Geographic and Genetic Diversity of Hepatitis B

Erik Alestig

Department of Infectious Medicine
Institute of Biomedicine
Sahlgrenska Academy at University of Gothenburg

UNIVERSITY OF GOTHENBURG

Göteborg 2011

Geographic and Genetic Diversity of Hepatitis B

erik.alestig@gu.se

ISBN 978-91-628-8182-5

Printed in Göteborg, Sweden 2010

An electronic version of this thesis is available on:
http://hdl.handle.net/2077/22932

Geographic and Genetic Diversity of Hepatitis B

Erik Alestig, Department of Infectious Medicine, Institute of Biomedicine
Sahlgrenska Academy at University of Gothenburg
Göteborg, Sweden

ABSTRACT

Hepatitis B Virus (HBV) infection is a global health problem and may lead to chronic hepatitis, liver cirrhosis and hepatocellular cancer. The ability of HBV to adapt to the host environment by genetic variation has lead to the evolution of 8 established (A-H) and 2 putative genotypes (I-J), each corresponding to a rather well-defined geographical distribution. The genotypes have clinical impact on natural course of infection, prognosis and treatment outcome. Genotypes A-D were identified in 1988, E-F in 1993, G in 2000 and H in 2002. The last decade the study of HBV phylogeny has been focused on identifying subgenotypes and recombinants, and their geographic distribution. The relatedness to non-human HBV has also become in focus but has remained unknown. Molecular methods have become increasingly used for epidemiological investigations of HBV infections, but the most appropriate genetic regions for such applications have not been established.

The aims of this thesis were to investigate the genotypes and genetic variants of HBV from different geographic areas in order to better understand the evolution and genetic variability of HBV and how it can be analysed.

HBV strains from Vietnam, Mongolia and Australia were amplified by polymerase chain reaction, subjected to direct sequencing and classified after phylogenetic analysis. In Vietnam 77% of all strains were of genotype B (mainly B4), 22% were of subgenotype C1 and one strain was a X/C recombinant (putative genotype I). Southeast Asian genotype C with C-1858, a variant that rarely develops the precore stop codon that abolishes synthesis of HBeAg, was shown to constitute a clade with two phylogenetic subgroups. Subgenotype D1 was found in strains from Mongolia, suggestive of a closer relation to Middle East and the Mediterranean area than with China. HBV from Australian Aborigines were found to represent a subgenotype (C4), with an S region that may originate from early recombination with an unknown genotype. Molecular clock rates were estimated by calculations based on genetic distances and data for human migration and were compared with mutations rates observed in patients. These rates, ranging from 2×10^{-5} to 10^{-6} substitutions per site per year suggest that HBV genotypes separated 2,000 – 40,000 years ago, while subgenotypes and the X/C recombinant would have evolved 1,000 – 20,000 years ago.

The HBV-D3 strains causing acute hepatitis B in Swedish injection drug users has changed at a low rate since the 1970ies, but three clades were shown to have circulated since 1975, and they were shown to be distinguishable by their pattern at 3 residues in the *a* determinant part of the S region. The S region appears to be favourable as primary target for subgenomic molecular epidemiology of HBV-D3.

In summary, results of this thesis contribute to explain the evolution of HBV and have a clinical impact on molecular epidemiology of acute HBV infections.

Keywords: hepatitis B virus, genotype, molecular epidemiology, phylogeny, sequence analysis, mutation

ISBN: 978-91-628-8182-5

SAMMANFATTNING PÅ SVENSKA

Hepatit B-virus (HBV) drabbar levern och kan orsaka kronisk leverinflammation, skrumplever och levercancer. Sjukdomen utgör ett globalt hälsoproblem med uppskattningsvis 360 miljoner kroniskt infekterade och mer än 600000 HBV-relaterade dödsfall årligen. Förekomsten av HBV-infektion är störst i östra Asien och subsahariska Afrika där många infekteras vid födseln eller i tidiga barnaår. I Sverige finns uppskattningsvis 20000 personer med kronisk HBV-infektion, varav de flesta härstammar från länder med hög förekomst. Akut hepatit B är ovanlig i Sverige, men årligen diagnostiseras ca 200 personer som till övervägande del har smittats genom injektionsmissbruk eller sexuellt. Screening av blodprodukter och riktad vaccination mot högriskgrupper har bidragit till att minska antalet nyinsjuknanden. För närvarande finns ingen behandling som släcker ut hepatit B-infektionen, men ett flertal läkemedel som hämmar virusets förmåga att föröka sig används framgångsrikt hos patienter med leverpåverkan och förhindrar fortsatt leverskada.

Leverskadan vid hepatit B-infektion orsakas inte av viruset i sig, utan anses istället bero på det immunförsvar som aktiveras för att bekämpa infektionen. Virus förändras genetiskt genom mutationer och rekombination, och genom naturligt urval selekteras de virusstammar som har bäst överlevnadsförmåga. Evolutionen har genom att virusstammar i olika delar av världen har fått utvecklas var för sig under lång tid, lett fram till uppkomsten av ett flertal varianter, s.k. genotyper. För HBV finns tio genotyper beskrivna (A-J), t ex dominerar genotyp A i Afrika och Europa, B och C i östra Asien och Oceanien samt D i Mellanöstern och Medelhavsområdet. Genotyperna har visat sig ha klinisk betydelse eftersom de påverkar behandlingssvar och risk för leverskadeutveckling inklusive cancer. Den ökande kunskapen om genotyper och deras betydelse har medfört ett ökande intresse för att kartlägga deras utveckling och geografiska utbredning. Dessa studier har lett till en bättre kunskap om HBV:s utbredning och har varit viktig för att förstå dess evolution. I arbetet används s.k. fylogenetiska metoder som genom att applicera matematiska kalkyler på DNA-sekvenser från olika virusstammar kan beräkna ett sannolikt släktträd.

HBV är ett av de smittämnen som omfattas av Smittskyddslagen. Det innebär att nya fall ska anmälas och att behandlande läkare och Smittskyddsmyndigheten ska försöka att klargöra smittvägar för att begränsa fortsatt spridning. I detta arbete används ibland genetisk typning av virus, i första hand genotypning men ibland mer detaljerad karaktärisering där DNA-sekvensen för en del av virusgenomet undersöks. En svårighet därvid är att de undersökta virusstammarna kan skilja sig mycket lite åt, och att kunskapen om molekylära epidemiologiska analyser av HBV är begränsad.

Det övergripande syftet med avhandlingen har varit att undersöka den genetiska variationen hos HBV-stammar avseende deras förekomst och ursprung. I delarbetena I-III analyserades och klassificerades virusstammar från Sydostasien, Mongoliet och Australien genom DNA-sekvensering och konstruktion av fylogenetiska träd. I Vietnam var tre fjärdedelar av alla virusstammar av genotyp B, en fjärdedel av genotyp C och en virusstam var X/C-rekombinant (även kallad genotyp I). I Mongoliet förekom genotyp D, vilket kan tyda på närmare kontakt med Mellanöstern och Medelhavsområdet där genotyp D dominerar än med Kina där genotyp B och C dominerar. Hos australiska aboriginer återfanns virusstammar av genotyp C, möjligen bärande på ett rekombinationssegment från en hittills okänd genotyp. Virusstammar med sydostasiatiskt ursprung av genotypen C med C-1858 (en variant som påverkar utvecklingen av en särskild mutation) visades utgöra en fylogenetisk enhet. Uppskattningsvis skedde förgreningen av C-1858 varianten för flera hundra år sedan eller mer, och separationen av de australiska aboriginerna för mer är 1000 år sedan.

I delarbete IV analyserades med samma metodik HBV-stammar av subgenotypen D3, vilka identifierats hos patienter med akut hepatit B i Göteborg under tre olika tidsperioder (1975, 1995-1996, 2007-2008). Vid jämförelse av hela HBV-genomet sågs en överraskande liten skillnad mellan stammarna, vilket försvårar epidemiologiska slutsatser. Trots att S-regionen inte är den mest variabla delen visade den sig vara den lämpligaste delen för den här typen av analyser, åtminstone i denna patientgrupp. Baserat på aminosyramönster i den så kallade *a*-determinanten kunde virusstammarna delas upp i tre olika varianter som sannolikt alla härstammar från en D3-stam som infördes i missbrukarkretsar i Västeuropa redan på 60-talet. Två av dessa varianter befanns fortfarande cirkulera.

Sammanfattningsvis har denna avhandling bidragit med ny kunskap om utbredningen och evolutionen av HBV-genotyper i ett längre perspektiv, samt till kunskap om hur dess förändring i ett kortare perspektiv påverkar tolkningen av molekylärepidemiologiska utredningar.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Alestig E, Hannoun C, Horal P, Lindh M.
Phylogenetic origin of hepatitis B virus strains with precore C-1858 variant.
Journal of Clinical Microbiology 2001.

II. Alestig E, Hannoun C, Horal P, Lindh M.
Hepatitis B virus genotypes in Mongols and Australian Aborigines.
Archives of Virology 2001.

III. Thuy PTB, Alestig E, Liem NT, Hannoun C, Lindh M.
Genotype X/C recombinant (putative genotype I) of hepatitis B virus is rare in Hanoi, Vietnam - genotypes B4 and C1 predominate.
Journal of Medical Virology 2010.

IV. Alestig E, Söderström A, Norkrans G, Lindh M.
Genetic diversity of genotype D3 in acute hepatitis B.
Submitted.

CONTENT

ABBREVIATIONS

aa	amino acid
APC	antigen-presenting cell
AHV	arctic squirrel hepatitis virus
BCP	basal core promotor
CID	core internal deletions
ESLD	end-stage liver disease
GSHV	ground squirrel hepatitis virus
HBV	hepatitis B virus
HBcAg	hepatitis B core antigen
HBeAg	hepatitis B e antigen
HBsAg	hepatitis B surface antigen
HBIG	hepatitis B immune globulin
HLA	human leukocyte antigen
IDU	injection drug user
IFN	interferon
IVDU	intravenous drug user
ML	maximum likelihood
MP	maximum parsimony
NA	nucleoside/nucleotide analog
NJ	neighbor joining
nt	nucleotide
ORF	open reading frame
PCR	polymerase chain reaction
pgRNA	pregenomic RNA
Th	T helper cell
UPMGA	unweighted pair-group method with arithmetic means
WHV	woodchuck hepatitis virus
WMHBV	woolly monkey hepatitis B virus

1 INTRODUCTION

Hepatitis B virus (HBV) infection of the liver may cause acute or chronic hepatitis, liver cirrhosis and hepatocellular carcinoma. About 2 billion people have been exposed to HBV and more than 360 million people worldwide are chronically infected. It has been estimated that 600 000 people die every year due to consequences of the disease, making HBV a serious global health problem.

The virus does not exert any direct cytopathic effects. Instead, activation of the host's immune system is considered to cause the liver damage. From the time of viral entry, a counteract interplay between viral replication and defence mechanisms of the immune system takes place. Attacks from the immune system force the virus to adapt. Modified strains emerge as the error prone replication of HBV easily gives rise to mutations, and if these strains are selected and transmitted to another individual, the evolution may continue.

The history and origin of human HBV is largely unknown. The genetic variability, in conjunction with the migration of people, has lead to the divergence of HBV into genetically different groups, so-called genotypes, with a distinct geographic distribution. The tracks of these genotypes can be traced by phylogenetic analysis, assessing the relations between sampled strains. In times of globalisation and frequent travel, these tracks slowly fade away.

This thesis focuses on the genetic diversity of HBV and explores the geographic distribution of genotypes. How and when these genotypes have evolved and their relation to mankind's population of the world is addressed.

1.1 The virion

Hepatitis B virus belongs to the *Hepadnaviridae* family. Its members are divided into two genera; *Orthohepadnaviruses* infecting mammals, and, *Avihepadnaviruses* affecting birds. Figure 1 shows representatives of species infected by the orthohepadnaviruses including humans and other primates. Aviahepadnaviruses have been found in ducks, herons and storks.

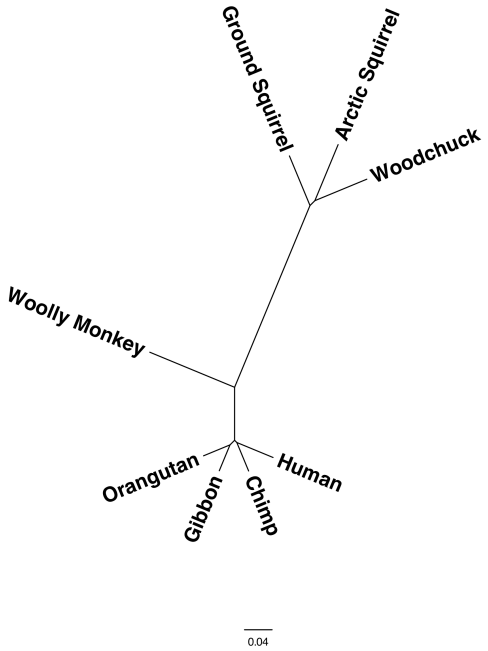

Figure 1. Unrooted Neighbor-joining tree of complete genome sequences of Orthohepadnaviridae from GenBank with accession number: Human (Genotype A, X02763), Chimp (Chimpanzee, D00220), Gibbon (U46935), Orangutan (AF193864), Woolly Monkey (WMHBV, AF046996), Ground Squirrel (GSHV, K02715), Arctic squirrel (AHV, AGU29144), American Woodchuck (WHV, M18752).

The genome of HBV is a small, circular, partially double-stranded DNA molecule of approximately 3.2 kb nucleotides. The virion (Figure 2), also known as the Dane particle, measures 45 nm in diameter by cryo-electron microscopy. An outer lipid bilayer containing three envelope proteins (HBsAg) encloses the icosahedral capsid. All three HBsAg molecules are translated from the same reading frame, but with different startcodons. They are named according to size, glycosylation and weight in Dalton; **S**HBs (p24/gp27); **M**HBs (gp33/gp36); **L**HBs (p39/gp42) and produced in the proportions 4:1:1. Besides being import proteins in the lipid bilayer of the virion, a large number of HBsAg molecules are secreted outside the cell as 22 nm spherical particles composed of the small HBsAg particles or as filamentous particles consisting of all three HBsAg sizes. The capsid is formed by the core protein (HBcAg) and contains a polymerase with reverse

transcriptase and ribonuclease H activity attached to the DNA genome. In addition to these proteins, HBV infected cells also produce e-antigen (HBeAg), a protein secreted into serum with immune regulating effects and a protein X with a putative and incompletely understood regulatory function. These are the seven proteins of HBV.

Of notice, HBsAg was first called the *Australian Antigen* as it was discovered by Blumberg in the serum of an Australian Aborigine 1965 [Blumberg et al., 1965], for which he was awarded the Nobel Prize 1976.

In comparison to mammalian HBV viruses, avian hepadnaviruses have major structural differences. For example, the DHBV (duck hepatitis B virus) DNA is almost fully double stranded, only has two surface proteins, lacks X-gene and is a several hundred bases shorter.

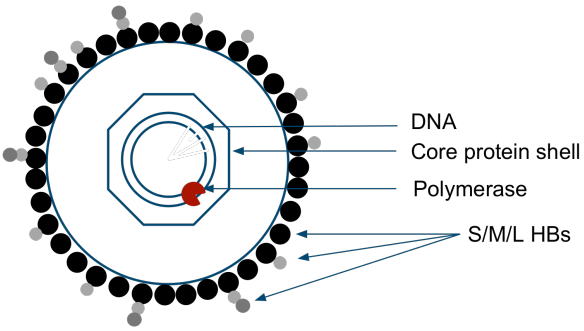

Figure 2. The hepatitis B virion.

1.2 In the hepatocyte

1.2.1 Replication

The replication of HBV is unique. Although a DNA virus, it encodes a reverse transcriptase and replicates through an RNA intermediate. The mechanism of entry into the hepatocyte is largely unknown, but probably HBV attaches to hepatocytes via the pre-S1 part of the HBsAg. This receptor-ligand binding may then, at least partly, account for the high specificity of liver cells, although HBV infection has also been found in lymphocytes, pancreas and the kidney. On penetration into the cell, the capsid is delivered to the nucleus where the partial DNA strand is synthesized to completion.

The HBV genome utilizes four overlapping reading frames (ORFs) to efficiently store the coding instructions for its proteins and regulatory sequences, as illustrated in figure 3. They are termed P (polymerase), S (surface), C (core) and X (HBx protein). As the genome is circular, the starting position is arbitrary, but conventionally (and in this thesis) the EcoR1 cleavage site is used for numbering. The minus strand is complete, whereas the plus strand is only partly synthesized and has a variable size at the 3' end. Two 11-nt direct repeats, DR1 and DR2, and a 225 nt long cohesive stretch enable the circulation of the genome [Tiollais et al., 1985].

In the nucleus, the DNA is transcribed into four mRNAs of different sizes (3.5 kb, 2.4 kb, 2.1 kb and 0.9 kb), which are transported into the cytoplasm. The large 3.5 kb mRNA (longer than the genome) encodes HBc and HBe antigens and the HBc mRNA also codes for the polymerase and serves as template for replication of the whole genome. HBcAg and HBeAg share 90% of their protein sequence but are translated differently, and HBeAg is secreted into serum and does not self-assemble like a capsid antigen. The 2.4 kb fragment encodes the large S protein, and the smaller M and L proteins are transcribed from the 2.1 kb fragment. The small 0.9 kb fragment encode the X protein, which probably is involved in transcription regulation and acts as a protein kinase. A new virion particle is assembled as the core proteins encapsulate the large 3.5 kb mRNA and negative-sense DNA is synthesized by a reverse transcription, RNA degradation and a positive-sense DNA synthesizes. Finally, the core is enveloped and the viron leaves the cell by exocytosis.

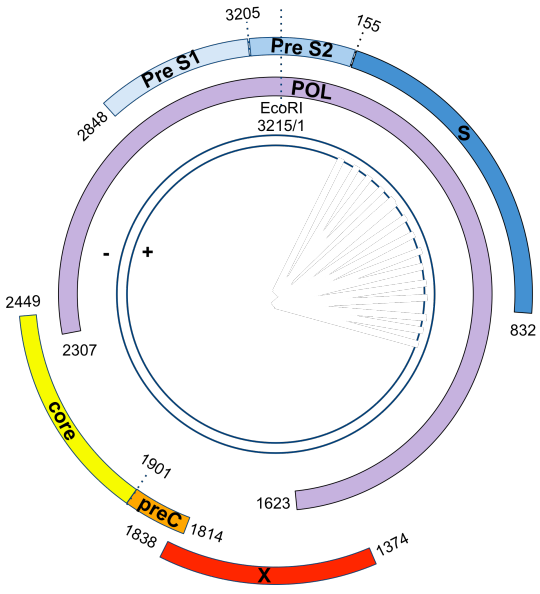

Figure 3. Genomic map of HBV (genotype B, C, F, H).

Replication and transcription are controlled by two enhancers (Enh I, Enh II) and four promoters (preC/C, preS1, S, X) initiating transcription of each ORF, as shown in figure 3. Transcription factors binding to Enh I and Enh II are more abundant in liver cells than in other cells, resulting in a higher viral replication and specificity for these cells.

1.3 The liver

1.3.1 Immune response

The HBV does not itself exert any known direct cytopathic effects in immune competent patients. This is supported by the fact that during the first phase of infection viral replication is very high without signs of liver damage. Hepatocellular injury is generally accepted to be the result of attacks from the immune system. The immune response to HBV infection is both B- and T-cell derived. Antibodies (anti-HBs, anti-HBc, anti-HBe) are produced and targeted towards the antigens HBsAg, HBcAg, HBeAg, and are used in the clinical diagnostic. Antibody activity of the pre-S domains of MHBs and LHBs [Klinkert et al., 1986], polymerase [Kann et al., 1993] and the X protein [Feitelson et al., 1990] have also been described. Antibodies against the so-called *a* determinant, a region of the S protein that is exposed on the virion, are neutralising and vaccines composed of (small) HBsAg are protective. Accordingly, it is presumed that the excess of HBsAg protein secreted as spheres and filament may function as immunological decoys.

Interferon (IFN) probably plays an important roll as it enhances human leukocyte antigen (HLA) class I expression on hepatocyte cell membranes [Pignatelli et al., 1986]. The epitopes presented on the cells evoke a T-cell response. Antigens such as HBc and HBe are displayed to cytotoxic T cells (CD8+), which in turn lyse the cell using perforins disrupting the cell structure [Ferrari et al., 1992]. T-cell response is also mediated by MHC class II expression on antigen-presenting cells (APCs) activating T helper cells (Th). It has been shown that cytotoxic T-cell response is dependent on HLA (MHC class I) type and that the Th response is genetically restricted [Milich et al., 1995]. Further studies have revealed an association of HLA polymorphism with outcomes of hepatitis B virus infection, which might have future clinical relevance [Ramezani et al., 2008; Thursz et al., 1995].

1.3.2 Liver damage

Liver biopsy is performed to assess liver injury and fibrosis and is by many considered the gold standard for this purpose [Feld et al., 2009]. It is a stable marker for disease progression, but is a costly, invasive procedure with the drawbacks of potential sampling error. Inflammation and fibrosis is histologically evaluated according to several scoring systems, most widely used are Ishak and METAVIR [Rozario and Ramakrishna, 2003]. Recently, elasticity measurements, an alternative non-invasive method for staging liver fibrosis in viral hepatitis has shown a relatively good correlation with histological fibrosis stage [Ogawa et al., 2007].

1.4 Clinical aspects

HBV is transmitted through percutaneous or mucosal contact with blood or body fluids. In high endemic areas the predominant routes of transmission are vertical, from mother to child during pregnancy and childbirth, and horizontal during the preschool years. In low endemic areas, sharing of syringes between injection drug abusers and unprotected sex are the main routes of transmission [Lavanchy, 2004]. Other risk factors for infection include transfusion of unscreened blood, tattooing, usage of non-sterilized instruments and multiple-dose vials in health care settings.

The scope of outcome in HBV infected patients varies widely, ranging from asymptomatic and self-limited infection to fulminant hepatitis and chronic disease. The risk of chronic infection is inversely proportional to age. Persistent infection has been reported in up to 90% of infants infected at birth and 20-50% at infection between ages 1-5 [Shapiro, 1993]. In adults the risk of chronic infection is less than 5% [Hyams, 1995].

1.4.1 Acute hepatitis

The majority of acute cases of HBV in the western countries are related to injection drug use (IDU) and sexual transmission in young adults. Besides symptoms common to all forms of acute hepatitis (jaundice, nausea, weight loss, flu-like illness etc), patients with acute HBV infection typically also suffer from fever, urticaria and arthralgia. These symptoms generally subside within a few weeks along with disappearance of HBV DNA and seroconversion from HBeAg to anti-HBe. A large proportion of the cases are asymptomatic and the infection may pass without notice. Most patients with acute hepatitis B are HBsAg positive at presentation, but the critical test is IgM anti-HBc, which confirms acute HBV infection. If HBsAg is not lost within 6 months, the patient is considered to be a chronic carrier (Figure 4). The risk for chronic disease in adults is low, and none of 126 individuals in a Swedish outbreak became chronic HBsAg carriers [Blackberg et al., 2000].

Acute hepatitis may in some cases progress to fulminant hepatitis leading to liver failure [Fagan and Williams, 1990], a state with high mortality. Antiviral treatment with lamivudine or other nucleoside analogs are often used in such cases, but still liver transplantation is often necessary [Wang and Tang, 2009].

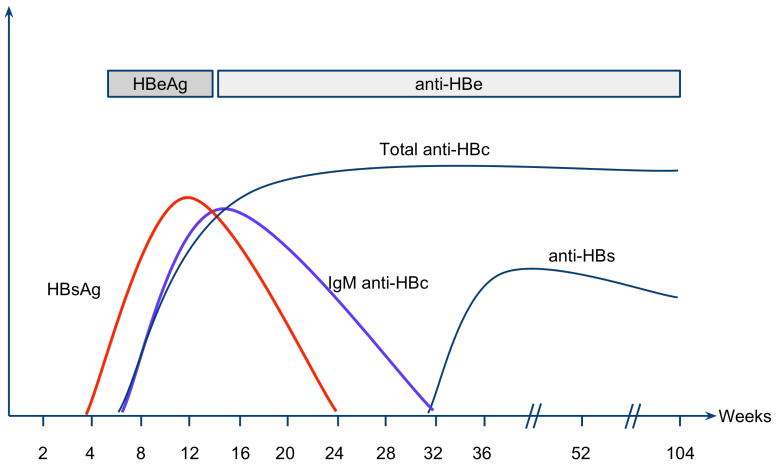

Figure 4. Serology at acute HBV infection.

1.4.2 Chronic hepatitis

The natural course of hepatitis B virus infection can be divided into four stages. The first stage, the *immune tolerance phase*, is characterized by active viral replication and immune system tolerance. In this initial phase, HBV DNA replicates at a high level and the HBs and HBe antigens are produced and detectable. ALT levels are normal, in this phase, which may last for 10-30 years. Next, in the *Immune clearance phase*, the immunologic response is causing inflammation and hepatic injury. As a result of viral clearance, ALT levels are elevated and moderate/severe necroinflammation on liver biopsy is observed. This usually occurs in the second or third decade of life in patients with perinatally acquired disease. At the third phase, *Inactive carrier state*, viral clearance is accompanied by seroconversion of HBeAg, resulting in relatively low HBV DNA level and normalized ALT levels. A few patients reach the *final stage*, when HBsAg is completely cleared and anti-HBs becomes detectable as a sign of immunity. Infected individuals not passing the second stages frequently develop progressive liver damage, which may cause cirrhosis or hepatocellular cancer. As mentioned, chronic infection rarely develops in adults presenting with acute HBV infection, but there is a great variation in the reported rates of persistence (0.2%-12%) [Ozasa et al., 2006].

Clinically, the e-antigen is important in chronic infection as it is regarded a marker for replication and indicative of ongoing infection. When seroconversion occurs, it normally reflects remission of liver disease and viral clearance.

1.4.3 Occult hepatitis

After resolved HBV infection, HBV DNA in serum or the liver may in some cases still be detectable in the absence of HBsAg [Blackberg and Kidd-Ljunggren, 2000; Michalak et al., 1994]. This is termed occult hepatitis B, and probably reflects that HBV persists life-long in a small proportion of the hepatocytes. The clinical importance of this is not completely understood, but occult hepatitis B has been associated with reactivation in the setting of immunosuppression, enhanced risk for liver cancer, interference with treatment response in patients with hepatitis C and a risk for transmission through blood transfusion [Michalak et al., 1994; Schmeltzer and Sherman, 2010; Yuki et al., 2003].

1.4.4 Treatment

The goal of treatment is to minimize the risk for complications and reduce viral replication. In general, two classes of drugs are available for the treatment of chronic hepatitis, interferon (IFN) and nucleoside/nucleotide analogs (NA). IFNs have antiviral effects and modulate the immune system. NAs directly interfere with the reverse transcription of hepatitis virus and thus have, strong antiviral effect on HBV.

A disadvantage of NAs over IFN is the development of resistance mutations during treatment, reducing and abolishing the antiviral effect. For lamivudine, which is still widely used in East Asia, the proportion of patients with lamivudine-resistant mutations (YMDD) increases rapidly for each year of treatment (the mutation rtM204V/I increased from 23% after one year to 65% after five years of treatment) [Lok et al., 2003]. Resistance mutations have also been associated with flare up-reactions. Entecavir and tenofovir have replaced lamivudine as preferred treatment due to a much lower risk for resistance and other NAs are under development and evaluation.

1.4.5 Hepatocellular cancer

Chronic HBV infection can lead to hepatocellular carcinoma (HCC). Worldwide more than 50% of HCC cases, and in highly endemic areas 70-80% of HCC cases, are attributable to HBV [Nguyen et al., 2009]. The risk for HCC has been reported to be 100 times higher in patients with persistent HBV infection than in non-infected individuals [Beasley et al., 1981]. In patients with chronic HBV infection, presence of HBeAg and higher levels of HBV DNA have been found to be strong risk factors for HCC [Chen et al., 2006], which mainly develops in patients with liver cirrhosis.

1.4.6 Immunoprophylaxis

Prevention of primary HBV infection by vaccination has proven to dramatically decrease the HBsAg prevalence and HCC incidence. The first report of an effective HBV vaccine, tested on chimpanzees, came 1976 [Buynak et al., 1976]. The first such vaccine, composed of highly purified HBsAg particles from human plasma donors was approved for general use in 1981. It was replaced by DNA-recombinant HBV vaccines when they became available in 1986. Vaccine efficacy studies reported 90-100% protection [Fitzsimons et al., 2005; Liao et al., 1999] and the vaccine also proved to

be effective in young children and infants, despite the fact that young children can be poorly responsive to vaccines due to their immature immune system. The currently used, second generation DNA recombinant vaccine, have been available for use since 1991. After a mass vaccination programme in Taiwan the overall prevalence rate of hepatitis B core antibody dropped from 26% in 1984 to 4.0% in 1994 [Chen et al., 1996], and lowered the HCC incidence [Chang et al., 2005]. Similar results have been reported from other highly endemic countries where routine vaccination have been implemented [Viviani et al., 1999]. The HBV vaccine is the world's first cancer prevention vaccine and part of the World Health Organization (WHO) childhood programme [Lavanchy, 2004].

Passive prophylaxis is given with hepatitis B immune globulin (HBIG), a sterile solution of antibodies against hepatitis B. It can be administered after needle stick and other types of accidental exposure, to liver transplant recipients and to newborns of mothers with high viral loads. Perinatal hepatitis B virus transmission can be reduced by more than 90% if HBIG are given in combination with recombinant vaccine [Stevens et al., 1987]. In utero transmission, high inoculum, surface gene escape mutations and HBeAg placental passage have been suggested to explain the failing proportion [Wang et al., 2003]. In a report the prevalence of *a* determinant mutants in HBV DNA-positive children from Taipei increased from 7.8 to 28.1% ten years after vaccination, and was significantly higher in those fully-vaccinated than in those unvaccinated [Hsu et al., 1999]. The most common vaccine mutants are sG145R and sT126A/S, estimated to account for almost half of breakthrough infections [Chang, 2010]. Concerns have been raised that these mutants could hamper the future effectiveness of hepatitis B immunization programmes, but their impact appears to be low and one study showed that these S region mutants usually were of maternal origin and not induced by the vaccine [Ngui et al., 1997].

1.5 Epidemiology

The prevalence of chronic HBV in the world is highly variable, ranging from more than 10% in some Asian and African countries to less than 0.5% in Northern Europe [Custer et al., 2004]. The HBV disease burden are generally classified as percentage of HBsAg carriers in the population and categorized as low (<2%), intermediate (2-7%) or high (>8%), as shown in figure 5.

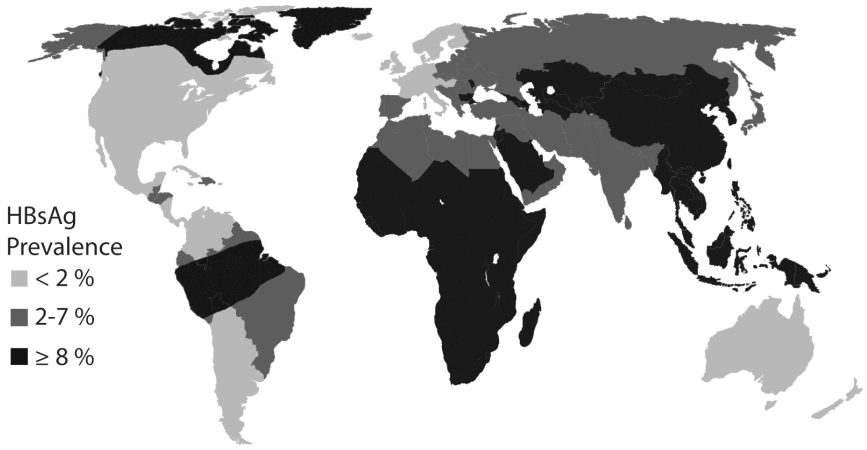

Figure 5. Worldwide prevalence of chronic HBV (Source: World Health Organization, adapted from www.who.int).

Sweden is a low endemic country in terms of HBV infections [Struve et al., 1992]. All known cases of HBV are reported to the Swedish Institute for Infectious Disease Control, who monitors the epidemiological situation. The incidence in Sweden of HBV is 12-21 per 100,000 habitants; the prevalence below 0.5% and the number of new reported cases has been around 1500 annually for the last decade. Most reported cases are chronic infections in immigrants arriving from endemic areas, but approximately 200 cases per year are acute infections (Figure 6), mainly acquired by injection drug use (IDU) or via the sexual route [Lindh et al., 2000]. Genotype D3 is common among IDU:s in Sweden as in the rest of the world, but other genotypes have been reported in this group too (England – genotype A2) [Sloan et al., 2009]. Genotypes A2 and D3 have been the prevailing genotypes in patients with acute HBV infection in Sweden like in the rest of Northern Europe, but increased travel has most likely contributed to the spread of other genotypes. In the county of Stockholm 154 cases of acute hepatitis were registered during the years 2004-2007, of which 65, 25, 18, 10 were genotype D3, D1,

C2, A2 respectively. The increase in subgenotypes A2 and D1 might indicate increased sexual transmission, because A2 has been more common among homosexual men and D1 more frequently affects women [Sundberg et al., 2009].

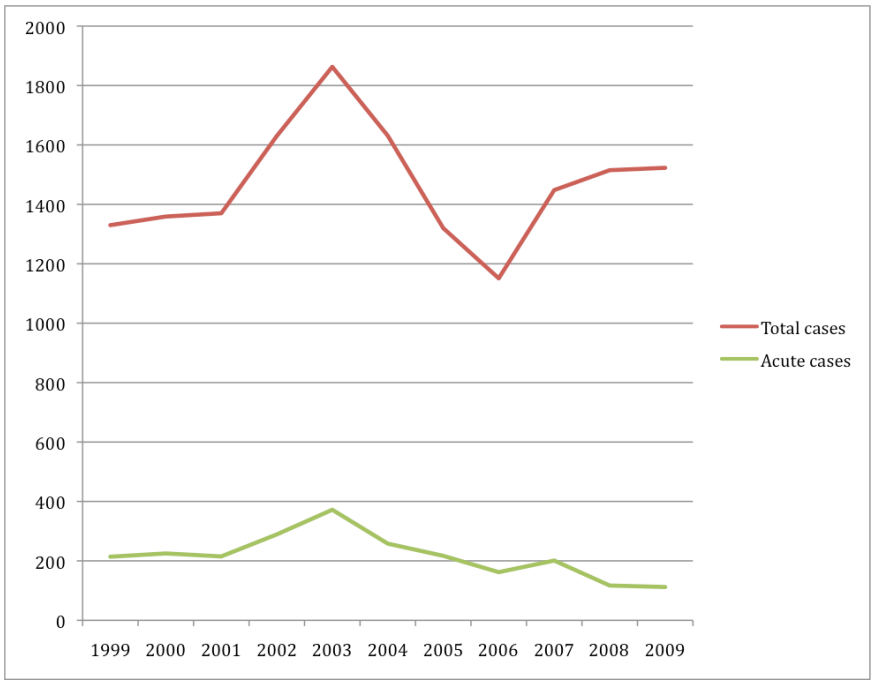

Figure 6. Reported cases of hepatitis B virus infection in Sweden. (source: Swedish Institute for Infectious Disease Control, www.smi.se)

Injection drug users constitute a subpopulation highly susceptible to HBV infection as they share infected injection tools. In Sweden, this group comprises an estimated 25000 individuals (75% males) and approximately 400 die of them of causes related to drug abuse [The Swedish Council for Information on Alcohol and Other Drugs, 2009].

1.5.1 Genotypes and subtypes

Based on a nucleotide difference of more than 8%, HBV has been classified into eight genotypes (A-H), reflecting their geographic and ethnic origin [Arauz-Ruiz et al., 2002; Norder et al., 1994; Okamoto et al., 1988; Stuyver et al., 2000]. In summary, genotype A is prevailing in European union, South America and Sub-Saharan Africa, B and C in Eastern Asia and Oceania, D in the Mediterranean and Middle East regions, E in Western Africa and F and H in the Americas [Sanchez et al., 2007]. Genotype G has no geographic association and has mainly been found in the U.S.A. and France [Kurbanov et al., 2010]. Intragenotype differences of >4% has divided genotype A, B, C, D and F into subgenotypes, as shown in table 1. Phylogenetic clustering within these subgenotypes has been proposed to be denoted clades, representing strains with less than 4% diversity [Kramvis et al., 2008]. The nomenclature defining genotypes and sub classification has lately been debated, and universal criteria pursued. A nucleotide divergence of 7.5% based on completed genomes has been suggested for genotypes, to better match phylogenetic clusters [Kramvis et al., 2008; Kurbanov et al., 2010; Schaefer et al., 2009].

Table 1. HBV genotypes and subgenotypes and their distinct geographical distribution.

Genotype/ subgenotype	Main geographical distribution	Reference
A1 (Aa)	South and East Africa, India	[Bowyer et al., 1997; Sugauchi et al., 2004a]
A2 (Ae)	Europe, USA, Canada	[Sugauchi et al., 2004a]
A3 (Ac)	Western Africa	[Hannoun et al., 2005; Kurbanov et al., 2005]
A4	Western Africa	[Olinger et al., 2006]
A5	Western Africa	[Olinger et al., 2006]
B1 (Bj)	Japan	[Sugauchi et al., 2004b; Sugauchi et al., 2002]
B2	China, Vietnam	[Sugauchi et al., 2004b; Sugauchi et al., 2002]
B3	Indonesia	[Norder et al., 2004]
B4	Vietnam	[Norder et al., 2004]
B5	Philippines	[Nagasaki et al., 2006; Sakamoto et al., 2006]
B6	Alaska, Canada, Greenland	[Sakamoto et al., 2007]
B7	Indonesia	[Nurainy et al., 2008]
B8	Indonesia	[Mulyanto et al., 2009]
C1 (Cs)	Southern Asia: Thailand, Vietnam, Laos, Bangladesh	[Chan et al., 2005; Huy et al., 2004; Norder et al., 2004]
C2 (Ce)	Eastern Asia: Japan, Korea, China	[Huy et al., 2004; Norder et al., 2004]
C3	Polynesia, Micronesia	[Norder et al., 2004]
C4	Australia	[Norder et al., 2004; Sugauchi et al., 2001]
C5	Philippines, Vietnam	[Sakamoto et al., 2006]
C6	Indonesia	[Utsumi et al., 2009]
C7	Philippines	[Cavinta et al., 2009]
C8	Indonesia (Nusa Tenggara)	[Mulyanto et al., 2009]
C9	Indonesia (Dili)	[Mulyanto et al., 2009]
C10	Indonesia (Nusa Tenggara)	[Mulyanto et al., 2010]
D1	Middle East, Mediterranean	[Norder et al., 2004]
D2	Russia	[Norder et al., 2004]
D3	Serbia, South Africa, Alaska	[Norder et al., 2004]
D4	Oceania, Somalia	[Norder et al., 2004]
D5	Eastern India	[Banerjee et al., 2006]
D6	Indonesia (Papua)	[Lusida et al., 2008]
D7	Tunisia	[Meldal et al., 2009]
D8	Niger	[Abdou Chekaraou et al., 2010]
E	West Africa	[Norder et al., 1994]
F1	South and Central America	[Norder et al., 2004]
F2	South America, Polynesia	[Norder et al., 2004]
F3	Venezuela and Colombia.	[Huy et al., 2006]
F4	Argentina, Bolivia	[Huy et al., 2006]
G	USA, France	[Stuyver et al., 2000]
H	South and Central America	[Arauz-Ruiz et al., 2002]

In addition to these established genotypes it has been proposed that the genotype X/C recombinant, first identified patients of indigenous Vietnamese origin living in Sweden [Andre et al., 2000] should be classified as genotype "I" [Tran et al., 2008]. Due to low diversity to the closest neighbour (7%) and evidence of recombination, it has however been debated whether it should be considered a new genotype [Kurbanov et al., 2008] or be regarded as a recombinant.

Recently, a new HBV strain (JRB34) differing by more than 10% from, and without evidence of recombination with, known genotypes was found in a Japanese patient with a history of residency in Borneo [Tatematsu et al., 2009]. It was proposed to be assigned genotype "J", and surprisingly, phylogenetical analyses of its complete genome indicated a closer relationship to HBV from apes than to human HBV genotypes.

Major structural differences between HBV genotypes are displayed in table 2.

Table 2. Nucleotide length and characteristic indels of HBV genotypes. Numbering of nucleotides according to EcoRI=0.

Genotype	Length	Indel
A	3221	Insertion core: 6 bp
B	3215	
C	3215	
D	3182	Deletion pre-S1: 33 bp
E	3212	Deletion pre-S1: 3 bp
F	3215	
G	3248	Insertion core: 36 bp Deletion pre-S1: 3 bp
H	3215	

1.5.2 Coinfection

HBV carriers with more than one genotype are frequently reported. Coinfection occurs to be more common between genotype B and C [Lin et al., 2007; Song et al., 2006], A and D [Hannoun et al., 2002], presumably due to the coexistence of these genotypes in the same regions. The clinical impact of coinfections is unclear, but viral loads have been reported to be higher in coinfected patients [Toan et al., 2006]. The frequency of coinfection may be associated with genotyping method as the reported frequency varies widely [Kurbanov et al., 2010].

1.5.3 Serotypes

Classically, before genotypes where defined, HBV strains were categorized into nine classes according to serological properties of the S protein: *adw2, adw4q-, adrq+, adrq-, ayw1, ayw2, ayw3, ayw4* and *ayr*. The molecular basis for this classification was variation at a few sites in the S region. The *a* determinant (aa 124-148) is the major antigenic determinant common for all serotypes. The *d/y* and *w/r* variations depend on Lys/Arg substitutions at residue 122 and 160 respectively [Okamoto et al., 1987b]. Determinants *w1/w2*, *w3* and *w4* are classified by Pro, Thr or Leu substitutions at residue 127 respectively, and *w1* variation is distinguished by Arg122, Phe134 and/or Ala159 [Norder et al., 1992]. Finally, *q* type is probably due to variation at residue at 177 and 178 [Norder et al., 1992]. To a large extent, genotypes and subgenotypes have replaced the usage of serotypes.

1.6 Diversity

Two opposing forces can be discerned in terms of diversity of HBV. Firstly, the error-prone polymerase producing a high rate of nucleotide substitutions. Secondly, the extreme compactness of the genome preventing a large degree of genetic variability from occurring, by the overlapping open reading frames [Mizokami et al., 1997]. The error rate of the HBV polymerase has been estimated to be $1.5\text{-}5 \times 10^{-5}$ per site per year in some reports [Okamoto et al., 1987a; Orito et al., 1989], although the results of these estimations vary [Girones and Miller, 1989]. Nevertheless, this mutation rate and an average daily production rate of up to 10^{11} virions per day, as in a highly infective patient, would imply that all possible mutations would arise each day in such an individual, and might spread to surrounding hepatocytes in case of selective advantage. This deduced large number of produced viral variants also gives support to the idea that HBV is present in the blood of infected individuals as several quasispecies [Shuhart et al., 2006].

The role of quasispecies for the evolution of HBV is unclear, but for estimating the molecular clock of HBV, quasispecies were analyzed in one report and did not influence the calculations in a general context (the dominate strain was sufficient for rate estimations) [Osiowy et al., 2006].

1.6.1 Precore mutations

It has been well known that some chronic HBV carriers with active liver disease are HBeAg negative and that HBeAg is not necessary for viral replication. In 1989 it was shown that a TAG stop codon (G1896A) in the precore region could explain the absence of HBeAg in patients with high viremia levels. Later it was found that this mutation was frequent also in HBeAg negative patients with low HBV DNA levels, and that other mutations in the precore region, e.g. in the start codon or codon 2, also may abolish the production of HBeAg [Lindh et al., 1996]. Further, the configuration of the nucleotides 1858 of HBV precore region was found to determine the development of the stop codon mutation at nt 1896 due to the requirements for stability of the stem-loop structure of the encapsidation signal [Li et al., 1993; Lindh et al., 1995; Lok et al., 1994]. Thus, the precore stop mutation (G1896A) develops only in strains having T at position 1858, explaining that the G1896A mutation was rare in genotype A, which always carries C-1858. This also explained the lack of stop codon mutations in part of the genotype C strains from East Asia, which were found to carry T-1858, either in combination C-1856 or T-1856. The latter variability has putative clinical importance because having TCC instead of CCC at position 1856 to 1858 has been associated with higher ALT levels and higher rate of livercirrosis [Chan et al., 2006].

Because precore mutations may abolish the production of HBeAg they are often considered as the cause of loss of HBeAg during the course of infection. In a longitudinal study, Moriyama et al. however observed that loss of detectable HBeAg preceded the appearance of precore mutations [Moriyama et al., 1994]. From this one may conclude that loss of HBeAg primarily is due to immune mediated reduction of HBV DNA levels, probably in combination with anti-HBe production, and that the emergence of precore mutants is a linked but parallel phenomenon resulting from immune selection. Still, a HBeAg negative status (“HBe minus phenotype”) is explained by the precore sequence for the proportion of patients who retain high HBV DNA levels (>1 million copies/mL) in the HBeAg negative phase.

The mechanism for the emergence of the precore stop codon mutation is not well understood. Probably, the abolished synthesis of HBeAg represents an escape from immunological attack. This escape may be effective by loss of MHC class I display of epitopes critical for recognition by cytotoxic T cells. Considering that most of these epitopes may be produced also by degradation of core, it seems likely that the basis for selection rather may be the loss of secreted HBeAg, because extracellular HBeAg in liver lobuli may be picked

up by antigen presenting cells, and serve to augment the immune response via Th activation.

1.6.2 Basal core promotor mutations

Basal core promoter (BCP) mutations frequently emerge in both HBeAg positive and negative patients. The most common is a double mutation at nts 1762-1764 (AGG to TGA) [Okamoto et al., 1994]. Several studies, both in vivo and in vitro, have shown that this double mutation reduces transcription of mRNA and the secretion of HBeAg, and enhances viral replication [Buckwold et al., 1996; Laras et al., 2002]. BCP mutations are more common in genotype C than genotype B [Lindh et al., 1999] and have been associated with an increased risk of developing severe inflammation [Lindh et al., 1999], cirrhosis and HCC [Fang et al., 2002; Kao et al., 2003]. BCP mutations have also been shown to predict HBeAg seroconversion [Chan et al., 1999; Yamaura et al., 2003]. The importance of BCP mutations have been less well studied in European patients, but one study of patients carrying genotype A and D, suggest that they are associated with a worse clinical course [Jardi et al., 2004]. In contrast, a report from India did not find any clinical impact of mutations in this region in subtypes of genotype D [Chandra et al., 2009].

1.6.3 Deletions

Deletions in the preS region have been associated with progressive liver disease, and, in comparison with genotype B, have been more frequently observed in genotype C, and associated with a more unfavourable clinical outcome for these patients [Sugauchi et al., 2003]. In the course of persistent infection, these preS deletions tend to accumulate with time. Deletions may occur anywhere in the preS region, and vary greatly in length (9-294 nt in the report of Sugauchi et al.).

Core deletions, in contrast to preS deletions, mainly appear in the HBeAg positive phase of patients who develop chronic active hepatitis [Ni et al., 2000; Tsubota et al., 1998]. This phase usually lacks mutations and other genetic changes that emerge in later stages. Mutants with deletions in the C gene may lead to non-functional core proteins, and hence, have been suggested only to co-exist with wild type strains that produce C-protein [Okamoto et al., 1987c; Yuan et al., 1998]. HBV with deletions in the C gene have been significantly associated with development of HBV-related end-stage liver disease (ESLD)

and death in renal transplant recipients [Gunther et al., 1996; Preikschat et al., 2002]. The deletions of the core gene can vary greatly in size and location, and are sometimes referred to as CID, core internal deletions [Marinos et al., 1996; Yuan et al., 1998].

Two specific deletions are characteristic for some genotypes. In the genomes of genotype D there is a 33-nucleotide deletion in the preS1 region, and for genotype E and G a short 3-nucleotide in the preS1 region exists.

1.6.4 Recombination

Recombination has a potential of changing genetic sequences and viral properties. The exchange of genome fragments from two co-infecting viruses may be a relatively frequent event in HBV, but probably recombinants rarely have selective advantages because only a small number of recombinants have been identified. However, recombination may be overlooked (and therefore underreported), because specific testing with Simplot analysis or similar methods often are needed to detect them.

Bollyky et al. reported recombination in 2 of 25 strains after constructing phylogenetic trees from the four different open reading frames separately, thus identifying intergenotypic recombination between genotypes D and A (genotype A fragment at nt 735-2365) and B and A (genotype A fragment at nt 2014-2203), respectively. Bowyer with colleagues examined partions and sub-partions of 65 complete published genome sequences of all genotypes and identified mosaic-like structures in 14 sequences, where no recombination previously was reported. Six sequences of genotype B had a core region more similar to genotype C and another six strains were D/A recombinations of different types [Bollyky et al., 1996; Bowyer and Sim, 2000]. Hannoun et al. examined five strains of HBV of Vietnamese origin with atypical RFLP genotyping patterns. Simplot and bootscanning indicated that in three sequences nt 1801 to 2865 were more similar to genotype C, while the rest of the genome resembled none of known genotypes (but were most similar to genotype A) [Hannoun et al., 2000b]. As mentioned above it is still under debate whether this recombinant should be considered a new genotype (I) [Kurbanov et al., 2008; Tran et al., 2008].

The most widely spread recombination is the one that is found in all B subgenotypes in mainland Asia and Indonesia (B2-B5, B7), and in which the precore and core regions originate from genotype C [Sugauchi et al., 2002]. This recombination appears to have an advantage because genotype B strains

without the recombination only remain in Japan and in the Arctic (subgenotypes B1 and B6). Possibly, presence of a genotype C core segment might explain the reportedly worse clinical outcome seen in patients carrying genotype B2 compared to B1 [Orito et al., 2005].

Because recombination requires co-infection with strains of different genotypes, recombinants most likely originate from geographic regions where multiple genotypes are circulating. Whether these co-infections become established through simultaneous or sequential transmission is not known. The capacity of recombinants to spread probably depends on their fitness and the susceptibility of the population. Some recombinants have become very prevalent, such as the mentioned B subtypes, and the C/D recombinant that is the prevailing in Tibet [Cui et al., 2002b]. It has been proposed that intergenotypic and intragenotypic recombinations may represent an important source of genetic variation, essential for the evolution of HBV [Simmonds and Midgley, 2005]. However, the number of identified recombinations is relatively small in comparison with point mutations and probably the circulating recombinants emerged a long time ago, suggesting that the importance of recombination for the genetic variability of HBV is relatively limited.

Interspecies recombination is also possible; evidence of this has been found in a chimpanzee HBV strain taken of a wild born chimpanzee in East Africa, which harboured a 500 nt genotype C segment (spanning the RT domain of the polymerase) [Magiorkinis et al., 2005], but is probably a rare event as no other case has been reported. Another example of possible interspecies recombination is genotype G which might be a result of a recombination involving ancestral strains of genotypes that have disappeared [Purdy et al., 2008; Simmonds and Midgley, 2005].

1.6.5 Genotype and clinical outcome

There have been substantial efforts to link genotypes to different clinical outcomes. In regions where both genotype B and C prevail, several reports have shown that genotype C infections have worse clinical outcome compared with genotype B in terms of severe inflammation, cirrhosis [Chan et al., 2002a; Kao et al., 2000a; Kao et al., 2002; Lindh et al., 1997; Lindh et al., 1999; Nakayoshi et al., 2003] and prevalence of hepatocellular carcinoma [Chan et al., 2004; Chan et al., 2003]. In accordance, genotype B has been associated with an earlier and a higher rate of HBeAg seroconversion than genotype C [Chu et al.,

2002; Chu and Liaw, 2005; Furusyo et al., 2002]. On the other hand, genotype B has also been associated with a higher rate of severe icteric flares as compared with patients in Hong Kong carrying genotype C [Chan et al., 2002b]. In comparison with C2, subgenotype C1 has been associated with a higher frequency of developing basal core promotor mutations [Chan et al., 2005].

Genotype A has been associated with a higher tendency to cause chronic infection, and for the better, transition into the inactive carrier state after HBeAg seroconversion, in comparison with genotype D [Rodriguez-Frias et al., 2006; Sanchez-Tapias et al., 2002]. However, in a European study comparing genotype A and D, no difference was found in the degree of liver damage [Rodriguez-Frias et al., 2006]. Subgenotype D1 have been associated with higher frequency of chronic liver disease compared to other D subgenotypes [Chandra et al., 2009].

The clinical impact of genotypes and on treatment efficiency has also been studied. Genotype A and B has been associated with a better response to interferon treatment compared to genotype C and D [Janssen et al., 2005; Kao et al., 2000b; Wai et al., 2002].

2 AIMS

The overall aim of this thesis was to investigate the genetic variability of hepatitis B virus, and more specifically:

- To phylogenetically characterise HBV from parts of Southeast Asia, Mongolia and Australia, geographical regions where no or relatively few strains have been reported.

- To evaluate the application of molecular epidemiology on acute HBV infections in western Sweden related to injection drug use.

- To explore the evolution of hepatitis B virus in a general context.

3 PATIENTS AND METHODS

3.1 Patients and samples

(I) Stored serum samples from nine HBeAg-positive HBV genotype C carriers residing in western Sweden were used for phylogenetical analysis of the complete HBV genome. The patients were of Southeast Asian origin (Vietnam, Malaysia and Thailand), eight carried HBV genotype C with C-1858 and one genotype C with T-1858.

(II) Serum samples from 14 HBV carriers were used for phylogenetical analysis of the complete HBV genome. Nine carriers were of Mongolian origin and five were Australian Aborigines. The samples originated from a previous investigation of HTLV (Human T Lymphotropic Virus) seroepidemiology and had previously been identified as HBsAg positive.

(III) Serum samples from 153 patients that had shown positive reaction for HBsAg at routine preoperative screening at the Department of Ophthalmology, Hanoi, Vietnam were used for sequencing of parts of the HBV genome. Amplification of the S region was successful in 87 of the samples and the X/core region was sequenced in a subset of 68 samples. Complete genome sequences were obtained from 12 of the samples.

(IV) Serum samples from three periodically separated cohorts (1975, 1995-1996 and 2007-2008) of injection drug users from western Sweden were analysed. The samples were serologically positive for HBc IgM and carried HBV genotype D3. Four samples originated from pairs that were epidemiologically related according to information obtained from contact tracing.

3.2 PCR and sequencing

PCR, polymerase chain reaction, is one of the most used techniques in molecular biology. It enables multiple duplication of a defined target region of DNA by a series of cycles involving denaturation of DNA strands, primer annealing and extension of annealed primers by Taq DNA polymerase. PCR is a sensitive method capable of detecting low levels of DNA (for HBV, the detection level is about 190 IU/mL). Overlapping segments covering the

entire HBV genome or parts of the genome were amplified by PCR with the primers described in Table 3 in this thesis. Primarily, eight overlapping segments covering the whole genome were used for complete genome sequencing in first round of PCR. Nesting was done as a semi-nest where needed, and in some cases additional primers listed in the table were used to obtain the sequences.

Prior to PCR, nucleic acid was extracted from serum samples and purified. Several methods are available. Samples of the paper I-II were mainly prepared with hot sodium extraction [Truett et al., 2000], a simple and inexpensive method. For a higher yield in some cases, a commercial kit QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) was used. Samples in paper III-IV were extracted with MagNA Pure LC Instrument (Roche Applied Science, Mannheim, Germany), an automated system using magnetic beads to isolate the nucleic acid.

The PCR set up used in this thesis where were mainly; denaturation at 94°C for 45 seconds for separation of the two DNA strands, annealing of primers at 50–65°C for 60 seconds to target sites (temperature depending on primers used) and elongation of the bases with DNA polymerase at 72°C for 90 seconds, extended by 3 seconds for every cycle. For the Taq DNA Polymerase (Boehringer Mannheim) used, an initialization step of heating the reaction to 94-96°C degrees for three minutes was necessary to activate the enzyme, so called “hotstart”.

Table 3. Primers used for PCR and direct sequencing.

Primer number	Sequence	Overlapping set*
74R	AACTGGAGCCACCAGCAGGAA	n8
80R	TTCCTGAACTGGAGCCACCA	7, 8
256F	GTGGTGGACTTCTCTCAATTTTC	2
476R	GACAAACGGGCAACATACC	1
635F	TTCCTATGGGAGTGGGCCTC	3
796R	CGGTA(A/T)AAAGGGACTCA(A/C)GAT	n2
986R	ACTTTCCAATCAATAGG	2
1022R	GCAGCAAADCCCAAAAGACC	
1164F	GCCAGGTCTGTGCCAAGTGTTTGCTGA	4
1170F	TCTGTGCCAAGTGTTTGCTGA	
1285R	CTAGGAGTTCCGCAGTATG	n3
1353R	GGAGAGCACGACAGAATTGTC	3
1603F	GTTGCATGGAGACCACCGTGAAC	5
1606F	GCATGGAGACCACCGTGAAC	

1680F	ATGTCGACAACCGACCTTGA	n5
1798R	ACCAATTTATGCCTACAGCC	n4
1801R	CAGACCAATTTATGCCTACAGCCT	
1865F	CAAGCCTCCAAGCTGTGCCTTGGGTGGCCTT	6
1888R	CCCAAGGCACAGCTTGGAGGCTT	4
2058R	GTATGGTGAGGTGAACAATG	5
2177F	ATGGGCCTAAAGTTCAGGCAA	
2251R	CTCAAGAACCGTTTCTCTTCCAA	
2381F	AACTCCCTCGCCTCGCAGAC	7
2467R	TGAGTCCAAGGAATACTAAC	n6
2488R	CCAGTAAAGTTCCCCACCTT	6
2823F	TCACCATATTCTTGGGAACAAGA	1, 8
3098R	GCAGGAGGCGGATTTGCT	n7
3199F	CATCCTCAGGCCATGCAG	n1

* Prefix *n* = used as inner primer in semi-nested PCR.

The amplified material was purified by Qiaex (Qiagen). Direct sequencing was performed with the same primers as were used for PCR. Each amplicon was analysed in both the sense and antisense directions after cycle sequencing reactions with fluorescent dye (Rhodamine; Applied Biosystems) terminators (ddNTPs). The sequences were read in an ABI Prism 310 or 3130 automated capillary sequence reader (Applied Biosystems) and processed using the Sequence Navigator software (Applied Biosystems) or Sequencher software (Genes Codes Generation).

3.3 RFLP and genotyping

Restriction fragment length polymorphism (RFLP) can be used to differentiate two or more homologous DNA molecules from each other. This method uses the genetic variability at particular positions in the DNA as restriction enzymes cleaves the molecules at predetermined sites resulting in fragments of different lengths. Visualization of the fragments with electrophoresis on agarose gels enables the identification of typical cleavage patterns, which can be used for genotyping or mutation analysis (as depicted in Figure 7). In paper I-II, an in-house developed method based on analysis of the preS or S gene was used for HBV genotyping [Lindh et al., 1997; Lindh et al., 1998]. Lately, direct sequencing in conjunction with phylogenetic analysis (paper III-IV) and real-time PCR, have to a large extend replaced RFLP for genotyping.

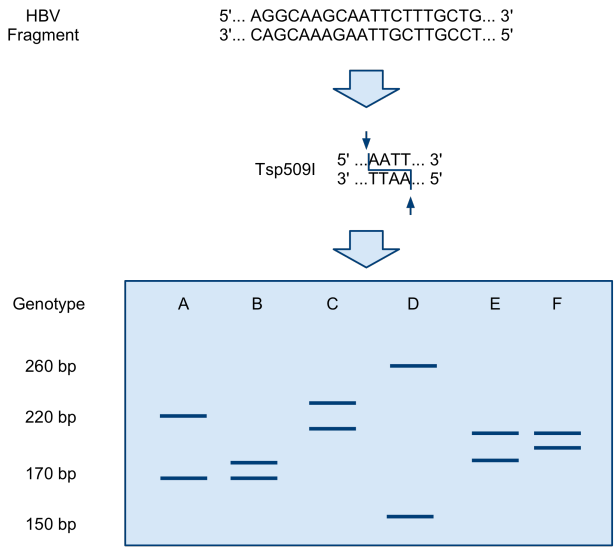

Figure 7. Estimated RFLP patterns of core region nt 1865-2384 cleaved by Tsp509I and association to genotypes.

3.4 Phylogenetic analysis

With phylogenetic analysis, the evolutionary history of related organisms are reconstructed. The relationships can be depicted in phylogenetic trees, composed of nodes and bifurcating branches. Some of the fundamental concepts of phylogenetic analysis are described below.

3.4.1 The genetic code

The genetic code is stored in a double stranded DNA for most organisms, and RNA for some viruses. This information can be coding for proteins or be non-coding (as for rRNA and regulatory sequences). The genetic code is made up of 4 bases (AGCT). Triplets of these four bases constitute a codon with 64 possible combinations. 61 sense codes encoding 20 amino acids, and 3 non-sense codes encoding stop codons.

During replication of the genome, errors in the replication sometimes occur, changing the sequence (mutation). If such a change results in an amino acid change, it is called a “non-synonymous” mutation. If it does not affect an amino acid, it is called a “silent” or “synonymous” mutation. The genetic code is degenerated in such a way that if the mutation occurs in the third codon position, an amino change is likely to happen in 30%, at second codon position always and at first position in 96% of the cases. Further, at the nucleotide level, a change from purine (A, G) to a purine or pyrimidine (C, T) to a pyrimidine is called a *transition*. *Transversion* is a change from a purine to pyrimidine or the reverse. There are four possible transitions (A<->G, C<->T) and eight tranversions (A<->C, A<->T, G<->C, G<->T). Transitions occur twice as often as transversions due to chemical properties of the bases and steric hindrance, but transversion infer more disruptive amino acid changes. These conditions can be taken into account when performing phylogenetic analysis by choosing the appropriate substitution model.

3.4.2 Evolution

Evolution can be defined as the consecutive fixation of mutations. First a mutation arises, then it forms a polymorphism and eventually it gets fixed when reaching 100% of the population. The alternative variants, alleles or mutants in terms of viruses, are lost and replaced by the new variant. Natural selection according to Charles Darwin is of course a key mechanism of viral evolution, as the fittest strain is selected over the other. Besides mutations and natural selection other major forces such as genetic drift, bottleneck effects, founder effects and recombination largely contribute to evolution.

Genetic drift reflects the stochastic process, not driven by environmental or adaptive pressures, by which genetic changes occur over time. The changes are independent of natural selection and are merely due to random sampling. In a large population genetic drift will have little effect since the sampling errors eventually will even itself out.

A *population bottleneck* can strongly add to genetic drift, when a significant part of the population for any reason is decreased, and chances for mutations to spread in a population are increased. Also, a population bottleneck can markedly increase inbreeding due to the reduced pool of possible mates.

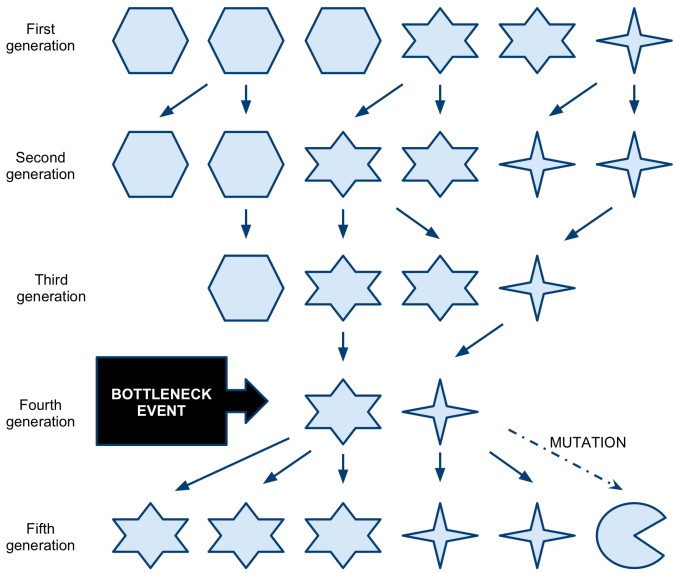

Figure 8. Bottleneck events. (Adapted from The Phylogenetic Handbook, 2nd Edition. Cambridge)

The founder effect is the loss of genetic variation that occurs when a small cohort of a population breaks off and forms a new population in another region. The new population may not have the same allele frequency (or the same polymorphism) as the parental population, and in addition, due to the small population, effects of genetic drifts mentioned above tend to increase impact on evolution.

Recombination should be mentioned as a large potential contributor to evolution. The exchange of genetic segments between two different viral strains in the same cell is a fast way to achieve new viral properties. For hepatitis B virus, recombination of homologous DNA sequences is most likely common to evolution, as evidence for intragenotypic, intergenotypic and interspecies recombination has been reported. Phylogenetic analysis generally does no take recombination events into account (since the recombinant fragments have different phylogenetic history), often leaving known recombinant strains out of the comparison.

3.4.3 Molecular clock

The *evolution rate* of the population is the rate of fixation of mutations and depends on several factors, partly mentioned above, including mutation rate (e.g. how error prone the is replication), generation time (e.g. chimpanzees reproduce at a younger age than humans), selective pressure and other evolutionary forces.

The number of substitutions per generation in a haploid population (applicable to most viruses) is $N\mu$ (where N is the population size and μ is the mutation rate). When substitutions are effectively neutral, they are fixed by random genetic drift only and with a probability of 1/(N), hence the fixation rate is then $N\mu * 1/(N) = \mu$. Selective pressure may influence the fixation rate, slow down or increase the rate. If this selective pressure is constant then the rate is constant and the molecular clock can be assumed. Molecular clock calculations to estimate the coalescence will only be exact if the clock is constant, otherwise it may only be approximate.

3.4.4 Alignment algorithms

Before phylogenetic analysis can be performed, the sequences must be properly aligned. Indels, defined as either an insertion or a deletion, cause gaps in alignments. In coding regions, gaps often occur as triplets, usually in loops where the three dimensional structure is less effected. To search for the best alignment, strategies aimed at reducing the length and the number of gaps are generally used. For each gap in the alignment, a penalty score is calculated and the alignment with the highest total score is chosen as the optimal alignment. Gap penalty (GP) score can be calculated as: $GP = g + hl$ (where g=gap opening penalty, l=length of gap, h=gap extension penalty)

For the alignment of protein sequences it is possible to take the biochemical properties of the amino acids into account. In these cases a special matrix recalculating the score is used, the most widely matrixes used are PAM (Point Accepted Mutation) and BLOSUM (Blocks Substitution Matrix).

Computers should be used for aligning sequences, but in the end of the process, visual inspection usually improves alignment since no algorithm is perfect in every sense. For divergent sequences, guiding the nucleotide alignment by amino acid alignment can yield a better alignment.

3.4.5 Substitution models

When two or more nucleotide or amino acid sequences are compared, the observed difference in substitutions may not reflect their true evolutionary history. A typical example is the mutation that reverts back to its original state of the parental strain, and hence, will not be taken into account when only observing differences. Likewise, the shift from one nucleotide to another when comparing two strains might be the result of several sequential mutations, not only one. Moreover, mutations may occur with different frequencies at different sites of a sequence, and, transversions and transitions may not occur equally. To correct for these factors and other types of errors essential to estimating evolutionary rates and phylogenetic analyses, several substitution models have been developed. Figure 9, displays the fundamental problem stated above. The expected difference is almost always larger than the observed dissimilarity.

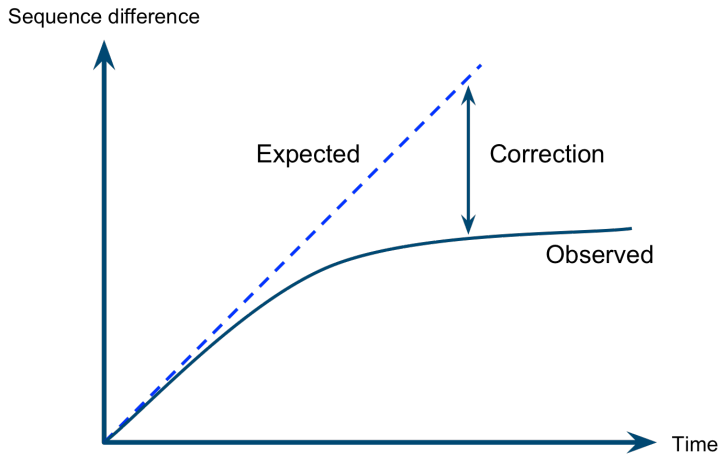

Figure 9. Expected and observed differences.

The distance between two taxa can be measure by the absolute difference or the uncorrected p-distance (the proportion of nucleotide sites at which the two sequences compared are different), but most often a substitution model is used. Jukes-Cantor is a simple model assuming that each base (A, C, T, G) in the DNA sequence has an equal chance of mutating [Jukes and Cantor, 1969]. Similar to this model is Kimura's K2P model [Kimura, 1980], which also take different frequencies of transversions and transitions into account. Further model have been developed, Tamura-Nei model has the features of

distinguishing between the two different types of transitions and assumes that all transversions occur at the same rate.

It is generally a good idea to try different models to find the best. There are special software that can be used for this purpose such as MODELTEST [Posada and Crandall, 1998].

3.4.6 Principles of tree construction

Phylogenetic trees depict the relationship between genes, species or other taxa, and can reflect their evolutionary history. The result from phylogenetic analysis is an inferred tree that may, or may not, be different from the true tree.

Phylogenetic analysis can be done with virtually any sequential data, but for viral strains nucleotides or amino acids are used. If the strains a very distantly related, nucleotide alignments may be too ambiguous and amino acid alignments are preferred. Opposingly, for closely related sequences, nucleotide alignment yields a higher phylogenetic signal than amino acids. In the phylogenetic analysis only alignments with homologous sequences should be used. Homologues sequences are two sequences sharing a common ancestor recent enough so that it is still detectable in their sequences. It should be noted that for nucleotide sequences, any two non-homologous sequences (i.e. not related) could have a similarity up to 50% just by chance, when allowing for gaps. For amino acids, two randomly selected sequences may have similarity of 20%.

The major concern in phylogeny is choosing method, and there are several. Maximum Parsimony (MP) evaluates all possible tree combinations and picks the tree requiring the smallest number of mutations. Using unweighted parsimony, all substitutions are handled equally without any regard of how frequently they appear. This is a slow method; already with as few as ten taxa, the number of tree combinations exceeds millions, setting a limitation even with today's computers. To speed up performance with MP, special algorithms decreasing the number of potential trees have been developed. The main advantage of MP is its ability to find the absolute optimal tree.

Maximum likelihood (ML) is also a method evaluating all possible tree combinations. It calculates the likelihood for each tree as the product of the likelihood of change at each position in the sequence, and seeks the one with

the maximum likelihood. ML is the slowest method and very computer intensive, but often selected when it is critical to find the best tree structure.

UPMGA (unweighted pair-group method with arithmetic means) is based on distance matrixes and is a fast and commonly used method. It does not predict all possible trees like the parsimony or maximum likelihood methods; instead it uses a sequential clustering algorithm and a stepwise approach to find the best tree. UPMGA does not take the possibility of unequal substitution rates along different branches into account, sometimes miscalculating the true distances between taxa in the trees. It should only be used when the evolutionary rate can be assumed to be constant.

Neighbor Joining (NJ) is method that uses distance matrixes, pairwise comparison and stepwise clustering to find the best tree [Saitou and Nei, 1987]. It is a fast method, not making any assumptions on equal amount of divergence for the lineages. NJ can therefore be used for datasets comprising lineages with largely varying rates of evolution; in conjunction with an appropriate substitution model.

To evaluate the robustness of a phylogenetic tree, bootstrapping can be used [Felsenstein, 1985]. Bootstrapping is a resampling technique where a hundred to a thousand of new datasets are created based on an original multiple alignment, creating a consensus tree and with frequencies of how many times each branch is placed between nodes. Bootstrap values above 70% are generally considered good. Bootstrapping gives an idea about the “reliability” of branches and clusters, like the statistical term ”significance”.

As already mentioned, recombination disturbs phylogenetic analysis since the two parts of the recombined gene have a different phylogenetic history. To scan for recombination, a tool called Simplot has been developed to detect if a single sequence is a recombinant given an array of non-recombinant references. It uses a sliding window of a user-specified size and plots the percent identity of the query to the panel of references. Evidence of recombination detected with Simplot should always be verified with separate phylogenetic trees for each part of the recombination [Lole et al., 1999]. Bootscan is another tool of the same software package, useful for recombinant detection as it uses a sliding window over a set of sequences (references and a target sequence), plotting the bootstrap values for each partial tree of the sliding window in a diagram, where cross-over points for recombination can be identified.

In addition to Simplot, a new method called GroupScanning has recently become available. This method are by some considered to provide a more robust result because it investigates the relation between the examined strain and clusters of reference strains, whereas Simplot only compares the sequences to a generated consensus [Simmonds and Midgley, 2005].

In summary, choosing the method for phylogenetic analysis should be done with care. One should always test the robustness of the tree topology by bootstrapping or similar methods, and be aware that trees may look very different depending on the chosen phylogeny method.

4 RESULTS AND DISCUSSION

4.1 Precore C-1858 variants (Paper I)

Precore mutations preventing the expression of HBe antigen have been extensively analysed. The seroconversion of HBe to anti-HBe is usually considered an indication of successful immune response, and vice versa, the continued HBe expression has been linked to high DNA replication and a worse clinical outcome. The C-1858 variant restricts the G1896A mutation, hence potentially reducing the probability of seroconversion. C-1858 predominates in genotype A and is frequently found in genotype C and F.

To determine if the C-1858 variant in genotype C was due to a sporadically evolving mutation (T1858C), or if it represented a distinct phylogenetic entity, the complete genomes of eight genotype C strains with C-1858 and Southeast Asian origin were sequenced. These strains had been identified in previous studies investigating HBV genotypes [Lindh et al., 1997].

Phylogenetic analysis comparing these strains with sequences representing HBV genotypes and strains identified in East Asia revealed that the C-1858 variant clustered separately on the genotype C branch. The sequences, originating from Thailand, Vietnam and Malaysia, were closely related to C-1858 sequences submitted to GenBank by others.

Since the publication of our study, several reports have described subgenotypes and clades of genotype C. Chan et al. reported that genotype C could be differentiated into two main subgroups, denoted C_s and C_e representing Southeast Asia and Far East Asia [Chan et al., 2005]. Later additional subtypes within genotype C were defined and currently 10 subgenotypes have been described (C1-C10).

In order to verify that genotype C with C-1858 constitutes a phylogenetic entity also when a greater number of complete HBV genotypes have been submitted to GenBank, we have repeated the phylogenetic analysis by comparing a sequence from paper I with a large number of similar complete HBV genomes from GenBank. These sequences were obtained using the blast tool at NCBI. The 250 sequences most similar to strain ea1 from paper I (accession nr AF223954) where selected (having an identity of 100%-96%). Out of these, 15 sequences were excluded for being duplicates or having large gaps. 76 (32%) had C-1858 (of which 29 had TCC at nt 1856-58) and 159 (68%) sequences had T-1858. A circular phylogenetic tree was

constructed with the 235 sequences and representatives of subgenotypes C1-C10, rooted by a genotype B sequence (Figure 10). Almost all C-1858 strains placed on the same branch within subgenotype C1 (C_s) in the phylogenetic tree, except for four sequences, of which two (GQ377631 and GQ377605) were from a study of drug resistant HBV strains in Chinese patients presenting with acute hepatitis [Xu et al., 2010] and two (AB112348 and AB112066) were described in a genotyping report from Myanmar [Huy et al., 2004]. C-1858 was not found in sequences of subgenotypes C2-C10. The subgenotype C1 contained 61 T-1858 strains, of which 58 clustered together on one branch, and the other 3 strains were dispersed in the C-1858 flank. As described in Paper I the C-1858 branch of subgenotype C1 could be broken down into two clades defined by nt 1856-58, where almost all sequences having TCC where found in one clade (25 out of 27) and the remaining strains with CCC in another clade.

This analysis verifies the accuracy of our findings in paper I. The observation that not a single genotype C sequence with C-1858 places outside the distinct sub-branch of C1 supports that spontaneous T1858C mutations are rare or at least rarely emerge and spread within the population of HBV carriers.

By comparing the genetic distance between the C-1858 and T-1858 sub-branches of the C1 branch it would be possible to estimate the time point of their divergence. The genetic distance between two representative strains, one from the C-1858 (DQ089766) and one from the T-1858 (FJ023639) was 2,10%. Assuming molecular clock rates (see section 4.5) in the range 1×10^{-6} to 2.0×10^{-5} substitutions per site and year, the divergence of these strains would have occurred 500 to 10,000 years ago.

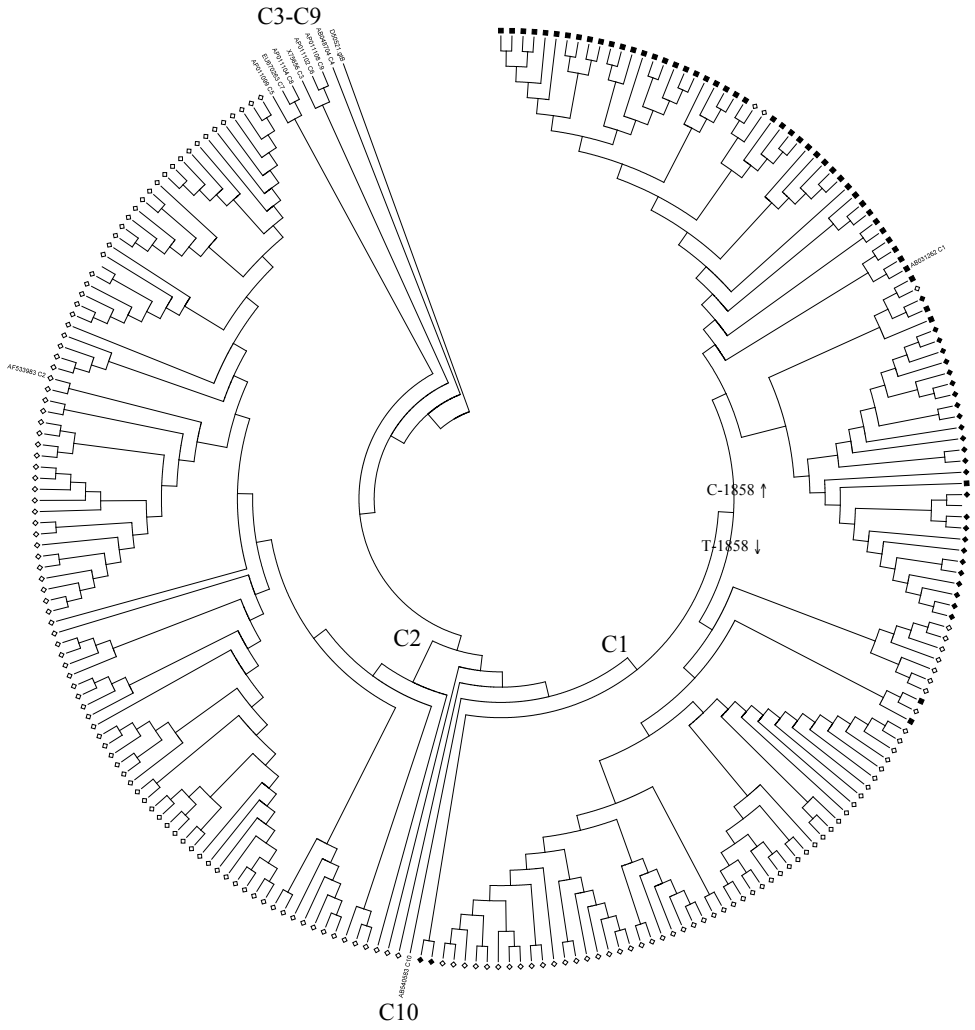

Figure 10. Phylogenetic tree created by Neighbor Joining method with Maximum Composite Likelihood model of 235 genotype C sequences similar to a C-1858 variant reported in paper I and representative strains of subgenotypes C1-C10 (C1, AB031262; C2, AF533983; C3, X75656; C4, AB048704; C5, AP011099; C6, AP011102; C7, EU670263; C8, AP011104 ; C9, AP011108; C10, AB540583). The strains with C-1858 are marked with black squares (CCC) or black diamonds (TCC), T-1858 strains are denoted with white diamonds (CCT).

4.2 Genotypes (Paper II-III)

Based on a genetic diversity of more than 8% of the entire genome, HBV have been divided into genotypes A-H, which have relatively distinct geographic distribution. It was shown early that genotypes B and C predominate in East Asia including China. A high HBsAg prevalence had been observed in Mongolia and in Australian Aborigines, but little was known of existing genotypes in HBV carriers in these regions. To explore this, we investigated the HBV genotype of strains from 9 Mongolian patients and 5 Australian Aborigines in paper II. During the preparation of the paper II, Sugauchi et al. reported a study of complete HBV genomes from 5 Australian Aborigines [Sugauchi et al., 2001]. They found that three of them were of genotype D, while two represented a novel variant of genotype C, distinct from strains of Polynesia.

Our phylogenetic analysis revealed that all sequences from Mongolian HBV carriers were of genotype D. This genotype is common in Central Asia and the Middle East, suggesting that the population in Mongolia has had a closer contact with or were more related to people in those regions than with China where genotype B and C dominates. This was also supported by a study reporting mitochondrial DNA from Central Asia to be phylogenetically closer to Mongolian than to East Asian sequences [Comas et al., 1998].

Since the publication of Paper II, additional HBV genomes have been submitted to GenBank. Four publications about HBV sequences from Mongolia were found by Pubmed search [Davaalkham et al., 2007; Elkady et al., 2008; Odgerel et al., 2006; Takahashi et al., 2004]. Genotyping data on these strains (where classification was possible) showed that 198/206 (96%) were of genotype D, 7/206 (3%) of genotype A and one strain was genotype F. Only the preS and S regions of the 9 Mongolian strains were sequenced in paper II, but phylogenetic analysis indicates that they belong to subgenotype D1. Subgenotyping is however more reliable when based on complete genomes [Schaefer et al., 2009]. To further explore the subgenotypes of Mongolian HBV strains, we performed phylogenetic analysis of 19 complete HBV genomes from the reports mentioned above (Figure 11). This showed that 12 Mongolian strains were of subgenotype D1, 4 of subgenotype D3 and one of genotype A. Two sequences appeared on the genotype C branch, but represent C/D recombinants [Elkady et al., 2008], common in Tibet [Cui et al., 2002a].

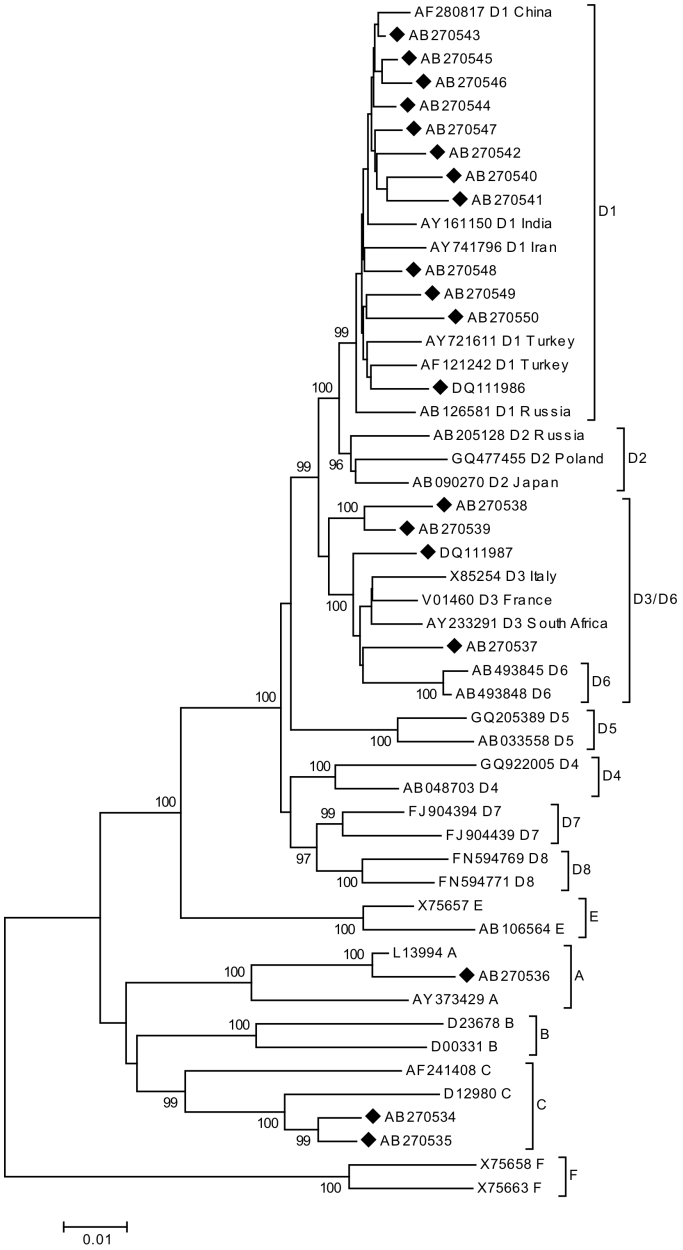

Figure 11. Phylogenetic tree created by Neighbor-Joining method with Tamura-Nei correction of 19 complete HBV genome sequences of Mongolian origin (marked with black diamonds) and representative references from common HBV genotypes.

The strains from Australian Aborigines were classified as genotype C by phylogenetic analysis of the preS region. In contrast, the S region differed from all known genotypes in the Simplot analysis, suggesting possible recombination with an unknown genotype. The sequences reported by Sugauchi, to which our strains are very similar, have been classified as subgenotype C4 based on analysis of complete genomes [Norder et al., 2004].

To our knowledge, no further reports on HBV sequences from Australian Aborigines have been published since our report. Accordingly, a BLAST search using the complete genome of the C4 sequences reported by Sugauchi et al. did not identify any additional C4 sequences in GenBank (the most similar was sequence GQ377533 from a Chinese carrier belonging to subgenotype C2). The possibility that the divergent S region of the C4 strains is a result of recombination is discussed in the next section. BLAST search was also performed with the S regions of one of our C4 sequences (AF360967), but no further C4 sequences were found: the closest hits (AB486012 and FJ715395) had a similarity of 95%; the first one was a C2 sequence of Chinese origin, the other has been assigned genotype J. BLAST analysis of the preS region of Abor2469 (AF360963) produced two C2 sequences with similarities of 97% as best hits: AF540585 from Indonesia (East Nusa Tenggara) and EU547562 from Malaysia. To conclude, no additional C4 sequences have been reported to GenBank after the submissions by Sugauchi et al. and by us 10 years ago.

The five C4 sequences from Australian Aborigines and sequences representing various C subgenotypes, including C3 from Polynesia and C6 from Hawaii [Sakurai et al., 2004] are shown in figure 12. The genetic distance between subgenotype C4 and subgenotype C2 (found in Indonesia), measured as group mean in MEGA (substitution model Tamura-Nei) was 5.32%.

The fact that distinct subgenotypes have been identified on islands in the Pacific (C3 in Polynesia, C4 in Australia, C5 in the Philippines, C6 on Hawaii) might be useful for understanding the evolution and molecular clock of HBV. This is discussed in section 4.6.

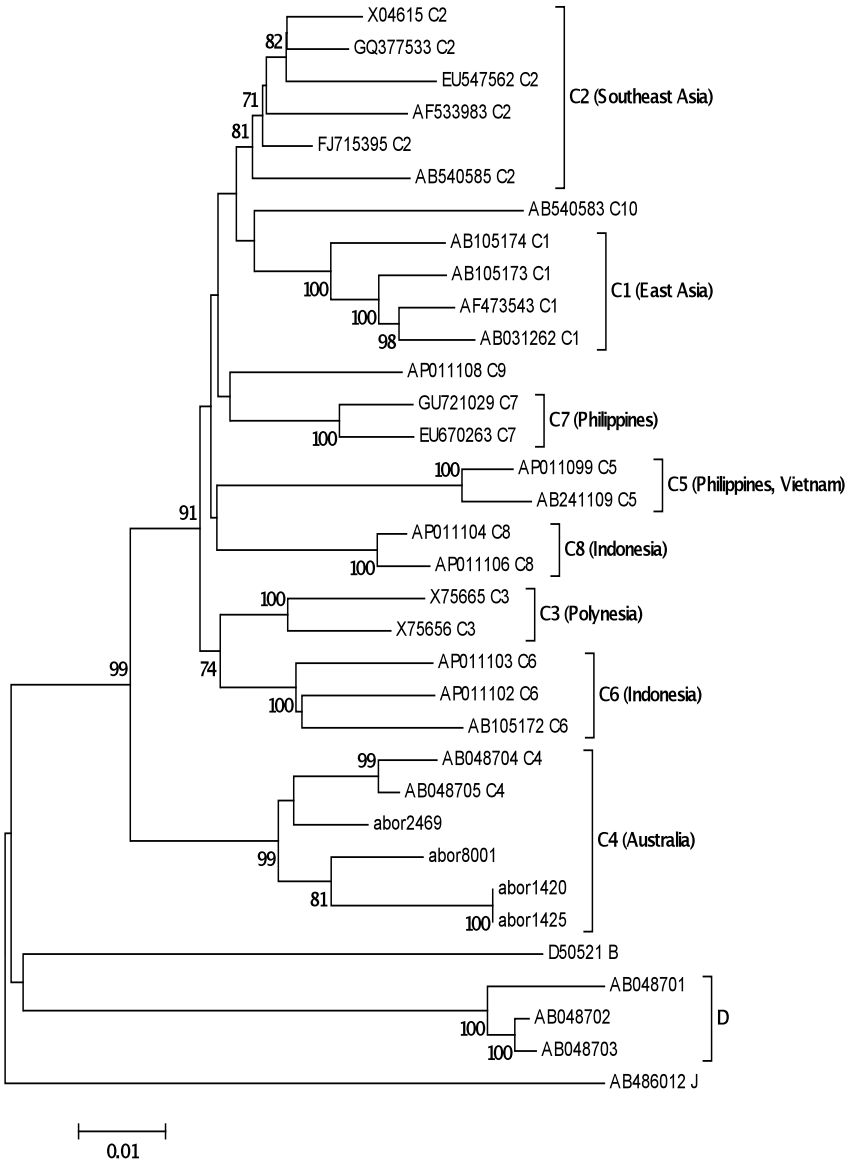

Figure 12. Phylogenetic tree of 609 nt of preS and S regions of HBV from Aborigines and representative strains from GenBank (Neighbor-Joining with model Maximum Composite Likelihood, pairwise deletion of gaps).

In paper III we explored HBV genotypes in Vietnam, a country with an HBV prevalence of more than 10% and a heavy burden of HBV complications. The aim was to investigate the relative frequencies of genotypes and subgenotypes, because this may be of importance for the risk of developing progressive liver damage and for response to treatment.

The S region of HBV from 87 samples were sequenced and phylogenetically compared to reference sequences, showing that 67 (77%) were of genotype B and 19 (22%) of genotype C. All genotype C sequences were of subtype C1, as previously described by others for HBV from that region. For genotype B, the bootstrap values for the S region phylogeny were not sufficient for reliable subgenotype classification. Instead, nucleotide differences at position 499 were assessed, as cytosine at this position is almost exclusively found in published sequences of subgenotype B4. Three sequences with A-499 where classified as subgenotype B2, and the remaining 16 with C-499 as subgenotype B4. No strains of the genotype C variant that we previously had identified in patients from Vietnam [Hannoun et al., 2000b] and later was assigned subgenotype C5 [Sakamoto et al., 2006], were found. One patient carried the recombinant [Hannoun et al., 2000b] that was later assigned genotype I [Tran et al., 2008], see below.

These findings agree with a previous study of 38 HBV carriers from Hanoi, in which genotype B was observed in 79%, genotype C in 18% and B/C co-infection in 3% [Thuy le et al., 2005]. The distribution was different in another study from Ho Chi Minh City in which only 43% were genotype B and 57% genotype C [Tran et al., 2003]. The higher rate of genotype C might to some extent reflect that most of the study patients were recruited from a unit for liver disease and that genotype C has been associated with more severe liver damage.

4.3 Recombination (Paper I-III)

No recombination was found in the C-1858 strains of genotype C sequenced by us in paper I. However, we found evidence of recombination with genotype B between nt 1260 and 1639 in one sequence previously submitted to GenBank (AB031265), which was used as reference.

In paper III we identified one patient carrying an X/C recombinant. This variant was identified 10 years ago in HBV carriers of northern Vietnamese origin residing in Sweden [Hannoun et al., 2000b]. Strains with the same recombination has later been identified in other Vietnamese carriers and suggested to be assigned genotype I [Tran et al., 2008]. It has also been found in Laos, and separated into 2 subgroups [Olinger et al., 2008], indicating that this strain may be widely spread in this region. We found this strain in one of 87 HBV carriers from Hanoi, Vietnam, suggesting that it is relatively rare but still may account for ≈ 1% of HBV infections in the area. Very recently it was reported that the same X/C recombinant is even more frequent (13%) in the Guangxi region in southern China, from where it may originate [Fang et al., 2010].

In Paper II, we could only, due to the limited amount of specimen, obtain sequences for preS and S regions of the C4 strains from the Australian Aborigines. Our analysis indicated potential recombination involving the S region. To further investigate this we have performed additional analyses of the complete C4 genomes submitted by Sugauchi et al., which were highly similar in the preS and S to the sequences investigated in paper II. A Simplot analysis of the consensus sequence of these complete genomes compared with representatives of HBV genotypes is shown in figure 13. Overall, a close relationship to genotype C is indicated (red line) for these strains. For the S gene (i.e. nucleotides 150-800), genotype J is more similar than any of the other genotypes. Possibly, an ancestor of the newly reported genotype J has been involved in a distant recombination with an ancestor of the Aborigine C4 strains described in Paper II. Bootscan analysis (Figure 14) gives some support for this suggestion.

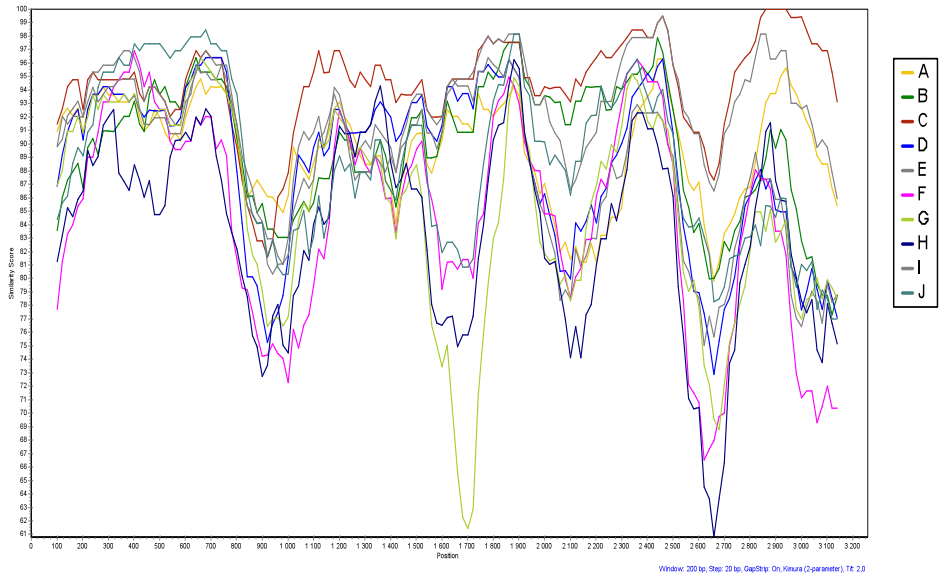

Figure 13. Simplot analysis of complete genomes of genotype C variant found in Australian aborigines sequenced by Sugautchi et al., compared to references representing genotypes A-J.

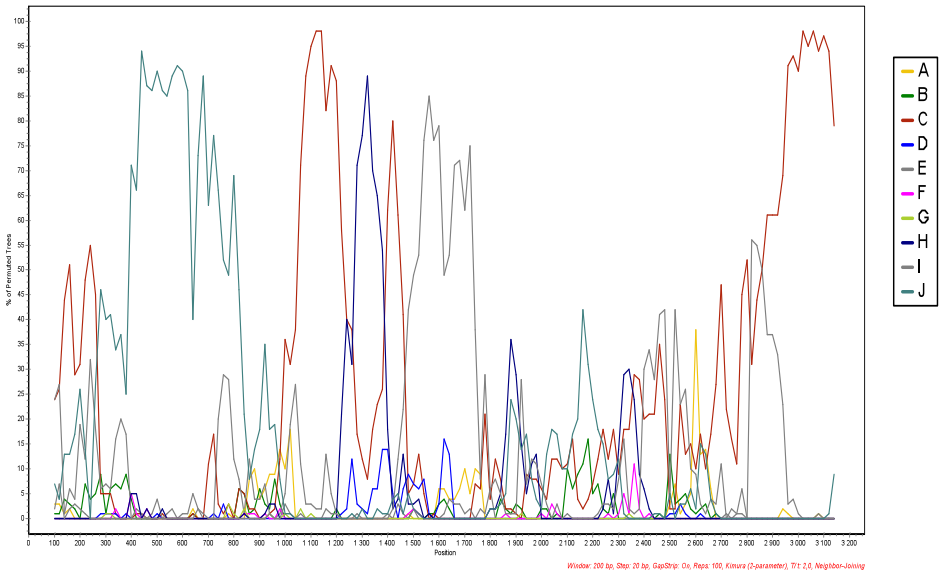

Figure 14. Bootscan analysis of complete genomes of genotype C variant found in Australian aborigines sequenced by Sugautchi et al., compared to references representing genotypes A-J.

A study comprising all published complete genome HBV sequences in 2004, including all human and ape HBV variants, identified 24 intergenotypic recombinant forms [Simmonds and Midgley, 2005]. The authors pointed out the difficulty in establishing recombination events when ancestral strains are missing (not yet discovered or have become extinct), like the case for the X/C recombinant reported in this thesis, and possibly also for the C4 strains. They also pointed out that some locations in the HBV genome probably are more favoured for recombinations, mainly the distal part of S, the start of pre-S2 and both ends of the core gene [Simmonds and Midgley, 2005].

4.4 Molecular epidemiology (Paper IV)

In paper IV we analysed sequences of HBV subgenotype D3 identified in 28 patients presenting with acute hepatitis B in Gothenburg, during three different time periods (1975, 1995-1996 and 2006-2007). The purpose was to study how the HBV strains in this group of patients changed over time and to evaluate molecular epidemiology strategies.

Several important observations were made. There was a low genetic diversity within the investigated strains (0.41% by complete genome comparison). A low genetic variability of D3 HBV strains has been observed also by others, in IDUs as well as other settings [Christensen et al., 2001; Kidd-Ljunggren et al., 1999; Panessa et al., 2009; Webster et al., 2000]. This indicates that HBV strains circulating in IDU communities may be genetically homogenous, with a lack of input of new strains. The latter may seem surprising, since sharing needles with drug users living elsewhere probably is rather frequent. However, such imports of new strains probably rarely become established because acute infections are time limited and therefore chains of transmission are likely to stop. Still, persistent infection evolves in approximately 1% of cases, and we indeed identified several IDUs with chronic HBV infection, and moreover, they carried strains that were similar to those found in acute cases. It is possible or even likely that the repeated outbreaks that occur in IDU populations originate from such chronic carriers, and appear when a sufficient number of susceptible persons have joined the IDU community. This model, i.e. the possibility that most acute HBV infections originate from a few chronic carriers may also explain the high degree of similarity between the analysed sequences, even over long time spans.

In our patients, 3 different variants or clades of D3 were found. These clades could be distinguished by comparison of the whole genome, but also by the variability at 3 positions within the *a* determinant. The only clade observed in 1975 had M125/S136/S143, which was seen in most of the reference sequences in GenBank, including V01460. Two different clades appeared in 1995-1996. One clade had M125/S136/L143, and was related to the AJ131956 sequence from German nosocomial outbreak starting 1986 and was causing a large number of acute HBV infections in Denmark in 1994 [Fisker et al., 2004]; the other had T125/Y136/S143 and was not represented by any reference sequence. In patients presenting in 2007-2008, only the first and third of these clades appeared. These observations indicate that at least three similar D3 clades have circulated in our population of IDUs and that at least two of them have been observed elsewhere and may be spread globally.

The high degree of similarity between the HBV-D3 strains circulating in IDUs has implications for molecular epidemiology investigations. Such investigations are performed to clarify sources of outbreaks and routes of transmission. From the results one may conclude that sequences are probably not related if they differ substantially (e.g. if they belong to different genotypes or subgenotypes). If sequences are very similar the interpretation might be that they were probably related. However, our results demonstrate that such a conclusion may be uncertain if D3 sequences are involved, because such sequences can be almost identical even if they are not (recently) related. For example, HBV of P11 from 1975 and P27 from 2008 differed by only 0.35% (11 nt) when whole genomes were compared, and by 0.15% (1 nt) in the HBsAg region. These findings suggest that choosing the best region may be of great importance. This issue has recently been debated, but there is no consensus regarding which region of the HBV genome that is best [Datta et al., 2007]. The complete genome analysis can be considered the gold standard, but this is not feasible. The core region has been suggested [Boot et al., 2008], but our data indicate that this is not the best choice, at least not for D3 strains, because almost no variability was seen.

Intuitively, the best choice should be the most variable region. Diversity analysis comparing the 10 complete HBV genomes with the reference strain V01460 identified the segment nt 832-1323, residing in a non-overlapping part of the genome, as the most variable. However, a phylogenetic tree based on this part of (Figure 15) provided a poorer resolution as compared with an S region tree. Sequencing of the S region has been proposed by others [Datta et al., 2006] and this choice is supported by our data, at least for D3 sequences, because it provided the most accurate phylogenetic tree compared to the tree of the complete genome sequences. This finding is interesting as substitutions

in this region are constrained by the overlap of the polymerase gene. Despite this, the sequence differences in S are sufficient for identification of genotype and subgenotype (not always). For D3, S region analysis could also separate clades, and this may apply also for other subgenotypes. One report from England found subgenotype A2 to prevail in IDUs, but clades were not discussed, as the number of studied patients was low [Sloan et al., 2009], so this issue requires further study.

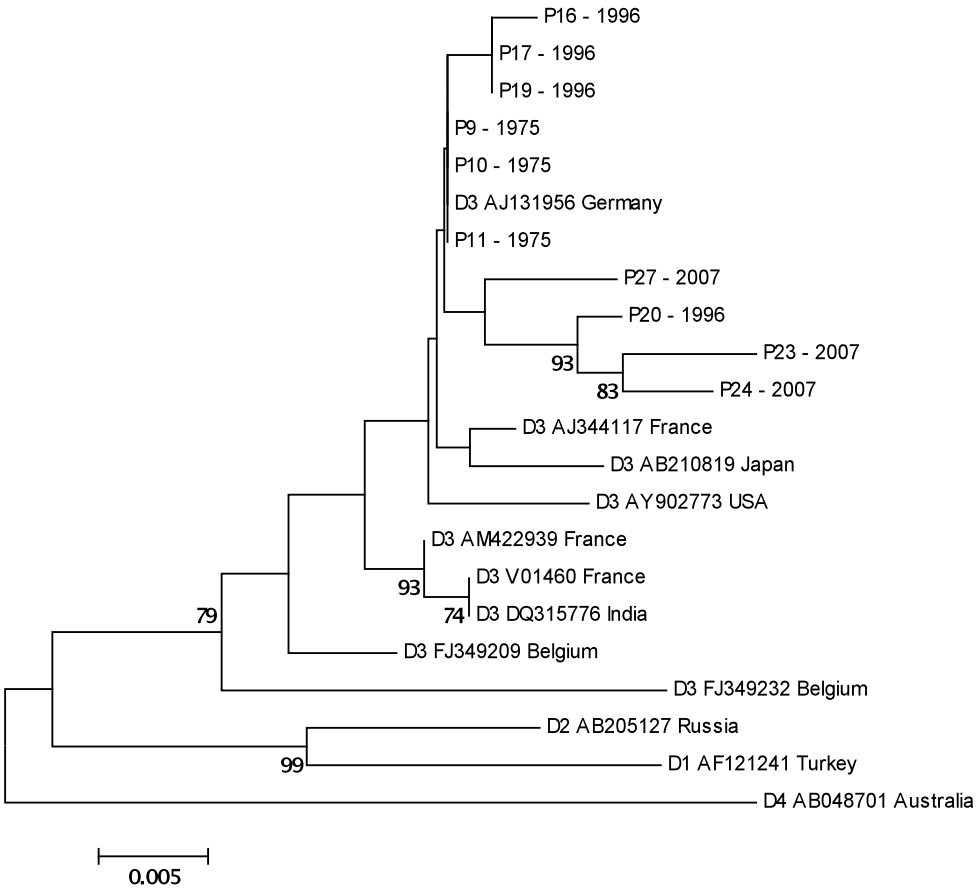

Figure 15. Phylogenetic tree of nt 832-1323 of 10 complete HBV genotype D3 genomes of paper IV (Neighbor-Joining tree with Tamura-Nei correction)

Our findings in paper IV indicate that analysis of a few signature residues in the S region is sufficient for separating the strains into clades (table 4). Key amino acids in the S region for IDU have been pointed out before. For example, the T126M substitution in HBsAg was reported to be characteristic

for subgenotype D3 strains circulating among IDUs in Europe [Norder et al., 2004]. This was seen in all our patients from the first period, 7 of 9 patients from the second, but only in 2 of 7 patients from the last period.

Table 4. Signature residues of the S gene of HBV genotype D3 strains for each period of time.

Period	Number of strains with amino acid variant at 125-136-143		
	M-S-S	M-S-L	T-Y-S
1975	10	0	0
1995-1996	0	6	2
2007-2008	2	0	5

In conclusion, the S region has both nucleotide variations useful for phylogenetic classification and amino acid signatures of importance, making it a preferred region for primary genomic analysis in molecular epidemiology. For patients carrying the same clade further investigation may or may not clarify if they are associated. The variation in the S region can be sufficient, but one has to keep in mind that even sequences with identical S regions may not be directly related. Conversely, patients with a few nt differences may still be closely related. These interpretation difficulties can be illustrated by the sequences from the clade with M125/S136/L143. In this group of strains (which all in some sense probably were related), the P20 complete genome from 1996 differed by 11 nt (0.35%) from each of P23 and P24 from 2007, but the difference between P23 and P24 was in the same range (8 nt, 0.25%). Thus, in such cases even complete genome comparison does not have enough resolution to reliably clarify epidemiological links.

4.5 Setting the molecular clock

Strictly, the molecular clock concept requires that the substitution rate is constant and that the rate depend on absolute time rather than generation time. If the rate of substitutions were variable, the molecular clock hypothesis would be rejected. Therefore, accurate molecular clock calculations for HBV may not be possible [Simmonds, 2001]. Still, estimations might be very useful for understanding the HBV evolution. It could for example serve to estimate the time points for events in the history of HBV, such as the separation of genotypes.

Two different strategies may be used for setting the clock. One is to analyse mutation rates in viruses from infected individuals, preferably in a longitudinal fashion. Another is to use knowledge about human migration, and calculate the mutation rate on the basis of sequence differences between viruses found in populations that separated at relatively well-known time points.

The first strategy was applied in a study that compared HBV clones from a 54 year-old Japanese woman who was presumed to be neonatally infected. Okamoto et al. thus estimated the molecular clock to 1.4 to 3.2 x 10^{-5} nucleotide substitutions per site per year [Okamoto et al., 1987a]. A similar value, 5 x 10^{-5} per site per year, was obtained by Orito et al. by calculating the rate of only silent substitutions in the P region [Orito et al., 1989]. Hannoun et al. compared complete genomes from mothers and their vertically infected adult children [Hannoun et al., 2000a] and found that the mutation rate was different in HBeAg positive and HBeAg negative patients, as also had been observed by others [Akarca and Lok, 1995]. Sequences from HBeAg positive patients without ALT elevations showed a mean of 1.6 mutations in the whole genome after an observation time of 24.4 years (mutation rate 2.0 x 10^{-5} substitutions per site and year), compared to sequences from HBeAg negative carriers who had a mean difference of 20 mutations after 27.6 years (mutation rate 2.3 x 10^{-4} substitutions per site and year). Further efforts to determine the longitudinal changes of HBV was done by Osiowy et al. who estimated the molecular clock to 7.9 x 10^{-5} substitutions per site per year by analysing paired samples with a time span of 25 years from treatment naïve, asymptomatic HBeAg-negative carriers. These patients were considered to be under a lower immune pressure compared with HBeAg-positive patients, and the authors argued that their estimate would better reflect the true mutation rate of HBV, resulting from errors of the viral polymerase and the selective outgrowths of fit variants [Osiowy et al., 2006]. An objection to this approach

may be that older patients are less likely to spread the infection, at least vertically, and that the mutation rate in HBeAg negative infections is not relevant for the evolution of HBV, because transmission from such cases rarely, if ever, results in chronic infection. It seems more likely that HBV from HBeAg positive women in tolerance phase, is involved in the propagation of HBV into the next generation, because vertical transmission from them is likely to result in chronic infection in the child. Still, the low mutation rate reported by Osiowy et al. is interesting because it supports that the mutation rate of HBV varies over the course of infection: low in the early *and* late phases, and high around the time of seroconversion. In a different type of study comparing the change of HBV over a shorter time span in an Italian population, of whom many presented with acute hepatitis, different mutation rates were observed for genotype A and D (0.9×10^{-4} and 3.67×10^{-4}) [Zehender et al., 2008], indicating that rates may vary between genotypes. The mutations rate has been estimated also for other hepadna viruses. For example, the rate in avian hepadnaviruses was calculated to be 0.8 to 4.5 x 10^{-5} per nucleotide per generation [Pult et al., 2001].

The other strategy has not been systematically applied for investigating the molecular clock of HBV. Table 5 shows some examples that might be utilised for this purpose. The calculations can be made either on the basis of the genetic distance between the types, e.g. between C4 and other C subgenotypes, or on the basis of the genetic distances within the type of interest, such as within European A2 sequences.

The genetic distances between C4 sequences and the other C subgenotypes are in the range 6.7% and 8.9% (pres and S, distance mean between groups). Archaeological evidence suggests that landmasses near Australia were settled by humans at least 40,000 years ago, when the sea levels were lower and the present Australia, Tasmania, and New Guinea was a single landmass known as the Sahul. Recent studies on human genome diversity of Australian Aborigines suggest that Australia was settled relatively soon after humans left Africa, and that there is a large human genome distance between Australian Aborigines also and surrounding islands [McEvoy et al., 2010]. These data suggest that the earliest separation between C4 and other C subtypes could be 40,000 years ago, which would match with a molecular clock rate of 1×10^{-6}. This is slower than the observed rates, but still almost in the same range. If the spread of the C4 ancestor occurred later than the first settlement of Australia, the matching with observed rates would be better. Still, a substitution rate of 2×10^{-5} corresponds to a separation of subgenotype C4 from the Asian and Polynesian C subgenotypes around 1000 years ago, a dating that intuitively seems much too late.

The genetic distance within European A2 sequences is 0.009. If the time point for spread was 500 years ago, this would correspond to 9×10^{-6} substitutions per site per year. However, A2 in Europe might have been transmitted later in life and changed relatively more than HBV genotypes in other regions, where HBV spreads vertically. Thus, even lower substitution rates may be expected in areas where vertical transmission is well established.

The mean distance between B6 sequences and the node separating B6 from B1 is 0.03 and the variation within B6 is 0.038. If the Eskimo ancestors crossed the Bering Strain 10,000 years ago and brought HBV-B along, the calculated mutation rate would be approximately $1.9\text{-}3 \times 10^{-6}$ per site per year.

Amerindian ancestors probably crossed Bering Strait earlier, probably 15,000 years ago. If they brought the ancestor of genotypes F and H along, a similar calculation as above would result in a slightly faster molecular clock.

In summary, available data indicate that the mutation rate of HBV is higher than for other DNA viruses and slower than for retroviruses [Mizokami and Orito, 1999]. The mentioned ways of estimating the molecular clock results in slower clock rates than calculations based on observed substitutions in patients, but it may be that estimates based on observed mutations gives too high rates, because the requirement to induce chronic infection over successive generations may constrain the evolution more than we imagine. In any case, the rates are fairly compatible and support that HBV may have been present in humans since the time of their exit from Africa.

Table 5. Molecular clock calculations.

	Putative event of spread	Years since event	Genetic distance from node separating (sub)genotype	Calculated mutation rate	Variability within (sub)genotype	Calculated mutation rate
Separation of A2	European sailors arriving in West Africa	500	NA	NA	0.009	9×10^{-6}
Separation of B1 and B6	Eskimo ancestors crossing the Bering Strait	10,000	0.03	3×10^{-6}	Max 0.038 (18 taxa B6)	$1{,}9 \times 10^{-6}$
Separation of C4	Aborigine ancestors arriving in Australia	40,000	0.04	1×10^{-6}	0.0091 (Only 2 taxa C4)	1.1×10^{-7}
Separation of F/H	Migration of Amerindian ancestors across Bering Strait	15,000	0.091	6.1×10^{-6}	Max 0.103 (23 taxa F, 22 taxa H)	6.9×10^{-6}

4.6 Evolution of genotypes

The origin of HBV is unclear. How and when the genotypes evolved and spread over the continents in humans and other species is not known. Several theories have been proposed, with both recent and distant emergence. The development of genotypes of HBV could be explained by:

1) *Co-species evolution*, meaning that all HBV strains share a very ancient common ancestor, before the split up of species into mammals and avians. Since this point in time, HBV has developed within each species, in parallel, resulting in different species-specific subtypes and genotypes in humans. This fits with the species-specific HBV forms in heron, duck and woodchuck. It is more difficult to understand and explain the difference between HBV in humans and in apes. The finding that HBV in apes are quite specific for the respective species (chimps, gorillas, orangutans and gibbons) suggests that HBV may have co-evolved with the hosts. If so, the phylogenetic trees of HBV and humans/apes should look similar. This cannot be judged, because the HBV tree does not have a sufficient resolution as regards the early separations. A main objection to the theory of co-species evolution is the fact that humans and apes separated 7-10 million years ago, and therefore this model would require a *much* slower molecular clock than has been observed.

The observation that HBV in apes and human genotypes A-E are more similar than are human genotypes F/H and genotypes A-E also speak against co-species evolution [Bollyky and Holmes, 1999; Fares and Holmes, 2002; MacDonald et al., 2000]. This puzzling finding might be explained if the mutation rate in HBV-F/H of some reason was higher in Amerindians. An alternative explanation might be that the ancestor of F/H was not brought to America by humans, but by an unidentified monkey species. This is however speculative, in particular as HBV has only been found in one monkey species, in the form of the much more distant woolly monkey HBV [Fares and Holmes, 2002]. Interestingly, there is a lack of evidence of HBV in Old World monkey species [Eichberg and Kalter, 1980; Michaels et al., 1996]. Further study of HBV infections in apes should be encouraged and might possibly serve to explain the evolution of human and ape HBV types.

2) Spread by *early human migration*, as discussed in the previous section. HBV genotypes may have evolved as a result of migration to separated geographic areas by humans carrying HBV (10,000-40,000 years ago). An example of such putative early spread of HBV is the hypothesis that subgenotype B6 evolved as a result of migration of Eskimo ancestors from East Asia [McMahon et al., 1998]. This idea is supported by the knowledge that

an ice or land connection over the Bering strait has joined present-day Alaska and eastern Siberia during several periods 10,000 to 25,000 years ago [O'Rourke and Raff, 2010].

In this model the spread of HBV to apes is explained by cross-species transmission. This idea might to some extent be supported by phylogenetic analyses indicating that HBV in apes are linked to geographical origin rather than to species [Starkman et al., 2003]. Possibly, the putative genotype J, which in phylogenetic trees based on whole genomes places closer to Gibbon and Orangutan HBV than to known human genotypes, might be a result of previous cross-species transmission of HBV [Tatematsu et al., 2009].
An objection to the cross-species transfer theory may be the simple fact that the distance to human HBV is the same for the four different ape species HBV types, which seems a bit remarkable as it implies that transmission to all these species would have occurred at the same time. As discussed above this model requires a 10 times slower molecular clock than the observed mutations rates.

3) Spread by *later human migration*. Divergence of HBV has occurred by migration of infected persons into already populated areas. This model fits better with observed mutation rates. The link between human migration and evolution of HBV genotype D was discussed by Jazayeri et al. [Jazayeri et al., 2010]. The most extreme theory was the proposal that HBV originated from the Americas, and spread into the Old World as recently as 400 years ago [Bollyky and Holmes, 1999].

5 CONCLUSION

HBV strains with C-1858 variant constitute a clade within genotype C, and likewise the CCC and TCC variants at precore codon 15 represent phylogenetic entities. Spontaneous T1858C mutations are rare. Subsequent analyses show that the C-1858 variant is characteristic for subtype C1 and not present in any other C subgenotypes.

Genotype D prevails in HBV carriers from Mongolia. Subsequent analyses show that D1 is the most common subgenotype.

Australian Aborigines carry genotype C of subgenotype C4 which has an S region that appears to be of non-C origin and possibly a result of a previous recombination with an unknown genotype.

In northern Vietnam genotype B was found in 77% of HBV carriers, with subgenotype B4 being most common, while genotype C of subgenotype C1 was found in 22% and the X/C recombinant in around 1%.

The HBV-D3 strains causing acute hepatitis B in Swedish IDUs has changed at a low rate since the 1970ies, but three clades can be distinguished. The S region appears to be favourable as primary target for subgenomic molecular epidemiology of HBV-D3.

The common ancestor for the genotypes of human HBV would be more than a million years old if one assumes co-species evolution. However, observed mutation rates and calculations based on genotype differences and human migration dates suggest that the HBV genome changes by a rate of $\approx 2 \times 10^{-5}$ to 10^{-6} substitutions per site per year. Based on such rates, the genotypes would have separated 2,000 – 40,000 years ago, while subgenotypes and the X/C recombinant would have evolved 1,000 – 20,000 years ago.

ACKNOWLEDGEMENT

I would like to thank all those who have supported me, in particular:

Magnus Lindh, my supervisor and friend. He has inspired me since I first entered Virologen as a medical school student some years ago. With his brilliant mind and expertise in the field, he has been an excellent tutor. We have had many interesting discussions and he has always, with a warm and welcoming attitude, helped me with all sorts of problems.

Ann-Sofie Tylö, *Annkatrin Gillbro* for skilful technical expertise and for always spreading a positive attitude around. They introduced me into the technique of PCR and helped me with a lot of the laboratory work.

Sebastian Malmström, my roommate for support and great company during many years.

Charles Hannoun, co-author, for sharing his knowledge and supporting me on any issue related to hepatitis B.

Peter Horal, Ann Söderström, *Phung Thi Bich Thuy*, *Nguyen Thanh Liem*, and *Gunnar Norkrans*, my co-authors for their contributions.

Tomas Bergström and *Örjan Strannegård* for a stimulating research facility and *Staffan Görander, Kristoffer Hellstrand*, *Jan-Åke Liljekvist*, *Bo Svennerholm*, *Sigvard Olofsson*, *Lars-Magnus Andersson*, *Johan Westin* for nice discussions and good laughs, particular at lunch time.

Gaby Helbok and *Sabina Wagner* for kindly helping me on all administrative tasks.

Martin Lagging, especially for the fruitful discussions on hepatitis C.

Nancy Nenonen for advice on PCR optimization.

Peter Norberg for helping me with phylogenetic problems.

Gunnar Byröd, at the clinic of Internal Medicine at Sahlgrenska Hospital in Mölndal for making it possible to combine clinical work and research.

PhD students and other scientists at the third floor of Virologen (*Joanna, Anette, Carolina, Maria, Nina, Ulrika, Rickard, Mona, Beata, Edward,*

Galia, Jonas, Lars, Elin, Anna) and colleagues, doing their residency at the clinic of Infectious Diseases, periodically sitting next to my desk: *Anders, Anna, Arash, Karin, Lisa, Martina, Tomas.*

Colleagues at the clinic of Internal Medicine at Sahlgrenska Hospital in Mölndal for being there when I wasn't (i.e. helping me out with patients and duty).

Steven Shev, for friendship and understanding for a long period when I was not available.

My dentist *Ulf* with staff, for an extremely quick appointment close to the deadline.

Robin Brittain-Long, for ensuring that anything is possible.

My *father*, *mother*, *sisters* and *brother* for the love and support they have given.

My dearest *Anna* for love and understanding during this hectic period, she supported me in many ways, I owe her a lot.

To anyone omitted I sincerely express my gratitude, thank you!

This work was supported by grants from The Göteborg Medical Society, Swedish Cancer Foundation and the Swedish Medical Association.

REFERENCES

Abdou Chekaraou M, Brichler S, Mansour W, Le Gal F, Garba A, Deny P, Gordien E. 2010. A novel hepatitis B virus (HBV) subgenotype D (D8) strain, resulting from recombination between genotypes D and E, is circulating in Niger along with HBV/E strains. J Gen Virol 91(Pt 6):1609-1620.

Akarca US, Lok AS. 1995. Naturally occurring hepatitis B virus core gene mutations. Hepatology 22(1):50-60.

Andre T, Balosso J, Louvet C, Hannoun L, Houry S, Huguier M, Colonna M, Lotz JP, De Gramont A, Bellaiche A, Parc R, Touboul E, Izrael V. 2000. Combined radiotherapy and chemotherapy (cisplatin and 5-fluorouracil) as palliative treatment for localized unresectable or adjuvant treatment for resected pancreatic adenocarcinoma: results of a feasibility study. Int J Radiat Oncol Biol Phys 46(4):903-911.

Arauz-Ruiz P, Norder H, Robertson BH, Magnius LO. 2002. Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. J Gen Virol 83(Pt 8):2059-2073.

Banerjee A, Kurbanov F, Datta S, Chandra PK, Tanaka Y, Mizokami M, Chakravarty R. 2006. Phylogenetic relatedness and genetic diversity of hepatitis B virus isolates in Eastern India. J Med Virol 78(9):1164-1174.

Beasley RP, Hwang LY, Lin CC, Chien CS. 1981. Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22 707 men in Taiwan. Lancet 2(8256):1129-1133.

Blackberg J, Braconier JH, Widell A, Kidd-Ljunggren K. 2000. Long-term outcome of acute hepatitis B and C in an outbreak of hepatitis in 1969-72. Eur J Clin Microbiol Infect Dis 19(1):21-26.

Blackberg J, Kidd-Ljunggren K. 2000. Occult hepatitis B virus after acute self-limited infection persisting for 30 years without sequence variation. J Hepatol 33(6):992-997.

Blumberg BS, Alter HJ, Visnich S. 1965. A "New" Antigen in Leukemia Sera. JAMA 191:541-546.

Bollyky PL, Holmes EC. 1999. Reconstructing the complex evolutionary history of hepatitis B virus. J Mol Evol 49(1):130-141.

Bollyky PL, Rambaut A, Harvey PH, Holmes EC. 1996. Recombination between sequences of hepatitis B virus from different genotypes. J Mol Evol 42(2):97-102.

Boot HJ, Cremer J, Koedijk FD, van Ballegooijen WM, Op de Coul EL. 2008. Improved tracing of hepatitis B virus transmission chains by phylogenetic analysis based on C region sequences. J Med Virol 80(2):233-241.

Bowyer SM, Sim JG. 2000. Relationships within and between genotypes of hepatitis B virus at points across the genome: footprints of recombination in certain isolates. J Gen Virol 81(Pt 2):379-392.

Bowyer SM, van Staden L, Kew MC, Sim JG. 1997. A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa. J Gen Virol 78 (Pt 7):1719-1729.

Buckwold VE, Xu Z, Chen M, Yen TS, Ou JH. 1996. Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication. J Virol 70(9):5845-5851.

Buynak EB, Roehm RR, Tytell AA, Bertland AU, 2nd, Lampson GP, Hilleman MR. 1976. Vaccine against human hepatitis B. JAMA 235(26):2832-2834.

Cavinta L, Sun J, May A, Yin J, von Meltzer M, Radtke M, Barzaga NG, Cao G, Schaefer S. 2009. A new isolate of hepatitis B virus from the Philippines possibly representing a new subgenotype C6. J Med Virol 81(6):983-987.

Chan HL, Hui AY, Wong ML, Tse AM, Hung LC, Wong VW, Sung JJ. 2004. Genotype C hepatitis B virus infection is associated with an increased risk of hepatocellular carcinoma. Gut 53(10):1494-1498.

Chan HL, Hussain M, Lok AS. 1999. Different hepatitis B virus genotypes are associated with different mutations in the core promoter and precore regions during hepatitis B e antigen seroconversion. Hepatology 29(3):976-984.

Chan HL, Tsang SW, Liew CT, Tse CH, Wong ML, Ching JY, Leung NW, Tam JS, Sung JJ. 2002a. Viral genotype and hepatitis B virus DNA levels are correlated with histological liver damage in HBeAg-negative chronic hepatitis B virus infection. Am J Gastroenterol 97(2):406-412.

Chan HL, Tsang SW, Wong ML, Tse CH, Leung NW, Chan FK, Sung JJ. 2002b. Genotype B hepatitis B virus is associated with severe icteric flare-up of chronic hepatitis B virus infection in Hong Kong. Am J Gastroenterol 97(10):2629-2633.

Chan HL, Tse CH, Ng EY, Leung KS, Lee KH, Tsui SK, Sung JJ. 2006. Phylogenetic, virological, and clinical characteristics of genotype C hepatitis B virus with TCC at codon 15 of the precore region. J Clin Microbiol 44(3):681-687.

Chan HL, Tsui SK, Tse CH, Ng EY, Au TC, Yuen L, Bartholomeusz A, Leung KS, Lee KH, Locarnini S, Sung JJ. 2005. Epidemiological and virological characteristics of 2 subgroups of hepatitis B virus genotype C. J Infect Dis 191(12):2022-2032.

Chan HL, Wong ML, Hui AY, Hung LC, Chan FK, Sung JJ. 2003. Hepatitis B virus genotype C takes a more aggressive disease course than hepatitis B virus genotype B in hepatitis B e antigen-positive patients. J Clin Microbiol 41(3):1277-1279.

Chandra PK, Biswas A, Datta S, Banerjee A, Panigrahi R, Chakrabarti S, De BK, Chakravarty R. 2009. Subgenotypes of hepatitis B virus genotype D (D1, D2, D3 and D5) in India: differential pattern of mutations, liver injury and occult HBV infection. J Viral Hepat 16(10):749-756.

Chang MH. 2010. Breakthrough HBV infection in vaccinated children in Taiwan: surveillance for HBV mutants. Antivir Ther 15(3 Pt B):463-469.

Chang MH, Chen TH, Hsu HM, Wu TC, Kong MS, Liang DC, Ni YH, Chen CJ, Chen DS. 2005. Prevention of hepatocellular carcinoma by universal vaccination against hepatitis B virus: the effect and problems. Clin Cancer Res 11(21):7953-7957.

Chen CJ, Yang HI, Su J, Jen CL, You SL, Lu SN, Huang GT, Iloeje UH. 2006. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. JAMA 295(1):65-73.

Chen HL, Chang MH, Ni YH, Hsu HY, Lee PI, Lee CY, Chen DS. 1996. Seroepidemiology of hepatitis B virus infection in children: Ten years of mass vaccination in Taiwan. JAMA 276(11):906-908.

Christensen PB, Krarup HB, Niesters HG, Norder H, Schaffalitzky de Muckadell OB, Jeune B, Georgsen J. 2001. Outbreak of Hepatitis B among injecting drug users in Denmark. J Clin Virol 22(1):133-141.

Chu CJ, Hussain M, Lok AS. 2002. Hepatitis B virus genotype B is associated with earlier HBeAg seroconversion compared with hepatitis B virus genotype C. Gastroenterology 122(7):1756-1762.

Chu CM, Liaw YF. 2005. Genotype C hepatitis B virus infection is associated with a higher risk of reactivation of hepatitis B and progression to cirrhosis than genotype B: a longitudinal study of hepatitis B e antigen-positive patients with normal aminotransferase levels at baseline. J Hepatol 43(3):411-417.

Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Martinez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D, Pettener D, Bertranpetit J. 1998. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. Am J Hum Genet 63(6):1824-1838.

Cui C, Shi J, Hui L, Xi H, Zhuoma, Quni, Tsedan, Hu G. 2002a. The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. J Gen Virol 83(Pt 11):2773-2777.

Cui X, Hong L, Lin X. 2002b. Electrocatalytic oxidation of hydroxylamine on glassy carbon electrodes modified by hybrid copper-cobalt hexacyanoferrate films. Anal Sci 18(5):543-547.

Custer B, Sullivan SD, Hazlet TK, Iloeje U, Veenstra DL, Kowdley KV. 2004. Global epidemiology of hepatitis B virus. J Clin Gastroenterol 38(10 Suppl 3):S158-168.

Datta S, Banerjee A, Chandra PK, Chakravarty R. 2007. Selecting a genetic region for molecular analysis of hepatitis B virus transmission. J Clin Microbiol 45(2):687; author reply 688.

Datta S, Banerjee A, Chandra PK, Chowdhury A, Chakravarty R. 2006. Genotype, phylogenetic analysis, and transmission pattern of occult hepatitis B virus (HBV) infection in families of asymptomatic HBsAg carriers. J Med Virol 78(1):53-59.

Davaalkham D, Ojima T, Uehara R, Watanabe M, Oki I, Endo K, Takahashi M, Okamoto H, Nakamura Y. 2007. Analysis of hepatitis B surface antigen mutations in Mongolia: molecular epidemiology and implications for mass vaccination. Arch Virol 152(3):575-584.

Eichberg JW, Kalter SS. 1980. Hepatitis A and B: serologic survey of human and nonhuman primate sera. Lab Anim Sci 30(3):541-543.

Elkady A, Tanaka Y, Kurbanov F, Oynsuren T, Mizokami M. 2008. Virological and clinical implication of core promoter C1752/V1753 and T1764/G1766 mutations in hepatitis B virus genotype D infection in Mongolia. J Gastroenterol Hepatol 23(3):474-481.

Fagan EA, Williams R. 1990. Fulminant viral hepatitis. Br Med Bull 46(2):462-480.

Fang ZL, Hue S, Sabin CA, Li GJ, Yang JY, Chen QY, Fang KX, Huang J, Wang XY, Harrison TJ. 2010. A complex hepatitis B virus (X/C) recombinant is common in Long An county, Guangxi and may have originated in southern China. J Gen Virol.

Fang ZL, Yang J, Ge X, Zhuang H, Gong J, Li R, Ling R, Harrison TJ. 2002. Core promoter mutations (A(1762)T and G(1764)A) and viral genotype in chronic hepatitis B and hepatocellular carcinoma in Guangxi, China. J Med Virol 68(1):33-40.

Fares MA, Holmes EC. 2002. A revised evolutionary history of hepatitis B virus (HBV). J Mol Evol 54(6):807-814.

Feitelson MA, Clayton MM, Phimister B. 1990. Monoclonal antibodies raised to purified woodchuck hepatitis virus core antigen particles demonstrate X antigen reactivity. Virology 177(1):357-366.

Feld JJ, Wong DK, Heathcote EJ. 2009. Endpoints of therapy in chronic hepatitis B. Hepatology 49(5 Suppl):S96-S102.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

Ferrari C, Penna A, Bertoletti A, Cavalli A, Valli A, Missale G, Pilli M, Marchelli S, Giuberti T, Fiaccadori F. 1992. Immune pathogenesis of hepatitis B. Arch Virol Suppl 4:11-18.

Fisker N, Pedersen C, Lange M, Nguyen NT, Nguyen KT, Georgsen J, Christensen PB. 2004. Molecular epidemiology of hepatitis B virus infections in Denmark. J Clin Virol 31(1):46-52.

Fitzsimons D, Francois G, Hall A, McMahon B, Meheus A, Zanetti A, Duval B, Jilg W, Bocher WO, Lu SN, Akarca U, Lavanchy D, Goldstein S, Banatvala J, Damme PV. 2005. Long-term efficacy of hepatitis B

vaccine, booster policy, and impact of hepatitis B virus mutants. Vaccine 23(32):4158-4166.
Furusyo N, Nakashima H, Kashiwagi K, Kubo N, Hayashida K, Usuda S, Mishiro S, Kashiwagi S, Hayashi J. 2002. Clinical outcomes of hepatitis B virus (HBV) genotypes B and C in Japanese patients with chronic HBV infection. Am J Trop Med Hyg 67(2):151-157.
Girones R, Miller RH. 1989. Mutation rate of the hepadnavirus genome. Virology 170(2):595-597.
Gunther S, Baginski S, Kissel H, Reinke P, Kruger DH, Will H, Meisel H. 1996. Accumulation and persistence of hepatitis B virus core gene deletion mutants in renal transplant patients are associated with end-stage liver disease. Hepatology 24(4):751-758.
Hannoun C, Horal P, Lindh M. 2000a. Long-term mutation rates in the hepatitis B virus genome. J Gen Virol 81(Pt 1):75-83.
Hannoun C, Krogsgaard K, Horal P, Lindh M. 2002. Genotype mixtures of hepatitis B virus in patients treated with interferon. J Infect Dis 186(6):752-759.
Hannoun C, Norder H, Lindh M. 2000b. An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam. J Gen Virol 81(Pt 9):2267-2272.
Hannoun C, Soderstrom A, Norkrans G, Lindh M. 2005. Phylogeny of African complete genomes reveals a West African genotype A subtype of hepatitis B virus and relatedness between Somali and Asian A1 sequences. J Gen Virol 86(Pt 8):2163-2167.
Hsu HY, Chang MH, Liaw SH, Ni YH, Chen HL. 1999. Changes of hepatitis B surface antigen variants in carrier children before and after universal vaccination in Taiwan. Hepatology 30(5):1312-1317.
Huy TT, Ushijima H, Quang VX, Win KM, Luengrojanakul P, Kikuchi K, Sata T, Abe K. 2004. Genotype C of hepatitis B virus can be classified into at least two subgroups. J Gen Virol 85(Pt 2):283-292.
Huy TT, Ushijima H, Sata T, Abe K. 2006. Genomic characterization of HBV genotype F in Bolivia: genotype F subgenotypes correlate with geographic distribution and T(1858) variant. Arch Virol 151(3):589-597.
Hyams KC. 1995. Risks of chronicity following acute hepatitis B virus infection: a review. Clin Infect Dis 20(4):992-1000.
Janssen HL, van Zonneveld M, Senturk H, Zeuzem S, Akarca US, Cakaloglu Y, Simon C, So TM, Gerken G, de Man RA, Niesters HG, Zondervan P, Hansen B, Schalm SW. 2005. Pegylated interferon alfa-2b alone or in combination with lamivudine for HBeAg-positive chronic hepatitis B: a randomised trial. Lancet 365(9454):123-129.
Jardi R, Rodriguez F, Buti M, Costa X, Valdes A, Allende H, Schaper M, Galimany R, Esteban R, Guardia J. 2004. Mutations in the basic core promoter region of hepatitis B virus. Relationship with precore

variants and HBV genotypes in a Spanish population of HBV carriers. J Hepatol 40(3):507-514.
Jazayeri SM, Alavian SM, Carman WF. 2010. Hepatitis B virus: origin and evolution. J Viral Hepat 17(4):229-235.
Jukes T, Cantor C. 1969. Evolution of protein molecules. In Munro HN, editor, Mammalian Protein Metabolism, pp 21-132, Academic Press, New York:pages 21–132.
Kann M, Kochel HG, Uy A, Thomssen R. 1993. Diagnostic significance of antibodies to hepatitis B virus polymerase in acutely and chronically HBV-infected individuals. J Med Virol 40(4):285-290.
Kao JH, Chen PJ, Lai MY, Chen DS. 2000a. Hepatitis B genotypes correlate with clinical outcomes in patients with chronic hepatitis B. Gastroenterology 118(3):554-559.
Kao JH, Chen PJ, Lai MY, Chen DS. 2002. Genotypes and clinical phenotypes of hepatitis B virus in patients with chronic hepatitis B virus infection. J Clin Microbiol 40(4):1207-1209.
Kao JH, Chen PJ, Lai MY, Chen DS. 2003. Basal core promoter mutations of hepatitis B virus increase the risk of hepatocellular carcinoma in hepatitis B carriers. Gastroenterology 124(2):327-334.
Kao JH, Wu NH, Chen PJ, Lai MY, Chen DS. 2000b. Hepatitis B genotypes and the response to interferon therapy. J Hepatol 33(6):998-1002.
Kidd-Ljunggren K, Broman E, Ekvall H, Gustavsson O. 1999. Nosocomial transmission of hepatitis B virus infection through multiple-dose vials. J Hosp Infect 43(1):57-62.
Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16(2):111-120.
Klinkert MQ, Theilmann L, Pfaff E, Schaller H. 1986. Pre-S1 antigens and antibodies early in the course of acute hepatitis B virus infection. J Virol 58(2):522-525.
Kramvis A, Arakawa K, Yu MC, Nogueira R, Stram DO, Kew MC. 2008. Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. J Med Virol 80(1):27-46.
Kurbanov F, Tanaka Y, Fujiwara K, Sugauchi F, Mbanya D, Zekeng L, Ndembi N, Ngansop C, Kaptue L, Miura T, Ido E, Hayami M, Ichimura H, Mizokami M. 2005. A new subtype (subgenotype) Ac (A3) of hepatitis B virus and recombination between genotypes A and E in Cameroon. J Gen Virol 86(Pt 7):2047-2056.
Kurbanov F, Tanaka Y, Kramvis A, Simmonds P, Mizokami M. 2008. When should "I" consider a new hepatitis B virus genotype? J Virol 82(16):8241-8242.
Kurbanov F, Tanaka Y, Mizokami M. 2010. Geographical and genetic diversity of the human hepatitis B virus. Hepatol Res 40(1):14-30.

Laras A, Koskinas J, Hadziyannis SJ. 2002. In vivo suppression of precore mRNA synthesis is associated with mutations in the hepatitis B virus core promoter. Virology 295(1):86-96.

Lavanchy D. 2004. Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. J Viral Hepat 11(2):97-107.

Li JS, Tong SP, Wen YM, Vitvitski L, Zhang Q, Trepo C. 1993. Hepatitis B virus genotype A rarely circulates as an HBe-minus mutant: possible contribution of a single nucleotide in the precore region. J Virol 67(9):5402-5410.

Liao SS, Li RC, Li H, Yang JY, Zeng XJ, Gong J, Wang SS, Li YP, Zhang KL. 1999. Long-term efficacy of plasma-derived hepatitis B vaccine: a 15-year follow-up study among Chinese children. Vaccine 17(20-21):2661-2666.

Lin CL, Liu CJ, Chen PJ, Lai MY, Chen DS, Kao JH. 2007. High prevalence of occult hepatitis B virus infection in Taiwanese intravenous drug users. J Med Virol 79(11):1674-1678.

Lindh M, Andersson AS, Gusdal A. 1997. Genotypes, nt 1858 variants, and geographic origin of hepatitis B virus--large-scale analysis using a new genotyping method. J Infect Dis 175(6):1285-1293.

Lindh M, Furuta Y, Vahlne A, Norkrans G, Horal P. 1995. Emergence of precore TAG mutation during hepatitis B e seroconversion and its dependence on pregenomic base pairing between nucleotides 1858 of 1896. J Infect Dis 172(5):1343-1347.

Lindh M, Gonzalez JE, Norkrans G, Horal P. 1998. Genotyping of hepatitis B virus by restriction pattern analysis of a pre-S amplicon. J Virol Methods 72(2):163-174.

Lindh M, Hannoun C, Dhillon AP, Norkrans G, Horal P. 1999. Core promoter mutations and genotypes in relation to viral replication and liver damage in East Asian hepatitis B virus carriers. J Infect Dis 179(4):775-782.

Lindh M, Horal P, Dhillon AP, Furuta Y, Norkrans G. 1996. Hepatitis B virus carriers without precore mutations in hepatitis B e antigen-negative stage show more severe liver damage. Hepatology 24(3):494-501.

Lindh M, Horal P, Norkrans G. 2000. Acute hepatitis B in Western Sweden--genotypes and transmission routes. Infection 28(3):161-163.

Lok AS, Akarca U, Greene S. 1994. Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal. Proc Natl Acad Sci U S A 91(9):4077-4081.

Lok AS, Lai CL, Leung N, Yao GB, Cui ZY, Schiff ER, Dienstag JL, Heathcote EJ, Little NR, Griffiths DA, Gardner SD, Castiglia M. 2003. Long-term safety of lamivudine treatment in patients with chronic hepatitis B. Gastroenterology 125(6):1714-1722.

Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 73(1):152-160.

Lusida MI, Nugrahaputra VE, Soetjipto, Handajani R, Nagano-Fujii M, Sasayama M, Utsumi T, Hotta H. 2008. Novel subgenotypes of hepatitis B virus genotypes C and D in Papua, Indonesia. J Clin Microbiol 46(7):2160-2166.

MacDonald DM, Holmes EC, Lewis JC, Simmonds P. 2000. Detection of hepatitis B virus infection in wild-born chimpanzees (Pan troglodytes verus): phylogenetic relationships with human and other primate genotypes. J Virol 74(9):4253-4257.

Magiorkinis EN, Magiorkinis GN, Paraskevis DN, Hatzakis AE. 2005. Re-analysis of a human hepatitis B virus (HBV) isolate from an East African wild born Pan troglodytes schweinfurthii: evidence for interspecies recombination between HBV infecting chimpanzee and human. Gene 349:165-171.

Marinos G, Torre F, Gunther S, Thomas MG, Will H, Williams R, Naoumov NV. 1996. Hepatitis B virus variants with core gene deletions in the evolution of chronic hepatitis B infection. Gastroenterology 111(1):183-192.

McEvoy BP, Lind JM, Wang ET, Moyzis RK, Visscher PM, van Holst Pellekaan SM, Wilton AN. 2010. Whole-genome genetic diversity in a sample of Australians with deep Aboriginal ancestry. Am J Hum Genet 87(2):297-305.

McMahon BJ, Lanier AP, Wainwright RB. 1998. Hepatitis B and hepatocellular carcinoma in Eskimo/Inuit population. Int J Circumpolar Health 57 Suppl 1:414-419.

Meldal BH, Moula NM, Barnes IH, Boukef K, Allain JP. 2009. A novel hepatitis B virus subgenotype, D7, in Tunisian blood donors. J Gen Virol 90(Pt 7):1622-1628.

Michaels MG, Lanford R, Demetris AJ, Chavez D, Brasky K, Fung J, Starzl TE. 1996. Lack of susceptibility of baboons to infection with hepatitis B virus. Transplantation 61(3):350-351.

Michalak TI, Pasquinelli C, Guilhot S, Chisari FV. 1994. Hepatitis B virus persistence after recovery from acute viral hepatitis. J Clin Invest 93(1):230-239.

Milich DR, Peterson DL, Schodel F, Jones JE, Hughes JL. 1995. Preferential recognition of hepatitis B nucleocapsid antigens by Th1 or Th2 cells is epitope and major histocompatibility complex dependent. J Virol 69(5):2776-2785.

Mizokami M, Orito E. 1999. Molecular evolution of hepatitis viruses. Intervirology 42(2-3):159-165.

Mizokami M, Orito E, Ohba K, Ikeo K, Lau JY, Gojobori T. 1997. Constrained evolution with respect to gene overlap of hepatitis B virus. J Mol Evol 44 Suppl 1:S83-90.

Moriyama K, Takada T, Tsutsumi Y, Fukada K, Ishibashi H, Niho Y, Maeda Y. 1994. Mutations in the transcriptional regulatory region of the precore and core/pregenome of a hepatitis B virus with defective HBeAg production. Fukuoka Igaku Zasshi 85(11):314-322.

Mulyanto, Depamede SN, Surayah K, Tjahyono AA, Jirintai, Nagashima S, Takahashi M, Okamoto H. 2010. Identification and characterization of novel hepatitis B virus subgenotype C10 in Nusa Tenggara, Indonesia. Arch Virol 155(5):705-715.

Mulyanto, Depamede SN, Surayah K, Tsuda F, Ichiyama K, Takahashi M, Okamoto H. 2009. A nationwide molecular epidemiological study on hepatitis B virus in Indonesia: identification of two novel subgenotypes, B8 and C7. Arch Virol 154(7):1047-1059.

Nagasaki F, Niitsuma H, Cervantes JG, Chiba M, Hong S, Ojima T, Ueno Y, Bondoc E, Kobayashi K, Ishii M, Shimosegawa T. 2006. Analysis of the entire nucleotide sequence of hepatitis B virus genotype B in the Philippines reveals a new subgenotype of genotype B. J Gen Virol 87(Pt 5):1175-1180.

Nakayoshi T, Maeshiro T, Nakasone H, Sakugawa H, Kinjo F, Orito E, 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. J Med Virol 70(3):350-354.

Ngui SL, O'Connell S, Eglin RP, Heptonstall J, Teo CG. 1997. Low detection rate and maternal provenance of hepatitis B virus S gene mutants in cases of failed postnatal immunoprophylaxis in England and Wales. J Infect Dis 176(5):1360-1365.

Nguyen VT, Law MG, Dore GJ. 2009. Hepatitis B-related hepatocellular carcinoma: epidemiological characteristics and disease burden. J Viral Hepat 16(7):453-463.

Ni YH, Chang MH, Hsu HY, Chen HL. 2000. Long-term follow-up study of core gene deletion mutants in children with chronic hepatitis B virus infection. Hepatology 32(1):124-128.

Norder H, Courouce AM, Coursaget P, Echevarria JM, Lee SD, Mushahwar IK, Robertson BH, Locarnini S, Magnius LO. 2004. Genetic diversity of hepatitis B virus strains derived worldwide: genotypes, subgenotypes, and HBsAg subtypes. Intervirology 47(6):289-309.

Norder H, Courouce AM, Magnius LO. 1992. Molecular basis of hepatitis B virus serotype variations within the four major subtypes. J Gen Virol 73 (Pt 12):3141-3145.

Norder H, Courouce AM, Magnius LO. 1994. Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the

hepatitis B virus, four of which represent two new genotypes. Virology 198(2):489-503.

Nurainy N, Muljono DH, Sudoyo H, Marzuki S. 2008. Genetic study of hepatitis B virus in Indonesia reveals a new subgenotype of genotype B in east Nusa Tenggara. Arch Virol 153(6):1057-1065.

O'Rourke DH, Raff JA. 2010. The human genetic history of the Americas: the final frontier. Curr Biol 20(4):R202-207.

Odgerel Z, Choi IK, Byun KS, Pak CY, Bennett S, Gu SH, Park K, Kee SH, Song KJ, Song JW. 2006. Complete genome sequence and phylogenetic analysis of hepatitis B virus (HBV) isolated from Mongolian patients with chronic HBV infection. Virus Genes 33(3):345-349.

Ogawa E, Furusyo N, Toyoda K, Takeoka H, Otaguro S, Hamada M, Murata M, Sawayama Y, Hayashi J. 2007. Transient elastography for patients with chronic hepatitis B and C virus infection: Non-invasive, quantitative assessment of liver fibrosis. Hepatol Res 37(12):1002-1010.

Okamoto H, Imai M, Kametani M, Nakamura T, Mayumi M. 1987a. Genomic heterogeneity of hepatitis B virus in a 54-year-old woman who contracted the infection through materno-fetal transmission. Jpn J Exp Med 57(4):231-236.

Okamoto H, Imai M, Tsuda F, Tanaka T, Miyakawa Y, Mayumi M. 1987b. Point mutation in the S gene of hepatitis B virus for a d/y or w/r subtypic change in two blood donors carrying a surface antigen of compound subtype adyr or adwr. J Virol 61(10):3030-3034.

Okamoto H, Tsuda F, Akahane Y, Sugai Y, Yoshiba M, Moriyama K, Tanaka T, Miyakawa Y, Mayumi M. 1994. Hepatitis B virus with mutations in the core promoter for an e antigen-negative phenotype in carriers with antibody to e antigen. J Virol 68(12):8102-8110.

Okamoto H, Tsuda F, Mayumi M. 1987c. Defective mutants of hepatitis B virus in the circulation of symptom-free carriers. Jpn J Exp Med 57(4):217-221.

Okamoto H, Tsuda F, Sakugawa H, Sastrosoewignjo RI, Imai M, Miyakawa Y, Mayumi M. 1988. Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. J Gen Virol 69 (Pt 10):2575-2583.

Olinger CM, Jutavijittum P, Hubschen JM, Yousukh A, Samountry B, Thammavong T, Toriyama K, Muller CP. 2008. Possible new hepatitis B virus genotype, southeast Asia. Emerg Infect Dis 14(11):1777-1780.

Olinger CM, Venard V, Njayou M, Oyefolu AO, Maiga I, Kemp AJ, Omilabu SA, le Faou A, Muller CP. 2006. Phylogenetic analysis of the precore/core gene of hepatitis B virus genotypes E and A in West Africa: new subtypes, mixed infections and recombinations. J Gen Virol 87(Pt 5):1163-1173.

Orito E, Mizokami M, Ina Y, Moriyama EN, Kameshima N, Yamamoto M, Gojobori T. 1989. Host-independent evolution and a genetic classification of the hepadnavirus family based on nucleotide sequences. Proc Natl Acad Sci U S A 86(18):7059-7062.

Orito E, Sugauchi F, Tanaka Y, Ichida T, Sata M, Tanaka E, Okanoue T, Sakugawa H, Watanabe H, Miyakawa H, Nishiguchi S, Kumada H, Ueda R, Mizokami M. 2005. Differences of hepatocellular carcinoma patients with hepatitis B virus genotypes of Ba, Bj or C in Japan. Intervirology 48(4):239-245.

Osiowy C, Giles E, Tanaka Y, Mizokami M, Minuk GY. 2006. Molecular evolution of hepatitis B virus over 25 years. J Virol 80(21):10307-10314.

Ozasa A, Tanaka Y, Orito E, Sugiyama M, Kang JH, Hige S, Kuramitsu T, Suzuki K, Tanaka E, Okada S, Tokita H, Asahina Y, Inoue K, Kakumu S, Okanoue T, Murawaki Y, Hino K, Onji M, Yatsuhashi H, Sakugawa H, Miyakawa Y, Ueda R, Mizokami M. 2006. Influence of genotypes and precore mutations on fulminant or chronic outcome of acute hepatitis B virus infection. Hepatology 44(2):326-334.

Panessa C, Hill WD, Giles E, Yu A, Harvard S, Butt G, Andonov A, Krajden M, Osiowy C. 2009. Genotype D amongst injection drug users with acute hepatitis B virus infection in British Columbia. J Viral Hepat 16(1):64-73.

Pignatelli M, Waters J, Brown D, Lever A, Iwarson S, Schaff Z, Gerety R, Thomas HC. 1986. HLA class I antigens on the hepatocyte membrane during recovery from acute hepatitis B virus infection and during interferon therapy in chronic hepatitis B virus infection. Hepatology 6(3):349-353.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14(9):817-818.

Preikschat P, Gunther S, Reinhold S, Will H, Budde K, Neumayer HH, Kruger DH, Meisel H. 2002. Complex HBV populations with mutations in core promoter, C gene, and pre-S region are associated with development of cirrhosis in long-term renal transplant recipients. Hepatology 35(2):466-477.

Pult I, Abbott N, Zhang YY, Summers J. 2001. Frequency of spontaneous mutations in an avian hepadnavirus infection. J Virol 75(20):9623-9632.

Purdy MA, Gonzales AC, Dimitrova Z, Khudyakov Y. 2008. Supragenotypic groups of the hepatitis B virus genome. J Gen Virol 89(Pt 5):1179-1183.

Ramezani A, Hasanjani Roshan MR, Kalantar E, Eslamifar A, Banifazl M, Taeb J, Aghakhani A, Gachkar L, Velayati AA. 2008. Association of human leukocyte antigen polymorphism with outcomes of hepatitis B virus infection. J Gastroenterol Hepatol 23(11):1716-1721.

Rodriguez-Frias F, Jardi R, Buti M, Schaper M, Hermosilla E, Valdes A, Allende H, Martell M, Esteban R, Guardia J. 2006. Hepatitis B virus genotypes and G1896A precore mutation in 486 Spanish patients with acute and chronic HBV infection. J Viral Hepat 13(5):343-350.

Rozario R, Ramakrishna B. 2003. Histopathological study of chronic hepatitis B and C: a comparison of two scoring systems. J Hepatol 38(2):223-229.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406-425.

Sakamoto T, Tanaka Y, Orito E, Co J, Clavio J, Sugauchi F, Ito K, Ozasa A, Quino A, Ueda R, Sollano J, Mizokami M. 2006. Novel subtypes (subgenotypes) of hepatitis B virus genotypes B and C among chronic liver disease patients in the Philippines. J Gen Virol 87(Pt 7):1873-1882.

Sakamoto T, Tanaka Y, Simonetti J, Osiowy C, Borresen ML, Koch A, Kurbanov F, Sugiyama M, Minuk GY, McMahon BJ, Joh T, Mizokami M. 2007. Classification of hepatitis B virus genotype B into 2 major types based on characterization of a novel subgenotype in Arctic indigenous populations. J Infect Dis 196(10):1487-1492.

Sakurai M, Sugauchi F, Tsai N, Suzuki S, Hasegawa I, Fujiwara K, Orito E, Ueda R, Mizokami M. 2004. Genotype and phylogenetic characterization of hepatitis B virus among multi-ethnic cohort in Hawaii. World J Gastroenterol 10(15):2218-2222.

Sanchez LV, Tanaka Y, Maldonado M, Mizokami M, Panduro A. 2007. Difference of hepatitis B virus genotype distribution in two groups of mexican patients with different risk factors. High prevalence of genotype H and G. Intervirology 50(1):9-15.

Sanchez-Tapias JM, Costa J, Mas A, Bruguera M, Rodes J. 2002. Influence of hepatitis B virus genotype on the long-term outcome of chronic hepatitis B in western patients. Gastroenterology 123(6):1848-1856.

Schaefer S, Magnius L, Norder H. 2009. Under construction: classification of hepatitis B virus genotypes and subgenotypes. Intervirology 52(6):323-325.

Schmeltzer P, Sherman KE. 2010. Occult Hepatitis B: Clinical Implications and Treatment Decisions. Dig Dis Sci.

Shapiro CN. 1993. Epidemiology of hepatitis B. Pediatr Infect Dis J 12(5):433-437.

Shuhart MC, Sullivan DG, Bekele K, Harrington RD, Kitahata MM, Mathisen TL, Thomassen LV, Emerson SS, Gretch DR. 2006. HIV infection and antiretroviral therapy: effect on hepatitis C virus quasispecies variability. J Infect Dis 193(9):1211-1218.

Simmonds P. 2001. The origin and evolution of hepatitis viruses in humans. J Gen Virol 82(Pt 4):693-712.

Simmonds P, Midgley S. 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. J Virol 79(24):15467-15476.

Sloan RD, Strang AL, Ramsay ME, Teo CG. 2009. Genotyping of acute HBV isolates from England, 1997-2001. J Clin Virol 44(2):157-160.
Song Y, Dai E, Wang J, Liu H, Zhai J, Chen C, Du Z, Guo Z, Yang R. 2006. Genotyping of hepatitis B virus (HBV) by oligonucleotides microarray. Mol Cell Probes 20(2):121-127.
Starkman SE, MacDonald DM, Lewis JC, Holmes EC, Simmonds P. 2003. Geographic and species association of hepatitis B virus genotypes in non-human primates. Virology 314(1):381-393.
Stevens CE, Taylor PE, Tong MJ, Toy PT, Vyas GN, Nair PV, Weissman JY, Krugman S. 1987. Yeast-recombinant hepatitis B vaccine. Efficacy with hepatitis B immune globulin in prevention of perinatal hepatitis B virus transmission. JAMA 257(19):2612-2616.
Struve J, Giesecke J, Olcen P, von Sydow M, Weiland O. 1992. Prevalence of hepatitis B virus markers in Sweden: a community-based serosurvey of 4,000 young Swedish adults. Am J Epidemiol 135(4):409-417.
Stuyver L, De Gendt S, Van Geyt C, Zoulim F, Fried M, Schinazi RF, Rossau R. 2000. A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. J Gen Virol 81(Pt 1):67-74.
Sugauchi F, Kumada H, Acharya SA, Shrestha SM, Gamutan MT, Khan M, Gish RG, Tanaka Y, Kato T, Orito E, Ueda R, Miyakawa Y, Mizokami M. 2004a. Epidemiological and sequence differences between two subtypes (Ae and Aa) of hepatitis B virus genotype A. J Gen Virol 85(Pt 4):811-820.
Sugauchi F, Kumada H, Sakugawa H, Komatsu M, Niitsuma H, Watanabe H, Akahane Y, Tokita H, Kato T, Tanaka Y, Orito E, Ueda R, Miyakawa Y, Mizokami M. 2004b. Two subtypes of genotype B (Ba and Bj) of hepatitis B virus in Japan. Clin Infect Dis 38(9):1222-1228.
Sugauchi F, Mizokami M, Orito E, Ohno T, Kato H, Suzuki S, Kimura Y, Ueda R, Butterworth LA, Cooksley WG. 2001. A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. J Gen Virol 82(Pt 4):883-892.
Sugauchi F, Ohno T, Orito E, Sakugawa H, Ichida T, Komatsu M, Kuramitsu T, Ueda R, Miyakawa Y, Mizokami M. 2003. Influence of hepatitis B virus genotypes on the development of preS deletions and advanced liver disease. J Med Virol 70(4):537-544.
Sugauchi F, Orito E, Ichida T, Kato H, Sakugawa H, Kakumu S, Ishida T, Chutaputti A, Lai CL, Ueda R, Miyakawa Y, Mizokami M. 2002. Hepatitis B virus of genotype B with or without recombination with genotype C over the precore region plus the core gene. J Virol 76(12):5985-5992.
Sundberg KM, Restorp K, Von Sydow M, Insulander M, Norder H, Magnius L. 2009. [New typing of hepatitis B virus improves transmission

monitoring. Prospective molecular typing shows increased sexual transmission]. Lakartidningen 106(17):1169-1171.

Takahashi M, Nishizawa T, Gotanda Y, Tsuda F, Komatsu F, Kawabata T, Hasegawa K, Altankhuu M, Chimedregzen U, Narantuya L, Hoshino H, Hino K, Kagawa Y, Okamoto H. 2004. High prevalence of antibodies to hepatitis A and E viruses and viremia of hepatitis B, C, and D viruses among apparently healthy populations in Mongolia. Clin Diagn Lab Immunol 11(2):392-398.

Tatematsu K, Tanaka Y, Kurbanov F, Sugauchi F, Mano S, Maeshiro T, Nakayoshi T, Wakuta M, Miyakawa Y, Mizokami M. 2009. A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. J Virol 83(20):10538-10547.

The Swedish Council for Information on Alcohol and Other Drugs. 2009. Drug Trends in Sweden 2009.

Thursz MR, Kwiatkowski D, Allsopp CE, Greenwood BM, Thomas HC, Hill AV. 1995. Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia. N Engl J Med 332(16):1065-1069.

Thuy le TT, Ryo H, Van Phung L, Furitsu K, Nomura T. 2005. Distribution of genotype/subtype and mutational spectra of the surface gene of hepatitis B virus circulating in Hanoi, Vietnam. J Med Virol 76(2):161-169.

Tiollais P, Pourcel C, Dejean A. 1985. The hepatitis B virus. Nature 317(6037):489-495.

Toan NL, Song le H, Kremsner PG, Duy DN, Binh VQ, Koeberlein B, Kaiser S, Kandolf R, Torresi J, Bock CT. 2006. Impact of the hepatitis B virus genotype and genotype mixtures on the course of liver disease in Vietnam. Hepatology 43(6):1375-1384.

Tran HT, Ushijima H, Quang VX, Phuong N, Li TC, Hayashi S, Xuan Lien T, Sata T, Abe K. 2003. Prevalence of hepatitis virus types B through E and genotypic distribution of HBV and HCV in Ho Chi Minh City, Vietnam. Hepatol Res 26(4):275-280.

Tran TT, Trinh TN, Abe K. 2008. New complex recombinant genotype of hepatitis B virus identified in Vietnam. J Virol 82(11):5657-5663.

Truett GE, Heeger P, Mynatt RL, Truett AA, Walker JA, Warman ML. 2000. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). Biotechniques 29(1):52, 54.

Tsubota A, Kumada H, Takaki K, Chayama K, Kobayashi M, Suzuki Y, Saitoh S, Arase Y, Murashima N, Ikeda K. 1998. Deletions in the hepatitis B virus core gene may influence the clinical outcome in hepatitis B e antigen-positive asymptomatic healthy carriers. J Med Virol 56(4):287-293.

Utsumi T, Lusida MI, Yano Y, Nugrahaputra VE, Amin M, Juniastuti, Soetjipto, Hayashi Y, Hotta H. 2009. Complete genome sequence

and phylogenetic relatedness of hepatitis B virus isolates in Papua, Indonesia. J Clin Microbiol 47(6):1842-1847.

Viviani S, Jack A, Hall AJ, Maine N, Mendy M, Montesano R, Whittle HC. 1999. Hepatitis B vaccination in infancy in The Gambia: protection against carriage at 9 years of age. Vaccine 17(23-24):2946-2950.

Wai CT, Chu CJ, Hussain M, Lok AS. 2002. HBV genotype B is associated with better response to interferon therapy in HBeAg(+) chronic hepatitis than genotype C. Hepatology 36(6):1425-1430.

Wang YM, Tang YZ. 2009. Antiviral therapy for hepatitis B virus associated hepatic failure. Hepatobiliary Pancreat Dis Int 8(1):17-24.

Wang Z, Zhang J, Yang H, Li X, Wen S, Guo Y, Sun J, Hou J. 2003. Quantitative analysis of HBV DNA level and HBeAg titer in hepatitis B surface antigen positive mothers and their babies: HBeAg passage through the placenta and the rate of decay in babies. J Med Virol 71(3):360-366.

Webster GJ, Hallett R, Whalley SA, Meltzer M, Balogun K, Brown D, Farrington CP, Sharma S, Hamilton G, Farrow SC, Ramsay ME, Teo CG, Dusheiko GM. 2000. Molecular epidemiology of a large outbreak of hepatitis B linked to autohaemotherapy. Lancet 356(9227):379-384.

Xu Z, Liu Y, Xu T, Chen L, Si L, Wang Y, Ren X, Zhong Y, Zhao J, Xu D. 2010. Acute hepatitis B infection associated with drug-resistant hepatitis B virus. J Clin Virol 48(4):270-274.

Yamaura T, Tanaka E, Matsumoto A, Rokuhara A, Orii K, Yoshizawa K, Miyakawa Y, Kiyosawa K. 2003. A case-control study for early prediction of hepatitis B e antigen seroconversion by hepatitis B virus DNA levels and mutations in the precore region and core promoter. J Med Virol 70(4):545-552.

Yuan TT, Lin MH, Qiu SM, Shih C. 1998. Functional characterization of naturally occurring variants of human hepatitis B virus containing the core internal deletion mutation. J Virol 72(3):2168-2176.

Yuki N, Nagaoka T, Yamashiro M, Mochizuki K, Kaneko A, Yamamoto K, Omura M, Hikiji K, Kato M. 2003. Long-term histologic and virologic outcomes of acute self-limited hepatitis B. Hepatology 37(5):1172-1179.

Zehender G, De Maddalena C, Giambelli C, Milazzo L, Schiavini M, Bruno R, Tanzi E, Galli M. 2008. Different evolutionary rates and epidemic growth of hepatitis B virus genotypes A and D. Virology 380(1):84-90.