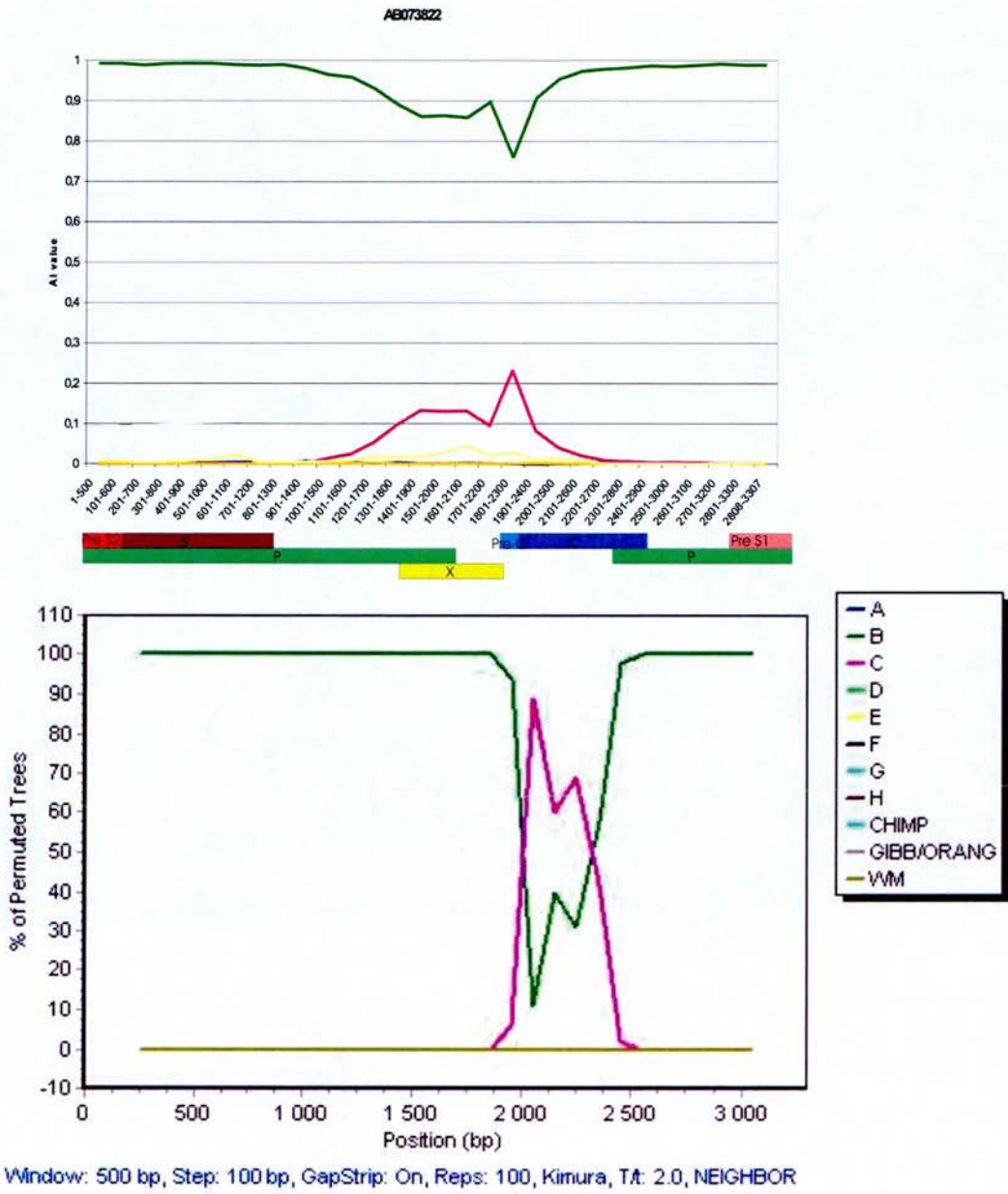


Figure D-13. Recombination analysis of HBV isolate AB073822 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

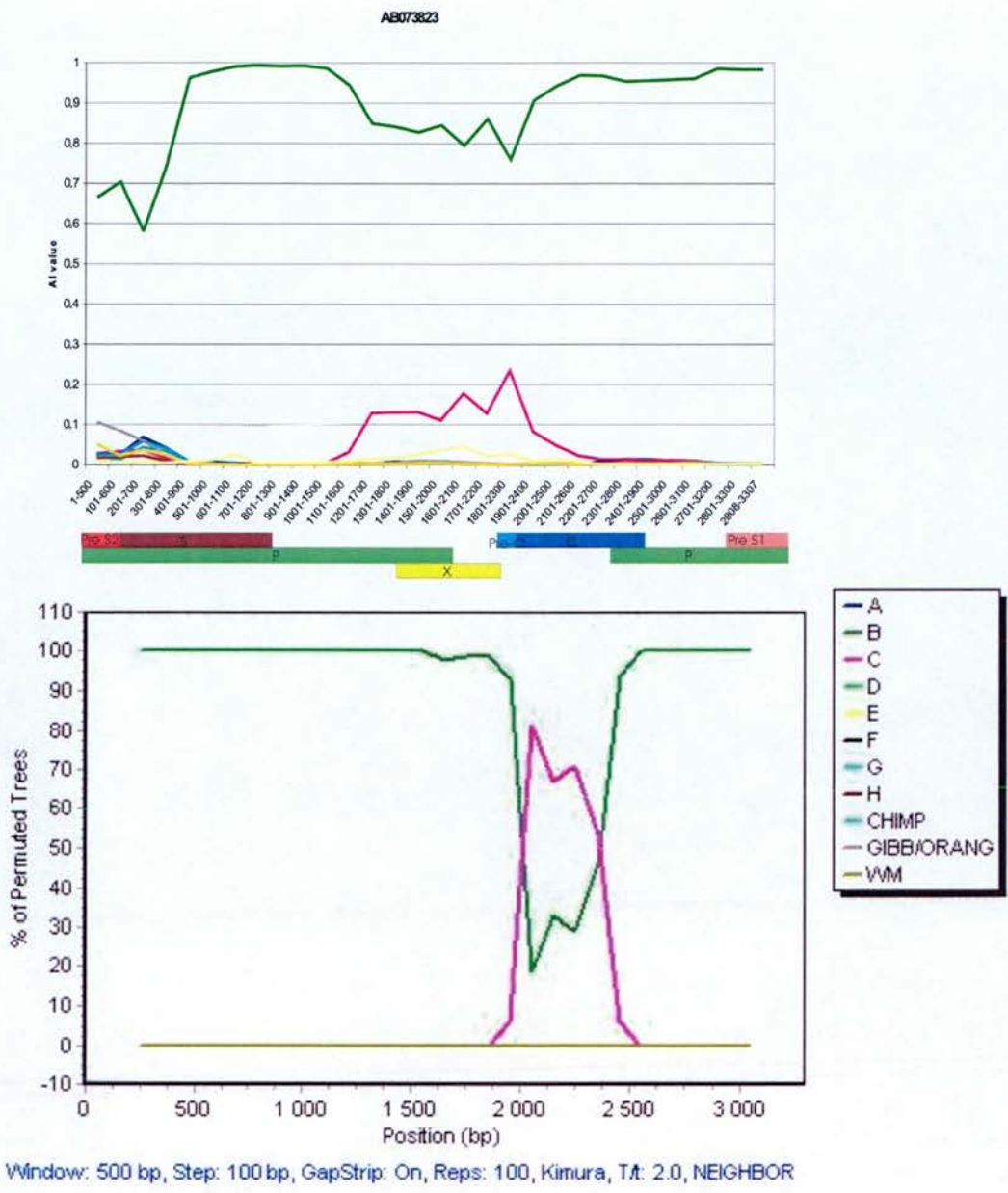


Figure D-14. Recombination analysis of HBV isolate AB073823 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant with a divergent PreS2 region, SimPlot – recombinant in the core region.

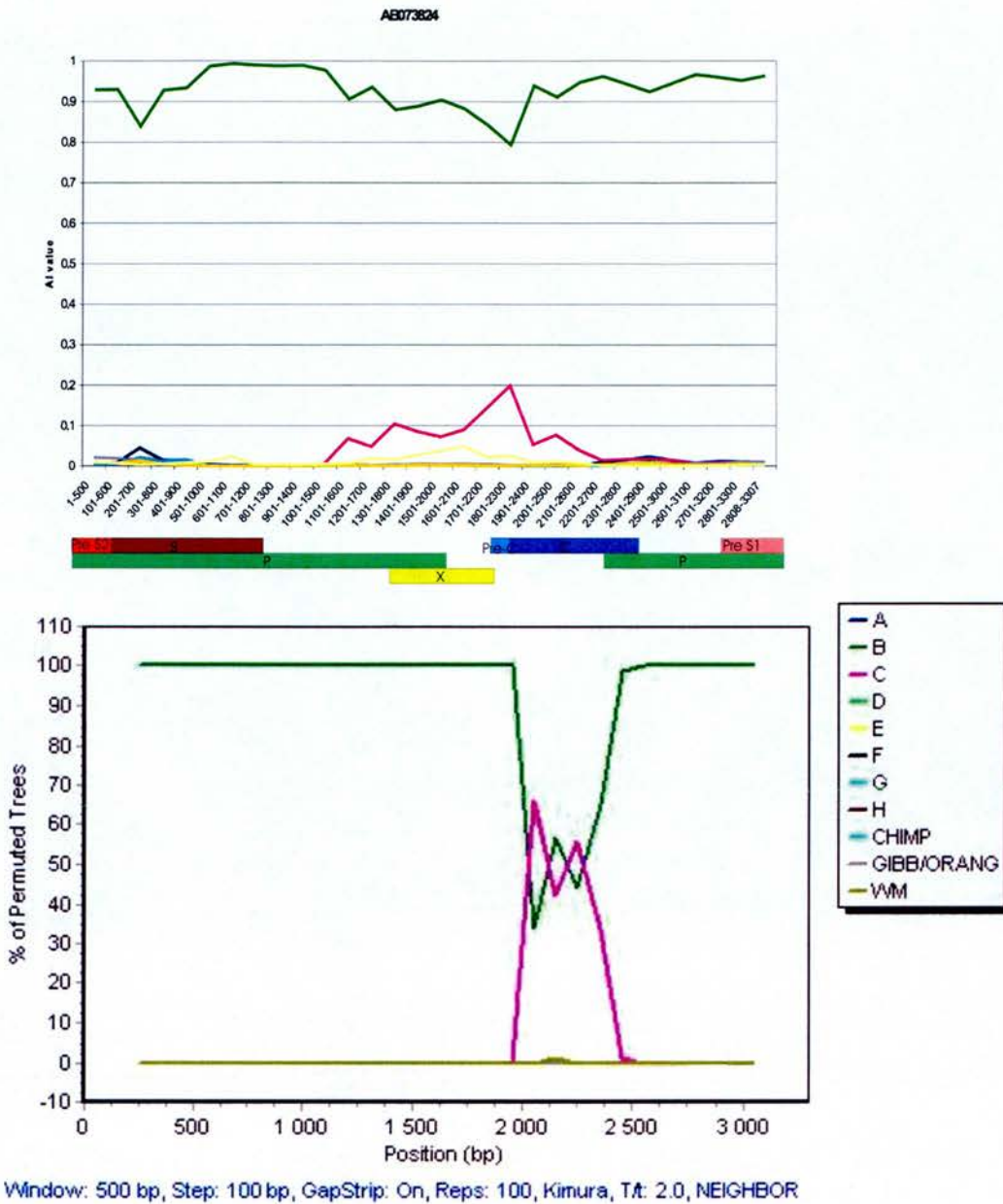


Figure D-15. Recombination analysis of HBV isolate AB073824 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

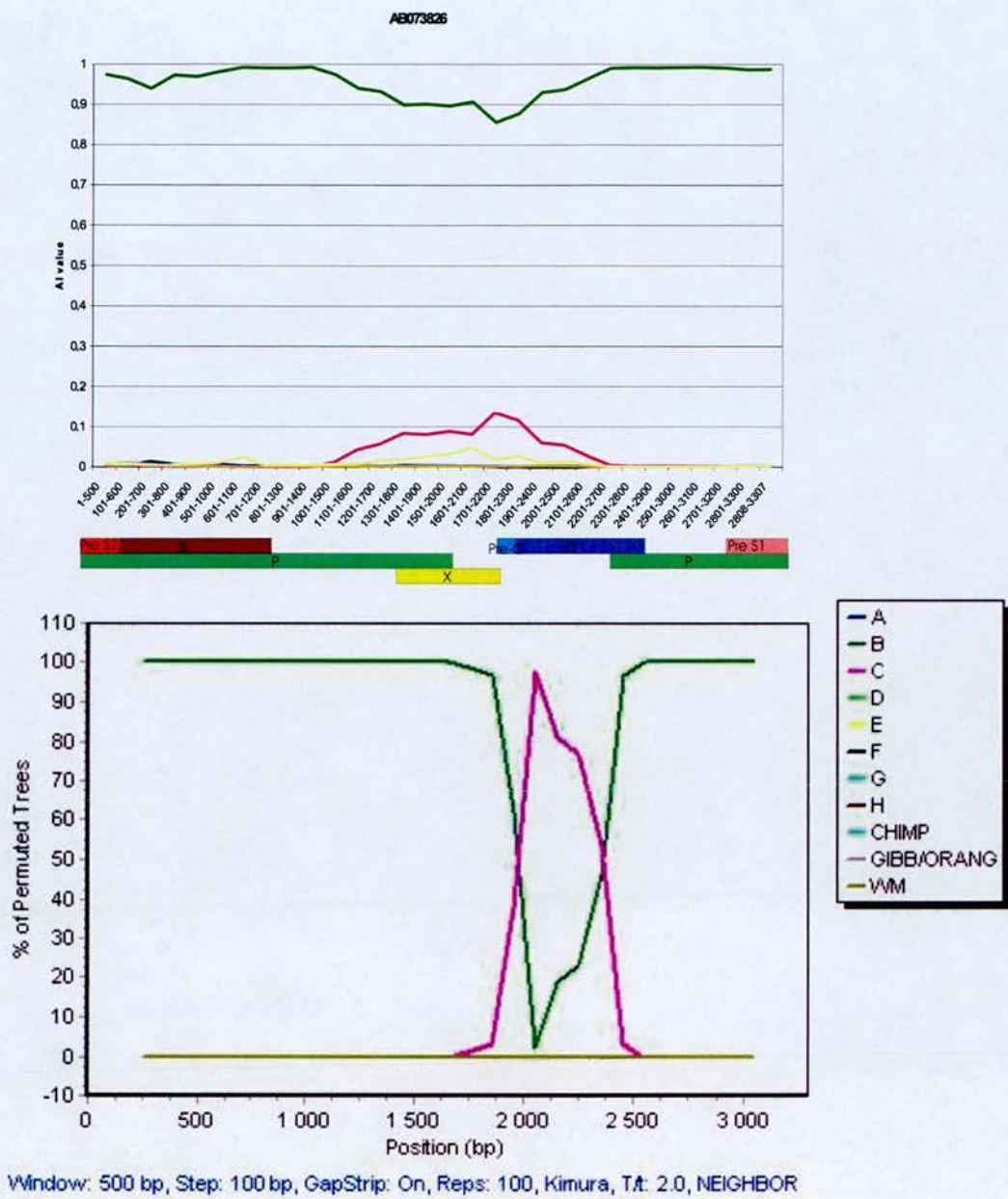


Figure D-16. Recombination analysis of HBV isolate AB073826 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

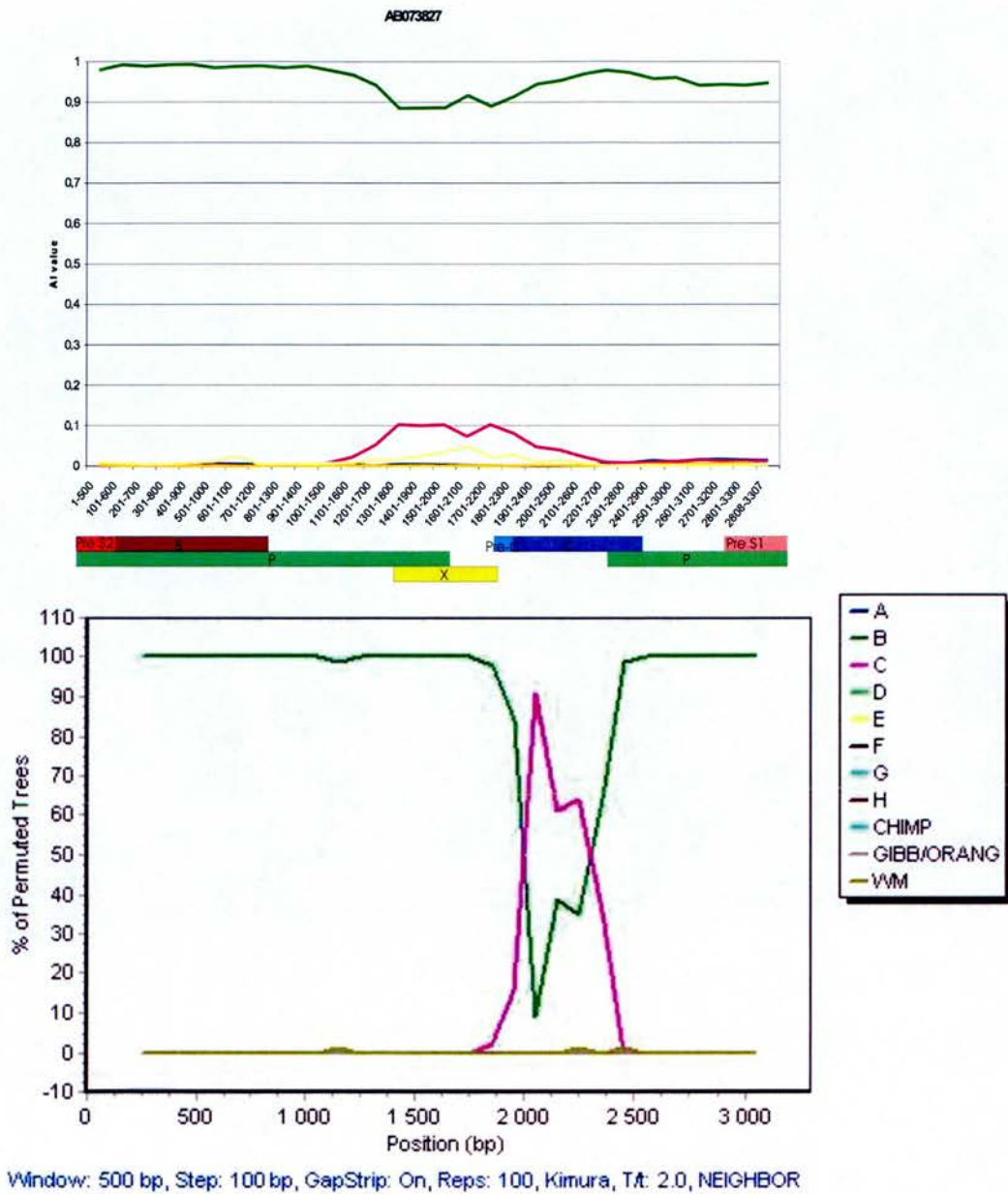


Figure D-17. Recombination analysis of HBV isolate AF073827 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

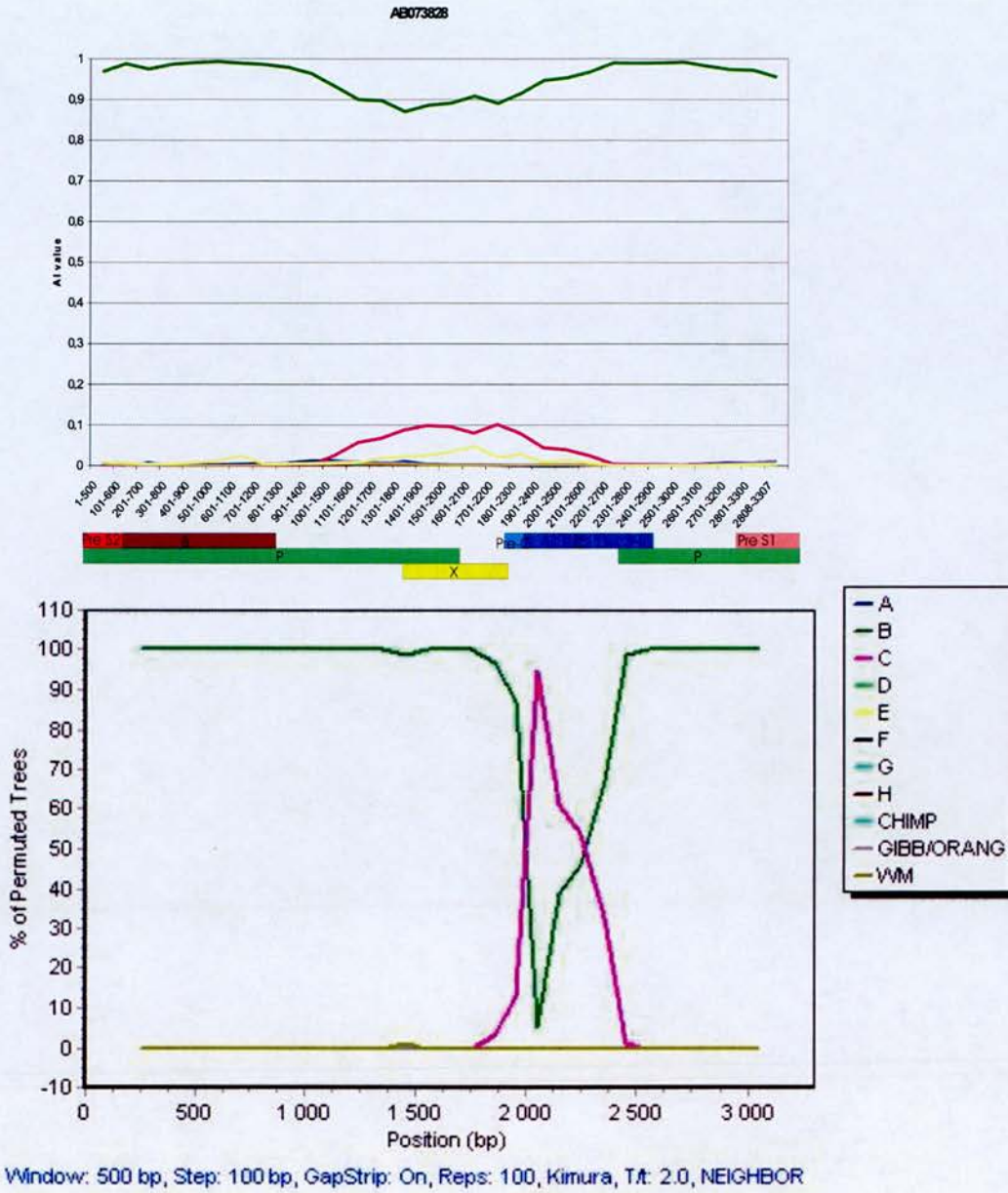


Figure D-18. Recombination analysis of HBV isolate AB073828 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

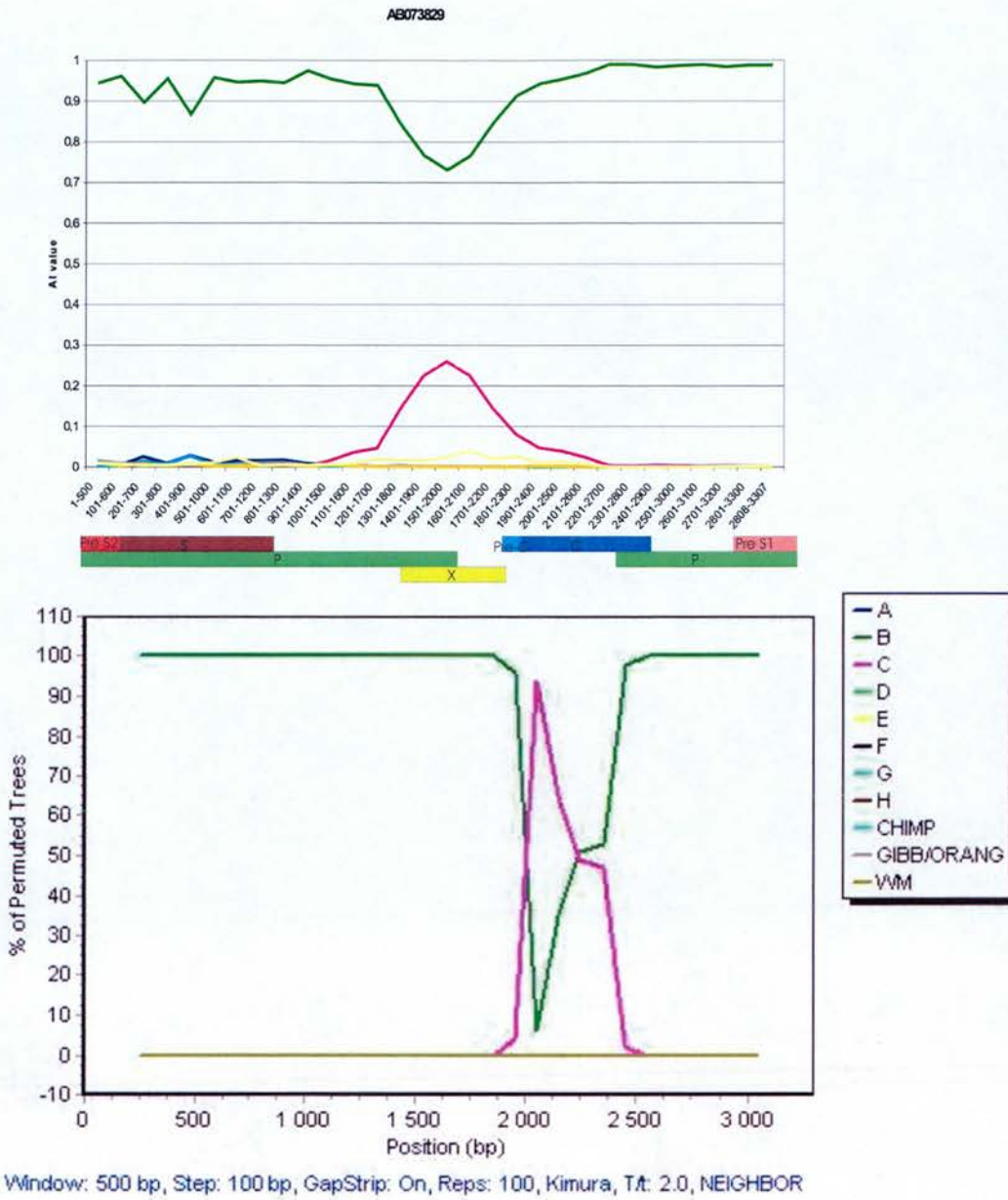


Figure D-19. Recombination analysis of HBV isolate AB073829 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

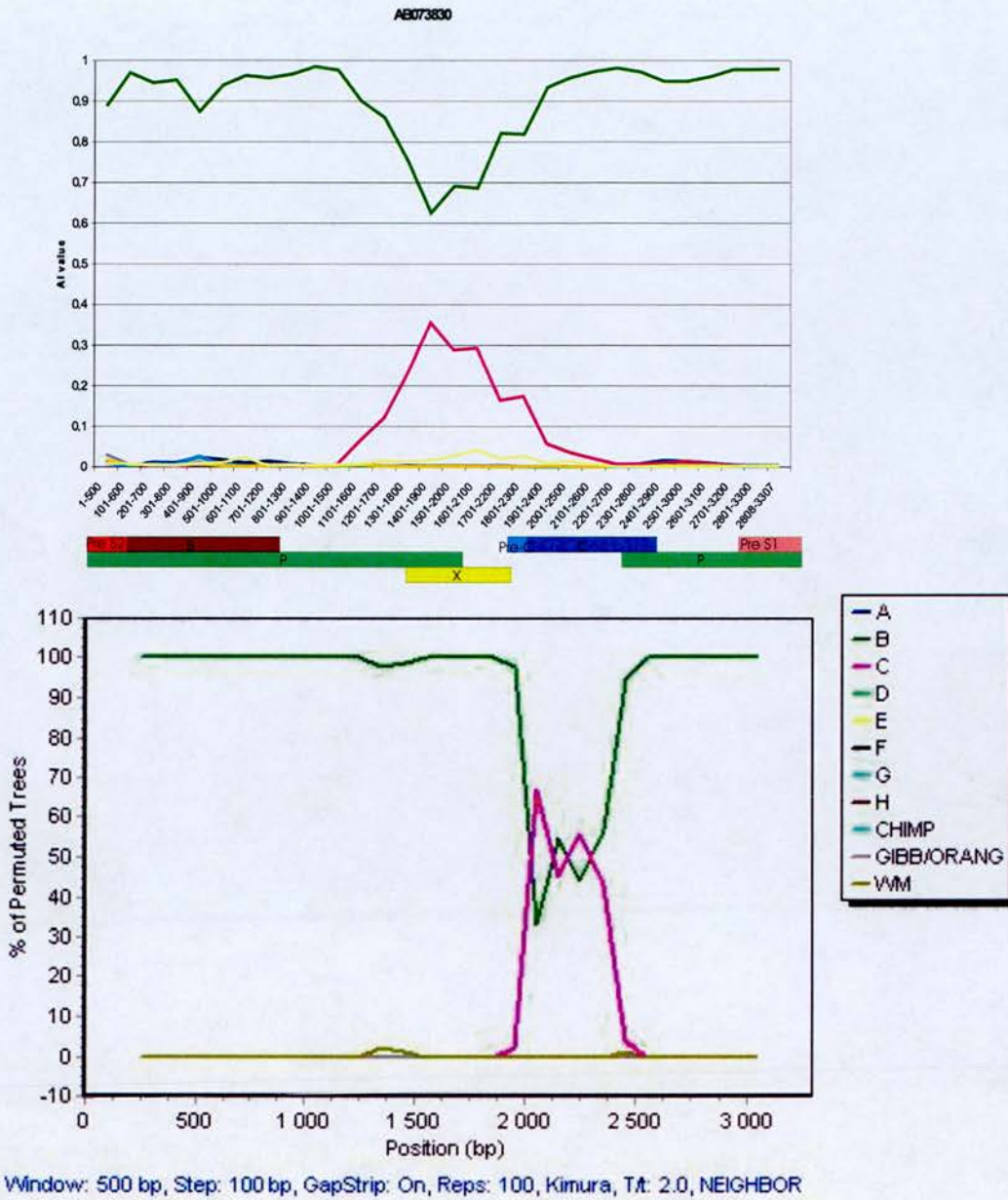


Figure D-20. Recombination analysis of HBV isolate AB073830 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant with a divergent core region, SimPlot – recombinant in the core region.

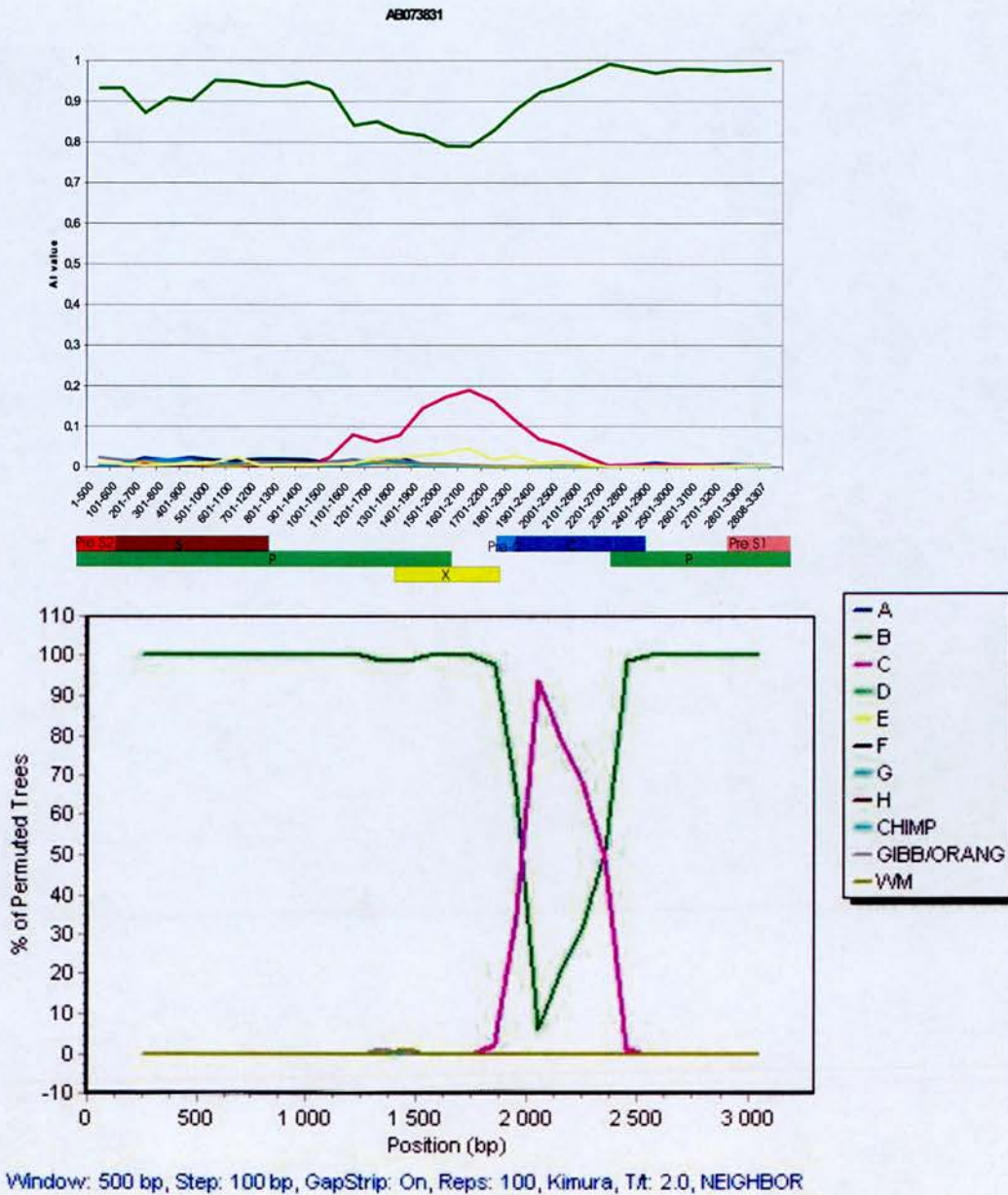


Figure D-21. Recombination analysis of HBV isolate AB073831 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

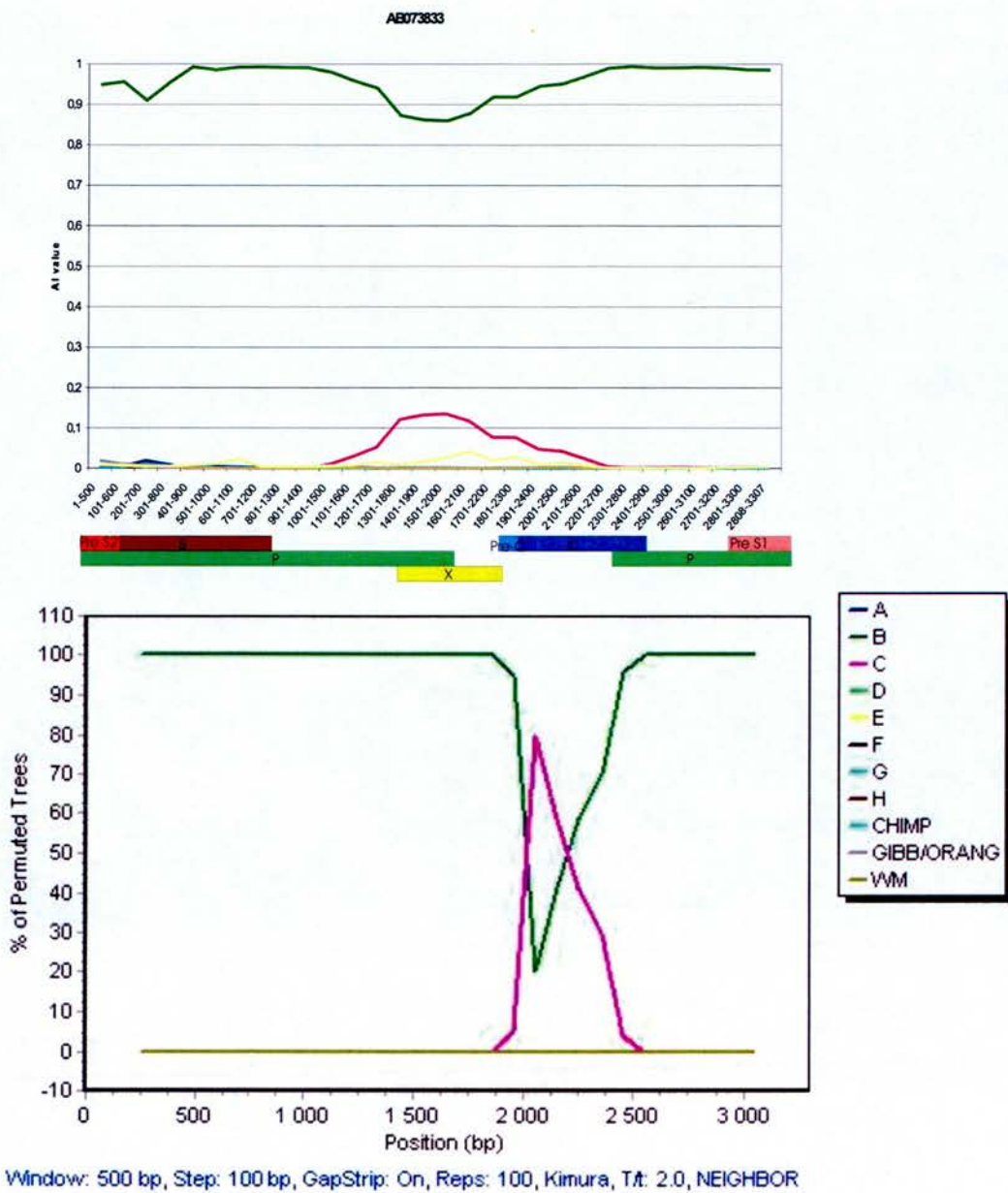


Figure D-22. Recombination analysis of HBV isolate AB073833 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

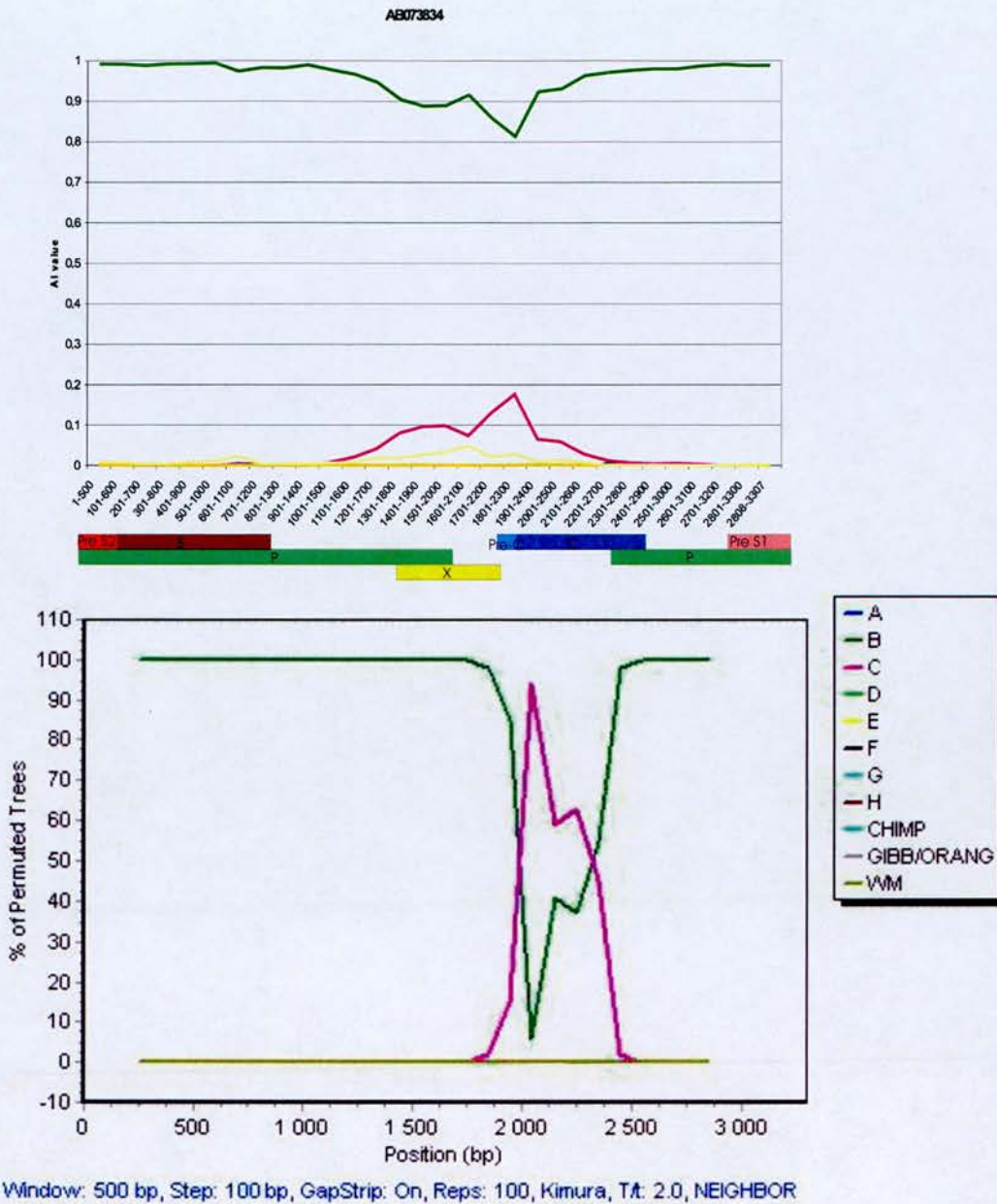


Figure D-23. Recombination analysis of HBV isolate AB073834 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

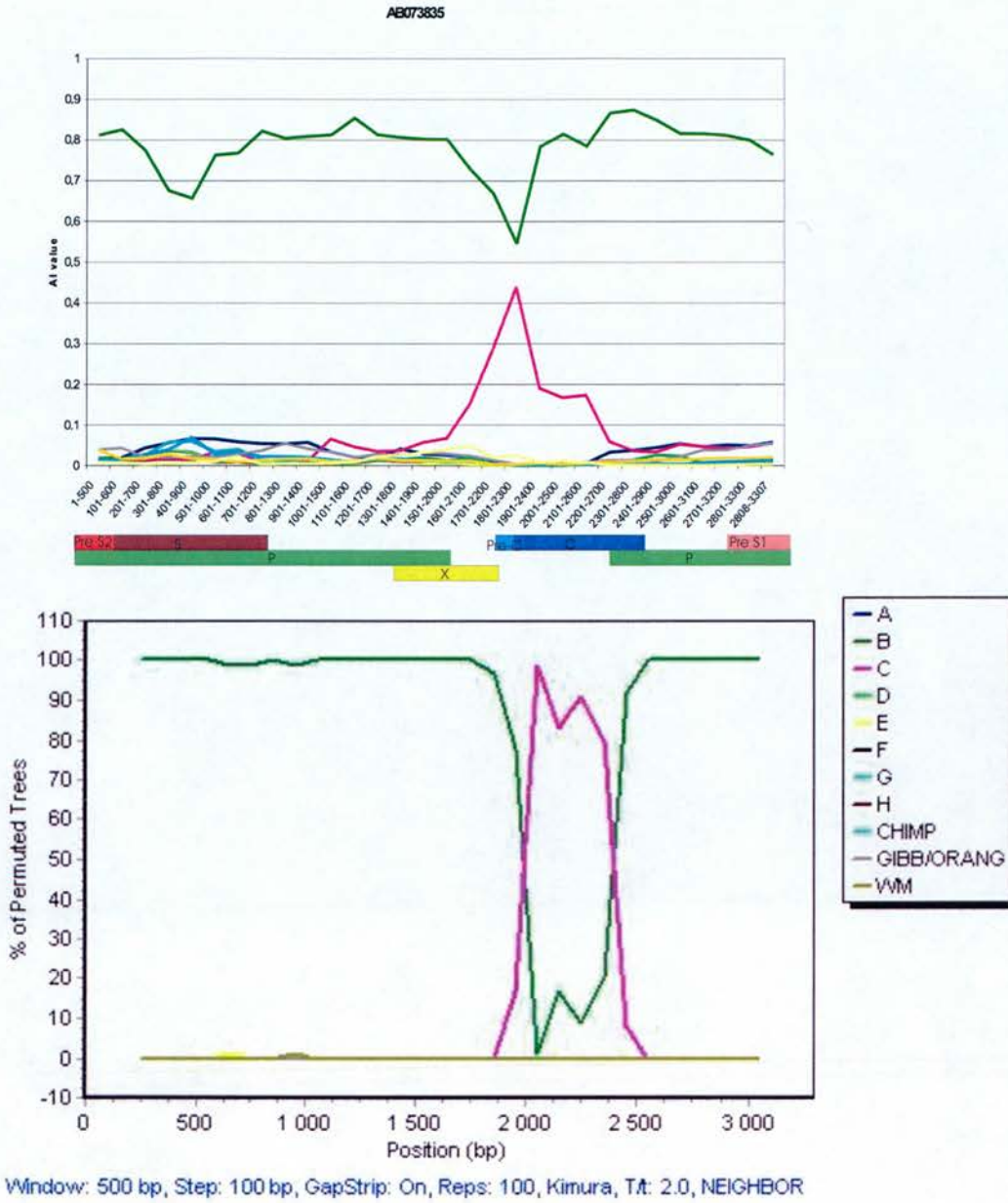


Figure D-24. Recombination analysis of HBV isolate AB073835 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant divergent across the entire genome, SimPlot – recombinant in the core region.

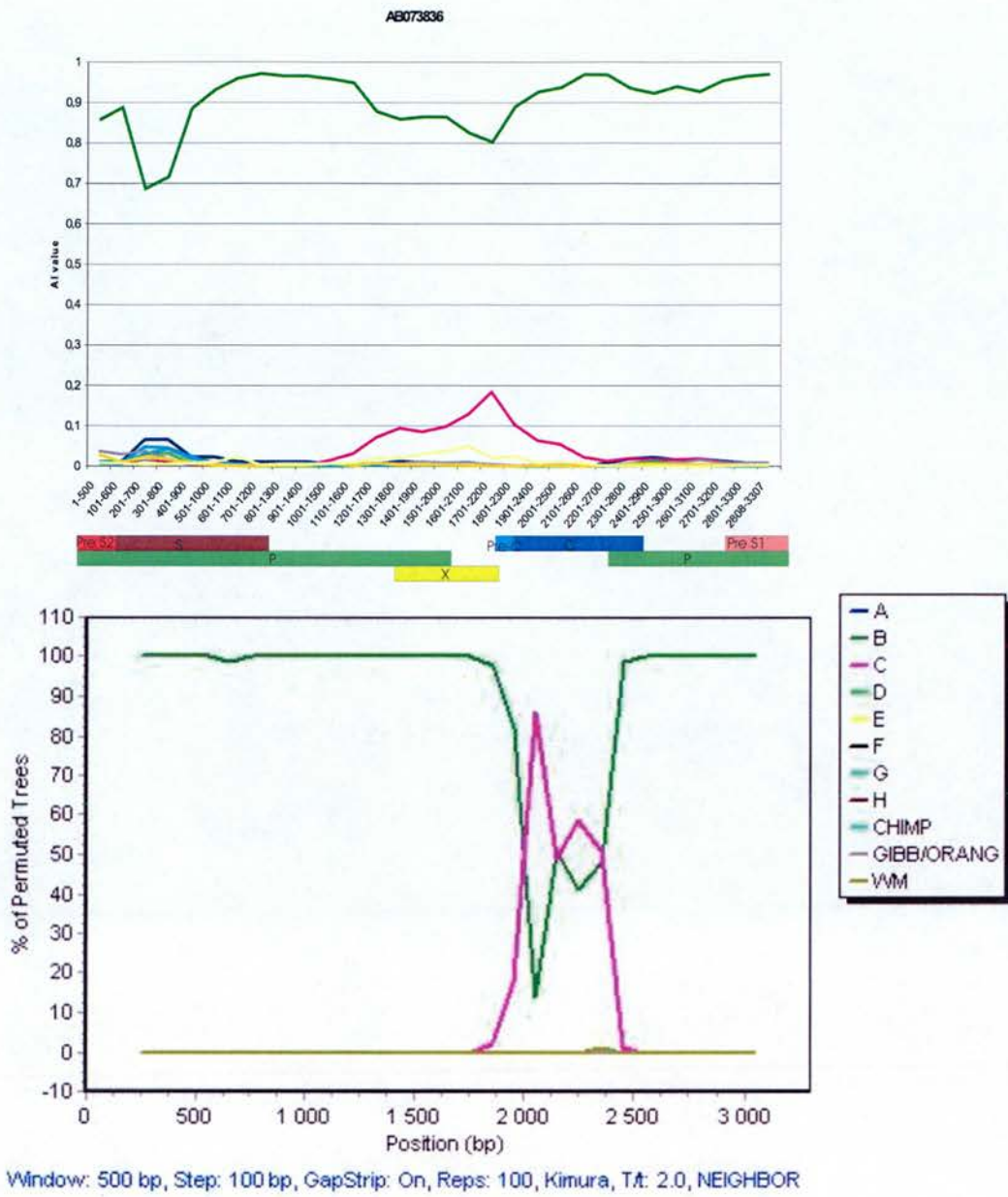


Figure D-25. Recombination analysis of HBV isolate AB073836 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot - recombinant in the core region.

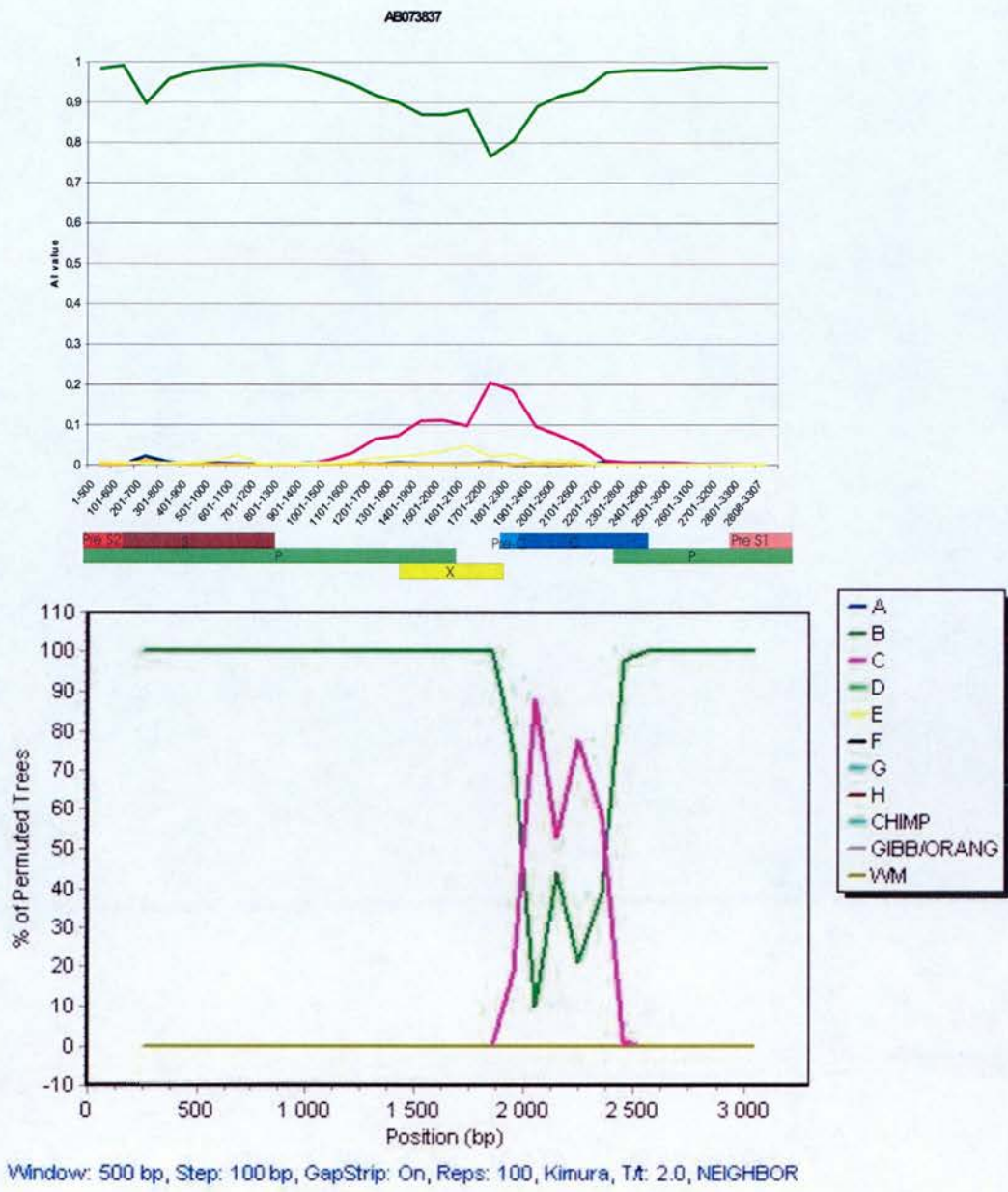


Figure D-26. Recombination analysis of HBV isolate AB073837 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

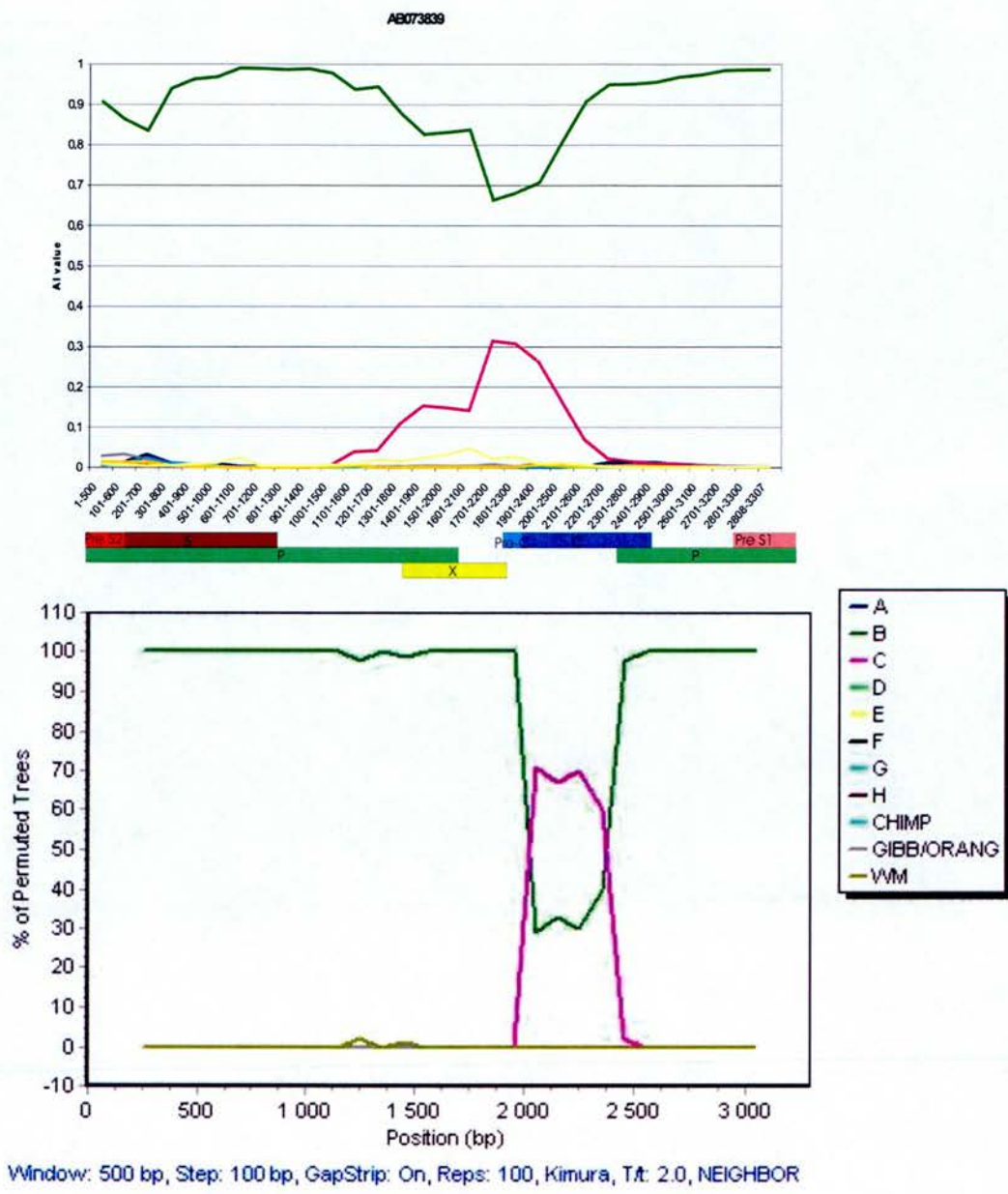


Figure D-27. Recombination analysis of HBV isolate AB073839 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant with a divergent core region, SimPlot – recombinant in the core region.

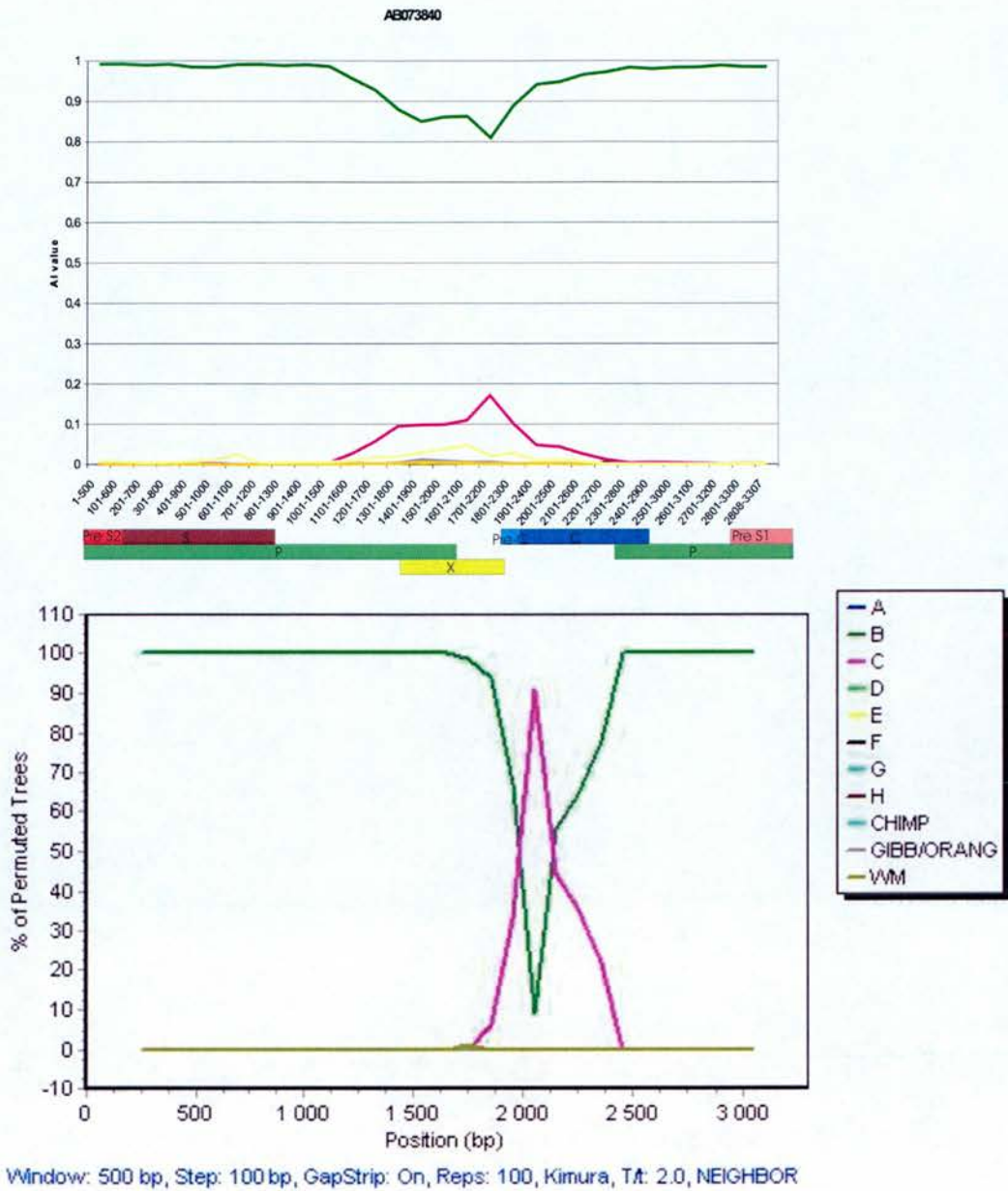


Figure D-28. Recombination analysis of HBV isolate AB073840 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

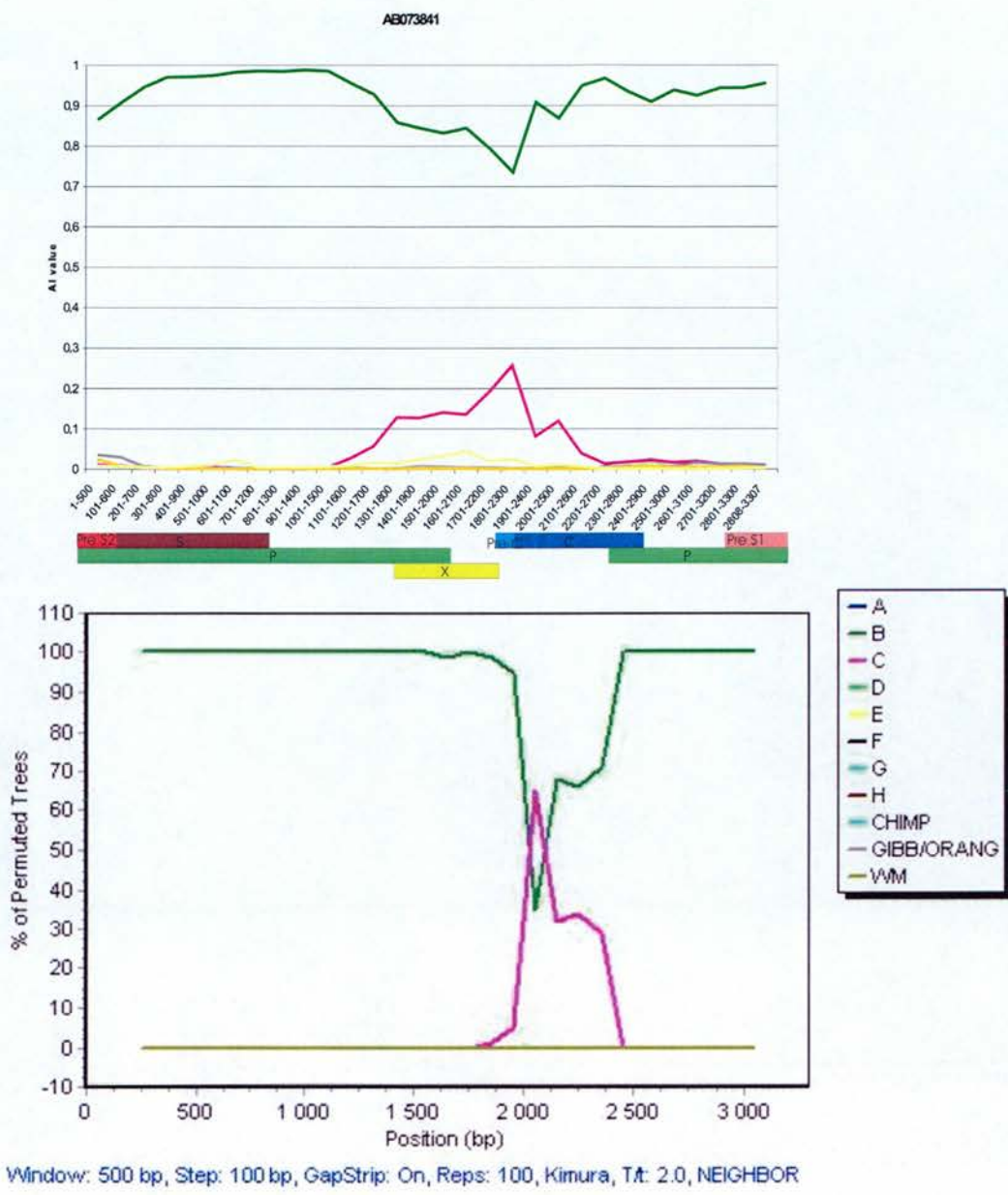
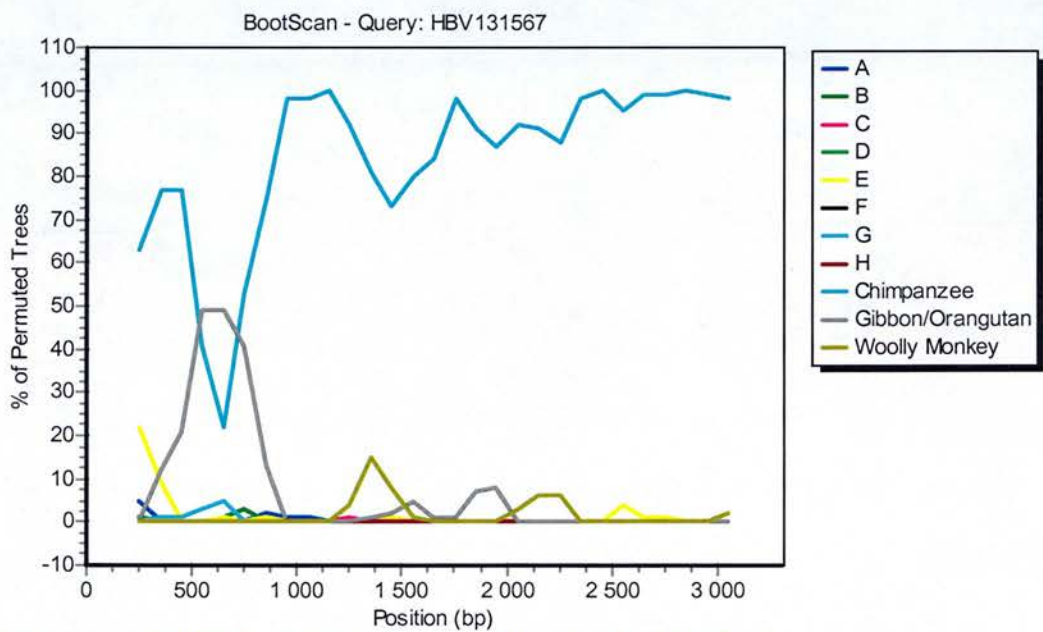
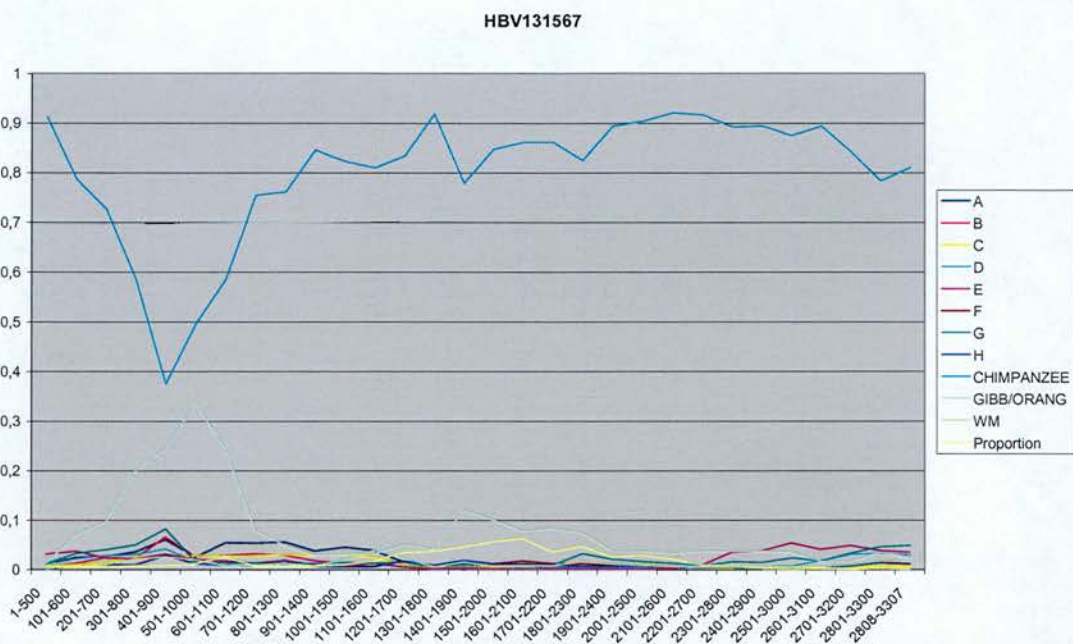


Figure D-29. Recombination analysis of HBV isolate AB073841 (Sugauchi et al., 2002b) using the AI (top) and SimPlot (bottom) methods. Interpretation: AI – non-recombinant, SimPlot – recombinant in the core region.

# **Appendix E**

## **AI and SimPlot Analysis of Primate isolates**

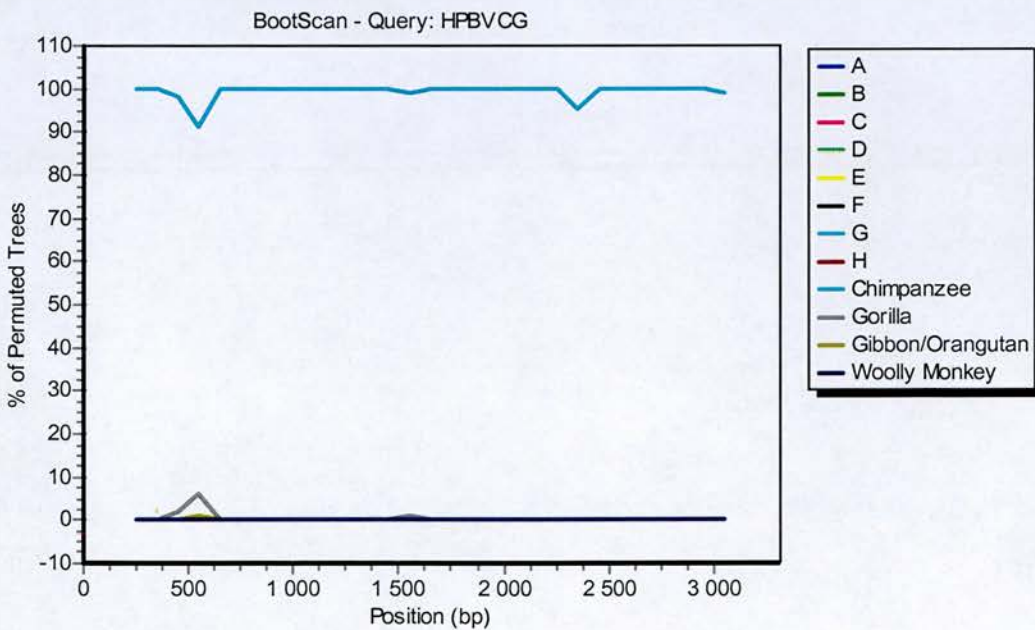
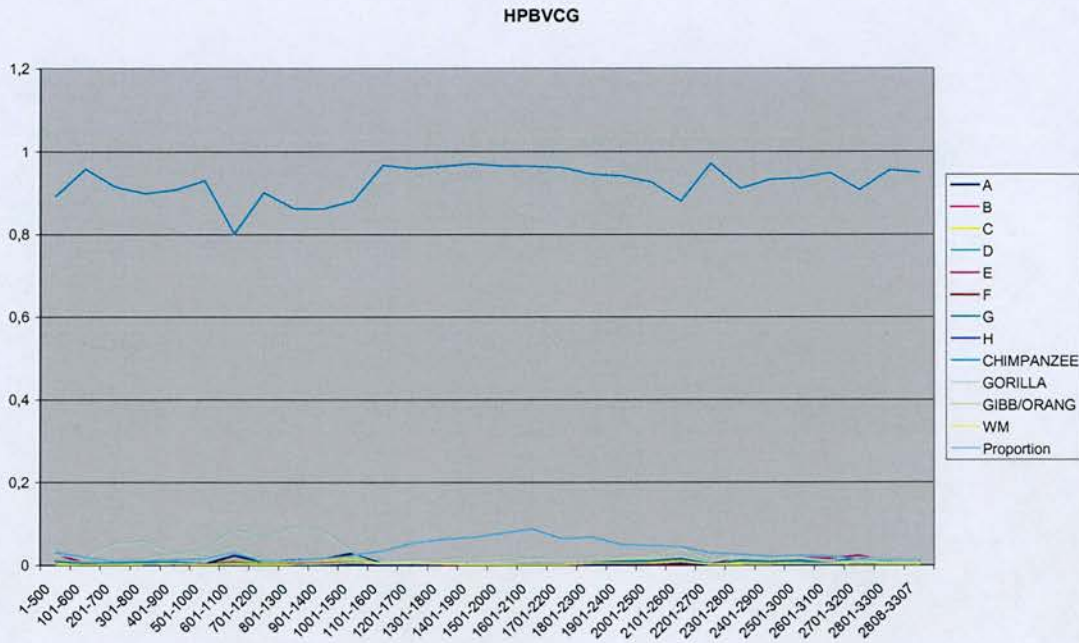
Figure E-1. AI and SimPlot analysis of the gorilla sequence, using the combined gibbon/orang-utan group.



Window : 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

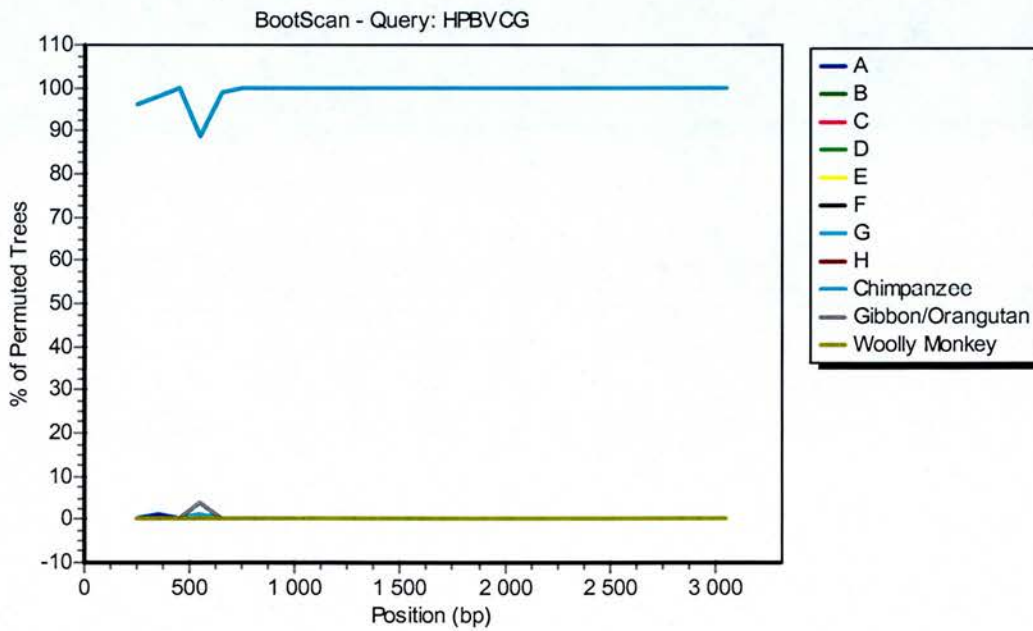
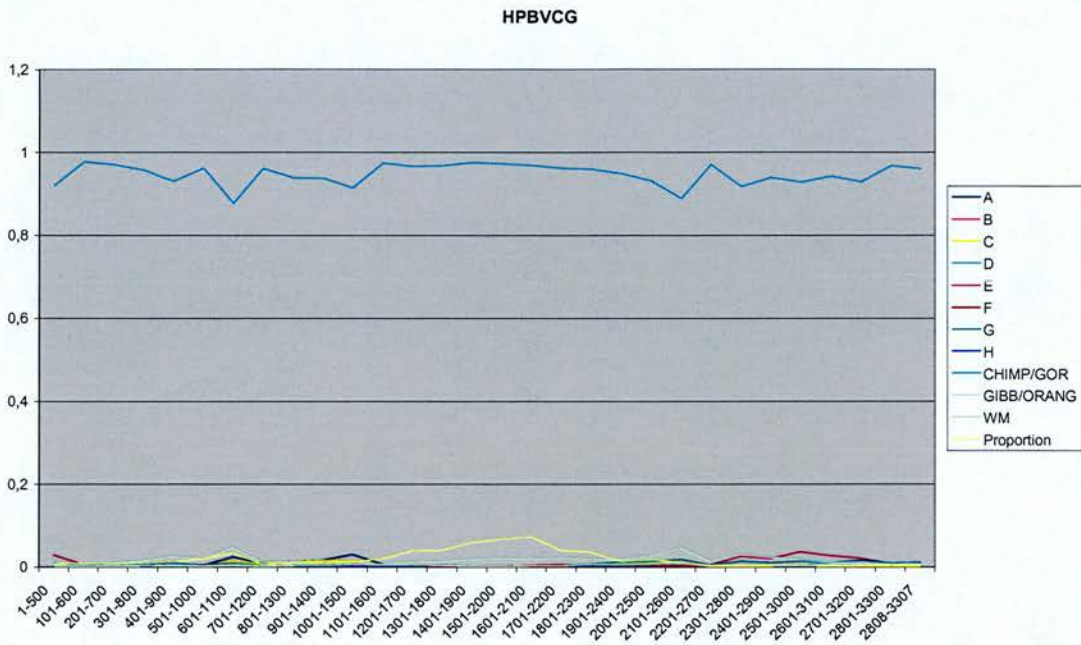
Figure E-2. AI and SimPlot analysis of a chimpanzee control sequence using separate groups for chimpanzee and gorilla (A) and the combined chimpanzee and gorilla group (B).

A



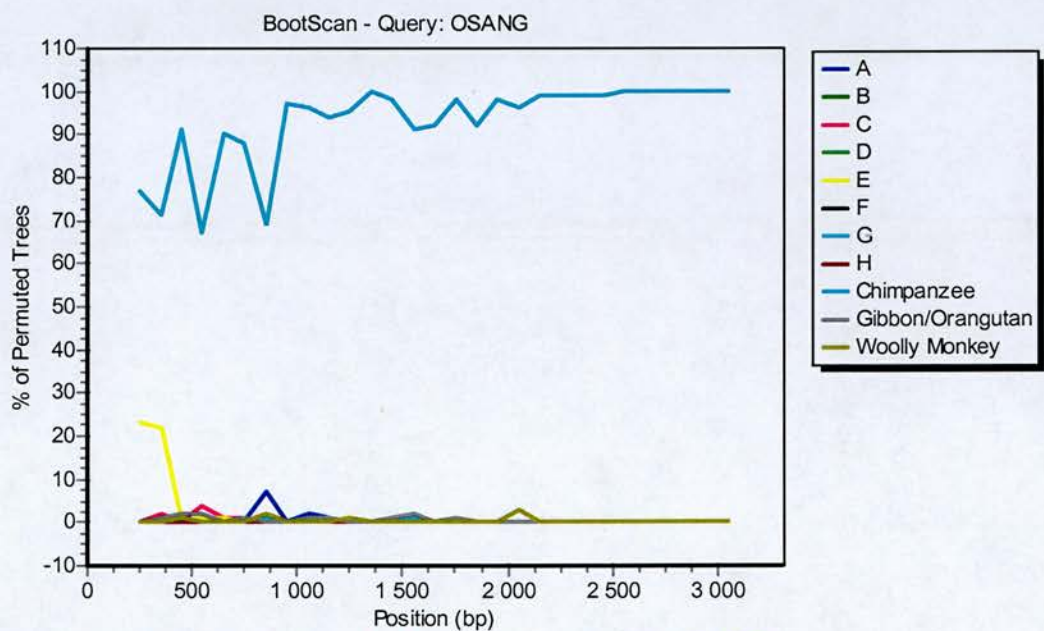
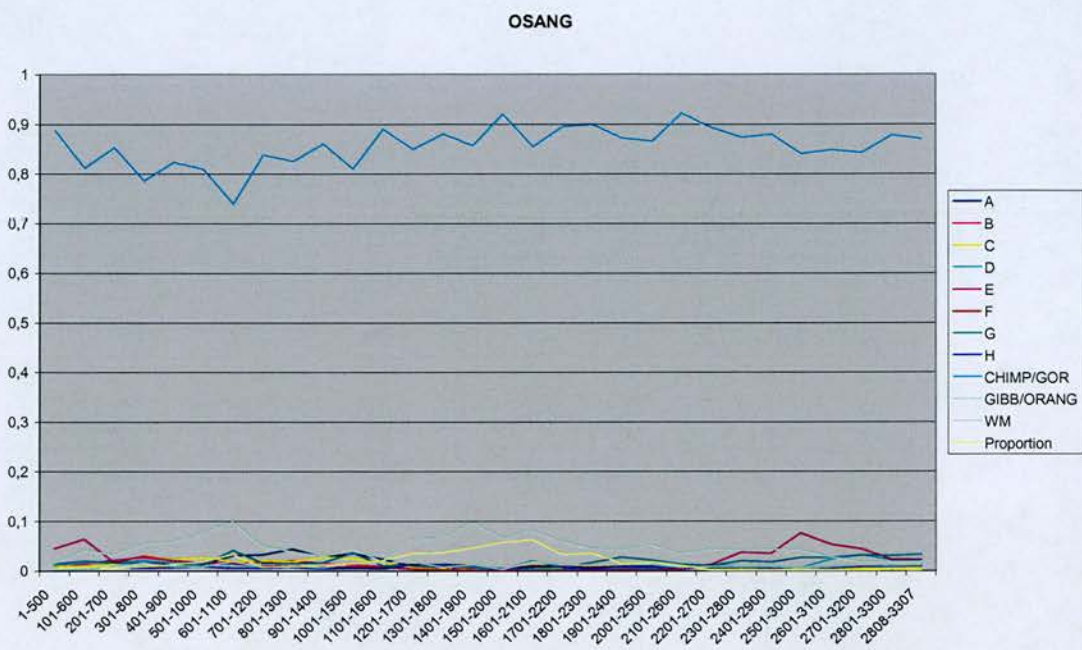
Window : 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

B



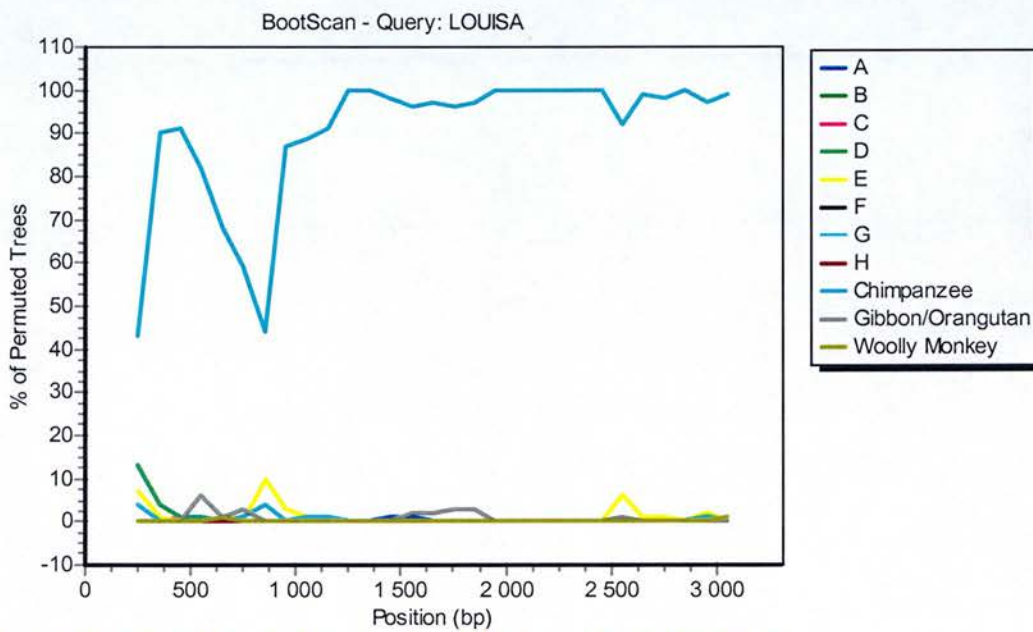
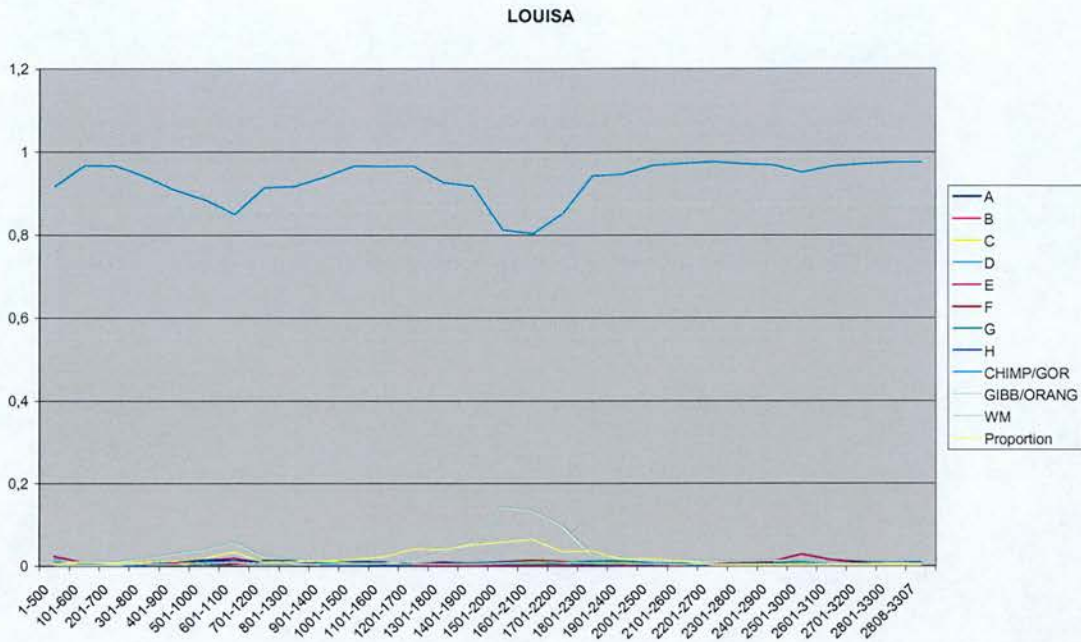
Window: 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

Figure E-3. AI and SimPlot analysis of the complete genome sequence of HBV isolated from chimpanzee Osang.



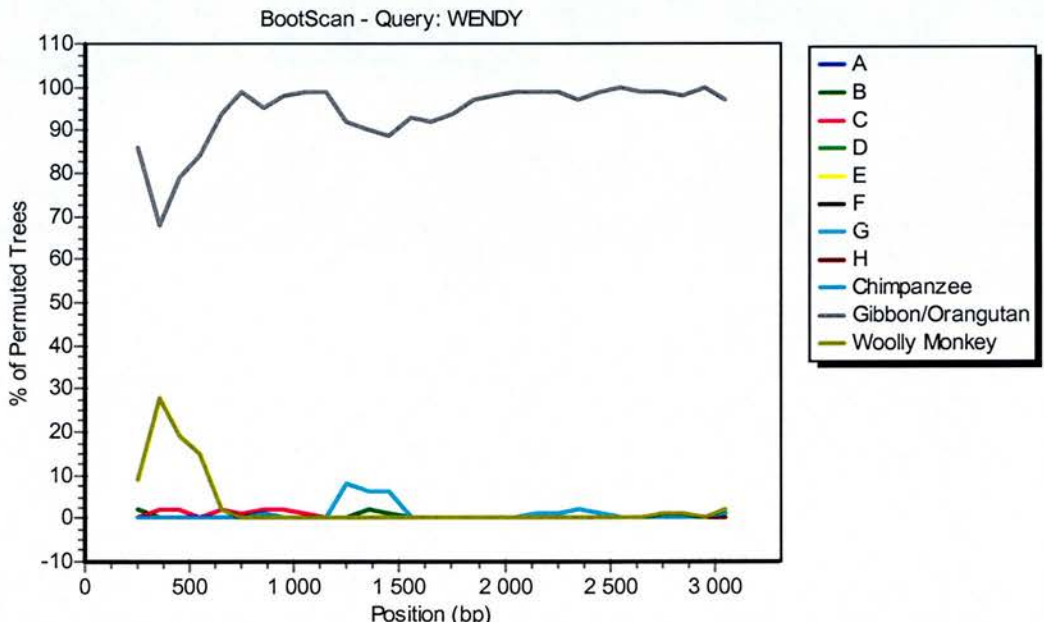
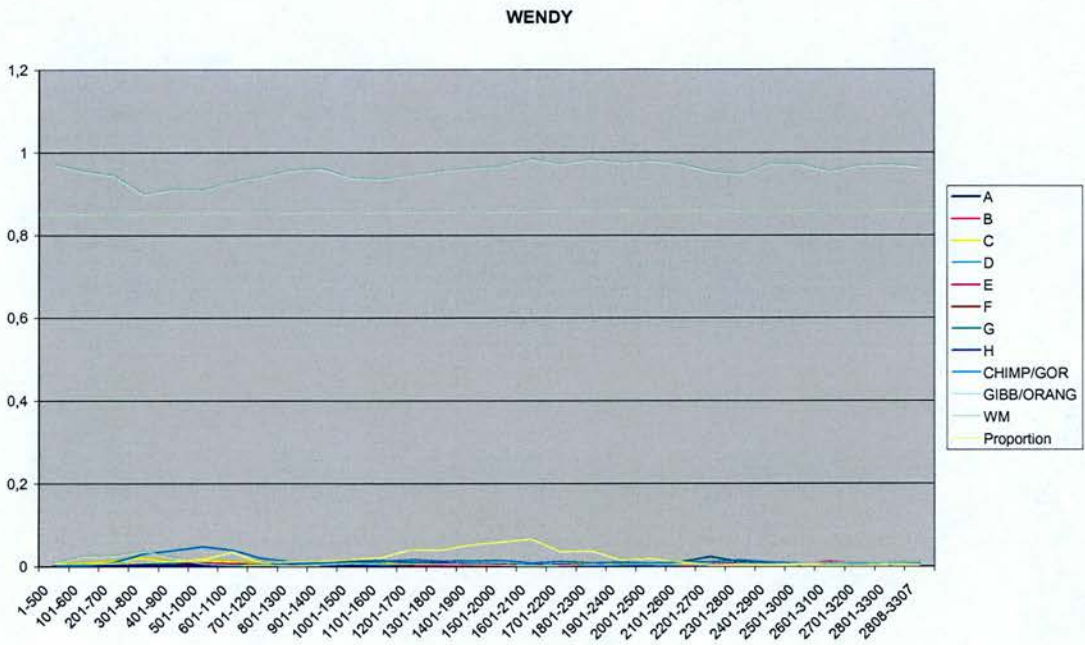
Window: 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

Figure E-4. AI and SimPlot analysis of the complete genome sequence of HBV isolated from chimpanzee Louisa.



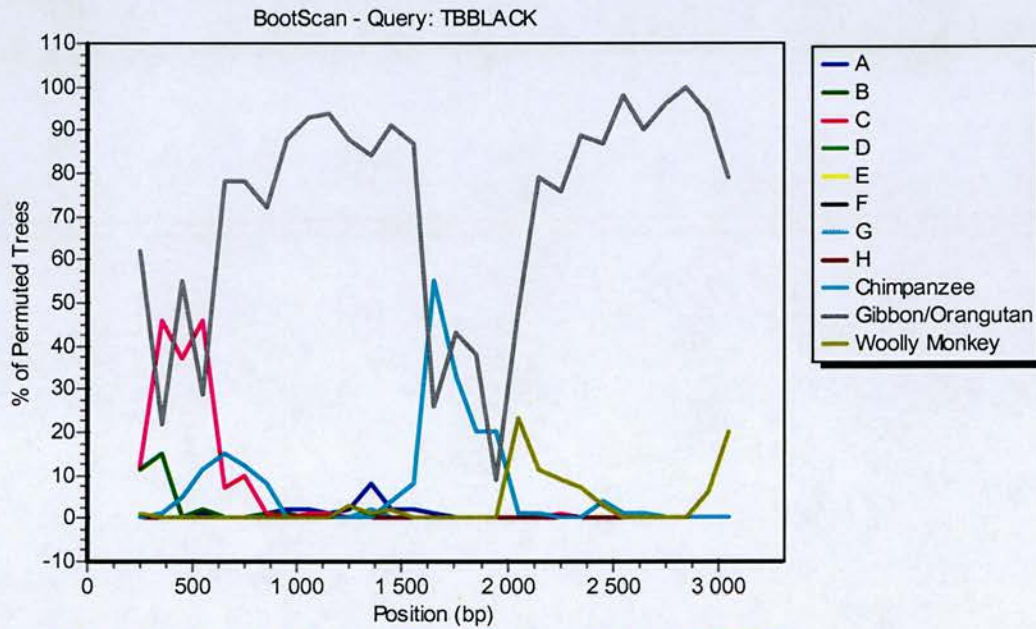
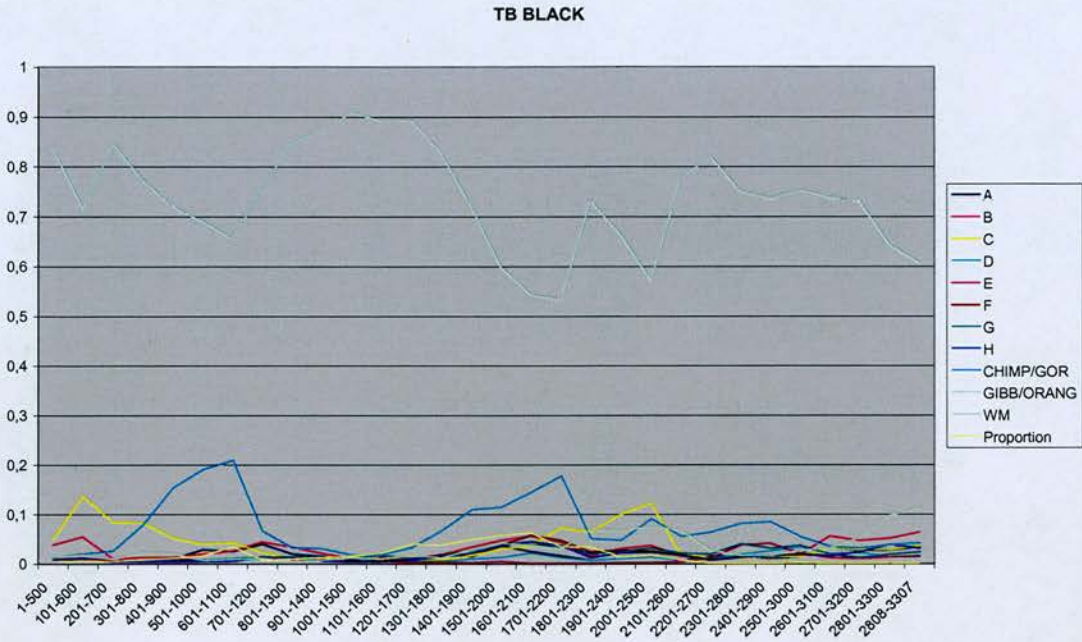
Window : 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

Figure E-5. AI and SimPlot analysis of the complete genome sequence of HBV isolated from gibbon Wendy.



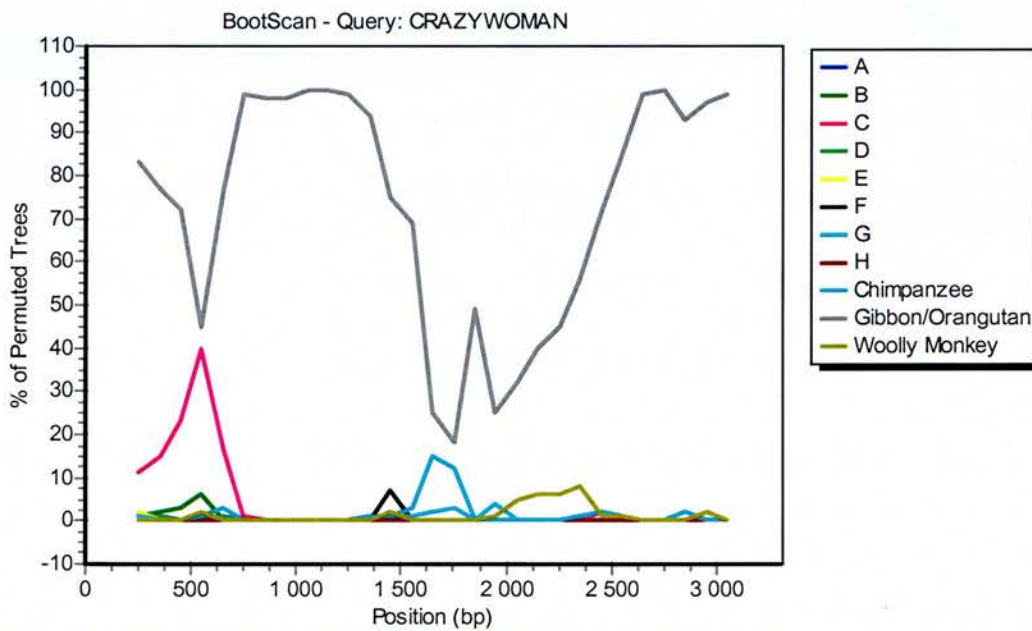
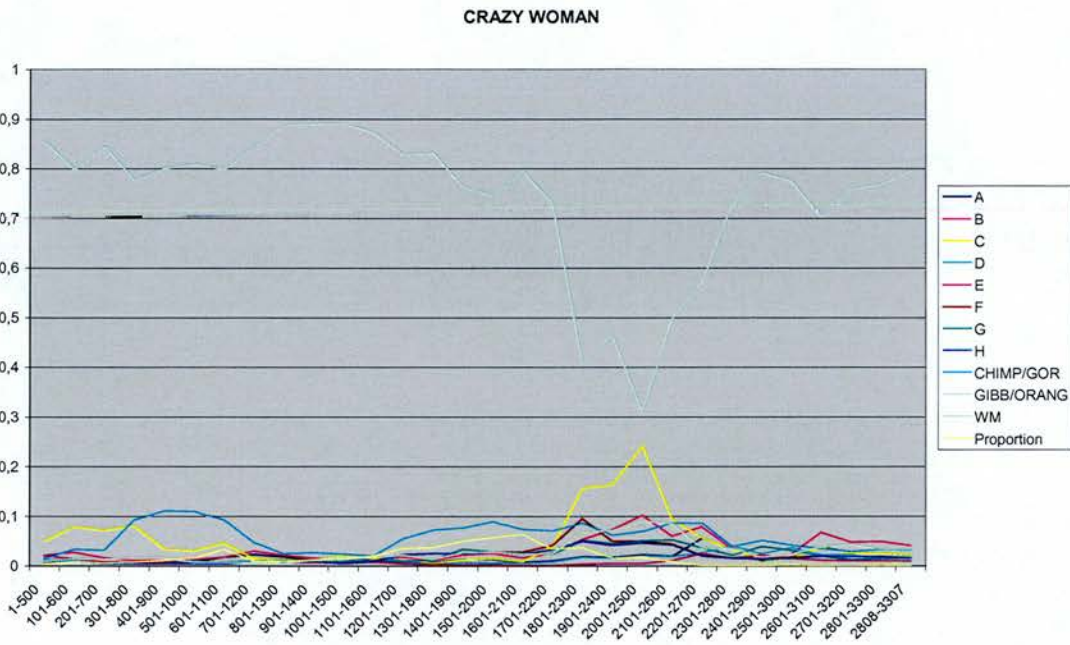
Window : 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

Figure E-6. AI and SimPlot analysis of the complete genome sequence of HBV isolated from gibbon TB Black.



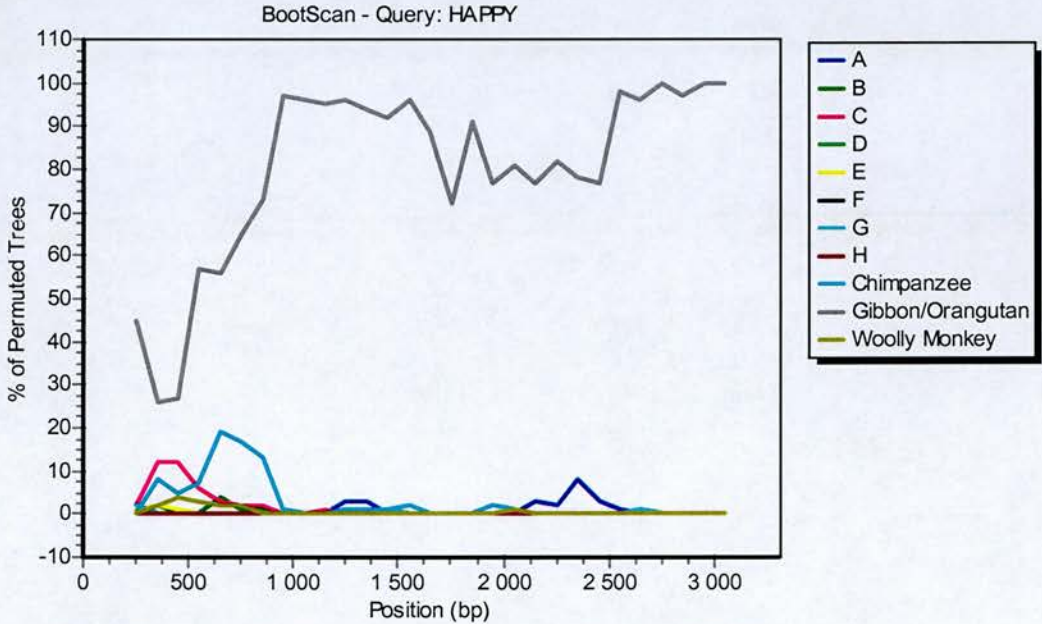
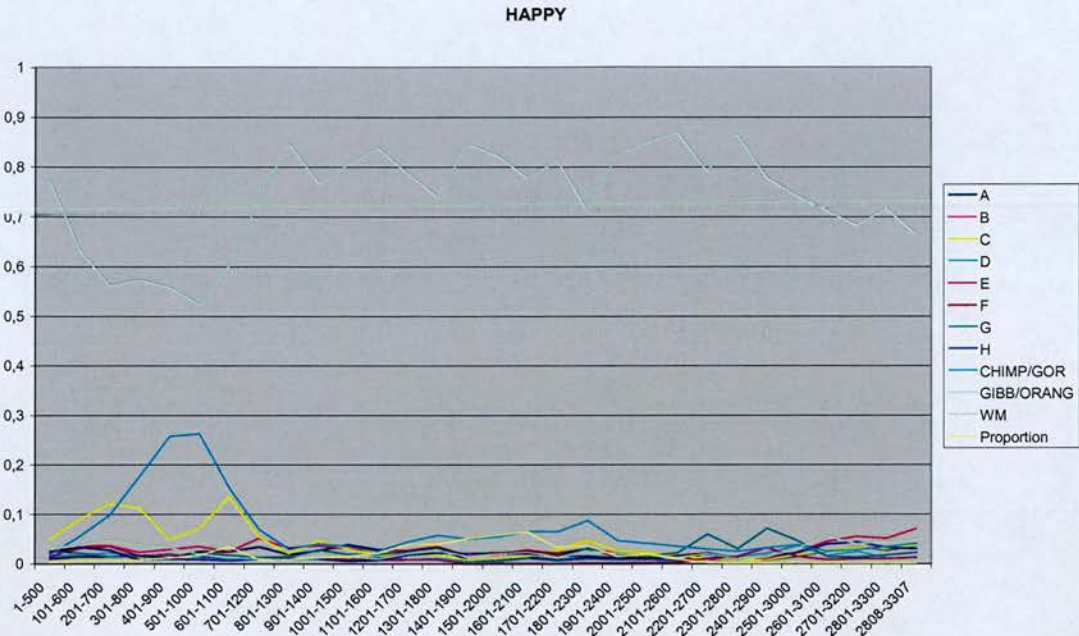
Window: 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

Figure E-7. AI and SimPlot analysis of the complete genome sequence of HBV isolated from gibbon Crazy Woman.



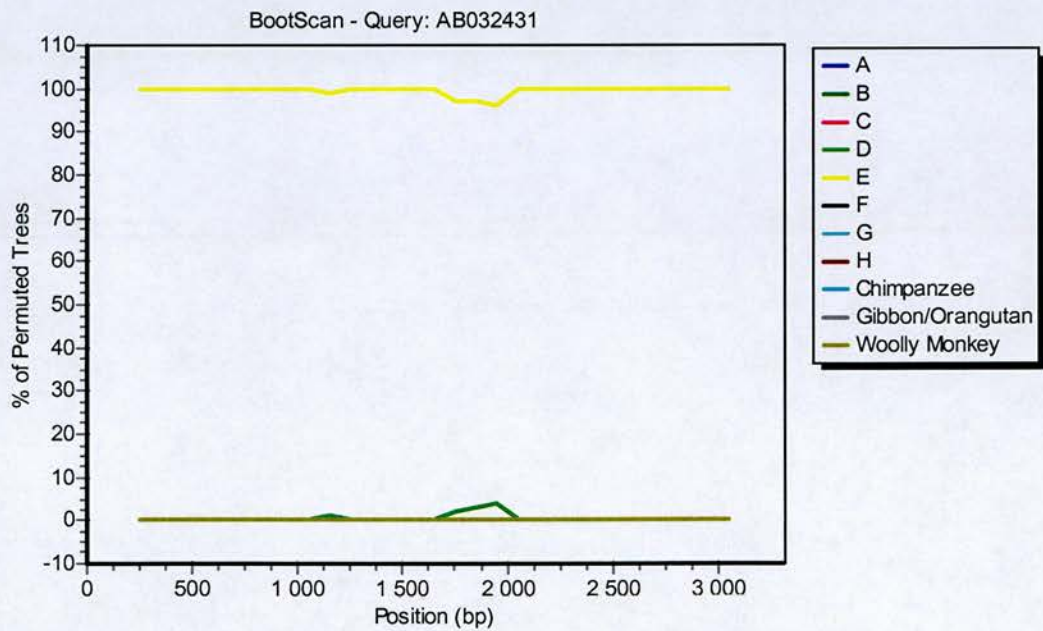
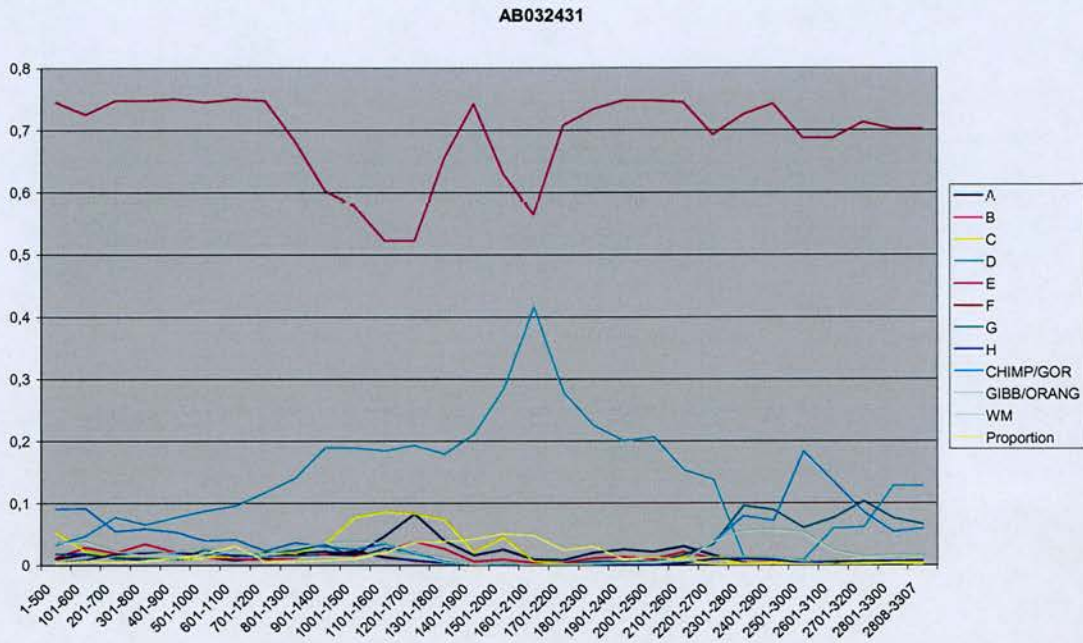
Window: 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

Figure E-8. AI and SimPlot analysis of the complete genome sequence of HBV isolated from gibbon Happy.



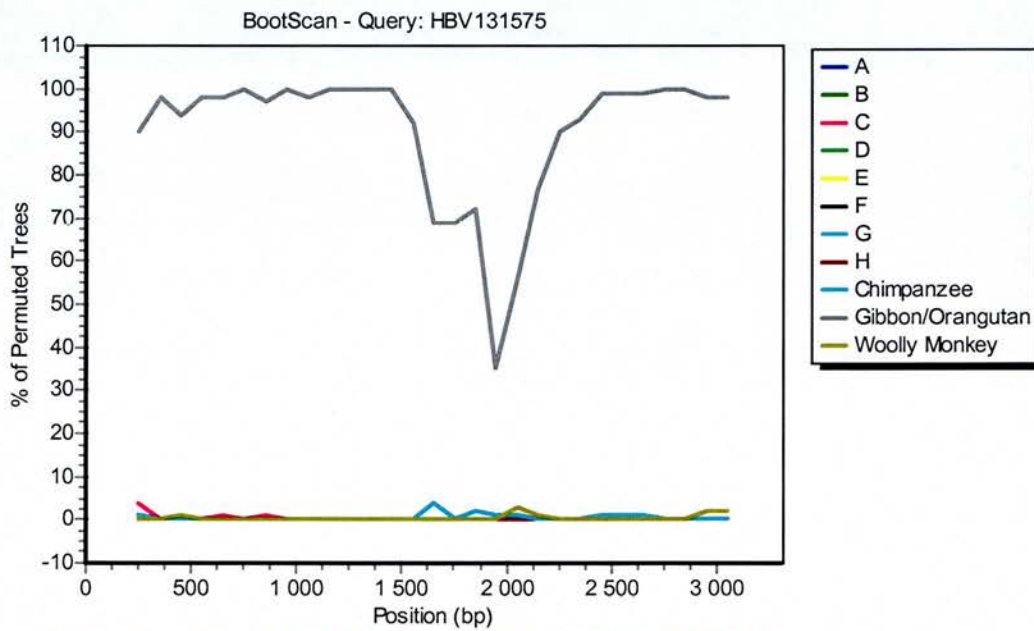
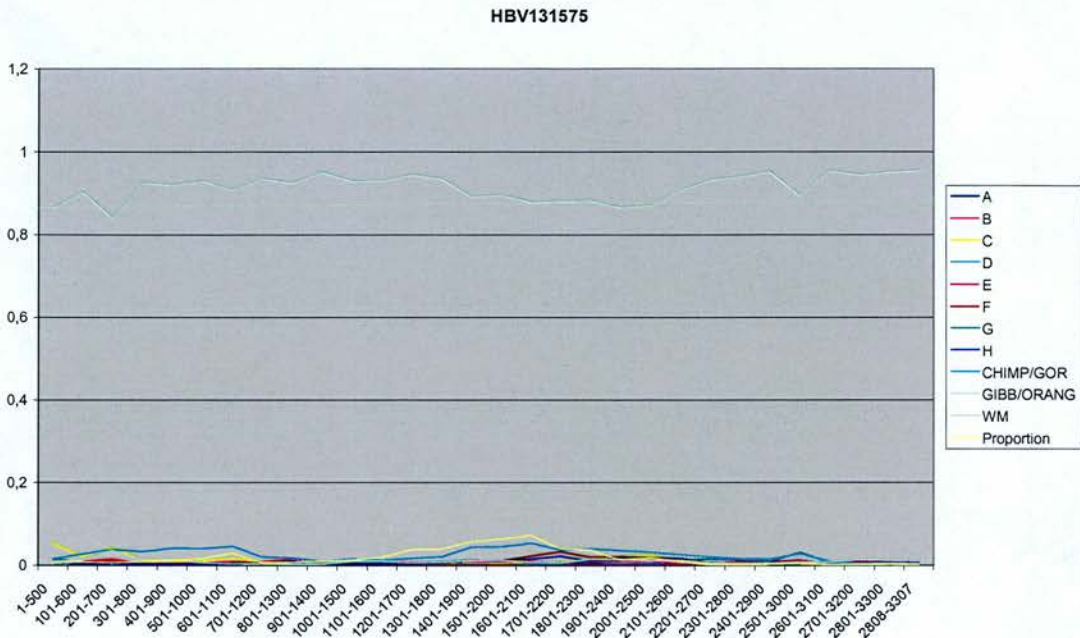
Window : 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

Figure E-9. AI and SimPlot analysis of the complete genome sequence of HBV isolated from a chimpanzee, accession number AB032431.



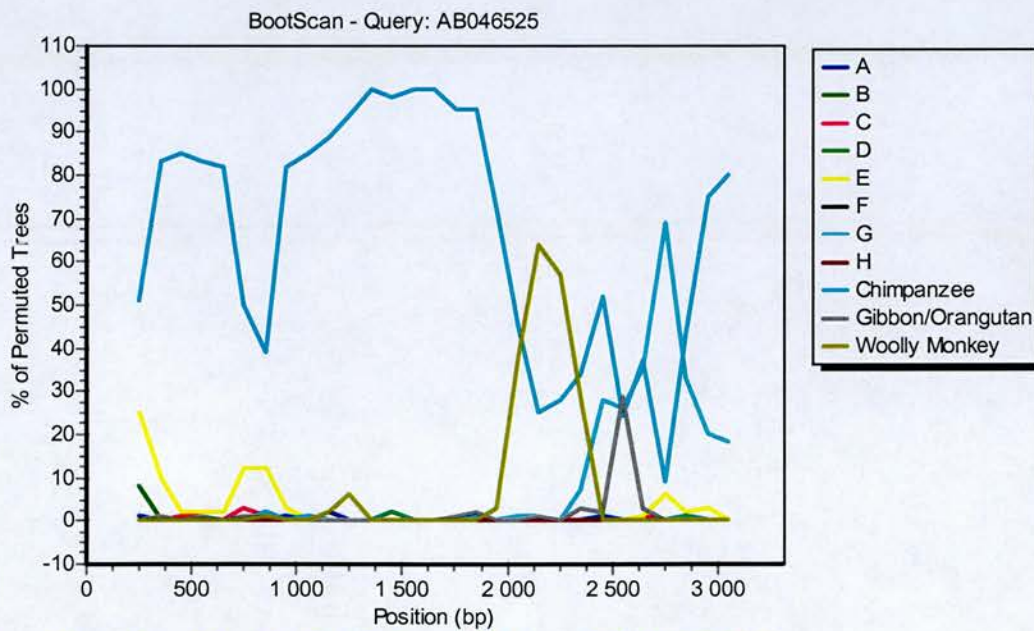
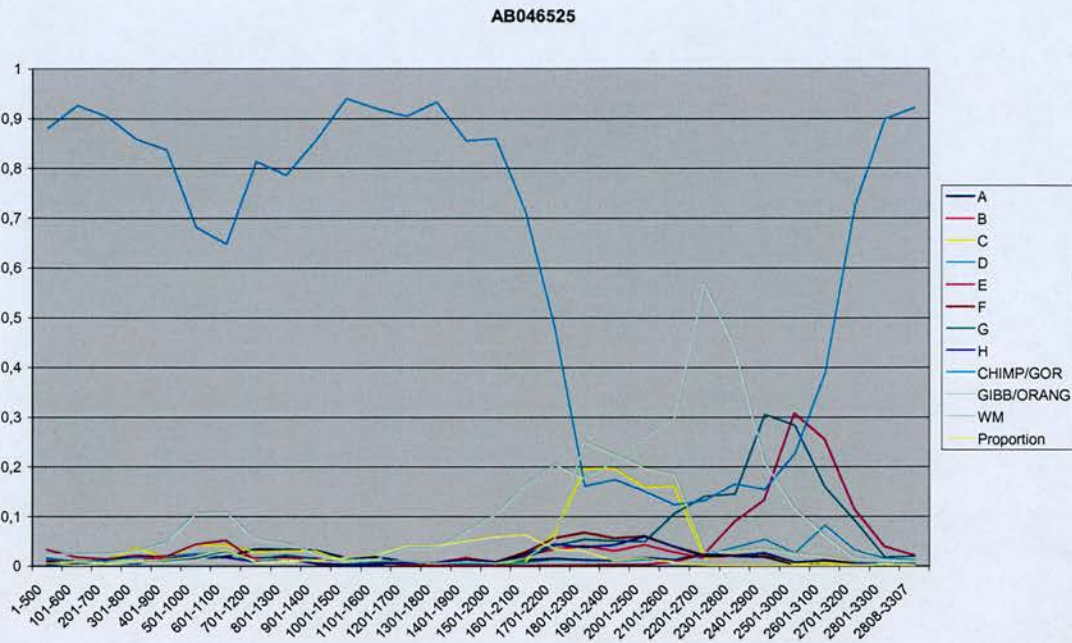
Window: 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

Figure E-10. AI and SimPlot analysis of the complete genome sequence of HBV isolated from a chimpanzee, accession number HBV131575.



/window : 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

Figure E-11. AI and SimPlot analysis of the complete genome sequence of HBV isolated from a chimpanzee, accession number AB046525.



/indow : 500 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura, T/t: 2.0, NEIGHBOR

# **Appendix F**

**(Starkman et al., 2003)**



ACADEMIC  
PRESS

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Virology 314 (2003) 381–393

VIROLOGY

[www.elsevier.com/locate/yviro](http://www.elsevier.com/locate/yviro)

## Geographic and species association of hepatitis B virus genotypes in non-human primates

S.E. Starkman,<sup>a</sup> D.M. MacDonald,<sup>a</sup> J.C.M. Lewis,<sup>b</sup> E.C. Holmes,<sup>c</sup> and P. Simmonds<sup>a,\*</sup>

<sup>a</sup> *Laboratory for Clinical and Molecular Virology, University of Edinburgh, Summerhall, Edinburgh, EH9 1QH, UK*

<sup>b</sup> *International Zoo Veterinary Group, Keighley, Yorkshire, BD21 1AG, UK*

<sup>c</sup> *Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK*

Received 26 March 2003; returned to author for revision 1 May 2003; accepted 15 May 2003

### Abstract

Infection with hepatitis B virus (HBV) has been detected in human populations throughout the world, as well as in a number of ape species (*Pan troglodytes*, *Gorilla gorilla*, gibbons [*Nomascus* and *Hylobates* species] and *Pongo pygmaeus*). To investigate the distribution of naturally occurring HBV infection in these species and other African Old World monkey species (*Cercopithecidae*), we screened 137 plasma samples from mainly wild caught animals by polymerase chain reaction (PCR) using several of highly conserved primers from the HB surface (HBs) gene, and for HBs antigen (HBsAg) by ELISA. None of the 93 *Cercopithecidae* screened (6 species) showed PCR or serology evidence for HBV infection; in contrast 2 from 8 chimpanzees and 5 from 22 gibbons were PCR-positive with each set of primers.

Complete genome sequences from each of the positive apes were obtained and compared with all previously published complete and surface gene sequences. This extended phylogenetic analysis indicated that HBV variants from orangutans were interspersed by with HBV variants from southerly distributed gibbon species (*H. agilis* and *H. moloch*) occupying overlapping or adjacent habitat ranges with orangutans; in contrast, HBV variants from gibbon species in mainland Asia were phylogenetically distinct. A geographical rather than (sub)species association of HBV would account for the distribution of HBV variants in different subspecies of chimpanzees in Africa, and explain the inlier position of the previously described lowland gorilla sequence in the chimpanzee clade. These new findings have a number of implication for understanding the origins and epidemiology of HBV infection in non-human primates.

© 2003 Elsevier Inc. All rights reserved.

### Introduction

Infection with hepatitis B virus (HBV) is a major global health problem, and is estimated to account for approximately one million deaths from chronic liver disease and hepatocellular carcinoma each year (Thomas and Jacyna, 1993). High frequencies of active infection, ranging from 8–15% are found in South and East Asia, sub-Saharan Africa and amongst indigenous peoples in Central and South America (Andre, 2000). In Asia this endemic pattern of HBV infection is primarily maintained through mother-to-child perinatal transmission and establishment of a highly infectious carrier state that transmits infection to the next

generation. In contrast, horizontal transmission is the predominant mechanism in Africa.

Although HBV (classified as a member of the *Hepadnaviridae*) contains a DNA genome, replication occurs through an RNA intermediate sequence analogous to the genomic RNA sequence of retroviruses. The copying of DNA from sequences from genomic DNA templates, and from the intermediate RNA transcript, is carried out by a virally encoded polymerase protein. The lack of proof-reading during viral transcription introduces a high frequency of mutations into the copied sequences (Hannoun et al., 2000). Indeed, HBV populations are characterized by a moderate degree of genetic diversity, with a total of 8 currently classified genotypes infecting human populations worldwide, differing from each other by nucleotide sequence distances of approximately 10–13%. Genotypes A, D and possibly G have global distributions,

\* Corresponding author. Fax: +44-131-650-6511.

E-mail address: [Peter.Simmonds@ed.ac.uk](mailto:Peter.Simmonds@ed.ac.uk) (P. Simmonds).

genotypes B and C are found predominantly in East and South East Asia, genotype E in West Africa, and genotypes F and H amongst various population groups, including indigenous peoples, in Central and South America (Norder et al., 1994; Arauz-Ruiz et al., 1997; Arauz-Ruiz et al., 2002).

Recently, we (MacDonald et al., 2000) and others (Takahashi et al., 2000; Hu et al., 2000; Vartanian et al., 2002; Hu et al., 2001) have documented the existence of HBV infection in chimpanzees in the wild, findings which add to other descriptions of frequent infection of gibbons and orangutans in South East Asia (Warren et al., 1999; Grethe et al., 2000; Verschoor et al., 2001; Noppornpanth et al., 2003). Nucleotide sequencing of HBV recovered from these ape species revealed the existence of new genotypes of HBV generally specific to each species, although with some exceptions, such as the detection of a human genotype E variant in a captive chimpanzee (Takahashi et al., 2000), the close relatedness of HBV recovered from a lowland gorilla to chimpanzee sequences (Grethe et al., 2000), and the detection of a gibbon-like HBV sequence in a captive chimpanzee (Grethe et al., 2000).

Apart from humans and non-human primates, far more divergent HBV-like viruses have been detected in New World rodents such as the woodchuck (*Marmota monax*), and squirrel species (*Spermophilus beecheyi*, *S. parryii*), and a range of bird species (ducks, geese, and grey heron). The evolutionary history of HBV in these various host species has been the subject of great debate over the past 5 years, principally fueled by the difficulty in reconciling the frequently interspersed genotype distributions of human and non-human primate sequences that seems to fit neither hypothesis for very recent or very ancient origins for HBV (Magnius and Norder, 1995; Norder et al., 1994; MacDonald et al., 2000) [reviewed in (Simmonds, 2001)]. In the current study, we have attempted to discover more about the distribution of HBV infection in apes and Old World monkey species (*Cercopithecidae*) using a large archive of stored serum and plasma samples from a wide range of different African and South East Asian primate species. Combining the new data obtained in this study with recently published sequences from gibbons, orangutans and chimpanzees (Vaudin et al., 1988; MacDonald et al., 2000; Takahashi et al., 2000; Hu et al., 2000; Hu et al., 2001; Warren et al., 1999; Grethe et al., 2000; Verschoor et al., 2001; Noppornpanth et al., 2003; Takahashi et al., 2001; Vartanian et al., 2002), we have been able to carry out comprehensive genetic comparisons of the expanded dataset of primate and human-derived HBV variants that indicates that geographical separation in primates explains genotype distributions of HBV better than species associations.

## Results

### *Detection of HBV infection in non-human primates*

To determine the frequency of active HBV infection in Apes and Old World primate species, we screened available plasma or serum samples collected from a range of species by PCR using previously described primers from the pre-S and S region (S1, S2). All samples had been screened for HBsAg by commercially available ELISAs. The pre-S and S primers were conserved between human genotypes A–G, the more divergent F and H genotypes, and all published HBV sequences recovered from non-human primates including the HBV variant obtained from New World woolly monkey species (Fig. 1A and B).

Results from the two sets of primers were concordant with each other and with the results of HBsAg screening with the exception of 1 HBsAg positive chimpanzee that was DNA negative. The frequency of detection of active HBV infection depended on the host species. Within the apes, we detected 2 HBV-positive samples from the 8 chimpanzees screened, 5 from the 22 gibbons species, and none from 14 orangutans (total prevalence of active infection in apes: 7/44 [15.9%]). In contrast, none of the 93 samples available from a total of 6 Old World Monkey species were positive in either PCR assay, nor confirmed positive for HBsAg. The observed restriction of HBV infection to apes and the frequency of active infection found in this study was consistent with previous surveys of HBV infection in primates (Table 1; see Discussion). Further evidence for a lack of HBV infection in these samples of Old World monkey species was provided by the lack of detectable antibodies to the HBV core protein (anti-HBc), used diagnostically as an indication of past resolved infection in samples negative for HBsAg.

Previous studies have demonstrated that non-human primates are generally infected with species-specific genotypes of HBV (Vaudin et al., 1988; MacDonald et al., 2000; Takahashi et al., 2000; Hu et al., 2000; Hu et al., 2001; Warren et al., 1999; Grethe et al., 2000; Verschoor et al., 2001; Noppornpanth et al., 2003; Takahashi et al., 2001; Vartanian et al., 2002). Therefore, one explanation for the apparent absence of HBV infection in Old World monkey species is that the putative variants of HBV perhaps present in these more evolutionarily distant species may be too divergent to be detectable by conventional HBV primers and hence too different antigenically to be detectable by conventional HBsAg or anti-HBc screening. To investigate this possibility, we developed a new set of primers (S3) from a highly conserved region of the surface gene that matched not only all human and non-human primate HBV variants, but also the sequences of each of the HBV-like viruses recovered from rodents (ground and arctic squirrels, woodchuck; Fig. 1C). The new primers showed equivalent sensitivity for human and primate HBV sequences as the S1 and S2 primers for human and primate genotypes of HBV

A) S1 (Pre-S region)

primer	Primer 41		Primer 42		Primer 43		Primer 19	
	CGTCG	CHGAAGATCC CAATCT	GTATYCCCT	GGACTCATAA GG	GTA	GGAGTCCGGT ACGTCACC	GACRACC	ACCGAGGTCA AGKC
	2469	2523	2531	2561	3287	3317	52	82
H-A	CGCCGCGTGC	CAGAAGATCT CAATCTGGG A	TGTT	AGTATTCCTT GGACTCATAA GGTGGGA	ACAGTCAT	CCTCAGGCCA TGCAGTGGAA TTC	TT	TTCTGCTGCT TGGCTCCAGT TCAGGAACA
H-B	.....	.....	.....	.....	.....	.....	.....	.....
H-C	.....	.....	.....	.....	.....	.....	.....	.....
H-D	.....	.....	.....	.....	.....	.....	.....	.....
H-E	.....	.....	.....	.....	.....	.....	.....	.....
H-F	.....	.....	.....	.....	.....	.....	.....	.....
H-G	.....	.....	.....	.....	.....	.....	.....	.....
H-H	.....	.....	.....	.....	.....	.....	.....	.....
CH	.....	.....	.....	.....	.....	.....	.....	.....
GN	.....	.....	.....	.....	.....	.....	.....	.....
OT	.....	.....	.....	.....	.....	.....	.....	.....
WM	.....	.....	.....	.....	.....	.....	.....	.....
GS	.....	.....	.....	.....	.....	.....	.....	.....
AGS	.....	.....	.....	.....	.....	.....	.....	.....
WC	.....	.....	.....	.....	.....	.....	.....	.....
OHVCGC	.....	.....	.....	.....	.....	.....	.....	.....

B) S2 (S region)

primer	Primer S1		Primer S3		Primer S6		Primer S5	
	CATCA	GGAYTCTAG GACCCCT	CGTGTACAG	GCGGGTCTT TCTTGT	CTACACAGA	GCGCGCAAAA TAGTA	GG	AGAMGTAGGA CGAGATAGC GAG
	163	223	371	432	371	432		
H-A	CATCACATCA	GGATTCCTAG GACCCCTGCT	CGTGTACAG	GCGGGTCTT TCTTGTGAC A	CGCT	GGATGTGCT GCGGGTCTT ATCATATCC	TCTTCATCT	GCTGCTATGC CTCATCTT
H-B	.....	.....	.....	.....	.....	.....	.....	.....
H-C	.....	.....	.....	.....	.....	.....	.....	.....
H-D	.....	.....	.....	.....	.....	.....	.....	.....
H-E	.....	.....	.....	.....	.....	.....	.....	.....
H-F	.....	.....	.....	.....	.....	.....	.....	.....
H-G	.....	.....	.....	.....	.....	.....	.....	.....
H-H	.....	.....	.....	.....	.....	.....	.....	.....
CH	.....	.....	.....	.....	.....	.....	.....	.....
GN	.....	.....	.....	.....	.....	.....	.....	.....
OT	.....	.....	.....	.....	.....	.....	.....	.....
WM	.....	.....	.....	.....	.....	.....	.....	.....
GS	.....	.....	.....	.....	.....	.....	.....	.....
AGS	.....	.....	.....	.....	.....	.....	.....	.....
WC	.....	.....	.....	.....	.....	.....	.....	.....
OHVCGC	.....	.....	.....	.....	.....	.....	.....	.....

C) S3 (Highly conserved region)

primer	Primer 21		Primer 22		Primer 23		Primer 24	
	GACTT	CTCTCARTTT TCYAGGGG	GATGTRI	CTGCGGCTT TTATCAT	GATACCCT	CACCCGGART CRGGCA	GGCAAAGAG	RACCGAGTYA AATGA
	257	289	371	404	639	672	659	692
H-A	TGGTGGACTT	CTCTCARTTT TCYAGGGGGA TCA	CG	CTGGATGTGT CTGCGGCTT TTATCATATT CC	AAT	ACCTATGGGA GTGGGCTCA GTCCGTTCT C	TCAG	TCCGTTTCTC TTGGCTCAGT TTACTAGTGC
H-B	.....	.....	.....	.....	.....	.....	.....	.....
H-C	.....	.....	.....	.....	.....	.....	.....	.....
H-D	.....	.....	.....	.....	.....	.....	.....	.....
H-E	.....	.....	.....	.....	.....	.....	.....	.....
H-F	.....	.....	.....	.....	.....	.....	.....	.....
H-G	.....	.....	.....	.....	.....	.....	.....	.....
H-H	.....	.....	.....	.....	.....	.....	.....	.....
CH	.....	.....	.....	.....	.....	.....	.....	.....
GN	.....	.....	.....	.....	.....	.....	.....	.....
OT	.....	.....	.....	.....	.....	.....	.....	.....
WM	.....	.....	.....	.....	.....	.....	.....	.....
GS	.....	.....	.....	.....	.....	.....	.....	.....
AGS	.....	.....	.....	.....	.....	.....	.....	.....
WC	.....	.....	.....	.....	.....	.....	.....	.....
OHVCGC	.....	.....	.....	.....	.....	.....	.....	.....

Fig. 1. Sequence conservation of HBV and HBV-like virus sequences from primates and rodents in regions of the genome targeted by S1, S2 and S3 primer sets used for PCR. Primer sequences were compared with consensus sequences from the human HBV genotypes A–H (H–A to H–H), consensus sequences from chimpanzee, gibbon, orangutan (CH, GN, OT), and the woolly monkey sequence (WM). The lower panel for each alignment contains HBV-like virus sequences from North American rodent species, ground squirrel (GS), arctic ground squirrel (AGS), woodchuck (WC, OHVCGC). For each alignment, differences from the consensus genotype A sequence was shown; primer sequences included degenerate bases (represented by the codes: M: A/C; Y: C/T; R: A/G; K: G/T) in variable regions. Alignment numberings refer to a minimal alignment of human and primate derived HBV sequences listed in the Methods section.

(data not shown). However, screening of this sample of Old World primate species failed to detect any additional cases of HBV infection (Table 1).

Species and sub-species identification of HBV-infected apes

Of the two chimpanzees positive for HBV DNA, one (Osang) originated from Central Africa and its subspecies identification as *P. troglodytes vellerosus* was determined by mitochondrial sequencing. The other infected chimpanzee (Louisa) originated from Cameroon and belonged to the troglodytes subspecies. The infected gibbons belonged to species *H. lar* (Tamang), *H. agilis* (Crazy Woman), *N. concolor* (Happy) and *N. gabriellae* (Wendy). This species

information was provided by the Welsh Mountain Zoo and the rescue centre respectively. It was not possible to obtain a species assignment for the final HBV-infected gibbon analyzed in the study (TB Black), although it was possible to identify it as belonging to the *Hylobates* genus and most likely belonging to the *agilis* or *moloch* subspecies as Mitochondrial 12S region sequences for these species were interspersed when compared phylogenetically (data not shown).

Sequence comparison of HBV sequences from non-human primates

Sequence relationships between the 7 (from 44) ape samples positive for HBV DNA by PCR with previously

Table 1  
Frequencies of detection of HBV DNA sequences, HBsAg and anti-HBc in current and previous studies of primates

Species	PCR*		Total		Active <sup>†</sup> infection (%)	anti-HBc	Reference
	S1/S2	S3	PCR <sup>†</sup>	HBsAg			
<b>APES</b>							
<i>Gorilla gorilla</i> spp.	—	—	7/18	2/15	7/18 (39%)	—	(Grethe et al., 2000; Zuckerman et al., 1978; Linnermann, Jr. et al., 1984; Thornton et al., 2001; Worley and Stalis, 2002)
<i>Pan troglodytes</i> spp.	2/8	—	28/38	38/496	54/517 (10.5%)	2/6	This study, (Takahashi et al., 2000; MacDonald et al., 2000; Hu et al., 2000; Takahashi et al., 2001; Hu et al., 2001; Grethe et al., 2000; Zuckerman et al., 1978) (Vaudin et al., 1988) (Deinhardt, 1976; Ogata et al., 1993; Eichberg and Kalter, 1980)
<i>Hylobates</i> spp.	5/22	—	53/177	35/191	55/213 (25.8%)	48/130	This study, (Grethe et al., 2000; Norder et al., 1996; Lanford et al., 2000; Vaudin et al., 1988; Deinhardt, 1976; Thornton et al., 2001; Noppornpanth et al., 2003)
<i>Pongo pygmaeus</i> spp.	0/14	—	32/104	58/141	58/297 (19.5%)	1/14	This study, (Warren et al., 1999; Davis et al., 2000; Vaudin et al., 1988; Deinhardt, 1976)
<b>Total Apes</b>	<b>7/45</b>	<b>—</b>	<b>120/337</b>	<b>133/843</b>	<b>174/1045 (16.7%)</b>	<b>51/150</b>	
<b>OLD WORLD MONKEYS</b>							
<i>Papio</i> spp.	—	—	0/4	0/168	0	0/103	(Deinhardt, 1976; Michaels et al., 1996; Eichberg and Kalter, 1980)
<i>Cercopithecus aetiops</i>	—	—	—	0/19	0	—	(Eichberg and Kalter, 1980)
<i>Cercopithecus torquatus</i>	0/4	0/4	0/4	0/4	0	0/4	This study
<i>Cebus albifrons</i>	—	—	—	0/10	0	—	(Eichberg and Kalter, 1980)
<i>C. mona</i>	0/3	0/3	0/3	0/3	0	0/3	This study
<i>C. nictitans</i>	0/2	0/2	0/2	0/2	0	0/2	This study
<i>C. erythrotis</i>	0/3	0/3	0/3	0/3	0	0/3	This study
<i>Mandrillus leucophaeus</i>	0/78	0/78	0/78	0/78	0	0/78	This study
<i>M. sphinx</i>	0/2	0/2	0/2	0/2	0	0/2	This study
<b>Total OWM</b>	<b>0/102</b>	<b>0/102</b>	<b>0/96</b>	<b>0/289</b>	<b>0</b>	<b>0/195</b>	
<b>NEW WORLD MONKEYS</b>							
<i>Lagothrix lagotricha</i>	—	—	10/35	7/31	10/33	—	(Lanford et al., 1998)
<i>Saimiri sciureus</i>	—	—	—	0/20	0	—	(Eichberg and Kalter, 1980)
<i>Callithrix jacchus</i>	—	—	—	0/6	0	—	(Eichberg and Kalter, 1980)
<i>Saguinus oedipus</i>	—	—	—	0/12	0	—	(Eichberg and Kalter, 1980)
<b>Total NWM</b>	<b>—</b>	<b>—</b>	<b>10/58</b>	<b>7/69</b>	<b>10/33 (30%)</b>	<b>—</b>	

\* PCR results from pre-S and S gene primers (S1, S2), and from highly conserved primers (S3; Fig. 1C) from this study.

<sup>†</sup> Combined PCR results from current survey and from previous studies (cited in last column).

<sup>‡</sup> Combined results for PCR-positivity and detection of HBsAg indicating active HBV infection.

described variants of HBV from human and non-human primates were investigated by determination of their complete genome sequences and phylogenetic analysis. HBV variants included in the analysis comprised all previously published full length HBV sequences from non-human primates and 10 representative sequences from each of the human genotypes A–H. The maximum likelihood tree of these data provided bootstrap support for each human genotype and primate-associated species, with the exception of gibbon and orangutan sequences (Fig. 2). However, other than the close relationships between genotypes D and E, and F and H, there was little support for any other inter-clade relationship. A separate comparison of S gene sequences was also carried out to enable several further partial genomic sequences from primates to be included in the analysis (Fig. 3; see below).

### Chimpanzee sequences

Both complete genome and surface gene sequences from the two HBV-infected chimpanzees (Louisa, Osang) grouped consistently with other previously published sequences (Fig. 2 and 3A). This together with the handling and testing history is consistent with infection in the wild. Within the divergent clade of chimpanzee sequences, individual sequences fell into a variety of groups, with a strong association with subspecies and/or geographical origin (Fig. 4A). For example, clusters were observed corresponding to verus subspecies from West Africa (whole genome sequences AB032433, HPBVCG, AB032432, AB032433, AF222323 and Chim 4), troglodytes from Central Africa (AF222322, Chim 2 and Louisa), vellerous from Central-West Africa (Osang, AF305327), and a separate clade for

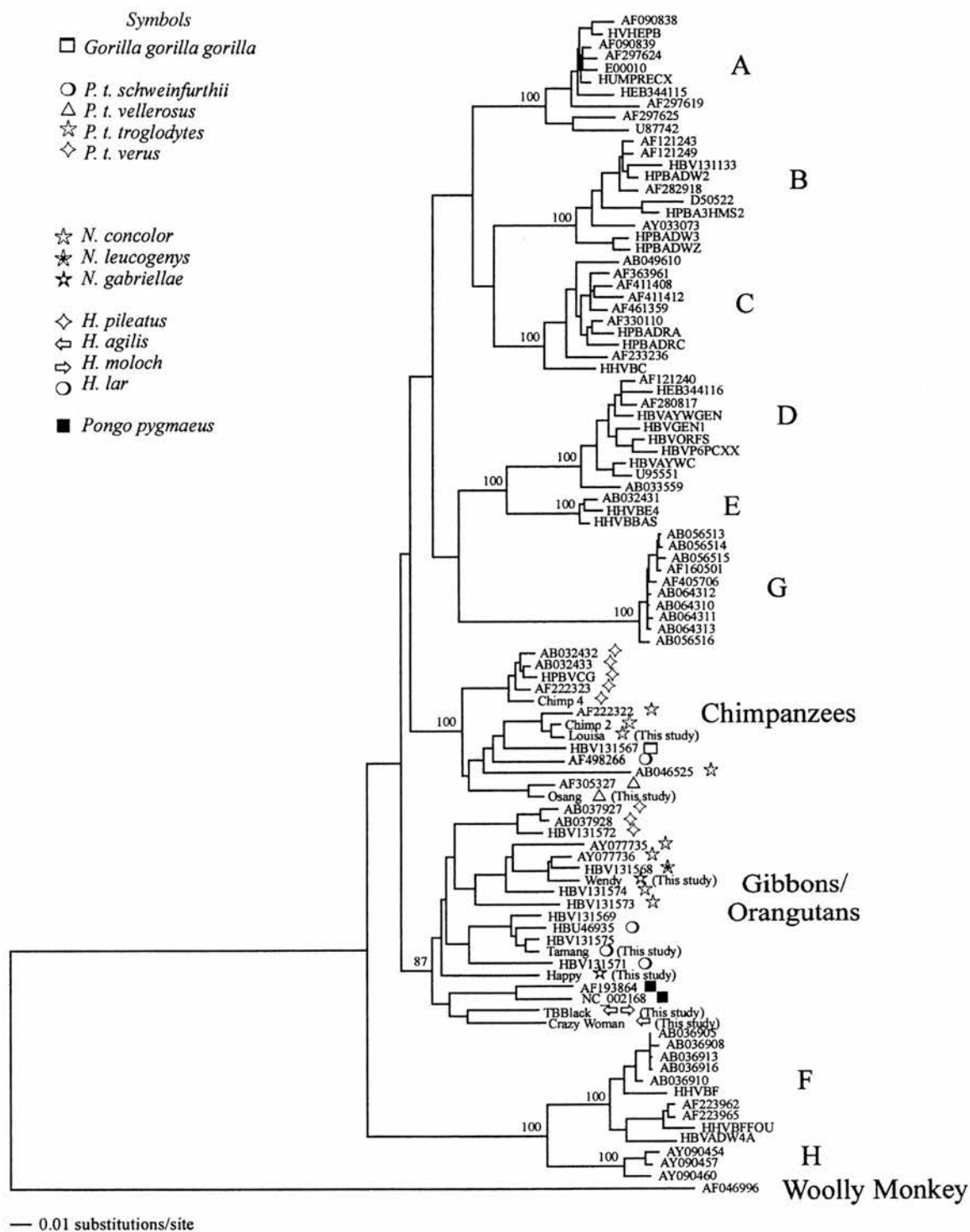


Fig. 2. Maximum likelihood phylogenetic tree for 99 complete HBV genomes (3342 bp) representing human genotypes A to H and all ape HBV sequences with the single woolly monkey HBV sequence as an outgroup. Neighbor-Joining bootstrap values (>75%) for major groupings are indicated on the relevant nodes. All horizontal branch lengths are drawn to scale.

the single Schweinfurthii (East African) HBV variant (AF498266). The only sequence that failed to group within its sub-species group was AB046525 obtained from a central African troglodytes species; this may be because it contains an unusual core gene sequence thought to have

originated through recombination with a highly divergent and hitherto undiscovered entirely separate genotype of HBV (Takahashi et al., 2001) (see below). Similar grouping of sequences into subspecies- or geographically-associated clades was found on analysis of partial S gene sequences.

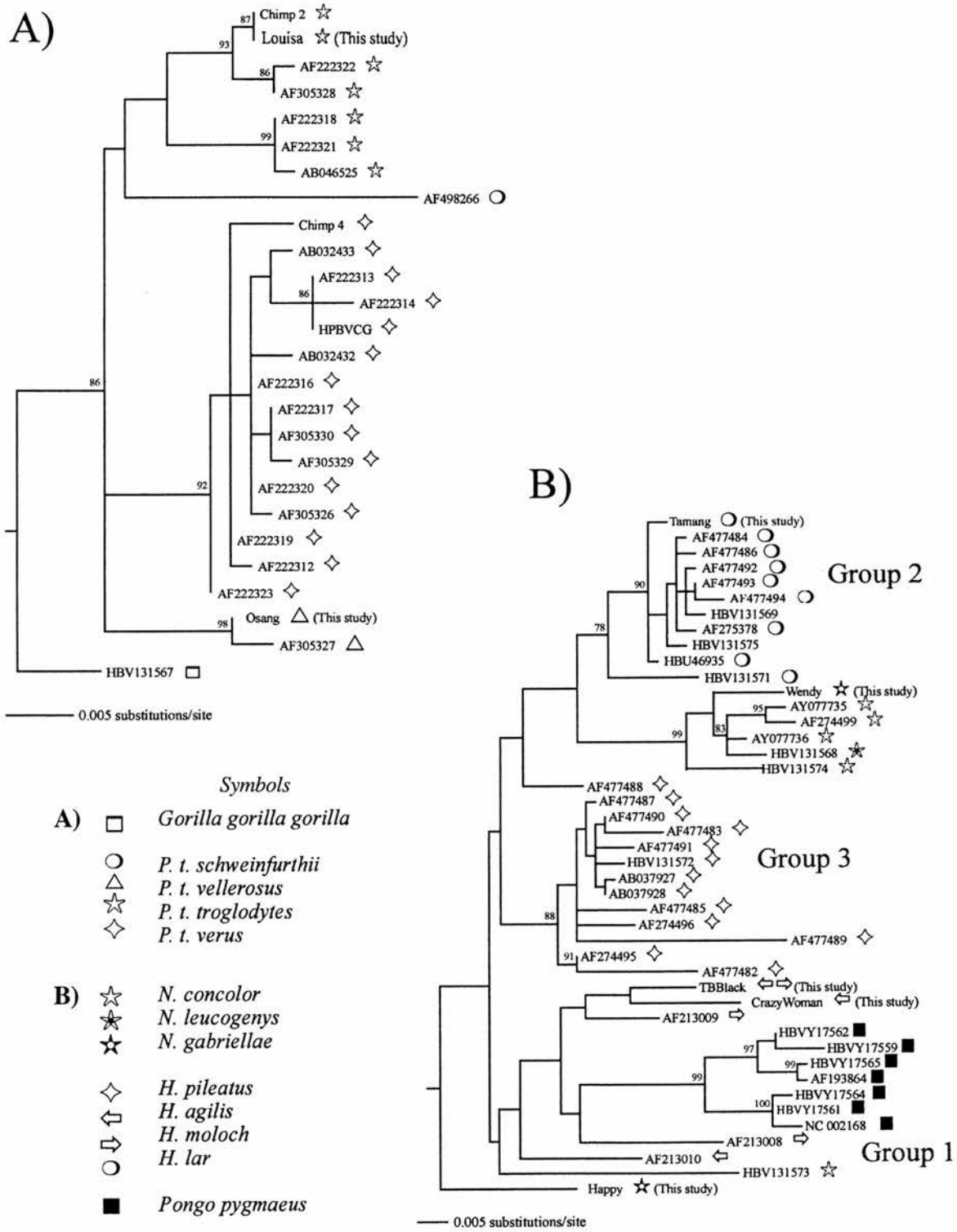
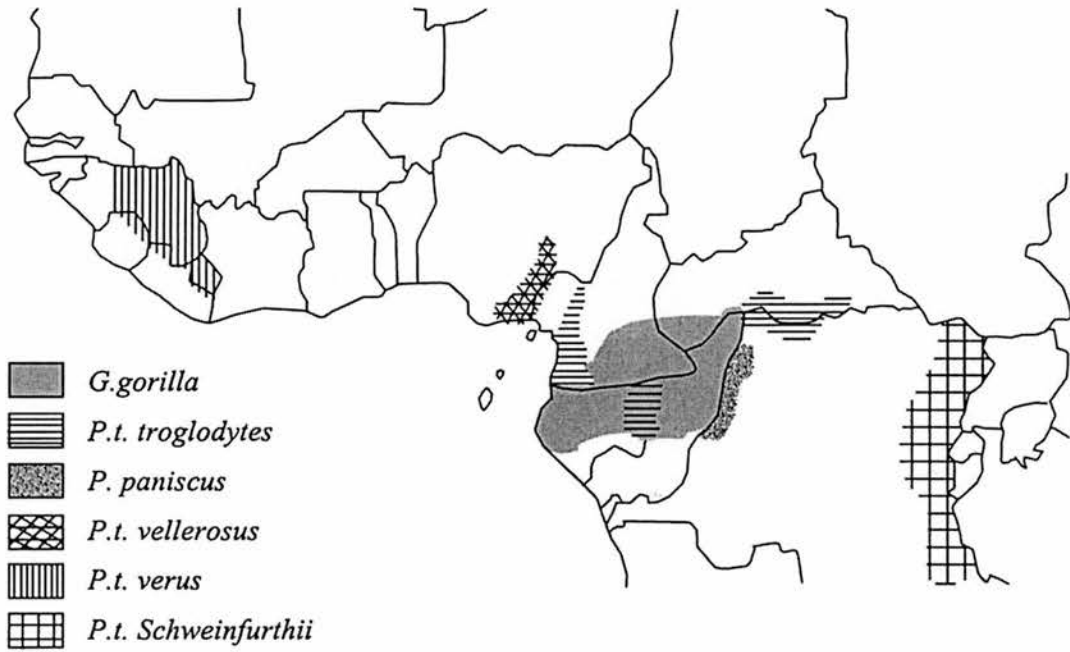


Fig. 3. Maximum likelihood phylogenetic trees of partial S gene sequences (689 bp) from (A) chimpanzees and (B) gibbons and orangutans. The chimpanzee tree is rooted using a single gibbon sequence (*Tamang*) and the gibbon/orangutan tree is rooted using a chimpanzee sequence (*Osang*). Neighbor-Joining bootstrap values (>75%) for major groupings are indicated on the relevant nodes. All horizontal branch lengths are drawn to scale.

Finally, the single available sequence from the gorilla (HBV131567) fell into its own unique clade within the chimpanzee genotype, an unexpected position if it is assumed that HBV variants are species (and sub-species)-specific, but not

inconsistent with the alternative hypothesis for a geographical basis for the observed sequence groups. As indicated (Fig. 4A), western lowland gorillas share their geographical range with the troglodytes subspecies of chimpanzee in Central Africa.

A)



B)

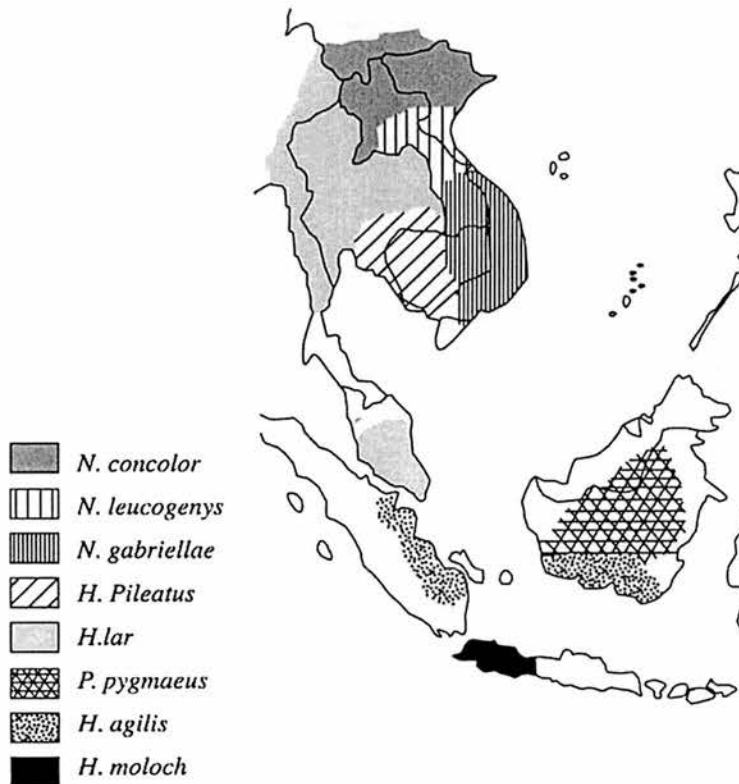


Fig. 4. Distribution of (A) chimpanzee subspecies and lowland gorillas in Africa, and (B) species of gibbons and orangutans in South East Asia.

Interestingly, the gorilla-derived sequence groups most closely with HBV sequences from this subspecies on comparison of complete genome sequences (Fig. 2), although this similarity does not lead to a bootstrap supported *G. gorilla/P.t.troglodytes* clade, nor is this similarity apparent upon sequence comparison of the surface gene fragment (Fig. 3A).

#### Gibbon sequences

Complete genome sequences from 5 gibbons were obtained and compared with the 12 previously published HBV sequences from gibbon species and with those from two orangutans, and the single chimpanzee sequence (Fig. 2). As with the chimpanzee sequences, there was evidence for marked clustering of HBV sequences according to their host species, although the total sequence divergence within the clade was greater than found amongst the chimpanzee subspecies, and much greater than within human genotypes. There was also greater lineage differentiation within the gibbon sequence group, including well defined sequence clusters of sequences from *H. lar*, nomascus species (*N. concolor*, *leucogenys* and *gabriellae*) and from *H. pileatus* and a series of divergent single sequences, including those from Happy, TBBlack and CrazyWoman obtained in the current study. Lack of strict species-specificity of HBV variants was demonstrated in the S gene fragment by the highly divergent sequences obtained from *H. moloch* and *agilis* (CrazyWoman, AF213010, AF213008 and AF213009), the relation of an HBV variant from *N. concolor* (HBV131573) with these sequences and its failure to group with the main clade of concolor sequences (Fig. 3). Other anomalies include the outlier positions of the *H. pileatus* variant (AF477488) and the highly divergent sequence we obtained from Happy (*N. gabriellae*).

If a broader division of gibbon species into the hylobates (*moloch*, *agilis*, *pileatus* and *lar* species) and nomascus (*concolor*, *leucogenys* and *gabriellae* species) genera is made, there is still no association between host and HBV phylogeny. However, the greatest discordance between HBV grouping and species of origin is demonstrated by the phylogenetic position of the HBV sequences obtained from orangutans. The two complete genome sequences (NC\_002168 and AF193864), and the further 5 partial S gene sequences (HBVY17559–HBVY17565) grouped closely together, but they consistently fell within the gibbon clade, in an analogous manner to the position of the lowland gorilla sequence within chimpanzee sequences (see above). A better correlation was observed between geographical origin and/or host range of the HBV-infected species with the phylogenetic clustering of HBV variants recovered from them. The gibbon and orangutan sequences can be divided into three main groups, the first of which contains orangutan HBV variants and those infecting two of the hylobates species, which inhabit the Southern region of SE Asia, and overlap in range with orangutans. The second group contains sequences from northerly distributed nomascus species

of gibbons and *H. lar*, distributed predominantly in Central Thailand and Laos. HBV variants infecting *H. pileatus*, a gibbon species inhabiting Cambodia and Central Thailand made up the third clade of sequences.

The main inconsistency with the phylogeny of gibbon-derived HBV sequences is the anomalous position of the chimpanzee-derived HBV variant, HBV131575. This sequence groups with a number of HBV variants infecting *H. lar*, including HBV131571 from the same study (Grethe et al., 2000). This is the only case where an HBV variant infecting a non-Asian primate groups with gibbon and orangutan sequences, and further information is required on whether a cross-species transmission of HBV could have occurred in captivity.

## Discussion

### *Host range and epidemiology of HBV in primates*

Our PCR-based and serological survey of a wide range of Old World primate species provided further evidence for a substantial restriction of HBV infection to apes (Table 1), and is consistent with previous observations for the absence of detectable HBV infection in Old World monkey species (Deinhardt, 1976; Michaels et al., 1996; Eichberg and Kalter, 1980). Combining data from this and previous surveys, HBV infection appears to be common in all species of apes, with combined rates of PCR-positivity and/or HBsAg carriage of 16.7% (Table 1). This rate of active infection is remarkably similarly to that of human populations in areas of endemic infection, such as Central Africa and South East Asia.

Further parallels between the epidemiology of HBV infection in human and non-human primates is provided by the evidence for efficient mother-to-child transmission of HBV in captive primate species (Noppornpanth et al., 2003; Lanford et al., 2000). In humans, such perinatally acquired infections generally lead to lifelong HBV high infectivity carriage associated with partial immunotolerance and the production of the HBV “e” antigen (HBeAg). Persistent infection of females perpetuates the infection cycle to each succeeding generation. Observations of frequent vertical transmission in captive gibbons (Noppornpanth et al., 2003) supports the hypothesis that this may also be an important mechanism for the maintenance of HBV infection in gibbons and potentially other ape species in the wild.

One explanation for the failure to detect HBV infection in Old World monkey species is that they were infected with HBV variants so divergent in sequence from previously described HBV variants infecting humans and other primates that they would be refractory to amplification with conventional primers, and be serologically non-cross-reactive with reagents used in HBsAg and anti-HBc assays. Such viruses would therefore have to be more divergent than the outlier HBV variant found in a captive woolly

monkey (Lanford et al., 1998). To address this issue we re-screened the Old World monkey species by PCR using primers that matched even the highly divergent HBV-like viruses infecting North American rodents. Our failure to detect a single additional case of HBV infection in the 93 African monkey samples available provided further evidence for an absence of infection of hepadnaviruses in these species of the *Cercopithecidae*, although it is possible that this picture may change with more extensive sampling in a greater range of species (see below).

Why HBV infection in the wild should be generally restricted to apes remains unclear at this stage. Given this highly selective distribution of HBV infection in Old World species, the existence of HBV infection in woolly monkeys (a New World primate species) becomes more difficult to account for. A total of 10 infected animals have been reported to be HBV-infected in a US zoo, although a previous serological survey of three other New World primate species (owl monkey, tamarind, and squirrel monkey) have failed to detect evidence of active or past HBV infection in this group [Table 1; (Eichberg and Kalter, 1980)]. Clearly further surveys would be required to determine whether HBV infection is present in wild caught monkeys in South America.

#### *HBV origins*

Reconstructing an evolutionary history of HBV using the data set of sequences of HBV variants infecting humans and other primates has proven to be both complex and difficult to reconcile with theories of either recent or remote times of origin (Bollyky et al., 1997; Magnius and Norder, 1995; Norder et al., 1994; MacDonald et al., 2000; Simmonds, 2001). The new sequence information obtained in this paper, and the combined analysis with the much larger dataset of HBV sequences now available from other studies, has provided strong evidence for a correlation between geographic origin of infection and sequence divergence that cuts across species barriers of non-human primates. Thus, HBV sequences from Central Africa group together, irrespective of their origin in chimpanzees or gorillas. A similar picture is observed in South Eastern Asian primates where gibbon and orangutan HBV sequences cluster together and their host species have overlapping geographic ranges.

Understanding the epidemiology of HBV infection in non-human primates is additionally complicated by the difficulty in constructing a time-scale for the divergence of the current distribution of variants found in non-human species. However, short-term rates of HBV sequence change have been determined for HBV infecting humans, with evidence for a higher rate of sequence change in individuals who mount an effective immune response to infection. Based on the HBeAg/anti-HBe status, mean frequencies of fixation of nucleotide substitution range from 2.1 to  $25 \times 10^{-5}$  nucleotide change per site per year (Hannoun et al., 2000). As HBeAg carriage (associated with a slower rate of sequence

change) is typically associated with early acquisition of HBV infection from vertical transmission, a mode of transmission that maintains HBV infection in human populations and potentially in primates (Noppornpanth et al., 2003), the sequence divergence of HBV variants infecting different subspecies of chimpanzee can be conservatively estimated as ranging at most from 1600 to 2200 years (mean sequence divergence 7%; rate of sequences change  $2.1 \times 10^{-5}$  nucleotide change per site per year). A similar time-scale for HBV evolution was inferred using non-overlapping regions of the HBV genome (Fares and Holmes, 2002).

The calculated time-scale of HBV evolution implied by measurement of rates of sequence change over short intervals indicates that the current wide distribution of HBV infection in apes must have arisen through several cross-species or subspecies transmissions in the relatively recent past. For example, HBV transmission following limited population contact between different ape species in South Eastern Asia (including between orangutans and gibbons), and between subspecies of chimpanzee and with gorillas with Africa may have provided the necessary conditions for both the introduction and subsequent differentiation of HBV genetic clusters into otherwise separate host populations.

In contrast to the co-speciation hypotheses discussed below, the recent emergence hypothesis for HBV infection in primates resolves many of the unexpected phylogenetic groupings evident from the comparison of the much larger datasets of ape-derived HBV sequences (Fig. 3). Despite the previous claim for the existence for an orangutan-specific genotype of HBV (Warren et al., 1999; Verschoor et al., 2001), the availability of a larger number of gibbon sequences, including those obtained in the current study, clearly places orangutan-derived variants deep into the gibbon clade of HBV sequences. Indeed, they show the closest genetic relationship with HBV variants infecting *Hylobates* species with geographically close or overlapping habitat ranges. In contrast, HBV variants infecting *H. lar* and the *Nomascus* gibbons from mainland Asia generally group separately.

The lack of sequence diversity of HBV sequences between separate infected orangutans further suggests that the introduction of HBV into this species may have been relatively recent. Indeed, the absence of HBV infection in our sample of 14 wild caught orangutans from Borneo suggests that HBV infection may not be so widely distributed geographically in this species as it evidently is in gibbons, and more consistent with recent introduction. Recent cross-species transmission events may also account for the close relationship between the single available gorilla-derived HBV variants with HBV infecting the *Troglodytes* subspecies of chimpanzee whose ranges overlap in Central Africa (Fig. 2 and 3A), although more sequence information of HBV infecting other gorillas in the wild is required to substantiate this further possible example of cross-species transmission in the wild.

An alternative explanation of the distribution of non-

human HBV genotypes in Africa and South East Asia proposes that the distinct genotypes of HBV in different primate species arose during the evolution of different ape species over the past 20 million years. However, the implied very low rate of nucleotide substitution of HBV over these extended periods of host diversification is incompatible with extrapolations of the previously calculated rate of HBV sequence change over short observation periods. For example, while we estimate that the sequence divergence of HBV infecting different subspecies of chimpanzees arose  $\approx 1600$ – $2200$  years ago, the times of host species divergence range from 0.8 million years for the divergence between subspecies of chimpanzees from Central Africa (troglodytes, vellerosus and schweinfurthii subspecies) and an estimated 1.5 million years between these subspecies and *P.t. verus* from West Africa (Morin et al., 1994). Similarly, extrapolating from the above rates of sequence change, the radiation of gibbon- and orangutan-associated HBV variants can be timed as no more than 3500 years, a time-scale difficult to equate with the species and/or geographic differentiation of HBV variants in the wide range of primates infected with HBV.

In defence of the co-speciation hypothesis, it could be argued that constraints on sequence change, such as the unusual and extensive use of overlapping reading frames for protein coding in the HBV genome, as well as the role of RNA structures in transcription and translation (Mizokami et al., 1997) may prevent the simple and relatively unconstrained fixation of neutral mutations. For example, the absence of true synonymous sites in over 65% of the genome through the use overlapping reading frames, and the existence of *cis*-acting RNA structures required for HBV transcription and translation may lead to a rapid saturation of sequence substitutions and homoplasy in the (relatively few) phenotypically unconstrained sites, although low rates of divergence were also observed in non-overlapping regions of the HBV genome (Fares and Holmes, 2002). The loss of the variable, phylogenetically informative sites through homoplasy would therefore explain the instability of any sequence grouping of HBV variants below the level of (human) genotype or of non-human primate genetic groupings. More importantly, extensive homoplasy would prevent the extrapolation of the short term rate of sequence change of HBV (over 10 years) to the longer periods underlying the differentiation of HBV into separate genotypes.

However, no permutation of the co-speciation hypothesis can explain the inlier position of orangutan-derived HBV variants within the clade containing gibbon sequences (Fig. 2 and 3B), or the anomalous position of the lowland gorilla HBV sequence in the chimpanzee HBV phylogeny. Even more seriously, it fails to account for the substantial sequence diversity of HBV variants infecting humans. While it might be possible to compress human genotypes A to E and G into a single group that reproduces the diversity of HBV sequences found in gibbons (and orangutans), this approach cannot account for the outlier position of geno-

types F and H. Nor can the diversity of HBV variants infecting humans, with a maximum population age of 150,000 years (Stringer, 2002), be reconciled with the much more restricted sequence diversity of HBV variants infecting the several subspecies of chimpanzee which diverged over a time-scale 10 times greater, or different species of gibbon that diverged even longer ago.

Secondly, the hypothetical restrictions on sequence divergence can not explain how HBV-like viruses infecting rodents and birds have become so divergent from human and primate HBV variants over maximum chronological times of 100 to 300 million years. Even against the yardstick of these much longer periods for sequence diversification, greater sequence divergence and an outlier position for gibbon-derived HBV sequences would be expected to be resolvable over the 15 million years of ape diversification.

In summary, the proposed scenario for the relatively recent spread of HBV among African and South East Asian primates and cross-species transmission between animals in adjacent or overlapping ranges accounts much better for the geographical rather than the species association of HBV genotypes in non-human primates than previous co-speciation theories (Magnius and Norder, 1995; Norder et al., 1994; MacDonald et al., 2000). However, while the transmission networks of HBV in Africa and South East Asia are clearly separate, one difficult remaining question is how HBV could have spread between the African, Asian and potentially South American continents in the previous few thousand years. Resolving this question will require a better understanding of the relationship between human and non-human primate HBV variants, and the potential role of humans in disseminating HBV infection over these much larger geographical distances in the more remote past.

#### *HBV divergence*

Current and previous studies of HBV genotype distributions in apes have demonstrated the existence of HBV variants with relatively limited sequence diversity, manifested in particular by the grouping of these variants interspersed with the human genotypes A–E and G (Fig. 2). There is no evidence so far for HBV variants in apes as divergent as the human genotypes F and H, nor the HBV variant obtained from a captive woolly monkey.

However, the core gene sequence of the HBV variant, AB046525, recovered from a *P.t. troglodytes* subspecies in Central Africa (Takahashi et al., 2001) is highly divergent from equivalent regions in other HBV variants obtained from chimpanzees, other apes or humans, and indeed is as divergent from viruses in these latter species as the woolly monkey sequence (Lanford et al., 1998). Strangely the rest of the genome of AB046525 groups with HBV variants recovered from the troglodytes subspecies of chimpanzees (eg., in the S gene; Fig. 3B). As previously suggested, this virus is likely to be a recombinant between a troglodytes-associated HBV variant and another highly divergent HBV

variant of unknown origin. Combined with the woolly monkey variant (again of obscure origin), these findings hint towards the existence of much more divergent HBV variants in nature that are refractory to current primate surveillance methods. Further work to discover the original species distribution of such variants will provide considerable insights into the ultimate origins and evolution of HBV in primates, and may help to resolve the many conflicting theories for the origin and past epidemiology of HBV infection in human populations.

## Materials and methods

### Primate samples

A total of 137 primate serum and plasma samples were available for screening. From Africa, samples from 71 drills (*Mandrillus leucophaeus*) were available from a rescue centre in Nigeria; in the study population, 48 were wild-born and 23 captive-born; all samples were negative for HBV surface antigen (HBsAg). A total of 14 samples were available from 8 drills, 2 mandrills (*Mandrillus sphinx*), 4 cherry-capped mangabeys (*Cercopithecus torquatus*) kept at the Limbe Wildlife Center in Cameroon, all of which were wild-born. From the same centre, a total of 7 samples from chimpanzees were also available, of which 2 were HBsAg positive, 3 were negative and 2 were unknown. Finally, one sample from a known HBsAg positive chimpanzee and 3 HBsAg-negative samples from mona monkeys (*Cercopithecus mona*), 2 samples from putty-nosed monkeys (*C. nictitans*) and 3 samples from red-eared monkeys (*C. erythrotis*) were provided for the study from a rescue centre in Nigeria.

From South East Asia, 20 gibbon samples from various wild-born gibbon species (*Hylobates* spp.; 4 HBsAg positive and 16 negative), and 14 samples from wild-born Bornean orangutan (*Pongo pygmaeus pygmaeus*; all HBsAg negative), were provided by the Pingtung Rescue Center in Taiwan. One male and one female lar gibbon (*H. lar*) from the Welsh Mountain Zoo were also included. They were HBsAg tested at the City Hospital, Edinburgh and the female was found to be positive whereas the male was negative.

All samples were stored at 4°C prior to shipping to the UK according to CITES regulations, and then stored at -25°C prior to analysis.

### Source of infection in HBV-positive primates

Louisa (*P. troglodytes*, subspecies *troglodytes*) was confiscated in Southwest Cameroon from local people, after which she was housed in the Limbe Wildlife Centre. Since confiscation she has had no contact with any other primate species. The exact details of how she was kept prior to confiscation are unknown. The second chimpanzee was

Osang (*P. troglodytes*, subspecies *vellerosus*), and was donated to a conservation project in South-Eastern Nigeria in 1996, at which time he was positive for HBsAg. Osang was also kept by local people after he was captured, and was subsequently kept with other chimpanzees at the conservation centre.

Of the 5 HBV-infected gibbons, 4 were wild-caught and one was captive-born. The captive-born was a female, Tamang (*Hylobates lar*) from a captive-born father (Paignton Zoo) with known wild-caught parents from Malaysia and Thailand, and a captive-born mother (Stuttgart Zoo) with unknown parental origins. At the Welsh Zoo she is kept with a male, Jake (*H. lar*) who was born at the Zoo, who was both HBsAg and HBV PCR negative. The remaining 4 positive gibbons consist of Crazy Woman (*H. agilis*) confiscated in Taiwan in 1993, TB Black (species not identified) confiscated in Taiwan in 1996, Happy (*H. concolor gabriellae*) confiscated in Taiwan in 1996, and Wendy (*H. concolor gabriellae*) confiscated in Taiwan in 1999. Their histories prior to confiscation are unknown, as are their origins. For the latter 4, all were HBsAg positive on first testing after confiscation. Tamang likely acquired infection from her mother peri- or post-natally. The subgenus and species identifications of the HBV-infected gibbons were provided by the sanctuary.

### PCR screening and sequencing of HBV DNA

HBV sequences were amplified by PCR as previously described (MacDonald et al., 2000) from 100 µl volumes of serum or plasma using pre-S (S1) and S (S2) gene primers (Fig. 1A and B). Nucleotide sequencing was carried out directly on second round amplification products using either the Sequenase version 2.0 kit (United States Biologicals), or ABI PRISM d-rhodamine or Big Dye kits (Applied Biosystems). Positive primate samples were sequenced using overlapping primers covering the entire genome as described previously (MacDonald et al., 2000) using primers S5, S3, 4, 22, 25, 26, 30, 33, 42 and the new primers 44 (5'-GTCCTTGWTTGGGRTTGAAGTCCCAATC-3'; positions 2959 to 2985), 45 (GGTCACMWTAYTCYTGGGAA-CAAGA; positions 2824 to 2848), 46 (ACCAWTTTATG-CYRCAGCCTCCTA; positions 1778 to 1801) and 47 (GGACGTCCTTTGKTACGTCCC; positions 1415 to 1442).

For investigation of infection with possibly more divergent HBV-like viruses in Old World primates, a further PCR assay was developed using highly conserved primers (set S3) in the surface gene that matched the following additional sequences recovered from HBV-like viruses recovered from North American rodent species: K02715 (ground squirrel), NC004107 and M19183 (woodchuck) and U29144 (arctic ground squirrel). Samples were amplified using the assay and cycling conditions as for the S1 and S2 gene primers. Nucleotide sequences obtained in this study have been submitted to GenBank and have been

assigned the following accession numbers: Ay330911 to Ay330917

#### Sequencing of the 12S mitochondrial region of primate samples

Primate samples were also amplified and sequenced using primers specific for the 12S region of the mitochondria to confirm the species and distinguish between subspecies and individual animals to allow comparisons of the primate and HBV phylogenies. 2  $\mu$ l nucleic acid was added to PCR mix as described before and amplified using a single round PCR reaction. Primers 12S-S (5'-CCATAAACAMAYAG-GYTTGGTCC-3'; positions 641 to 664) and 12S-AS (CAGGGTTTGGCTGAAGATGGCGGTATATA; positions 1270 to 1298) were used. PCR conditions were 40 cycles of 94°C for 60 sec, 55°C for 60 sec and 72°C for 60 sec, followed by a final extension at 72°C for 6 min. Amplified DNA was directly sequenced using the ABI PRISM Big Dye kit (Applied Biosystems).

Mitochondrial sequencing was used to determine the sub-species of the 2 HBV-infected chimpanzees, from which the following assignments were made: Osang: *P. troglodytes vellerosus*, Louisa: *P. troglodytes troglodytes*. Mitochondrial sequencing was unable to determine the species for gibbon TB Black because sequences for subspecies agilis and moloch were interspersed, and published sequences for gibbons were mainly determined in a different region of the mitochondrial genome. Mitochondrial sequencing was also performed on samples from the previously described HBV-infected chimpanzees from whom complete genome sequences were published [Chimp 2, Chimp 4; (MacDonald et al., 2000)]. Although both were originally classified as being versus subspecies on the basis of morphology, mitochondrial sequencing indicated that Chimp 2 was the troglodytes subspecies; this designation has subsequently been confirmed by the sanctuary.

#### Sequence analysis

Sequence data obtained using the ABI PRISM kits was viewed using the CHROMAS sequence viewer and directly imported into the SIMMONICS sequence alignment program and edited. HBV sequences were aligned with up to 10 representative sequences of each human genotype: Genotype A—HUMPRECX, AF090839, AF297624, E00010, HVHEPB, HEB344115, AF090838, U87742, AF297625, AF297619; Genotype B—AY033073, HPBA3HMS2, D50522, HPBADW3, HPBADWZ, AF282918, HBV131133, HPBADW2, AF121249, AF121243; Genotype C—AF233236, HHVBC, AB049610, AF411412, HPBADRA, AF330110, AF461359, HPBADRC, AF411408, AF363964; Genotype D—HEB344116, AB033559, AF280817, AF121240, U95551, HBVAYWC, HBVAYWGEN, HBVGEN1, HBVORFS, HBVP6PCXX; Genotype E—HHVBE4, HHVB-

BAS; Genotype F—HBVADW4A, HHVBFFOU, AF223965, AF223962, HHVBF, AB036910, AB036908, AB036916, AB036905, AB036913; Genotype G—AB056515, AF405706, AB056516, AB064312, AB064310, AB064311, AB056514, AF160501, AB056513, AB064313 and Genotype H—AY090454, AY090457, AY090460. Sequences were also compared with every available complete genome sequence obtained from primates: Chimpanzee—AB032431, HBV131575, AF242585, AF222322, AF498266, AB046525, AF305327, AF222323, AF242586, AB032432, HPBVCG, AB032433; Gorilla—HBV131567; Gibbon—AY077736, AY077735, HBV131568, HBV131574, HBV131571, HBV131569, HBV46935, HBV131572, ABO37927, AB037928; Orangutan—AF193864, NC\_002168 and woolly monkey—AF046996.

Sequence comparisons in the S-gene included the following additional further partial sequences from primates: Chimpanzees—AF305328, AF222318, AF222321, AF222312, AF222319, AF222313, AF222314, AF305326, AF222316, AF222320, AF305329, AF222317, AF305330; Gibbons—AF213009, AF213008, AF21210, AF274495, AF274496, AF274499, AF477482, AF477483, AF477484, AF477485, AF477486, AF477487, AF477488, AF477489, AF477490, AF477491, AF477492, AF477493, AF477494; Orangutans—HBVY17565, HBVY17562, HBVY17559, HBVY-17564, HBVY17561.

Phylogenetic trees for these data sets were estimated using a maximum likelihood (ML) method. To undertake as robust an analysis as possible, we employed the most complex GTR+I+ $\Gamma$  model of nucleotide substitution available; this allowed each type of nucleotide change to occur at a different rate (the general time-reversible substitution model, GTR), a proportion of nucleotide sites to be invariant (I) and a gamma ( $\gamma$ ) distribution of among-site rate variation with the  $\alpha$  shape parameter (with 4 categories) estimated from the empirical data. The maximum likelihood base composition was also estimated from the data. All parameter values are available from the authors on request. An heuristic search procedure was used to find the ML tree using successive rounds of TBR branch-swapping, optimizing the ML substitution parameters at each stage. To determine the support for key nodes on the tree we conducted a bootstrap resampling analysis using 1000 replicate neighbor-joining trees constructed under the substitution model as defined above. All these analyses were performed using the PAUP\* package (Swofford, 1998). In all cases the single woolly monkey sequence (AF046996) was used as an outgroup to root the phylogenetic trees.

#### Acknowledgments

We would like to thank Andrew Greenwood at the International Zoo Veterinary Group for providing samples and background information for gibbons Tamang and Jake; the

Limbe Wildlife Centre, Limbe, Cameroon, the Drill Rehabilitation and Breeding Centre, Nigeria, and the Pingtung Rescue Centre, Pingtung University, Taiwan for providing the remainder of the primate samples.

## References

- Andre, F., 2000. Hepatitis B epidemiology in Asia, the Middle East and Africa. *Vaccine* 18 (Suppl. 1), S20–S22.
- Arauz-Ruiz, P., Norder, H., Robertson, B.H., Magnius, L.O., 2002. Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. *Journal of General Virology* 83, 2059–2073.
- Arauz-Ruiz, P., Norder, H., Visona, K.A., Magnius, L.O., 1997. Genotype F prevails in HBV infected patients of hispanic origin in Central America and may carry the precore stop mutant. *J. Med. Virol.* 51, 305–312.
- Bollyky, P.L., Rambaut, A., Grassly, N., Carman, W.F., Holmes, E.C., 1997. Hepatitis B virus has a new world evolutionary origin. *Hepatology* 26, 765.
- Davis H.L., Suparto, I.I., Weeratna, R.R., Jumintarto, Iskandriati, D.D., Chamzah, S.S., Ma'ruf, A.A., Nente, C.C., Pawitri, D.D., Krieg, A.M., Heriyanto, Smits, W., and Sajuthi, D.D., 2000. CpG DNA overcomes hyporesponsiveness to hepatitis B vaccine in orangutans. *Vaccine* 18, 1920–1924.
- Deinhardt, F., 1976. Hepatitis in primates. *Adv. Virus Res.* 20, 113–157.
- Eichberg, J.W., Kalter, S.S., 1980. Hepatitis A and B: serologic survey of human and nonhuman primate sera. *Lab Anim Sci.* 30, 541–543.
- Fares, M.A., Holmes, E.C., 2002. A revised evolutionary history of hepatitis B virus (HBV). *J. Mol. Evol.* 54, 807–814.
- Grethe, S., Heckel, J.O., Rietschel, W., Hufert, F.T., 2000. Molecular epidemiology of hepatitis B virus variants in nonhuman primates. *Journal of Virology* 74, 5377–5381.
- Hannoun, C., Horal, P., Lindh, M., 2000. Long-term mutation rates in the hepatitis B virus genome. *Journal of General Virology* 81, 75–83.
- Hu, X., Javadian, A., Gagneux, P., Robertson, B.H., 2001. Paired chimpanzee hepatitis B virus (ChHBV) and mtDNA sequences suggest different ChHBV genetic variants are found in geographically distinct chimpanzee subspecies. *Virus Res.* 79, 103–108.
- Hu, X., Margolis, H.S., Purcell, R.H., Ebert, J., Robertson, B.H., 2000. Identification of hepatitis B virus indigenous to chimpanzees. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1661–1664.
- Lanford, R.E., Chavez, D., Brasky, K.M., Burns, R.B.3., Rico-Hesse, R., 1998. Isolation of a hepadnavirus from the woolly monkey, a New World primate. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5757–5761.
- Lanford, R.E., Chavez, D., Rico-Hesse, R., Mootnick, A., 2000. Hepadnavirus infection in captive gibbons. *Journal of Virology* 74, 2955–2959.
- Linnemann Jr., C.C., Kramer, L.W., Askey, P.A., 1984. Familial clustering of hepatitis B infections in gorillas. *Am. J. Epidemiol.* 119, 424–430.
- MacDonald, D.M., Holmes, E.C., Lewis, J.C., 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.
- Magnius, L.O., Norder, H., 1995. Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. *Intervirology* 38, 24–34.
- Michaels, M.G., Lanford, R., Demetris, A.J., Chavez, D., Brasky, K., Fung, J., Starzl, T.E., 1996. Lack of susceptibility of baboons to infection with hepatitis B virus. *Transplantation* 61, 350–351.
- Mizokami M., Orito, E., Ohba, K., Ieko, K., Lau, J.Y., Gojoberi, T., 1997. Constrained evolution with respect to gene overlap of hepatitis B virus. *J. Mol. Evol.* 44 Suppl 1:S83–90, S83–90.
- Morin, P.A., Moore, J.J., Chakraborty, R., Jin, L., Goodall, J., Woodruff, D.S., 1994. Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* 265, 1193–1201.
- Noppornpanth, S., Haagmans, B.L., Bhattarakosol, P., Ratanakorn, P., Niesters, H.G., Osterhaus, A.D., Poovorawan, Y., 2003. Molecular epidemiology of gibbon hepatitis B virus transmission. *Journal of General Virology* 84, 147–155.
- Norder, H., Courouze, A.M., Magnius, L.O., 1994. Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. *Virology* 198, 489–503.
- Norder, H., Ebert, J.W., Fields, H.A., Mushahwar, I.K., Magnius, L.O., 1996. Complete sequencing of a gibbon hepatitis B virus genome reveals a unique genotype distantly related to the chimpanzee hepatitis B virus. *Virology* 218, 214–223.
- Ogata, N., Ostberg, L., Ehrlich, P.H., Wong, D.C., Miller, R.H., Purcell, R.H., 1993. Markedly prolonged incubation period of hepatitis B in a chimpanzee passively immunized with a human monoclonal antibody to the a determinant of hepatitis B surface antigen. *Proc. Natl. Acad. Sci. U.S.A.* 90, 3014–3018.
- Simmonds, P., 2001. 2000 Fleming Lecture. The origin and evolution of hepatitis viruses in humans. *Journal of General Virology* 82, 693–712.
- Stringer, C., 2002. Modern human origins: progress and prospects. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 357, 563–579.
- Swofford, D.L., 1998. PAUP\*. Phylogenetic analysis using parsimony (\*and other methods). Version 4.
- Takahashi, K., Brotman, B., Usuda, S., Mishiro, S., Prince, A.M., 2000. Full-genome sequence analyses of hepatitis B virus (HBV) strains recovered from chimpanzees infected in the wild: implications for an origin of HBV. *Virology* 267, 58–64.
- Takahashi, K., Mishiro, S., Prince, A.M., 2001. Novel hepatitis B virus strain from a chimpanzee of Central Africa (*Pan troglodytes troglodytes*) with an unusual antigenicity of the core protein. *Intervirology* 44, 321–326.
- Thomas, H.C., Jacyna, M.R., 1993. Hepatitis B virus: pathogenesis and treatment of chronic infection, in: Zuckerman, A.J., Thomas, H.C. (Eds.), *Viral Hepatitis*, Churchill Livingstone, Edinburgh, pp. 185–207.
- Thornton, S.M., Walker, S., Zuckerman, J.N., 2001. Management of hepatitis B virus infections in two gibbons and a western lowland gorilla in a zoological collection. *Vet. Rec.* 149, 113–115.
- Vartanian, J.P., Pineau, P., Henry, M., Hamilton, W.D., Muller, M.N., Wrangham, R.W., Wain-Hobson, S., 2002. Identification of a hepatitis B virus genome in wild chimpanzees (*Pan troglodytes schweinfurthi*) from East Africa indicates a wide geographical dispersion among equatorial African primates. *Journal of Virology* 76, 11155–11158.
- Vaudin, M., Wolstenholme, A.J., Tsiquaye, K.N., Zuckerman, A.J., Harrison, T.J., 1988. The complete nucleotide sequence of the genome of a hepatitis B virus isolated from a naturally infected chimpanzee. *Journal of General Virology* 69, 1383–1389.
- Verschoor E.J., Warren, K.S., Langenhuijzen, S., Heriyanto, Swan, R.A., Heeney, J.L., 2001. Analysis of two genomic variants of orang-utan hepadnavirus and their relationship to other primate hepatitis B-like viruses. *Journal of General Virology* 82, 893–897.
- Warren, K.S., Heeney, J.L., Swan, R.A., Heriyanto Verschoor, E.J., 1999. A new group of hepadnaviruses naturally infecting orangutans (*Pongo pygmaeus*). *Journal of Virology* 73, 7860–7865.
- Worley, M.B., Stalis, I.H., 2002. Detection of virus-like particles in the liver of black and white ruffed lemurs with hepatitis. *J. Wildl. Dis.* 38, 258–265.
- Zuckerman, A.J., Thornton, A., Howard, C.R., Tsiquaye, K.N., Jones, D.M., Brambell, M.R., 1978. Hepatitis B outbreak among chimpanzees at the London Zoo. *Lancet* 2, 652–654.